About This Manual

Congratulations on your purchase of the PASS package! PASS offers:

- Easy parameter entry.
- A comprehensive list of power analysis routines that are accurate and verified, yet are quick and easy to learn and use.
- Straightforward procedures for creating paper printouts and file copies of both the numerical and graphical reports.

Our goal is that with the help of these user's guides, you will be up and running on PASS quickly. After reading the quick start manual (at the front of User's Guide I) you will only need to refer to the chapters corresponding to the procedures you want to use. The discussion of each procedure includes one or more tutorials that will take you step-by-step through the tasks necessary to run the procedure.

I believe you will find that these user’s guides provides a quick, easy, efficient, and effective way for first-time PASS users to get up and running.

I look forward to any suggestions you have to improve the usefulness of this manual and/or the PASS system. Meanwhile, good computing!

Jerry Hintze, President
Important: The enclosed Power Analysis and Sample Size software program (PASS) is licensed by NCSS to customers for their use only on the terms set forth below. Your purchase and use of the PASS system indicates your acceptance of these terms.

1. **LICENSE.** NCSS hereby agrees to grant you a non-exclusive license to use the accompanying PASS program subject to the terms and restrictions set forth in this License Agreement.

2. **COPYRIGHT.** PASS and its documentation are copyrighted. You may not copy or otherwise reproduce any part of PASS or its documentation, except that you may load PASS into a computer as an essential step in executing it on the computer and make backup copies for your use on the same computer.

3. **BACKUP POLICY.** PASS may be backed up by you for your use on the same machine for which PASS was purchased.

4. **RESTRICTIONS ON USE AND TRANSFER.** The original and any backup copies of PASS and its documentation are to be used only in connection with a single user. This user may load PASS onto several machines for his/her convenience (such as a desktop and laptop computer), but only for use by the licensee. You may physically transfer PASS from one computer to another, provided that PASS is used in connection with only one user. You may not distribute copies of PASS or its documentation to others. You may transfer this license together with the original and all backup copies of PASS and its documentation, provided that the transferee agrees to be bound by the terms of this License Agreement. PASS licenses may not be transferred more frequently than once in twelve months. Neither PASS nor its documentation may be modified or translated without written permission from NCSS.

5. **NO WARRANTY OF PERFORMANCE.** NCSS does not and cannot warrant the performance or results that may be obtained by using PASS. Accordingly, PASS and its documentation are licensed "as is" without warranty as to their performance, merchantability, or fitness for any particular purpose. The entire risk as to the results and performance of PASS is assumed by you. Should PASS prove defective, you (and not NCSS nor its dealer) assume the entire cost of all necessary servicing, repair, or correction.

6. **LIMITED WARRANTY ON CD.** To the original licensee only, NCSS warrants the medium on which PASS is recorded to be free from defects in materials and faulty workmanship under normal use and service for a period of ninety days from the date PASS is delivered. If, during this ninety-day period, a defect in a CD should occur, the CD may be returned to NCSS at its address, or to the dealer from which PASS was purchased, and NCSS will replace the CD without charge to you, provided that you have sent a copy of your receipt for PASS. Your sole and exclusive remedy in the event of a defect is expressly limited to the replacement of the CD as provided above.

   Any implied warranties of merchantability and fitness for a particular purpose are limited in duration to a period of ninety (90) days from the date of delivery. If the failure of a CD has resulted from accident, abuse, or misapplication of the CD, NCSS shall have no responsibility to replace the CD under the terms of this limited warranty. This limited warranty gives you specific legal rights, and you may also have other rights which vary from state to state.

7. **LIMITATION OF LIABILITY.** Neither NCSS nor anyone else who has been involved in the creation, production, or delivery of PASS shall be liable for any direct, incidental, or consequential damages, such as, but not limited to, loss of anticipated profits or benefits, resulting from the use of PASS or arising out of any breach of any warranty. Some states do not allow the exclusion or limitation of direct, incidental, or consequential damages, so the above limitation may not apply to you.

8. **TERM.** The license is effective until terminated. You may terminate it at any time by destroying PASS and documentation together with all copies, modifications, and merged portions in any form. It will also terminate if you fail to comply with any term or condition of this License Agreement. You agree upon such termination to destroy PASS and documentation together with all copies, modifications, and merged portions in any form.

9. **YOUR USE OF PASS ACKNOWLEDGES** that you have read this customer license agreement and agree to its terms. You further agree that the license agreement is the complete and exclusive statement of the agreement between us and supersedes any proposal or prior agreement, oral or written, and any other communications between us relating to the subject matter of this agreement.

Dr. Jerry L. Hintze & NCSS, Kaysville, Utah
Preface

PASS (Power Analysis and Sample Size) is an advanced, easy-to-use statistical analysis software package. The system was designed and written by Dr. Jerry L. Hintze and other members of the development team over the last twenty years. Dr. Hintze drew upon his experience both in teaching statistics at the university level and in various types of statistical consulting.

The present version, written for 32-bit and 64-bit versions of Microsoft Windows (7, Vista, XP) computer systems, is the result of several iterations. Experience over the years with several different types of users has helped the program evolve into its present form.

NCSS maintains a website at www.ncss.com where we make the latest edition of PASS available for free downloading. The software is password protected, so only users with valid serial numbers may use this downloaded edition. We hope that you will download the latest edition routinely and thus avoid any bugs that have been corrected since you purchased your copy.

We believe PASS to be an accurate, exciting, easy-to-use program. If you find any portion which you feel needs to be changed, please let us know. Also, we openly welcome suggestions for additions and enhancements.

Verification

All calculations used in this program have been extensively tested and verified. First, they have been verified against the original journal article or textbook that contained the formulas. Second, they have been verified against second and third sources when these exist.
# User’s Guide II
## Table of Contents

### Control Charts
- 290 Control Charts for Process Means (Simulation)
- 295 Control Charts for Process Variation (Simulation)

### Means
- **One Mean**
  - 400 Inequality Tests
  - 405 Inequality Tests (Exponential Data)
  - 410 Inequality Tests (Simulation)
  - 412 Inequality Tests (Poisson Data)
  - 413 Non-Zero Null Tests
  - 415 Non-Inferiority Tests
  - 420 Confidence Intervals
  - 421 Confidence Intervals with Tolerance Probability
- **Two Independent Means**
  - 430 Inequality Tests using Differences (Two-Sample T-Test)
  - 431 Inequality Tests (Repeated Measures Design)
  - 435 Inequality Tests (Exponential Data)
  - 437 Inequality Tests (Poisson Data)
  - 440 Inequality Tests (Simulation)
  - 445 Inequality Tests using Ratios
  - 448 Non-Zero Null Tests using Differences
  - 450 Non-Inferiority Tests using Differences
  - 453 Non-Unity Null Tests using Ratios
  - 455 Non-Inferiority Tests using Ratios
  - 460 Equivalence Tests using Differences
  - 465 Equivalence Tests (Simulation)
  - 470 Equivalence Tests using Ratios
  - 471 Confidence Intervals for the Difference
  - 472 Confidence Intervals for the Difference with Tolerance Probability
  - 475 Group-Sequential Tests
  - 476 Group-Sequential Tests (Simulation)
  - 477 Group-Sequential Tests assuming Normality (Simulation)
  - 478 Group-Sequential Non-Inferiority Tests (Simulation)
- 480 Inequality Tests (Cluster-Randomized Design)

### Two Correlated (Paired) Means
- 485 Inequality Tests
- 490 Inequality Tests (Simulation)
- 495 Equivalence Tests (Simulation)
- 496 Confidence Intervals
- 497 Confidence Intervals with Tolerance Probability

### Two Independent Means in a 2x2 Cross-Over Design
- 500 Inequality Tests using Differences
- 505 Inequality Tests using Ratios
- 508 Non-Zero Null Tests using Differences
- 510 Non-Inferiority Tests using Differences
- 513 Non-Unity Null Tests using Ratios
- 515 Non-Inferiority Tests using Ratios
- 520 Equivalence Tests using Differences
- 525 Equivalence Tests using Ratios

### Two Independent Means in a Higher-Order Cross-Over Design
- 528 Non-Zero Null Tests using Differences
- 530 Non-Inferiority Tests using Differences
- 533 Non-Unity Null Tests using Ratios
- 535 Non-Inferiority Tests using Ratios
- 540 Equivalence Tests using Differences
- 545 Equivalence Tests using Ratios

### Many Means (ANOVA)
- 550 One-Way Analysis of Variance
- 551 Analysis of Covariance (ANCOVA)
- 555 One-Way Analysis of Variance (Simulation)
- 560 Fixed Effects Analysis of Variance
- 565 Randomized Block Analysis of Variance
- 570 Repeated Measures Analysis of Variance

### References and Index
- References
- Index
## User’s Guide III
### Table of Contents

### Means (Continued)

<table>
<thead>
<tr>
<th>Page</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>571</td>
<td>Mixed Models</td>
</tr>
<tr>
<td>575</td>
<td>Multiple Comparisons</td>
</tr>
<tr>
<td>580</td>
<td>Pair-Wise Multiple Comparisons (Simulation)</td>
</tr>
<tr>
<td>585</td>
<td>Multiple Comparisons of Treatments vs. a Control (Simulation)</td>
</tr>
<tr>
<td>590</td>
<td>Multiple Contrasts (Simulation)</td>
</tr>
<tr>
<td>595</td>
<td>Williams Test for the Minimum Effective Dose</td>
</tr>
</tbody>
</table>

### Mixed Models

<table>
<thead>
<tr>
<th>Page</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>571</td>
<td>Mixed Models</td>
</tr>
</tbody>
</table>

### Multiple Comparisons

<table>
<thead>
<tr>
<th>Page</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>575</td>
<td>Multiple Comparisons</td>
</tr>
<tr>
<td>580</td>
<td>Pair-Wise Multiple Comparisons (Simulation)</td>
</tr>
<tr>
<td>585</td>
<td>Multiple Comparisons of Treatments vs. a Control (Simulation)</td>
</tr>
<tr>
<td>590</td>
<td>Multiple Contrasts (Simulation)</td>
</tr>
<tr>
<td>595</td>
<td>Williams Test for the Minimum Effective Dose</td>
</tr>
</tbody>
</table>

### Multivariate Means

<table>
<thead>
<tr>
<th>Page</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>600</td>
<td>Hotelling’s T2</td>
</tr>
<tr>
<td>605</td>
<td>Multivariate Analysis of Variance (MANOVA)</td>
</tr>
</tbody>
</table>

### Microarrays

<table>
<thead>
<tr>
<th>Page</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>610</td>
<td>One-Sample or Paired T-Test</td>
</tr>
<tr>
<td>615</td>
<td>Two-Sample T-Test</td>
</tr>
</tbody>
</table>

### Standard Deviations

<table>
<thead>
<tr>
<th>Page</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>640</td>
<td>Confidence Intervals for One Standard Deviation using Standard Deviation</td>
</tr>
<tr>
<td>641</td>
<td>Confidence Intervals for One Standard Deviation with Tolerance Probability</td>
</tr>
<tr>
<td>642</td>
<td>Confidence Intervals for One Standard Deviation using Relative Error</td>
</tr>
</tbody>
</table>

### Variances

#### One Variance

<table>
<thead>
<tr>
<th>Page</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>650</td>
<td>Inequality Tests</td>
</tr>
<tr>
<td>651</td>
<td>Confidence Intervals using Variance</td>
</tr>
<tr>
<td>652</td>
<td>Confidence Intervals with Tolerance Probability</td>
</tr>
<tr>
<td>653</td>
<td>Confidence Intervals using Relative Error</td>
</tr>
</tbody>
</table>

### Two Variances

<table>
<thead>
<tr>
<th>Page</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>655</td>
<td>Inequality Tests</td>
</tr>
<tr>
<td>656</td>
<td>Confidence Intervals for the Ratio of Two Variances using Variances</td>
</tr>
<tr>
<td>657</td>
<td>Confidence Intervals for the Ratio of Two Variances using Relative Error</td>
</tr>
</tbody>
</table>

### Normality Tests

<table>
<thead>
<tr>
<th>Page</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>670</td>
<td>Normality Tests (Simulation)</td>
</tr>
</tbody>
</table>

### Survival Analysis

<table>
<thead>
<tr>
<th>Page</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>700</td>
<td>Logrank Tests (Freedman)</td>
</tr>
<tr>
<td>705</td>
<td>Logrank Tests (Lachin and Foulkes)</td>
</tr>
<tr>
<td>706</td>
<td>Logrank Tests for Non-Inferiority</td>
</tr>
<tr>
<td>710</td>
<td>Group-Sequential Logrank Tests</td>
</tr>
<tr>
<td>711</td>
<td>Group-Sequential Logrank Tests (Simulation)</td>
</tr>
<tr>
<td>715</td>
<td>Logrank Tests (Lakatos)</td>
</tr>
</tbody>
</table>

### Correlations

<table>
<thead>
<tr>
<th>Page</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>800</td>
<td>Inequality Tests for One Correlation</td>
</tr>
<tr>
<td>801</td>
<td>Confidence Intervals for One Correlation</td>
</tr>
<tr>
<td>805</td>
<td>Inequality Tests for Two Correlations</td>
</tr>
<tr>
<td>810</td>
<td>Inequality Tests for Intraclass Correlation</td>
</tr>
<tr>
<td>811</td>
<td>Kappa Test for Agreement Between Two Raters</td>
</tr>
<tr>
<td>815</td>
<td>Inequality Tests for One Coefficient Alpha</td>
</tr>
<tr>
<td>820</td>
<td>Inequality Tests for Two Coefficient Alphas</td>
</tr>
</tbody>
</table>

### Regression

<table>
<thead>
<tr>
<th>Page</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>850</td>
<td>Cox Regression</td>
</tr>
<tr>
<td>855</td>
<td>Linear Regression</td>
</tr>
<tr>
<td>856</td>
<td>Confidence Intervals for Linear Regression Slope</td>
</tr>
<tr>
<td>860</td>
<td>Logistic Regression</td>
</tr>
<tr>
<td>865</td>
<td>Multiple Regression</td>
</tr>
<tr>
<td>870</td>
<td>Poisson Regression</td>
</tr>
</tbody>
</table>
User’s Guide III
Table of Contents (Continued)

Design of Experiments
880  Randomization Lists
881  Two-Level Designs
882  Fractional Factorial Designs
883  Balanced Incomplete Block Designs
884  Latin Square Designs
885  Response Surface Designs
886  Screening Designs
887  Taguchi Designs
888  D-Optimal Designs
889  Design Generator

Tools, Helps, and Aids
900  Chi-Square Effect Size Estimator
903  Survival Parameter Conversion Tool
905  Standard Deviation Estimator

910  Odds Ratio and Proportions Estimator
912  Standard Deviation of Means Estimator
915  Probability Calculator
920  Data Simulator
925  Spreadsheets
930  Macros

Plots
940  Scatter Plots
945  Histograms

References and Index
References
Index
Chapter 290

Control Charts for Process Means (Simulation)

Introduction

This procedure allows you to study the run length distribution of Shewhart (Xbar), Cusum, FIR Cusum, and EWMA process control charts for means using simulation. This procedure can also be used to study charts with a single observation at each sample. The in-control mean and standard deviation can be input directly or a specified number of in-control preliminary samples can be simulated based on a user-determined in-control distribution. The out-of-control distribution is flexible in terms of distribution type and distribution parameters. The Shewhart, Cusum, and EWMA parameters can also be flexibly input. This procedure can also be used to determine the necessary sample size to obtain a given run length.

Simulation Details

If the in-control mean and in-control standard deviation are assumed to be known, the steps to the simulation process are as follows (assume a sample consists of n observations).

1. An out-of-control sample of size \( n \) is generated according to the specified distribution parameters of the out-of-control distribution.

2. The average of the sample is produced and, if necessary for the particular type of control chart, the standard deviation.

3. Based on the control chart criteria, it is determined whether this sample results in an out-of-control signal.

4. If the sample results in an out-of-control signal, the sample number is recorded as the run length for that simulation. If the sample does not result in an out-of-control signal, return to Step 1.

5. Steps 1 through 4 are repeated until the number of simulations \( N_{sim} \) is reached. The result is \( N_{sim} \) run lengths.

6. The average or median or specified percentile of the run length distribution is reported.
If the in-control mean and in-control standard deviation are to be simulated based on in-control preliminary samples \((NP_{\text{Prelim}})\), the steps to the simulation process are as follows (assume a sample consists of \(n\) observations).

1. \(NP_{\text{Prelim}}\) in-control samples of size \(n\) are generated according to the specified distribution parameters of the in-control distribution.
2. The in-control average and standard deviation are calculated based on the \(NP_{\text{Prelim}}\) simulated in-control samples.
3. An out-of-control sample of size \(n\) is generated according to the specified distribution parameters of the out-of-control distribution.
4. The average of the sample is produced and, if necessary for the particular type of control chart, the standard deviation.
5. Based on the control chart criteria, it is determined whether this sample results in an out-of-control signal.
6. If the sample results in an out-of-control signal, the sample number is recorded as the run length for that simulation. If the sample does not result in an out-of-control signal, return to Step 3.
7. Steps 1 through 6 are repeated until the number of simulations \((N_{\text{sim}})\) is reached. The result is \(N_{\text{sim}}\) run lengths.
8. The average or median or specified percentile of the run length distribution is reported.

**Data Distributions**

A wide variety of distributions may be studied. These distributions can vary in skewness, elongation, or other features such as bimodality. A detailed discussion of the distributions that may be used in the simulation is provided in the chapter ‘Data Simulator’.

**Formulas for Constructing Control Charts**

Suppose we have \(k\) subgroups, each of size \(n\). Let \(x_{ij}\) represent the measurement in the \(j\)th sample of the \(i\)th subgroup. Three statistics that are routinely computed (depending on the type of control chart) for each subgroup are:

The subgroup mean

\[
\bar{x}_i = \frac{1}{n} \sum_{j=1}^{n} x_{ij}
\]

the subgroup range

\[
R_i = x_{(n)} - x_{(1)}
\]

and/or the subgroup standard deviation

\[
s_i = \sqrt{\frac{1}{n-1} \sum_{j=1}^{n} (x_{ij} - \bar{x}_i)^2}
\]
Estimating Sigma

Control limits vary according to the type of control chart used. These require an estimate of the process mean, $\mu_x$ (mu), and the process variability, $\sigma_x$ (sigma). A known estimate of $\mu_x$ may be supplied by the user, or it may be estimated by the average of the averages of a number of in-control preliminary samples, ‘x double bar’ (also known as the grand mean):

$$\bar{x}_i = \sum_{i=1}^{k} \bar{x}_{ij} / k$$

When sigma is not input directly, there are two methods available for estimating $\sigma_x$.

**Method 1: Estimating Sigma from the Ranges**

$$\hat{\sigma}_x = \frac{\bar{R}}{d_2}$$

where

$$\bar{R} = \frac{\sum_{i=1}^{k} R_i}{k}$$

$$d_2 = \frac{E(R)}{\sigma_x} = \frac{\mu_R}{\sigma_x}$$

Making the assumption that the $x_{ij}$’s follow the normal distribution with constant mean and variance, we can derive values for $d_2$ through the use of numerical integration.

**Method 2: Estimating Sigma from the Standard Deviations**

$$\hat{\sigma}_x = \frac{\bar{s}}{c_4}$$

where

$$\bar{s} = \frac{\sum_{i=1}^{k} s_i}{k}$$

$$c_4 = \frac{E(s)}{\sigma_x} = \frac{\mu_s}{\sigma_x}$$

Making the assumption that the $x_{ij}$’s follow the normal distribution with constant mean and variance, we can derive values for $c_4$ from the following formula.

$$c_4 = \sqrt{\frac{2}{n-1}} \Gamma\left(\frac{n}{2}\right) / \Gamma\left(\frac{n-1}{2}\right)$$
Estimating Sigma when \( n = 1 \)

When \( n \) is one, we cannot calculate \( R_i \) or \( s_i \) since these require at least two measurements. In this case, we use the standard deviation of all \( k \) measurements. Unfortunately, this method does not approximate the within-subgroup variation. Rather, it combines the within and the between subgroup variation.

**Xbar Chart Limits**

The lower and upper control limits for the Xbar chart are calculated using the formula

\[
LCL = \bar{x} - z \left( \frac{\hat{\sigma}_x}{\sqrt{n}} \right)
\]

\[
UCL = \bar{x} + z \left( \frac{\hat{\sigma}_x}{\sqrt{n}} \right)
\]

where \( z \) is a multiplier (often set to three) chosen to reduce the possibility of false alarms (signaling an out-of-control situation when the process is in control).

**Cusum and FIR Cusum Charts**

The Cusum chart has been shown to detect small shifts in the process average much quicker than the Xbar chart.

In **PASS** we use the Cusum procedure presented by Ryan (1989). This procedure may be summarized as follows:

1. Calculate all statistics as if you were going to generate an Xbar chart.
2. Calculate the \( z_i \) using the formula

\[
z_i = \frac{\bar{x}_i - \bar{x}}{\hat{\sigma}_x}
\]

3. Calculate the lower and upper cumulative sums as follows

\[
S_{L_i} = -\max[0,(-z_i - K) + S_{L_{i-1}}]
\]

\[
S_{H_i} = \max[0,(z_i - K) + S_{H_{i-1}}]
\]

4. The control limits are chosen as plus or minus \( h \). Often, \( K \) is set to 0.5 (for detecting one-sigma shifts in the mean) and \( h \) is set to 5.

5. Usually, the starting value for \( S_{L_i} \) and \( S_{H_i} \) is zero. Occasionally, however, a “fast initial response” (FIR) value of \( h/2 \) is used.
**EWMA Chart Limits**

The lower and upper control limits for the exponentially weighted moving-average (EWMA) chart are calculated using the formula

\[
LCL_t = \bar{x} - L \left( \frac{\hat{\sigma}_x}{\sqrt{n}} \right) \sqrt{\frac{R}{2 - R} \left[ 1 - \left( R \right)^{-2} \right]}
\]

\[
UCL_t = \bar{x} + L \left( \frac{\hat{\sigma}_x}{\sqrt{n}} \right) \sqrt{\frac{R}{2 - R} \left[ 1 - \left( R \right)^{-2} \right]}
\]

where \( L \) is a multiplier (usually set to three) and \( R \) is smoothing constant.

**Procedure Options**

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

**Data Tab**

**Solve For**

**Find (Solve For)**

Solve for either the run length distribution or the sample size. This is the parameter displayed on the vertical axis of the plot.

Note: The search for the sample size may take several minutes to complete. You may find it quicker and more informative to solve for run length distribution for a range of sample sizes.

**Run Length**

**Run Length Summary 1 and 2**

Select the first and second run length distribution summary value to be reported.

If a value between 0 and 1 (not inclusive) is entered, the corresponding distribution percentile is reported.

For example, if 0.40 is entered, the run length such that 40% of simulated run lengths are shorter is reported.

**Percentile or Average**

When solving for \( N \), the value entered for "Percentile or Average" indicates whether the search is based on the average run length or some percentile, such as the median.

If a value between 0 and 1 (not inclusive) is entered, the corresponding distribution percentile is used in the search for \( N \).

For example, if ARL (Average Run Length) is entered, and "Run Length" is 10, the program will search for the smallest \( N \) for which the average run length is 10. If 0.40 is entered, and "Run
Length" is 10, the program will search for the smallest N for which 40% of simulated run lengths are shorter than 10.

**Run Length**

When solving for N, this is run length that is searched for based on the specified "Percentile or Average".

For example, if ARL (Average Run Length) is entered, and "Run Length" is 10, the program will search for the smallest N for which the average run length is 10. If 0.40 is entered, and "Run Length" is 10, the program will search for the smallest N for which 40% of simulated run lengths are shorter than 10.

**Sample Size**

**N (Sample Size)**

Enter a value for the sample size, N. This is the number of observations in each sample. This sample size is also used when preliminary samples are used.

When a sample size of one is used, the preliminary sample standard deviation is estimated with all the preliminary data jointly.

You may enter a range such as 2 to 10 by 2 or a list of values separated by commas or blanks. You might try entering the same number two or three times to get an idea of the variability in your results. For example, you could enter "10 10 10".

**Simulations**

**Simulations**

Specify the number of Monte Carlo iterations. Each simulation consists of the full process required to obtain a run length. That is, the number of simulations here is the number of run lengths that are used to form the run length distribution.

A 2007 article in Communication in Statistics-Simulation and Computation suggests that 5000 to 10000 simulations is usually sufficient. For estimating percentiles near 0 or 1, more simulations may be needed.

For searches for N, you may wish to use fewer simulations to determine the ballpark for N and then larger numbers of simulations for fine-tuning.
Control Charts for Process Means (Simulation) 290-7

**Maximum Run Length**
For each simulation, if the run length reaches the maximum run length, the maximum run length is used for that simulation. This keeps simulations from going on indefinitely when run lengths are extremely long.

For example, if the average run length is close to the maximum run length, the maximum run length should be increased.

**Distributions Tab**

**In-Control and Out-of-Control Distributions**

**In-Control Distributions Specified By**
Specify whether a known mean and standard deviation will be used for the centerline and standard deviation for boundaries, or if the centerline and standard deviation for boundaries will be based on simulated preliminary samples.

**In-Control and Out-of-Control Distributions - Known**

**In-Control Mean (Center Line)**
This is the centerline mean from which boundaries and the in-control standard deviation are based.

**In-Control Standard Deviation**
This is an assumed known standard deviation from which the boundaries are calculated.

**In-Control and Out-of-Control Distributions – Preliminary Samples**

**In-Control Distribution**
This is the distribution from which in-control preliminary samples are simulated. The mean and standard deviation of the in-control samples are used to define the centerline and the control boundaries for that simulation.

The parameters of the distribution can be specified using numbers or letters. If letters are used, their values are specified in the boxes below. The value "M0" is reserved for the value of the mean under the null hypothesis.

**M0 (In-Control Mean)**
This is the mean of the in-control distribution. Preliminary samples are simulated based on a distribution with this mean.

These values are substituted for the "M0" in the distribution specification given above.

You can enter a list of values using the syntax 0 2 3 or 0 to 3 by 1.

Note that whether "M0" is the mean of the simulated distribution depends on the formula you have entered. For example, "N(M0 S)" does have M0 as its mean,
but "N(M0 S)-N(M0 S)" has a mean of zero.

**Number of Preliminary Samples**

For each run length simulation, this is the number of samples that are simulated to determine the centerline and standard deviation for the control chart.

For example, if a control chart uses 50 samples with \( N = 5 \) to establish the in control centerline and standard deviation, 50 should be entered here. The program will generate 50 samples of size 5 each, obtain means and standard deviations (or ranges) for each sample and use those to produce a centerline and standard deviation estimate as a basis for the simulated control chart.

Commonly recommended numbers of in-control preliminary samples are 30, 50, and 60.

Varying this number allows one to see the effect on the distribution of run lengths.

**Standard Deviation Estimation Method**

Specify whether the overall control chart standard deviation for each simulation is to be estimated from the ranges of the preliminary samples or the standard deviations of the preliminary samples.

If \( N \) is one, then this selection is ignored and the standard deviation is calculated based on all the preliminary samples jointly.

**In-Control and Out-of-Control Distributions – Out-of-Control**

**Out-of-Control Distribution**

This is the distribution from which out-of-control samples are simulated.

The parameters of the distribution can be specified using numbers or letters. If letters are used, their values are specified in the boxes below. The value "M1" is reserved for the value of the mean under the alternative hypothesis.

**M1 (Out-of-Control Mean)**

This is the mean of the out-of-control distribution. Out-of-control samples are simulated based on a distribution with this mean.

These values are substituted for the "M1" in the distribution specification given above.

You can enter a list of values using the syntax \( 0 \ 2 \ 3 \) or \( 0 \ to \ 3 \ by \ 1 \).

Note that whether "M1" is the mean of the simulated distribution depends on the formula you have entered. For example, "N(M1 S)" does have M1 as its mean, but "N(M1 S)-N(M1 S)" has a mean of zero.

**In-Control and Out-of-Control Distributions – S and Other Parameters**

**Parameter Values (S, A, B, C)**

Enter the numeric value(s) of parameter listed above. These values are substituted for the corresponding letter in the distribution specifications.

You can enter a list of values using syntax such as \( 0 \ 1 \ 2 \ 3 \) or \( 0 \ to \ 3 \ by \ 1 \).

You can also change the letter that is used as the name of this parameter.
Tests Tab

Include Tests
Specify whether to include this test in the Reports and Plots. The specific options for each test are specified below.

Shewhart (Xbar) Options

Use Z-Multiplier or Probability
Specify whether the Z-multiplier will be input directly or calculated based on the two-sided probability.

The Z-multiplier is the value of Z in the Shewhart Xbar chart formula for the limits:
meanline +/- Z * SDmean

Z-Multiplier
The Z-multiplier is the value of Z in the Shewhart Xbar chart formula for the limits:
meanline +/- Z * SDmean

Two-Sided Probability
This probability is used to calculate the Z-multiplier to create the Shewhart limits.
The Z-multiplier is the value of Z in the Shewhart Xbar chart formula for the limits:
meanline +/- Z * SDmean

Cusum and FIR Cusum Options

K
K is the value that is subtracted from the z-score in the CUSUM procedure formula.

H
H is the out-of-control threshold for the CUSUM procedure.
Typical choices are 4 and 5.

FIR
FIR or Fast Initial Response is a value that may be used to shorten the run lengths of the CUSUM procedure.
FIR is often set to H/2.

EWMA Options

R
R is the weight assigned to the terms of the moving average chart.
It is sometimes known as lambda or w.
Values in the range of 0.10 to 0.30 are common.
L

L defines the distance to the boundary. It is essentially a Z-multiplier used to create L-sigma limits.

It is sometimes known as q.

A typical value is 3.

---

**Example 1 – Run Length Distribution**

A researcher wishes to examine the run length distribution for a process monitored by a Shewhart (Xbar) chart. Six observations are to make up the sample examined at each hour. The in-control mean and standard deviation are known to be 5.2 and 3.1, respectively. The researcher would like see the run length distribution if the out-of-control mean and standard deviation are 6.2 and 3.1, respectively. A Z-Multiplier of 3.0 is to be used in the control chart for the boundaries.

---

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Control Charts for Means (Simulation)** procedure window by clicking on **Quality Control**, and then clicking on **Control Charts for Means (Simulation)**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Tab</strong></td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Run Length Distribution</td>
</tr>
<tr>
<td>Run Length Summary 1</td>
<td>ARL (Average Run Length)</td>
</tr>
<tr>
<td>Run Length Summary 2</td>
<td>MRL (Median Run Length)</td>
</tr>
<tr>
<td>Percentile or Average</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Run Length</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>6</td>
</tr>
<tr>
<td>Simulations</td>
<td>5000</td>
</tr>
<tr>
<td>Maximum Run Length</td>
<td>5000</td>
</tr>
<tr>
<td><strong>Distributions Tab</strong></td>
<td></td>
</tr>
<tr>
<td>In-Control Distributions Specified By</td>
<td>Mean, Standard Deviation Directly</td>
</tr>
<tr>
<td>In-Control Mean</td>
<td>5.2</td>
</tr>
<tr>
<td>In-Control Standard Deviation</td>
<td>3.1</td>
</tr>
<tr>
<td>Out-of-Control Distribution</td>
<td>N(M1 S)</td>
</tr>
<tr>
<td>M1 (Out-of-Control Mean)</td>
<td>6.2</td>
</tr>
<tr>
<td>S</td>
<td>3.1</td>
</tr>
<tr>
<td><strong>Tests Tab</strong></td>
<td></td>
</tr>
<tr>
<td>Shewhart (Xbar)</td>
<td>Checked</td>
</tr>
<tr>
<td>All Other Tests</td>
<td>Unchecked</td>
</tr>
<tr>
<td>Use Z-Multiplier or Probability</td>
<td>Z-Multiplier</td>
</tr>
</tbody>
</table>
Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Chart Type</th>
<th>Average Run Length</th>
<th>Median Run Length</th>
<th>N</th>
<th>In-Control Mean</th>
<th>Z</th>
<th>SD</th>
<th>LCL</th>
<th>UCL</th>
<th>Mean</th>
<th>Out-of-Control Mean</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shewhart</td>
<td>72.3</td>
<td>50.0</td>
<td>6</td>
<td>5.20</td>
<td>3.00</td>
<td>3.10</td>
<td>1.40</td>
<td>9.00</td>
<td>6.20</td>
<td>3.10</td>
<td></td>
</tr>
</tbody>
</table>

Notes

Simulations: 5000. Run Time: 12.16 seconds.

References


Report Definitions

Average Run Length is the mean of the run lengths across all simulations.
Median Run Length is the median of the run lengths across all simulations.
N is the number of units measured in each sample.
In-control Mean is the assumed known value of the center line of the control chart.
Z is the Z-multiplier which corresponds to the two-sided probability of a single sample mean outside the control limits.
In-control SD is the assumed known standard deviation that is used in the calculation of limits.
LCL and UCL are the lower and upper control chart limits, respectively.
Out-of-Control Mean (M1) is the mean of the distribution from which out-of-control samples are drawn. It's difference from the Mean Line is the shift to detect.
Other parameters, often S (standard deviation), define the distribution from which out-of-control samples are drawn.

Summary Statements

For a Xbar control chart with a mean line at 5.20 and lower and upper Shewhart control limits of 1.40 and 9.00, respectively, samples of size 6 from the distribution Normal(M1 S) with mean 6.20 have an average run length of 72.3 and a median run length of 50.0. These results are based on 5000 Monte Carlo samples.

Individual Summaries

Out-of-Control Distribution (determines size of shift): Normal(M1 S)

<table>
<thead>
<tr>
<th>Chart Type</th>
<th>Average Run Length</th>
<th>Median Run Length</th>
<th>N</th>
<th>In-Control Mean</th>
<th>Z</th>
<th>SD</th>
<th>LCL</th>
<th>UCL</th>
<th>Mean</th>
<th>Out-of-Control Mean</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shewhart</td>
<td>72.3</td>
<td>50.0</td>
<td>6</td>
<td>5.20</td>
<td>3.00</td>
<td>3.10</td>
<td>1.40</td>
<td>9.00</td>
<td>6.20</td>
<td>3.10</td>
<td></td>
</tr>
</tbody>
</table>

Average Run Length 95% CI: (70.3, 74.3)
Median Run Length 95% CI: (48.0, 52.0)

Average Run Length and Percentiles

<table>
<thead>
<tr>
<th>Avg</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>90%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>72.3</td>
<td>1.0</td>
<td>4.0</td>
<td>8.0</td>
<td>21.0</td>
<td>50.0</td>
<td>97.0</td>
<td>170.0</td>
<td>221.0</td>
<td>331.0</td>
</tr>
</tbody>
</table>
Plots Section

This plot shows the distribution of run lengths of 5000 simulated runs.

Example 2 – Comparing Tests

Continuing with the Example 1 parameters, the researchers would like to compare the various control chart tests available.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Control Charts for Means (Simulation) procedure window by clicking on Quality Control, and then clicking on Control Charts for Means (Simulation). You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

Option Value

Data Tab
Find (Solve For) Run Length Distribution
Run Length Summary 1 ARL (Average Run Length)
Run Length Summary 2 MRL (Median Run Length)
Percentile or Average Ignored since this is the Find setting
Ignored since this is the Find setting
N (Sample Size) 6
Simulations 5000
Maximum Run Length 5000
Control Charts for Process Means (Simulation) 290-13

Distributions Tab
In-Control Distributions Specified By ...... Mean, Standard Deviation Directly
In-Control Mean ............................................. 5.2
In-Control Standard Deviation ................... 3.1
Out-of-Control Distribution ..................... N(M1 S)
M1 (Out-of-Control Mean) ......................... 6.2
S .............................................................. 3.1

Tests Tab
All Tests .................................................. Checked
Use Z-Multiplier or Probability ................. Z-Multiplier
K .............................................................. 0.5
H .............................................................. 5
FIR ........................................................... 2.5
R .............................................................. 0.25
L .............................................................. 3

Output
Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Chart Type</th>
<th>Average Run Length</th>
<th>Median Run Length</th>
<th>N</th>
<th>In-Cntrl Mean</th>
<th>Z</th>
<th>In-Cntrl SD</th>
<th>LCL</th>
<th>UCL</th>
<th>Out-Cntrl Mean</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shewhart</td>
<td>74.3</td>
<td>53.0</td>
<td>6</td>
<td>5.20</td>
<td>3.00</td>
<td>3.10</td>
<td>1.40</td>
<td>9.00</td>
<td>6.20</td>
<td>3.10</td>
<td></td>
</tr>
<tr>
<td>Cusum</td>
<td>15.6</td>
<td>13.0</td>
<td>6</td>
<td>5.20</td>
<td>3.00</td>
<td>3.10</td>
<td>1.40</td>
<td>9.00</td>
<td>6.20</td>
<td>3.10</td>
<td></td>
</tr>
<tr>
<td>Cus+Shew</td>
<td>14.5</td>
<td>12.0</td>
<td>6</td>
<td>5.20</td>
<td>3.00</td>
<td>3.10</td>
<td>1.40</td>
<td>9.00</td>
<td>6.20</td>
<td>3.10</td>
<td></td>
</tr>
<tr>
<td>FIR Cusum</td>
<td>10.0</td>
<td>7.0</td>
<td>6</td>
<td>5.20</td>
<td>3.00</td>
<td>3.10</td>
<td>1.40</td>
<td>9.00</td>
<td>6.20</td>
<td>3.10</td>
<td></td>
</tr>
<tr>
<td>FIR+Shew</td>
<td>10.0</td>
<td>7.0</td>
<td>6</td>
<td>5.20</td>
<td>3.00</td>
<td>3.10</td>
<td>1.40</td>
<td>9.00</td>
<td>6.20</td>
<td>3.10</td>
<td></td>
</tr>
<tr>
<td>EWMA</td>
<td>17.1</td>
<td>13.0</td>
<td>6</td>
<td>5.20</td>
<td>3.00</td>
<td>3.10</td>
<td>1.40</td>
<td>9.00</td>
<td>6.20</td>
<td>3.10</td>
<td></td>
</tr>
<tr>
<td>EWMA+Shew</td>
<td>16.6</td>
<td>13.0</td>
<td>6</td>
<td>5.20</td>
<td>3.00</td>
<td>3.10</td>
<td>1.40</td>
<td>9.00</td>
<td>6.20</td>
<td>3.10</td>
<td></td>
</tr>
</tbody>
</table>

Notes

The FIR tests show the process is out-of-control much sooner than the other tests.
Example 3 – Validation Using Montgomery

A table in Montgomery (1991), page 298, gives the average run lengths for a mean shift of one standard deviation to be Cusum (10.4), Cusum with Shewhart (10.20), Cusum with FIR (6.35), Cusum with FIR and Shewhart (6.32).

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Control Charts for Means (Simulation)** procedure window by clicking on **Quality Control**, and then clicking on **Control Charts for Means (Simulation)**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Tab</strong></td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Run Length Distribution</td>
</tr>
<tr>
<td>Run Length Summary 1</td>
<td>ARL (Average Run Length)</td>
</tr>
<tr>
<td>Run Length Summary 2</td>
<td>MRL (Median Run Length)</td>
</tr>
<tr>
<td>Percentile or Average</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Run Length</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>1</td>
</tr>
<tr>
<td>Simulations</td>
<td>100000</td>
</tr>
<tr>
<td>Maximum Run Length</td>
<td>5000</td>
</tr>
<tr>
<td><strong>Distributions Tab</strong></td>
<td></td>
</tr>
<tr>
<td>In-Control Distributions Specified By</td>
<td>Mean, Standard Deviation Directly</td>
</tr>
<tr>
<td>In-Control Mean</td>
<td>0</td>
</tr>
<tr>
<td>In-Control Standard Deviation</td>
<td>1</td>
</tr>
<tr>
<td>Out-of-Control Distribution</td>
<td>N(M1 S)</td>
</tr>
<tr>
<td>M1 (Out-of-Control Mean)</td>
<td>1</td>
</tr>
<tr>
<td>S</td>
<td>1</td>
</tr>
<tr>
<td><strong>Tests Tab</strong></td>
<td></td>
</tr>
<tr>
<td>Cusum</td>
<td>Checked</td>
</tr>
<tr>
<td>Cusum + Shewhart</td>
<td>Checked</td>
</tr>
<tr>
<td>FIR Cusum</td>
<td>Checked</td>
</tr>
<tr>
<td>FIR Cusum + Shewhart</td>
<td>Checked</td>
</tr>
<tr>
<td>All Other Tests</td>
<td>Unchecked</td>
</tr>
<tr>
<td>Use Z-Multiplier or Probability</td>
<td>Z-Multiplier</td>
</tr>
<tr>
<td>Z-Multiplier</td>
<td>3.5</td>
</tr>
<tr>
<td>K</td>
<td>0.5</td>
</tr>
<tr>
<td>H</td>
<td>5</td>
</tr>
<tr>
<td>FIR</td>
<td>2.5</td>
</tr>
</tbody>
</table>
Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Chart Type</th>
<th>Average Run Length</th>
<th>Median Run Length</th>
<th>N</th>
<th>In-Cntrl Mean</th>
<th>Z</th>
<th>In-Cntrl SD</th>
<th>LCL</th>
<th>UCL</th>
<th>Out-of-Cntrl Mean</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cusum</td>
<td>10.36</td>
<td>9.00</td>
<td>1</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cus+Shew</td>
<td>10.26</td>
<td>9.00</td>
<td>1</td>
<td>0.00</td>
<td>3.50</td>
<td>-3.50</td>
<td>3.50</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIR Cusum</td>
<td>6.35</td>
<td>5.00</td>
<td>1</td>
<td>0.00</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FIR+Shew</td>
<td>6.33</td>
<td>5.00</td>
<td>1</td>
<td>0.00</td>
<td>3.50</td>
<td>-3.50</td>
<td>3.50</td>
<td>1.00</td>
<td>1.00</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes

The average run lengths are very close to those presented in Montgomery (1991).
290-16 Control Charts for Process Means (Simulation)
Chapter 295

Control Charts for Process Variation (Simulation)

Introduction

This procedure allows you to study the run length distribution of R, S, and S with probability limits process control charts using simulation. This procedure can also be used to study charts with two or more observations at each sample. The in-control standard deviation can be input directly or a specified number of in-control preliminary samples can be simulated based on a user-determined in-control distribution. This procedure can also be used to determine the necessary sample size to obtain a given run length.

Simulation Details

If the in-control standard deviation is assumed to be known (standard), the steps to the simulation process are as follows (assume a sample consists of \( n \) observations).

1. An out-of-control sample of size \( n \) is generated from a normal distribution with the specified out-of-control standard deviation.
2. The standard deviation of the sample is calculated and compared to the control limits.
3. If the sample results in an out-of-control signal, the sample number is recorded as the run length for that simulation. If the sample does not result in an out-of-control signal, return to Step 1.
4. Steps 1 through 3 are repeated until the number of simulations (\( N_{sim} \)) is reached. The result is \( N_{sim} \) run lengths.
5. The average or median or specified percentile of the run length distribution is reported.

If the in-control standard deviation is to be simulated based on in-control preliminary samples (\( N_{Prelim} \)), the steps to the simulation process are as follows (assume a sample consists of \( n \) observations).

1. \( N_{Prelim} \) in-control samples of size \( n \) are generated from a normal distribution with specified standard deviation of the in-control distribution.
2. The in-control standard deviation is calculated based on the \textit{NPrelim} simulated in-control samples.

3. An out-of-control sample of size \( n \) is generated from a normal distribution with the specified out-of-control standard deviation.

4. The standard deviation of the sample is calculated and compared to the control limits.

5. If the sample results in an out-of-control signal, the sample number is recorded as the run length for that simulation. If the sample does not result in an out-of-control signal, return to Step 3.

6. Steps 1 through 5 are repeated until the number of simulations (\( N_{\text{sim}} \)) is reached. The result is \( N_{\text{sim}} \) run lengths.

7. The average or median or specified percentile of the run length distribution is reported.

---

**Formulas for Constructing Control Charts**

The formula used for each set of control chart limits depends on whether the standard deviation is assumed to be a known, standard value, or estimated from in-control preliminary samples.

---

**R Chart Limits**

If the standard deviation is a known, standard value, the lower and upper control limits for the R chart are calculated (see Montgomery, 1991) using the formulae

\[
LCL = d_2 \bar{R} - z_{1-\alpha/2} d_3 \sigma \\
UCL = d_2 \bar{R} + z_{1-\alpha/2} d_3 \sigma
\]

where \( d_2 = R / \sigma \) and \( d_3 = \sigma / \bar{R} \).

If the limits are to be created from in-control preliminary samples, the lower and upper control limits for the R chart are calculated using the formulae

\[
LCL = \bar{R} - z_{1-\alpha/2} d_3 \frac{\bar{R}}{d_2} \\
UCL = \bar{R} + z_{1-\alpha/2} d_3 \frac{\bar{R}}{d_2}
\]

The statistic that is compared to the limits at each subgroup is the subgroup range

\[
R_i = x_{(n)} - x_{(1)}
\]

---

**S Chart Limits**

If the standard deviation is a known, standard value, the lower and upper control limits for the S chart are calculated (see Montgomery, 1991) using the formulae

\[
LCL = \frac{S}{d_3 \sigma} \\
UCL = \frac{S}{d_3 \sigma}
\]
\[ LCL = c_4 \sigma - z_{1-\alpha/2} \sigma \sqrt{1 - c_4^2} \]
\[ UCL = c_4 \sigma + z_{1-\alpha/2} \sigma \sqrt{1 - c_4^2} \]

where

\[ c_4 = \sqrt{\frac{2}{n-1}} \Gamma\left(\frac{n}{2}\right) \Gamma\left(\frac{n-1}{2}\right) \]

If the limits are to be created from in-control preliminary samples, the lower and upper control limits for the S chart are calculated using the formulae

\[ LCL = \bar{S} - z_{1-\alpha/2} \frac{\bar{S}}{c_4} \sqrt{1 - c_4^2} \]
\[ UCL = \bar{S} + z_{1-\alpha/2} \frac{\bar{S}}{c_4} \sqrt{1 - c_4^2} \]

The statistic that is compared to the limits at each subgroup is the subgroup standard deviation

\[ s_i = \sqrt{\frac{\sum_{j=1}^{n}(x_{ij} - \overline{x}_i)^2}{n-1}} \]

**S Chart with Probability Limits**

If the standard deviation is a known, standard value, the lower and upper control limits for the S chart with probability limits are calculated (see Ryan, 1989) using the formulae

\[ LCL = \sigma \sqrt{\frac{\chi^2_{a/2}}{n-1}} \]
\[ UCL = \sigma \sqrt{\frac{\chi^2_{1-a/2}}{n-1}} \]

If the limits are to be created from in-control preliminary samples, the lower and upper control limits for the S chart with probability limits are calculated using the formulae
The statistic that is compared to the limits at each subgroup is the same as that of the previous S Chart, namely, the subgroup standard deviation

\[
LCL = \frac{\bar{s}}{c_4} \sqrt{\frac{X^2_{\alpha/2}}{n-1}} \\
UCL = \frac{\bar{s}}{c_4} \sqrt{\frac{X^2_{1-\alpha/2}}{n-1}}
\]

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

Solve For

Find (Solve For)

Solve for either the run length distribution or the sample size. This is the parameter displayed on the vertical axis of the plot.

Note: The search for the sample size may take several minutes to complete. You may find it quicker and more informative to solve for run length distribution for a range of sample sizes.

Run Length

Run Length Summary 1 and 2

Select the first and second run length distribution summary value to be reported.

If a value between 0 and 1 (not inclusive) is entered, the corresponding distribution percentile is reported.

For example, if 0.40 is entered, the run length such that 40% of simulated run lengths are shorter is reported.
Percentile or Average
When solving for N, the value entered for "Percentile or Average" indicates whether the search is based on the average run length or some percentile, such as the median.

If a value between 0 and 1 (not inclusive) is entered, the corresponding distribution percentile is used in the search for N.

For example, if ARL (Average Run Length) is entered, and "Run Length" is 10, the program will search for the smallest N for which the average run length is 10. If 0.40 is entered, and "Run Length" is 10, the program will search for the smallest N for which 40% of simulated run lengths are shorter than 10.

Run Length
When solving for N, this is run length that is searched for based on the specified "Percentile or Average".

For example, if ARL (Average Run Length) is entered, and "Run Length" is 10, the program will search for the smallest N for which the average run length is 10. If 0.40 is entered, and "Run Length" is 10, the program will search for the smallest N for which 40% of simulated run lengths are shorter than 10.

Sample Size

N (Sample Size)
Enter a value for the sample size, N. This is the number of observations in each sample. This sample size is also used when preliminary samples are used.

When a sample size of one is used, the preliminary sample standard deviation is estimated with all the preliminary data jointly.

You may enter a range such as 2 to 10 by 2 or a list of values separated by commas or blanks.

You might try entering the same number two or three times to get an idea of the variability in your results. For example, you could enter "10 10 10".

Simulations

Simulations
Specify the number of Monte Carlo iterations. Each simulation consists of the full process required to obtain a run length. That is, the number of simulations here is the number of run lengths that are used to form the run length distribution.

A 2007 article in Communication in Statistics-Simulation and Computation suggests that 5000 to 10000 simulations is usually sufficient. For estimating percentiles near 0 or 1, more simulations may be needed.

For searches for N, you may wish to use fewer simulations to determine the ballpark for N and then larger numbers of simulations for fine-tuning.
Maximum Run Length
For each simulation, if the run length reaches the maximum run length, the maximum run length is used for that simulation. This keeps simulations from going on indefinitely when run lengths are extremely long.

For example, if the average run length is close to the maximum run length, the maximum run length should be increased.

Distributions, Tests Tab

In-Control and Out-of-Control Distributions

In-Control Distributions Specified By
Specify whether a standard known standard deviation will be used for the standard deviation for control limits, or if the standard deviation for control limits will be based on simulated preliminary samples.

In-Control and Out-of-Control Distributions - Known

In-Control Standard Deviation
This is an assumed known standard deviation from which the boundaries are calculated.

In-Control and Out-of-Control Distributions – Preliminary Samples

In-Control Distribution Standard Deviation
This is the standard deviation of the normal distribution that is used to simulate preliminary samples.

You can enter a single value or a list of values using the syntax 1 2 3 or 1 to 5 by 1.

Number of Preliminary Samples
For each run length simulation, this is the number of samples that are simulated to determine the range or standard deviation for calculating the limits of the control chart.

For example, if a control chart uses 50 samples with N = 5 to establish the in-control standard deviation, 50 should be entered here. The program will generate 50 samples of size 5 each, obtain standard deviations (or ranges) for each sample and use those to produce a standard deviation (or range) estimate as a basis for the simulated control chart.

Commonly recommended numbers of in-control preliminary samples are 30, 50, and 60.

Varying this number allows one to see the effect on the distribution of run lengths.
In-Control and Out-of-Control Distributions – Out-of-Control

Out-of-Control Standard Deviation
This is the standard deviation of the out-of-control distribution. Out-of-control samples are simulated based on a normal distribution with this standard deviation.
You can enter a single value or a list of values using the syntax 1 2 3 or 1 to 5 by 1.

Tests

Include Tests
Specify whether to include this test in the Reports and Plots. The specific options for each test are specified below.

Tests – Z or Probability Options

Specify Probability Using
For R and S charts, specify whether the Z-multiplier will be input directly or calculated based on the probability.
The Z-multiplier is the value of Z in the chart formula for the limits:
Rbar +/- Z * SD_Rbar for two-sided R charts, and
Sbar +/- Z * SD_Sbar for two-sided S charts.
For S with Probability Limits, if a Z-multiplier is specified, the probability corresponding to the Z-multiplier is used to create the limits.

Z-Multiplier
The Z-multiplier is the value of Z in the chart formula for the limits:
Rbar +/- Z * SD_Rbar for two-sided R charts, and
Sbar +/- Z * SD_Sbar for two-sided S charts.
For S with Probability Limits, the probability corresponding to the Z-multiplier is used to create the limits.
You can enter a list of values using the syntax 2 3 4 or 2 to 5 by 0.5.

Probability
For R and S charts, the probability is used to determine the Z-multiplier for the limits:
Rbar +/- Z * SD_Rbar for two-sided R charts, and
Sbar +/- Z * SD_Sbar for two-sided S charts.
For S with Probability Limits, the probability is used directly to create the limits using the chi-square distribution.
For two-sided tests, the probability is divided by two (for each side).
You can enter a list of values using the syntax 0.001 0.002 0.003 or 0.001 to 0.005 by 0.001.
Example 1 – Run Length Distribution

A researcher wishes to examine the run length distribution for a process monitored by an R chart. Ten observations are to make up the sample examined at each hour. The in-control standard deviation is known to be 3.8. The researcher would like to see the run length distribution if the out-of-control standard deviation is 4.6. A probability of 0.00135 is to be used in the control chart for the boundaries.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Control Charts for Variability (Simulation) procedure window by clicking on Quality Control, and then clicking on Control Charts for Variability (Simulation). You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td>Run Length Distribution</td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Run Length Summary 1</td>
</tr>
<tr>
<td></td>
<td>ARL (Average Run Length)</td>
</tr>
<tr>
<td>Run Length Summary 2</td>
<td>MRL (Median Run Length)</td>
</tr>
<tr>
<td>Percentile or Average</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Run Length</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>10</td>
</tr>
<tr>
<td>Simulations</td>
<td>5000</td>
</tr>
<tr>
<td>Maximum Run Length</td>
<td>5000</td>
</tr>
<tr>
<td>Distributions, Tests Tab</td>
<td>In-Control Distributions Specified By</td>
</tr>
<tr>
<td></td>
<td>Standard Deviation Directly</td>
</tr>
<tr>
<td>In-Control Standard Deviation</td>
<td>3.8</td>
</tr>
<tr>
<td>Out-of-Control Standard Deviation</td>
<td>4.6</td>
</tr>
<tr>
<td>R</td>
<td>Checked</td>
</tr>
<tr>
<td>All Other Tests</td>
<td>Unchecked</td>
</tr>
<tr>
<td>Two-Sided or Upper Limit Only</td>
<td>Upper Limit Only</td>
</tr>
<tr>
<td>Specify Probability Using</td>
<td>Probability</td>
</tr>
<tr>
<td>Probability</td>
<td>0.00135</td>
</tr>
</tbody>
</table>
Output

Click the Run button to perform the calculations and generate the following output.

**Numeric Results**

<table>
<thead>
<tr>
<th>Chart Type</th>
<th>Average Run Length</th>
<th>Median Run Length</th>
<th>N</th>
<th>In-Control SD</th>
<th>One-Sided Prob</th>
<th>Z</th>
<th>LCL</th>
<th>UCL</th>
<th>Out-of-Control SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>22.0</td>
<td>15.0</td>
<td>10</td>
<td>3.8000</td>
<td>0.00135</td>
<td>3.205</td>
<td>0.0000</td>
<td>20.7821</td>
<td>4.6000</td>
</tr>
</tbody>
</table>

**Notes**


**References**


**Report Definitions**

Average Run Length is the mean of the run lengths across all simulations.
Median Run Length is the median of the run lengths across all simulations.
N is the number of units measured in each sample.
In-control SD is the assumed known standard value of the standard deviation.
One-Sided Prob is the probability of a sample range or standard deviation outside the control limits.
Z is the Z-multiplier which corresponds to the One-sided probability of a single sample range or standard deviation outside the control limits.
In-control SD is the assumed known standard deviation that is used in the calculation of limits.
LCL and UCL are the lower and upper control chart limits, respectively.
Out-of-Control SD is the standard deviation of the normal distribution from which out-of-control samples are drawn.

**Summary Statements**

For an R control chart with lower and upper control limits of 0.0000 and 20.7821, respectively, samples of size 10 from a normal distribution with standard deviation 4.6000 have an average run length of 22.0 and a median run length of 15.0. These results are based on 5000 Monte Carlo samples.

**Individual Summaries**

Control limits were determined by specifying a standard in-control standard deviation directly.
Out-of-Control Distribution: Normal with Out-of-Control SD

<table>
<thead>
<tr>
<th>Chart Type</th>
<th>Average Run Length</th>
<th>Median Run Length</th>
<th>N</th>
<th>In-Control SD</th>
<th>One-Sided Prob</th>
<th>Z</th>
<th>LCL</th>
<th>UCL</th>
<th>Out-of-Control SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>22.0</td>
<td>15.0</td>
<td>10</td>
<td>3.8000</td>
<td>0.00135</td>
<td>3.205</td>
<td>0.0000</td>
<td>20.7821</td>
<td>4.6000</td>
</tr>
</tbody>
</table>

Average Run Length 95% CI: (21.4, 22.6)
Median Run Length 95% CI: (15.0, 16.0)

**Average Run Length and Percentiles**

<table>
<thead>
<tr>
<th>Avg</th>
<th>1%</th>
<th>5%</th>
<th>10%</th>
<th>25%</th>
<th>50%</th>
<th>75%</th>
<th>90%</th>
<th>95%</th>
<th>99%</th>
</tr>
</thead>
<tbody>
<tr>
<td>22.0</td>
<td>1.0</td>
<td>2.0</td>
<td>3.0</td>
<td>7.0</td>
<td>15.0</td>
<td>30.0</td>
<td>50.0</td>
<td>66.0</td>
<td>103.0</td>
</tr>
</tbody>
</table>
This plot shows the distribution of run lengths of 5000 simulated runs.

Example 2 – Comparing Tests

Continuing with the Example 1 parameters, the researchers would like to compare the various control chart tests available.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Control Charts for Means (Simulation) procedure window by clicking on Quality Control, and then clicking on Control Charts for Means (Simulation). You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Find (Solve For)</td>
<td>Run Length Distribution</td>
</tr>
<tr>
<td>Run Length Summary 1</td>
<td>ARL (Average Run Length)</td>
</tr>
<tr>
<td>Run Length Summary 2</td>
<td>MRL (Median Run Length)</td>
</tr>
<tr>
<td>Percentile or Average</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>10</td>
</tr>
<tr>
<td>Simulations</td>
<td>5000</td>
</tr>
<tr>
<td>Maximum Run Length</td>
<td>5000</td>
</tr>
</tbody>
</table>
Distributions, Tests Tab
In-Control Distributions Specified By: Standard Deviation Directly
In-Control Standard Deviation: 3.8
Out-of-Control Standard Deviation: 4.6
All Tests: Checked
Two-Sided or Upper Limit Only: Upper Limit Only
Specify Probability Using: Probability
Probability: 0.00135

Output
Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Chart Type</th>
<th>Average Run Length</th>
<th>Median Run Length</th>
<th>N</th>
<th>In-Control SD</th>
<th>One-Sided Prob</th>
<th>Z</th>
<th>LCL</th>
<th>UCL</th>
<th>Out-of-Control SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>22.2</td>
<td>16.0</td>
<td>10</td>
<td>3.8000</td>
<td>0.00135</td>
<td>3.205</td>
<td>0.0000</td>
<td>20.7821</td>
<td>4.6000</td>
</tr>
<tr>
<td>S</td>
<td>21.5</td>
<td>15.0</td>
<td>10</td>
<td>3.8000</td>
<td>0.00135</td>
<td>3.205</td>
<td>0.0000</td>
<td>6.3436</td>
<td>4.6000</td>
</tr>
<tr>
<td>S Prob</td>
<td>33.7</td>
<td>24.0</td>
<td>10</td>
<td>3.8000</td>
<td>0.00135</td>
<td></td>
<td>0.0000</td>
<td>6.5931</td>
<td>4.6000</td>
</tr>
</tbody>
</table>

Notes

The R and S control chart tests show the process is out-of-control a bit sooner than the S test with probability limits.

Example 3 – Validation Using Ryan

Ryan (1989), page 93, indicates that in-control processes with a sample size of 4 at each look will have a false signal on average after 222 samples for the S test, while the S with Probability Limits test will only give a false positive after 741 (1/0.00135) samples.

Setup
This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Control Charts for Means (Simulation) procedure window by clicking on Quality Control, and then clicking on Control Charts for Means (Simulation). You may then make the appropriate entries as listed below, or open Example 3 by going to the File menu and choosing Open Example Template.
Data Tab
Find (Solve For) ...................................... Run Length Distribution
Run Length Summary 1 .......................... ARL (Average Run Length)
Run Length Summary 2 .......................... MRL (Median Run Length)
Percentile or Average......................... Ignored since this is the Find setting
Run Length .............................................. Ignored since this is the Find setting
N (Sample Size) ...................................... 4
Simulations .............................................. 10000
Maximum Run Length ............................ 5000

Distributions, Tests Tab
In-Control Distributions Specified By ...... Standard Deviation Directly
In-Control Standard Deviation .................. 3.8
Out-of-Control Standard Deviation......... 3.8
R .............................................................. Unchecked
S and S with Probability Limits .............. Checked
Two-Sided or Upper Limit Only .............. Upper Limit Only
Specify Probability Using .................... Probability
Probability ................................................ 0.00135

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Chart Type</th>
<th>Average Run Length</th>
<th>Median Run Length</th>
<th>N</th>
<th>In-Control SD</th>
<th>One-Sided Prob</th>
<th>Z</th>
<th>LCL</th>
<th>UCL</th>
<th>Out-of-Control SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>S</td>
<td>221.7</td>
<td>152.0</td>
<td>4</td>
<td>3.8000</td>
<td>0.00135</td>
<td>3.205</td>
<td>0.0000</td>
<td>7.9334</td>
<td>3.8000</td>
</tr>
<tr>
<td>S Prob</td>
<td>737.3</td>
<td>501.0</td>
<td>4</td>
<td>3.8000</td>
<td>0.00135</td>
<td>8.6738</td>
<td>0.0000</td>
<td>8.6738</td>
<td>3.8000</td>
</tr>
</tbody>
</table>

Notes
Simulations: 10000. Run Time: 3.65 minutes.

The average run lengths are very close to those presented in Ryan (1989).
Chapter 400

Inequality Tests for One Mean

Introduction

The one-sample \( t \)-test is used to test whether the mean of a population is greater than, less than, or not equal to a specific value. Because the \( t \) distribution is used to calculate critical values for the test, this test is often called the one-sample \( t \)-test. If the standard deviation is known, the normal distribution is used instead of the \( t \) distribution and the test is officially known as the \( z \) test.

When the data are differences between paired values, this test is known as the paired \( t \)-test.

This module also calculates the power of the nonparametric analog of the \( t \)-test, the Wilcoxon test.

Test Procedure

1. **Find the critical value.** Assume that the true mean is \( M_0 \). Choose a value \( T_a \) so that the probability of rejecting \( H_0 \) when \( H_0 \) is true is equal to a specified value called \( \alpha \). Using the \( t \) distribution, select \( T_a \) so that \( \Pr(t > T_a) = \alpha \). This value is found using a \( t \) probability table or a computer program (like PASS).

2. **Select a sample of \( n \) items from the population and compute the \( t \) statistic.** Call this value \( T \). If \( T > T_a \) reject the null hypothesis that the mean equals \( M_0 \) in favor of an alternative hypothesis that the mean equals \( M_1 \) where \( M_1 > M_0 \).

Following is a specific example. Suppose we want to test the hypothesis that a variable, \( X \), has a mean of 100 versus the alternative hypothesis that the mean is greater than 100. Suppose that previous studies have shown that the standard deviation, \( \sigma \), is 40. A random sample of 100 individuals is used.

We first compute the critical value, \( T_a \). The value of \( T_a \) that yields \( \alpha = 0.05 \) is 106.6. If the mean computed from a sample is greater than 106.6, reject the hypothesis that the mean is 100. Otherwise, do not reject the hypothesis. We call the region greater than 106.6 the Rejection Region and values less than or equal to 106.6 the Acceptance Region of the significance test.
Now suppose that you want to compute the power of this testing procedure. In order to compute the power, we must specify an alternative value for the mean. We decide to compute the power if the true mean were 110. Figure 2 shows how to compute the power in this case.

The power is the probability of rejecting $H_0$ when the true mean is 110. Since we reject $H_0$ when the calculated mean is greater than 106.6, the probability of a Type-II error (called $\beta$) is given by the dark, shaded area of the second graph. This value is 0.196. The power is equal to $1 - \beta$ or 0.804.

Note that there are six parameters that may be varied in this situation: two means, standard deviation, alpha, beta, and the sample size.
Assumptions

This section describes the assumptions that are made when you use one of these tests. The key assumption relates to normality or non-normality of the data. One of the reasons for the popularity of the t-test is its robustness in the face of assumption violation. However, if an assumption is not met even approximately, the significance levels and the power of the t-test are invalidated. Unfortunately, in practice it often happens that several assumptions are not met. This makes matters even worse! Hence, take the steps to check the assumptions before you make important decisions based on these tests.

One-Sample T-Test Assumptions

The assumptions of the one-sample t-test are:

1. The data are continuous (not discrete).
2. The data follow the normal probability distribution.
3. The sample is a simple random sample from its population. Each individual in the population has an equal probability of being selected in the sample.

Paired T-Test Assumptions

The assumptions of the paired t-test are:

1. The data are continuous (not discrete).
2. The data, i.e., the differences for the matched-pairs, follow a normal probability distribution.
3. The sample of pairs is a simple random sample from its population. Each individual in the population has an equal probability of being selected in the sample.

Wilcoxon Signed-Rank Test Assumptions

The assumptions of the Wilcoxon signed-rank test are as follows (note that the difference is between a data value and the hypothesized median or between the two data values of a pair):

1. The differences are continuous (not discrete).
2. The distribution of each difference is symmetric.
3. The differences are mutually independent.
4. The differences all have the same median.
5. The measurement scale is at least interval.

Limitations

There are few limitations when using these tests. Sample sizes may range from a few to several hundred. If your data are discrete with at least five unique values, you can often ignore the continuous variable assumption. Perhaps the greatest restriction is that your data come from a random sample of the population. If you do not have a random sample, your significance levels will probably be incorrect.
Technical Details

Standard Deviation Known
When the standard deviation is known, the power is calculated as follows for a directional alternative (one-tailed test) in which $M_1 > M_0$.

1. Find $z_\alpha$ such that $1 - \Phi(z_\alpha) = \alpha$, where $\Phi(x)$ is the area under the standardized normal curve to the left of $x$.
2. Calculate: $X_a = M_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$.
3. Calculate: $z_a = \frac{X_a - M_1}{\frac{\sigma}{\sqrt{n}}}$.
4. Power = $1 - \Phi(z_a)$.

Standard Deviation Unknown
When the standard deviation is unknown, the power is calculated as follows for a directional alternative (one-tailed test) in which $M_1 > M_0$.

1. Find $t_\alpha$ such that $1 - T_{df}(t_\alpha) = \alpha$, where $T_{df}(t_\alpha)$ is the area under a central-$t$ curve to the left of $x$ and $df = n - 1$.
2. Calculate: $x_a = M_0 + t_\alpha \frac{\sigma}{\sqrt{n}}$.
3. Calculate the noncentrality parameter: $\lambda = \frac{M_1 - M_0}{\frac{\sigma}{\sqrt{n}}}$.
4. Calculate: $t_a = \frac{x_a - M_1}{\frac{\sigma}{\sqrt{n}}} + \lambda$.
5. Calculate: Power = $1 - T'_{df,\lambda}(t_a)$, where $T'_{df,\lambda}(x)$ is the area under a noncentral-$t$ curve with degrees of freedom $df$ and noncentrality parameter $\lambda$ to the left of $x$. 
Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains most of the parameters and options that you will be concerned with.

Solve For

Find (Solve For)

This option specifies the parameter to be calculated from the values of the other parameters. Under most conditions, you would select either Power and Beta or N.

Select N when you want to determine the sample size needed to achieve a given power and alpha error level.

Select Power and Beta when you want to calculate the power of an experiment that has already been run.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.
Sample Size

N (Sample Size)
This option specifies one or more values of the sample size, the number of individuals in the study. This value must be an integer greater than one. Note that you may enter a list of values using the syntax 50,100,150,200,250 or 50 to 250 by 50.

Effect Size – Means

Mean0 (Null or Baseline)
This option specifies one or more values of the mean corresponding to the null hypothesis. If you are analyzing a paired t-test, this value should be zero.

Only the difference between Mean0 and Mean1 is used in the calculations.

Means1 (Alternative)
This option specifies one or more values of the mean corresponding to the alternative hypothesis. If you are analyzing a paired t-test, this value represents the mean difference that you are interested in.

Only the difference between Mean0 and Mean1 is used in the calculations.

Effect Size – Standard Deviation

Standard Deviation
This option specifies one or more values of the standard deviation. This must be a positive value. Be sure to use the standard deviation of X and not the standard deviation of the mean (the standard error).

When this value is not known, you must supply an estimate of it. PASS includes a special module for estimating the standard deviation. This module may be loaded by pressing the SD button. Refer to the Standard Deviation Estimator chapter for further details.

Known Standard Deviation
This option specifies whether the standard deviation (sigma) is known or unknown. In almost all experimental situations, the standard deviation is not known. However, great calculation efficiencies are obtained if the standard deviation is assumed to be known.

When this box is checked, the program performs its calculations assuming that the standard deviation is known. This results in the use of the normal distribution in all probability calculations. Calculations using this option will be much faster than for the unknown standard deviation case. The results for either case will be close when the sample size is over 30.

When this box is not checked, the program assumes that the standard deviation is not known and will be estimated from the data when the t-test is run. This results in probability calculations using the noncentral-t distribution. This distribution requires a lot more calculations than does the normal distribution.

The calculation speed comes into play whenever the Find option is set to something besides Beta. In these cases, the program uses a special searching algorithm which requires numerous iterations. You will note a real difference in calculation speed depending on whether this option is checked.
A reasonable strategy would be to leave this option checked while you are experimenting with the parameters and then turn it off when you are ready for your final results.

**Test**

**Alternative Hypothesis**

This option specifies the alternative hypothesis. This implicitly specifies the direction of the hypothesis test. The null hypothesis is always $H_0: \text{Mean}_0 = \text{Mean}_1$.

Note that the alternative hypothesis enters into power calculations by specifying the rejection region of the hypothesis test. Its accuracy is critical.

Possible selections are:

- **Ha: Mean0 <> Mean1**
  This is the most common selection. It yields the *two-tailed t-test*. Use this option when you are testing whether the means are different but you do not want to specify beforehand which mean is larger. Many scientific journals require two-tailed tests.

- **Ha: Mean0 < Mean1**
  This option yields a *one-tailed t-test*. Use it when you are only interested in the case in which Mean1 is greater than Mean0.

- **Ha: Mean0 > Mean1**
  This options yields a *one-tailed t-test*. Use it when you are only interested in the case in which Mean1 is less than Mean0.

**Nonparametric Adjustment**

This option makes appropriate sample size adjustments for the Wilcoxon test. Results by Al-Sunduqchi and Guenther (1990) indicate that power calculations for the Wilcoxon test may be made using the standard $t$-test formulations with a simple adjustment to the sample size. The size of the adjustment depends upon the actual distribution of the data. They give sample size adjustment factors for four distributions. These are 1 for the uniform distribution, $2/3$ for the double exponential distribution, $9 / \pi^2$ for the logistic distribution, and $\pi / 3$ for the normal distribution.

The options are as follows:

- **Ignore**
  Do not make a Wilcoxon adjustment. This indicates that you want to analyze a $t$-test, not the Wilcoxon test.

- **Uniform**
  Make the Wilcoxon sample size adjustment assuming the uniform distribution. Since the factor is one, this option performs the same as Ignore. It is included for completeness.

- **Double Exponential**
  Make the Wilcoxon sample size adjustment assuming that the data actually follow the double exponential distribution.
Logistic
Make the Wilcoxon sample size adjustment assuming that the data actually follow the logistic distribution.

Normal
Make the Wilcoxon sample size adjustment assuming that the data actually follow the normal distribution.

Population Size
This is the number of subjects in the population. Usually, you assume that samples are drawn from a very large (infinite) population. Occasionally, however, situations arise in which the population of interest is of limited size. In these cases, appropriate adjustments must be made.

When a finite population size is specified, the standard deviation is reduced according to the formula:

$$\sigma_1^2 = \left(1 - \frac{n}{N}\right)\sigma^2$$

where $n$ is the sample size, $N$ is the population size, $\sigma$ is the original standard deviation, and $\sigma_1$ is the new standard deviation.

The quantity $n/N$ is often called the sampling fraction. The quantity $\left(1 - \frac{n}{N}\right)$ is called the finite population correction factor.

Example 1 – Power after a Study

This example will cover the situation in which you are calculating the power of a $t$-test on data that have already been collected and analyzed. For example, you might be playing the role of a reviewer, looking at the power of $t$-test from a study you are reviewing. In this case, you would not vary the means, standard deviation, or sample size since they are given by the experiment. Instead, you investigate the power of the significance tests. You might look at the impact of different alpha values on the power.

Suppose an experiment involving 100 individuals yields the following summary statistics:

<table>
<thead>
<tr>
<th>Hypothesized mean (M0)</th>
<th>100.0</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sample mean (M1)</td>
<td>110.0</td>
</tr>
<tr>
<td>Sample standard deviation</td>
<td>40.0</td>
</tr>
<tr>
<td>Sample size</td>
<td>100</td>
</tr>
</tbody>
</table>

Given the above data, analyze the power of a $t$-test which tests the hypothesis that the population mean is 100 versus the alternative hypothesis that the population mean is 110. Consider the power at significance levels 0.01, 0.05, 0.10 and sample sizes 20 to 120 by 20.

Note that we have set $M1$ equal to the sample mean. In this case, we are studying the power of the $t$-test for a mean difference the size of that found in the experimental data.
Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for One Mean procedure window by expanding Means, then One Mean, then clicking on Test (Inequality), and then clicking on Tests for One Mean. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.01 0.05 0.10</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>20 to 120 by 20</td>
</tr>
<tr>
<td>Mean0 (Null or Baseline)</td>
<td>100</td>
</tr>
<tr>
<td>Mean1 (Alternative)</td>
<td>110</td>
</tr>
<tr>
<td>S (Standard Deviation)</td>
<td>40</td>
</tr>
<tr>
<td>Known Standard Deviation</td>
<td>Not checked</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Ha: Mean0 ≠ Mean1</td>
</tr>
<tr>
<td>Nonparametric Adjustment</td>
<td>Ignore</td>
</tr>
<tr>
<td>Population Size</td>
<td>Infinite</td>
</tr>
</tbody>
</table>

Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>Alpha</th>
<th>Beta</th>
<th>Mean0</th>
<th>Mean1</th>
<th>S</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.06051</td>
<td>20</td>
<td>0.01000</td>
<td>0.93949</td>
<td>100.0</td>
<td>110.0</td>
<td>40.0</td>
<td>0.250</td>
</tr>
<tr>
<td>0.14435</td>
<td>40</td>
<td>0.01000</td>
<td>0.85565</td>
<td>100.0</td>
<td>110.0</td>
<td>40.0</td>
<td>0.250</td>
</tr>
<tr>
<td>0.24401</td>
<td>60</td>
<td>0.01000</td>
<td>0.75599</td>
<td>100.0</td>
<td>110.0</td>
<td>40.0</td>
<td>0.250</td>
</tr>
<tr>
<td>0.34953</td>
<td>80</td>
<td>0.01000</td>
<td>0.65047</td>
<td>100.0</td>
<td>110.0</td>
<td>40.0</td>
<td>0.250</td>
</tr>
<tr>
<td>0.45316</td>
<td>100</td>
<td>0.01000</td>
<td>0.54684</td>
<td>100.0</td>
<td>110.0</td>
<td>40.0</td>
<td>0.250</td>
</tr>
<tr>
<td>0.54958</td>
<td>120</td>
<td>0.01000</td>
<td>0.45042</td>
<td>100.0</td>
<td>110.0</td>
<td>40.0</td>
<td>0.250</td>
</tr>
<tr>
<td>0.18590</td>
<td>20</td>
<td>0.05000</td>
<td>0.81410</td>
<td>100.0</td>
<td>110.0</td>
<td>40.0</td>
<td>0.250</td>
</tr>
<tr>
<td>0.33831</td>
<td>40</td>
<td>0.05000</td>
<td>0.66169</td>
<td>100.0</td>
<td>110.0</td>
<td>40.0</td>
<td>0.250</td>
</tr>
<tr>
<td>0.47811</td>
<td>60</td>
<td>0.05000</td>
<td>0.52189</td>
<td>100.0</td>
<td>110.0</td>
<td>40.0</td>
<td>0.250</td>
</tr>
<tr>
<td>0.59828</td>
<td>80</td>
<td>0.05000</td>
<td>0.40172</td>
<td>100.0</td>
<td>110.0</td>
<td>40.0</td>
<td>0.250</td>
</tr>
<tr>
<td>0.68698</td>
<td>100</td>
<td>0.05000</td>
<td>0.30302</td>
<td>100.0</td>
<td>110.0</td>
<td>40.0</td>
<td>0.250</td>
</tr>
<tr>
<td>0.77532</td>
<td>120</td>
<td>0.05000</td>
<td>0.22468</td>
<td>100.0</td>
<td>110.0</td>
<td>40.0</td>
<td>0.250</td>
</tr>
<tr>
<td>0.28873</td>
<td>20</td>
<td>0.10000</td>
<td>0.71127</td>
<td>100.0</td>
<td>110.0</td>
<td>40.0</td>
<td>0.250</td>
</tr>
<tr>
<td>0.46435</td>
<td>40</td>
<td>0.10000</td>
<td>0.53565</td>
<td>100.0</td>
<td>110.0</td>
<td>40.0</td>
<td>0.250</td>
</tr>
<tr>
<td>0.60636</td>
<td>60</td>
<td>0.10000</td>
<td>0.39364</td>
<td>100.0</td>
<td>110.0</td>
<td>40.0</td>
<td>0.250</td>
</tr>
<tr>
<td>0.71639</td>
<td>80</td>
<td>0.10000</td>
<td>0.28361</td>
<td>100.0</td>
<td>110.0</td>
<td>40.0</td>
<td>0.250</td>
</tr>
<tr>
<td>0.79900</td>
<td>100</td>
<td>0.10000</td>
<td>0.20100</td>
<td>100.0</td>
<td>110.0</td>
<td>40.0</td>
<td>0.250</td>
</tr>
<tr>
<td>0.85952</td>
<td>120</td>
<td>0.10000</td>
<td>0.14048</td>
<td>100.0</td>
<td>110.0</td>
<td>40.0</td>
<td>0.250</td>
</tr>
</tbody>
</table>
Report Definitions
Power is the probability of rejecting a false null hypothesis. It should be close to one.
N is the size of the sample drawn from the population. To conserve resources, it should be small.
Alpha is the probability of rejecting a true null hypothesis. It should be small.
Beta is the probability of accepting a false null hypothesis. It should be small.
Mean0 is the value of the population mean under the null hypothesis. It is arbitrary.
Mean1 is the value of the population mean under the alternative hypothesis. It is relative to Mean0.
Sigma is the standard deviation of the population. It measures the variability in the population.
Effect Size, |Mean0-Mean1|/Sigma, is the relative magnitude of the effect under the alternative.

Summary Statements
A sample size of 20 achieves 6% power to detect a difference of -10.0 between the null hypothesis mean of 100.0 and the alternative hypothesis mean of 110.0 with an estimated standard deviation of 40.0 and with a significance level (alpha) of 0.01000 using a two-sided one-sample t-test.

This report shows the values of each of the parameters, one scenario per row. The values of power and beta were calculated from the other parameters.
The definitions of each column are given in the Report Definitions section.

Plots Section

This plot shows the relationship between sample size and power for various values of alpha.
Example 2 – Finding the Sample Size

This example will consider the situation in which you are planning a study that will use the one-sample t-test and want to determine an appropriate sample size. This example is more subjective than the first because you now have to obtain estimates of all the parameters. In the first example, these estimates were provided by the data.

In studying deaths from SIDS (Sudden Infant Death Syndrome), one hypothesis put forward is that infants dying of SIDS weigh less than normal at birth. Suppose the average birth weight of infants is 3300 grams with a standard deviation of 663 grams. Use an alpha of 0.05 and power of both 0.80 and 0.90. How large a sample of SIDS infants will be needed to detect a drop in average weight of 25%? Of 10%? Of 5%? Note that applying these percentages to the average weight of 3300 yields 2475, 2970, and 3135.

Although a one-sided hypothesis is being considered, sample size estimates will assume a two-sided alternative to keep the research design in line with other studies.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for One Mean procedure window by expanding Means, then One Mean, then clicking on Test (Inequality), and then clicking on Tests for One Mean. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N</td>
</tr>
<tr>
<td>Power</td>
<td>0.80 0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Mean0 (Null or Baseline)</td>
<td>3300</td>
</tr>
<tr>
<td>Mean1 (Alternative)</td>
<td>2475 2970 3135</td>
</tr>
<tr>
<td>S (Standard Deviation)</td>
<td>663</td>
</tr>
<tr>
<td>Known Standard Deviation</td>
<td>Not checked</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Ha: Mean0 ≠ Mean1</td>
</tr>
<tr>
<td>Nonparametric Adjustment</td>
<td>Ignore</td>
</tr>
<tr>
<td>Population Size</td>
<td>Infinite</td>
</tr>
</tbody>
</table>
**Output**

Click the Run button to perform the calculations and generate the following output.

**Numeric Results**

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>Alpha</th>
<th>Beta</th>
<th>Mean0</th>
<th>Mean1</th>
<th>S</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.90307</td>
<td>9</td>
<td>0.05000</td>
<td>0.09693</td>
<td>3300.0</td>
<td>2475.0</td>
<td>663.0</td>
<td>1.244</td>
</tr>
<tr>
<td>0.85339</td>
<td>8</td>
<td>0.05000</td>
<td>0.14661</td>
<td>3300.0</td>
<td>2475.0</td>
<td>663.0</td>
<td>1.244</td>
</tr>
<tr>
<td>0.90409</td>
<td>45</td>
<td>0.05000</td>
<td>0.09591</td>
<td>3300.0</td>
<td>2970.0</td>
<td>663.0</td>
<td>0.498</td>
</tr>
<tr>
<td>0.80426</td>
<td>34</td>
<td>0.05000</td>
<td>0.19574</td>
<td>3300.0</td>
<td>2970.0</td>
<td>663.0</td>
<td>0.498</td>
</tr>
<tr>
<td>0.90070</td>
<td>172</td>
<td>0.05000</td>
<td>0.09930</td>
<td>3300.0</td>
<td>3135.0</td>
<td>663.0</td>
<td>0.249</td>
</tr>
<tr>
<td>0.80105</td>
<td>129</td>
<td>0.05000</td>
<td>0.19895</td>
<td>3300.0</td>
<td>3135.0</td>
<td>663.0</td>
<td>0.249</td>
</tr>
</tbody>
</table>

This report shows the values of each of the parameters, one scenario per row. Since there were three values of Mean1 and two values of beta, there are a total of six rows in the report.

We were solving for the sample size, $N$. Notice that the increase in sample size seems to be most directly related to the difference between the two means. The difference in beta values does not seem to be as influential, especially at the smaller sample sizes.

Note that even though we set the beta values at 0.1 and 0.2, these are not the beta values that were achieved. This happens because $N$ can only take on integer values. The program selects the first value of $N$ that gives at least the values of alpha and beta that were desired.

**Example 3 – Finding the Minimum Detectable Difference**

This example will consider the situation in which you want to determine how small of a difference between the two means can be detected by the $t$-test with specified values of the other parameters.

Continuing with the previous example, suppose about 50 SIDS deaths occur in a particular area per year. Using 50 as the sample size, 0.05 as alpha, and 0.20 as beta, how large of a difference between the means is detectable?

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for One Mean** procedure window by expanding **Means**, then **One Mean**, then clicking on **Test (Inequality)**, and then clicking on **Tests for One Mean**. You may then make the appropriate entries as listed below, or open Example 3 by going to the **File** menu and choosing **Open Example Template**.
Option | Value
---|---
Data Tab | 
Find (Solve For) | Mean1 (Search<Mean0)
Power | 0.80
Alpha | 0.05
N (Sample Size) | 50
Mean0 (Null or Baseline) | 3300
Mean1 (Alternative) | Ignored since this is the Find setting
S (Standard Deviation) | 663
Known Standard Deviation | Not checked
Alternative Hypothesis | Ha: Mean0 ≠ Mean1
Nonparametric Adjustment | Ignore
Population Size | Infinite

Output

Click the Run button to perform the calculations and generate the following output.

**Numeric Results**

<table>
<thead>
<tr>
<th>Numeric Results for One-Sample T-Test</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Null Hypothesis: Mean0=Mean1</td>
<td>Alternative Hypothesis: Mean0&lt;&gt;Mean1</td>
</tr>
<tr>
<td>Unknown standard deviation.</td>
<td></td>
</tr>
<tr>
<td>Power</td>
<td>N</td>
</tr>
<tr>
<td>0.80000</td>
<td>50</td>
</tr>
</tbody>
</table>

With a sample of 50, a difference of 3300 - 3032 = 268 would be detectable. This difference represents about an 8% decrease in weight.

**Example 4 – Paired T-Test**

Usually, a researcher designs a study to compare two or more groups of subjects, so the one sample case described in this chapter occurs infrequently. However, there is a popular research design that does lead to the single mean test: **paired observations**.

For example, suppose researchers want to study the impact of an exercise program on the individual’s weight. To do so they randomly select $N$ individuals, weigh them, put them through the exercise program, and weigh them again. The variable of interest is not their actual weight, but how much their weight changed.

In this design, the data are analyzed using a one-sample $t$-test on the differences between the paired observations. The null hypothesis is that the average difference is zero. The alternative hypothesis is that the average difference is some nonzero value.

To study the impact of an exercise program on weight loss, the researchers decide to conduct a study that will be analyzed using the paired $t$-test. A sample of individuals will be weighed before and after a specified exercise program that will last three months. The difference in their weights will be analyzed.
Past experiments of this type have had standard deviations in the range of 10 to 15 pounds. The researcher wants to detect a difference of 5 pounds or more. Alpha values of 0.01 and 0.05 will be tried. Beta is set to 0.20 so that the power is 80%. How large of a sample must the researchers take?

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for One Mean procedure window by expanding Means, then One Mean, then clicking on Test (Inequality), and then clicking on Tests for One Mean. You may then make the appropriate entries as listed below, or open Example 4 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N</td>
</tr>
<tr>
<td>Power</td>
<td>0.80</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.01 0.05</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>Ignored since this is the Find setting.</td>
</tr>
<tr>
<td>Mean0 (Null or Baseline)</td>
<td>0</td>
</tr>
<tr>
<td>Mean1 (Alternative)</td>
<td>-5</td>
</tr>
<tr>
<td>S (Standard Deviation)</td>
<td>10 12.5 15</td>
</tr>
<tr>
<td>Known Standard Deviation</td>
<td>Not checked</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Ha: Mean0 ≠ Mean1</td>
</tr>
<tr>
<td>Nonparametric Adjustment</td>
<td>Ignore</td>
</tr>
<tr>
<td>Population Size</td>
<td>Infinite</td>
</tr>
</tbody>
</table>

**Output**

Click the Run button to perform the calculations and generate the following output.

**Numeric Results and Plots**

<table>
<thead>
<tr>
<th>Numeric Results for One-Sample T-Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null Hypothesis: Mean0=Mean1</td>
</tr>
<tr>
<td>Alternative Hypothesis: Mean0&lt;&gt;Mean1</td>
</tr>
<tr>
<td>Unknown standard deviation.</td>
</tr>
<tr>
<td>Power</td>
</tr>
<tr>
<td>0.80939</td>
</tr>
<tr>
<td>0.80778</td>
</tr>
<tr>
<td>0.80434</td>
</tr>
<tr>
<td>0.80779</td>
</tr>
<tr>
<td>0.80252</td>
</tr>
<tr>
<td>0.80230</td>
</tr>
</tbody>
</table>
The report shows the values of each of the parameters, one scenario per row. We were solving for the sample size, \( N \).

Note that depending on our choice of assumptions, the sample size ranges from 34 to 109. Hence, the researchers have to make a careful determination of which standard deviation and significance level should be used.

**Example 5 – Wilcoxon Test**

The Wilcoxon test, a nonparametric analog of the paired comparison \( t \)-test, is recommended when the distribution of the data is symmetrical, but not normal. A study by Al-Sunduqchi (1990) showed that sample size and power calculations for the Wilcoxon test can be made using the standard \( t \)-test results with a simple adjustment to the sample size.

Suppose the researchers in Example 4 want to compare sample size requirements of the \( t \)-test with those of the Wilcoxon test. They would use the same values, only this time the Nonparametric Adjustment would be set to *double exponential*. The double exponential was selected because it requires the largest adjustment of the distributions available in *PASS* and they wanted to know what the largest adjustment was.

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the *Tests for One Mean* procedure window by expanding *Means*, then *One Mean*, then clicking on *Test (Inequality)*, and then clicking on *Tests for One Mean*. You may then make the appropriate entries as listed below, or open Example 5 by going to the *File* menu and choosing *Open Example Template*. 
Option | Value
---|---
Data Tab | 
Find (Solve For) | N
Power | 0.80
Alpha | 0.01 0.05
N (Sample Size) | Ignored since this is the Find setting.
Mean0 (Null or Baseline) | 0
Mean1 (Alternative) | -5
S (Standard Deviation) | 10 12.5 15
Known Standard Deviation | Not checked
Alternative Hypothesis | Ha: Mean0 ≠ Mean1
Nonparametric Adjustment | Double Exponential
Population Size | Infinite

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results and Plots

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>Alpha</th>
<th>Beta</th>
<th>Mean0</th>
<th>Mean1</th>
<th>S</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.80939</td>
<td>34</td>
<td>0.01000</td>
<td>0.19061</td>
<td>0.0</td>
<td>-5.0</td>
<td>10.0</td>
<td>0.500</td>
</tr>
<tr>
<td>0.80778</td>
<td>23</td>
<td>0.05000</td>
<td>0.19222</td>
<td>0.0</td>
<td>-5.0</td>
<td>10.0</td>
<td>0.500</td>
</tr>
<tr>
<td>0.81069</td>
<td>52</td>
<td>0.01000</td>
<td>0.18931</td>
<td>0.0</td>
<td>-5.0</td>
<td>12.5</td>
<td>0.400</td>
</tr>
<tr>
<td>0.80779</td>
<td>35</td>
<td>0.05000</td>
<td>0.19221</td>
<td>0.0</td>
<td>-5.0</td>
<td>12.5</td>
<td>0.400</td>
</tr>
<tr>
<td>0.80252</td>
<td>73</td>
<td>0.01000</td>
<td>0.19748</td>
<td>0.0</td>
<td>-5.0</td>
<td>15.0</td>
<td>0.333</td>
</tr>
<tr>
<td>0.80230</td>
<td>49</td>
<td>0.05000</td>
<td>0.19770</td>
<td>0.0</td>
<td>-5.0</td>
<td>15.0</td>
<td>0.333</td>
</tr>
</tbody>
</table>

If you compare these sample size values with those of Example 4, you will find that these are about two-thirds of those required for the *t*-test. This is the value of the adjustment factor for the Wilcoxon test when the underlying distribution is the double exponential.
Example 6 – Validation using Zar

Zar (1984) pages 111-112 presents an example in which Mean0 = 0.0, Mean1 = 1.0, S = 1.25, alpha = 0.05, and N = 12. Zar obtains an approximate power of 0.72.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for One Mean procedure window by expanding Means, then One Mean, then clicking on Test (Inequality), and then clicking on Tests for One Mean. You may then make the appropriate entries as listed below, or open Example 6 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting.</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>12</td>
</tr>
<tr>
<td>Mean0 (Null or Baseline)</td>
<td>0</td>
</tr>
<tr>
<td>Mean1 (Alternative)</td>
<td>1</td>
</tr>
<tr>
<td>S (Standard Deviation)</td>
<td>1.25</td>
</tr>
<tr>
<td>Known Standard Deviation</td>
<td>Not checked</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Ha: Mean0 ≠ Mean1</td>
</tr>
<tr>
<td>Nonparametric Adjustment</td>
<td>Ignore</td>
</tr>
<tr>
<td>Population Size</td>
<td>Infinite</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>Alpha</th>
<th>Beta</th>
<th>Mean0</th>
<th>Mean1</th>
<th>S</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.71366</td>
<td>12</td>
<td>0.0500</td>
<td>0.28634</td>
<td>0.0</td>
<td>1.0</td>
<td>1.3</td>
<td>0.800</td>
</tr>
</tbody>
</table>

The difference between the power computed by PASS of 0.71366 and the 0.72 computed by Zar is mostly due to Zar’s use of an approximation to the noncentral t distribution.
Example 7 – Validation using Machin

Machin, Campbell, Fayers, and Pinol (1997) page 37 presents an example in which Mean0 = 0.0, Mean1 = 0.2, S = 1.0, alpha = 0.05, and beta = 0.20. They obtain a sample size of 199.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for One Mean procedure window by expanding Means, then One Mean, then clicking on Test (Inequality), and then clicking on Tests for One Mean. You may then make the appropriate entries as listed below, or open Example 7 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N</td>
</tr>
<tr>
<td>Power</td>
<td>0.80</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Mean0 (Null or Baseline)</td>
<td>0</td>
</tr>
<tr>
<td>Mean1 (Alternative)</td>
<td>0.2</td>
</tr>
<tr>
<td>S (Standard Deviation)</td>
<td>1.0</td>
</tr>
<tr>
<td>Known Standard Deviation</td>
<td>Not checked</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Ha: Mean0 ≠ Mean1</td>
</tr>
<tr>
<td>Nonparametric Adjustment</td>
<td>Ignore</td>
</tr>
<tr>
<td>Population Size</td>
<td>Infinite</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>Alpha</th>
<th>Beta</th>
<th>Mean0</th>
<th>Mean1</th>
<th>S</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.80169</td>
<td>199</td>
<td>0.05000</td>
<td>0.19831</td>
<td>0.0</td>
<td>0.2</td>
<td>1.0</td>
<td>0.200</td>
</tr>
</tbody>
</table>

The sample size of 199 matches Machin’s result.
Chapter 405

Inequality Tests for One Exponential Mean

Introduction
This program module designs studies for testing hypotheses about the mean of the exponential distribution. Such tests are often used in reliability acceptance testing, also called reliability demonstration testing.

Results are calculated for plans that are time censored or failure censored, as well as for plans that use with replacement or without replacement sampling. We adopt the basic methodology outlined in Epstein (1960), Juran (1979), Bain and Engelhardt (1991), and Schilling (1982).

Technical Details
The test procedures described here make the assumption that lifetimes follow the exponential distribution. The density of the exponential distribution is written as

$$f(t) = \frac{1}{\theta} \exp\left(-\frac{t}{\theta}\right)$$

The parameter $\theta$ is interpreted as average failure time, mean time to failure (MTTF), or mean time between failures (MTBF). Its reciprocal is the failure rate.

The reliability, or probability that a unit continues running beyond time $t$, is

$$R(t) = e^{-\frac{t}{\theta}}$$

Hypothesis Test
The relevant statistical hypothesis is $H_0: \theta_0 = \theta_1$ versus the one-sided alternative $H_1: \theta_0 > \theta_1$. Here, $\theta_0$ represents an acceptable (high) mean life usually set from the point of view of the producer and $\theta_1$ represents some unacceptable (low) mean life usually set from the point of view of the
Inequality Tests for One Exponential Mean

consumer. The test procedure is to reject the null hypothesis if the observed mean life \( \hat{\theta} \) is larger than a critical value selected to meet the error rate criterion.

The error rates are often interpreted in reliability testing as risks. The consumer runs the risk that the study will fail to reject products that have a reliability less than they have specified. This consumer risk is \( \beta \). Similarly, the producer runs the risk that the study will reject products that actually meet the consumer’s requirements. This producer risk is \( \alpha \).

Fixed-Failure Sampling Plans

Fixed failure plans are those in which a specified number of items, \( n \), are observed until a specified number of items, \( r_0 \), fail. The length of the study \( t_0 \) is random. Failed items may, or may not, be immediately replaced (with replacement versus without replacement).

The test statistic is the observed mean life \( \hat{\theta} \) which is computed using

\[
\hat{\theta} = \frac{\sum t_i}{r_0}
\]

where \( t_i \) is the elapsed time that the \( i \)th item is tested, whether measured until failure or until the study is completed.

For both with-replacement and without-replacement sampling, \( \hat{\theta} \) follows the two-parameter gamma distribution with density

\[
g(y|r_0, \theta) = \frac{1}{(r_0 - 1)!} \left( \frac{r_0}{\theta} \right)^{r_0-1} y^{r_0-1} e^{-r_0 y/\theta}
\]

This may be converted to a standard, one-parameter gamma using the transformation

\[x = r_0 y / \theta\]

However, because chi-square tables were more accessible, and because the gamma distribution may be transformed to the chi-square distribution, most results in the statistical literature are based on the chi-square distribution. That is, \( 2r_0 \hat{\theta} / \theta \) is distributed as a chi-square random variable with \( 2r_0 \) degrees of freedom.

Assuming that the testing of all \( n \) items begins at the same instant, the expected length of time needed to observe the first \( r_0 \) failures is

\[
E(t_0) = \begin{cases} 
\theta \sum_{i=1}^{r_0} \frac{1}{n-i+1} & \text{without replacement} \\
\frac{\theta r_0}{n} & \text{with replacement}
\end{cases}
\]

If you choose to solve the without replacement equation for \( n \), you can make use of the approximation

\[
\sum_{i=1}^{r} \frac{1}{n-i+1} \approx \log_e \left( \frac{n + 0.5}{n-r + 0.5} \right)
\]
Using the above results, sampling plans that meet the specified producer and consumer risk values may be found using the result (see Epstein (1960) page 437) that \( r_0 \) is the smallest integer such that

\[
\frac{\chi^2_{\alpha,2r_0}}{\chi^2_{1-\beta,2r_0}} \geq \frac{\theta_1}{\theta_0}
\]

for testing \( H_1: \theta_0 > \theta \)

and

\[
\frac{\chi^2_{\alpha,2r_0}}{\chi^2_{1-\beta,2r_0}} \geq \frac{\theta_0}{\theta_1}
\]

for testing \( H_1: \theta_0 < \theta_1 \)

Note that the above formulation depends on \( r_0 \) but not \( n \). An appropriate value of \( n \) can be found by considering \( E(t_0) \). Two options are available.

1. The value of \( n \) is set (perhaps on economic grounds) and the value of \( E(t_0) \) is calculated.
2. The value of \( E(t_0) \) is set and the value of \( n \) is calculated.

**Fixed-Time Sampling Plans**

*Fixed Time* plans refer to those in which a specified number of items \( n \) are observed for a fixed length of time \( t_0 \). The number of items failing \( r \) is recorded. Sampling can be with or without replacement. The accept/reject decision can be based on \( r \) or the observed mean life \( \hat{\theta} \) which is computed using

\[
\hat{\theta} = \frac{\sum t_i}{n}
\]

where \( t_i \) is the time that the \( i \)th item is being tested, whether measured until failure or until the study is completed.

**With Replacement Sampling**

If failed items are immediately replaced with additional items, the distribution of \( r \) (and \( \hat{\theta} \), since \( \hat{\theta} = nt_0 / r \)) follows the Poisson distribution. The probability distribution of \( r \) is given by the Poisson probability formula

\[
P(r \leq r_0 | \theta) = \sum_{i=0}^{r_0} \frac{(nt_0 / \theta)^i}{i!} e^{-nt_0 / \theta}
\]

Thus, values of \( n \) and \( t_0 \) can be found which meet the \( \alpha \) and \( \beta \) requirements.

**Without Replacement Sampling**

If failed items are not replaced, the distributions of \( r \) and \( \hat{\theta} \) are different and thus the power and sample size calculations depend on which statistic will be used. The probability distribution of \( r \) is given by the binomial formula
405-4 Inequality Tests for One Exponential Mean

\[
P(r \leq r_0 | r, \theta) = \sum_{i=0}^{c} \binom{n}{i} p^i (1-p)^{n-i}
\]

where

\[ p = 1 - e^{-t_0/\theta} \]

Thus, values of \( n \) and \( t_0 \) can be found which meet the \( \alpha \) and \( \beta \) requirements. Note that this formulation ignores the actual failure times.

If \( \hat{\theta} \) will be used as the test statistic, power calculations must be based on it. Bartholomew (1963) gave the following results for the case \( r > 0 \).

\[
\Pr(\hat{\theta} \geq \theta_c) = \frac{1}{1 - e^{-m_0/\theta}} \sum_{i=1}^{n} \binom{n}{i} \sum_{k=0}^{i} \binom{k}{i} (-1)^{i} \exp\left\{- \frac{t_0}{\theta} (n-k+i)\right\} \int_{W} g(x) \, dx
\]

where \( g(x) \) is the chi-square density function with \( 2k \) degrees of freedom and

\[
W = \frac{2k}{\theta} \left( \frac{\theta - t_0}{k} (n-k+i) \right)
\]

\[
\left\langle X \right\rangle = \begin{cases} \text{if } X > 0 \\ \text{0 otherwise} \end{cases}
\]

The above equation is numerically unstable for large values of \( N \), so we use the following approximation also given by Bartholomew (1961). This approximation is used when \( N > 30 \) or when the exact equation cannot be calculated. Bain and Engelhardt (1991) page 140 suggest that this normal approximation can be used when \( p > 0.5 \)

\[
z = \frac{u \sqrt{np}}{\sqrt{1 - 2u(1-p) \log_e(1-p) + (1-p)u^2}}
\]

where

\[
u = \frac{\hat{\theta} - \theta}{\theta}
\]

\[ p = 1 - e^{-t_0/\theta} \]
Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains most of the parameters and options that you will be concerned with.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for from the other parameters. Under most situations, you will select either Power and Beta for a power analysis or N for sample size determination.

Select N when you want to calculate the sample size needed to achieve a given power and alpha level. Select Power and Beta when you want to calculate the power of an experiment or test.

Error Rates

Power or Beta (Beta is Consumer’s Risk)

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta (consumer’s risk) is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of equal probabilities of the event of interest when in fact they are different.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

Alpha (Producer’s Risk)

This option specifies one or more values for the probability of a type-I error (alpha), also called the producer’s risk. A type-I error occurs when you reject the null hypothesis of equal probabilities when in fact they are equal.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.
Sample Size

N (Sample Size)
Enter one or more values for the sample size N, the number of items in the study. Note that the sample size is arbitrary for sampling plans that are terminated after a fixed number of failures are observed.

You may enter a range such as 10 to 100 by 10 or a list of values separated by commas or blanks.

Test

Alternative Hypothesis
Specify the alternative hypothesis of the test. Since the null hypothesis is equality (a difference between theta0 and theta1 of zero), the alternative is all that needs to be specified. Usually, a one-tailed option is selected for these designs. In fact, the two-tailed options are only available for time terminated experiments.

Effect Size

Theta0 (Baseline Mean Life)
Enter one or more values for the mean life under the null hypothesis. This is sometimes called the producer’s mean life. This value is usually scaled in terms of elapsed time such as hours, days, or years. Of course, all time values must be on the same time scale.

Note that the value of theta may be calculated from the estimated probability of failure using the relationship

\[ P(Failure) = 1 - e^{-\frac{t}{\theta}} \]

so that

\[ \theta = -\frac{t_0}{\ln(1 - P(Failure))} \]

Only positive values are valid. You may enter a range of values such as ‘10 20 30’ or ‘100 to 1000 by 100.’

Because the exponential function is used in the calculations, try to scale the numbers so they are less than 100. For example, instead of 720 days, use 7.2 hundreds of days. This will help to avoid numerical problems during the calculations.

Theta1 (Alternative Mean Life)
Enter one or more values for the mean life under the alternative hypothesis. This is sometimes called the consumer’s mean life. This value is usually scaled in terms of elapsed time such as hours, days, or years. Of course, all time values must be on the same time scale.

Note that the value of theta may be calculated from the estimated probability of failure using the relationship

\[ P(Failure) = 1 - e^{-\frac{t}{\theta}} \]
so that

\[ \theta = -\frac{t_0}{\ln(1 - P(\text{Failure}))} \]

Any positive values are valid. You may enter a range of values such as ‘10 20 30’ or ‘100 to 1000 by 100.’

Because the exponential function is used in the calculations, try to scale the numbers so they are less than 100. For example, instead of 720 days, use 7.2 hundreds of days. This will help to avoid numerical problems during the calculations.

**Sampling Plan**

**Replacement Method**
When failures occur, they may be immediately replaced (With Replacement) with new items or not (Without Replacement). One of the assumptions of the exponential distribution is that the probability of failure does not depend on the previous running time. That is, it is assumed that there is no wear-out. Adopting ‘with replacement’ sampling will shorten the elapsed time of an experiment that is failure terminated.

**Termination Criterion**
This option specifies the method used to terminate the study or experiment. There are two basic choices:

- **Fixed failures (r)**
  Terminate after \( r \) failures occur. This is also called failure terminated or Type-II Censoring.

- **Fixed time (t0)**
  Terminate after an elapsed time of \( t_0 \). This is also called time terminated or Type-I Censoring. This is the most common.

In fixed failure sampling, \( N \) may be fixed while \( t_0 \) varies or \( t_0 \) may be fixed while \( N \) varies. All that matters is the product of these two quantities.

In fixed time sampling, two test statistics are available: \( r \) and theta-hat. When sampling is without replacement, tests based on theta-hat are more powerful (require smaller sample size).

**r (Number of Failures)**
Enter one or more values for the rejection number of the test. If \( r \) or more items fail, the null hypothesis that Theta0 = Theta1 is rejected in favor of the alternative the Theta0 > Theta1.

Note that this value is ignored for time terminated experiments, because the appropriate value is calculated. This value is also ignored in some situations in failure terminated experiments.

**t0 (Test Duration Time)**
Enter one or more values for the duration of the test. This value may be interpreted as the exact duration time, \( t_0 \), or the expected duration time, \( E(t0) \), depending on the Termination Criterion and Replacement Method selected.

These values must be positive and in the same time units as Theta0 and Theta1.
E(t0) based on Theta1

When the experiment is failure terminated, the expected waiting time until r failures are observed, E(t0), is calculated. This value depends on the value of theta, the mean life. When checked, E(t0) calculations are based on Theta1. When unchecked, E(t0) calculations are based on Theta0. Either choice may be reasonable in a given situation.

Example 1 – Power for Several Sample Sizes

This example will calculate power for a time terminated, without replacement study in which the results will be analyzed using theta-hat. The study will be used to test the alternative hypothesis that Theta0 > Theta1, where Theta0 = 2.0 days and Theta1 = 1.0 days. The test duration is 1.0 days. Funding for the study will allow for a sample size of up to 40 test items. The researchers decide to look at sample sizes of 10, 20, 30, and 40. Significance levels of 0.01 and 0.05 will be considered.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for One Exponential Mean procedure window by expanding Means, then One Mean, then clicking on Test (Inequality), and then clicking on Tests for One Exponential Mean. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.01 0.05</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>10 to 40 by 10</td>
</tr>
<tr>
<td>Theta0 (Baseline Mean Life)</td>
<td>2</td>
</tr>
<tr>
<td>Theta1 (Alternative Mean Life)</td>
<td>1</td>
</tr>
<tr>
<td>Replacement Method</td>
<td>Without Replacement</td>
</tr>
<tr>
<td>Termination Criterion</td>
<td>Fixed Time using Theta-hat</td>
</tr>
<tr>
<td>t0 (Test Duration Time)</td>
<td>1</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Ha: Theta0 &gt; Theta1</td>
</tr>
</tbody>
</table>
Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>Time</th>
<th>Theta0</th>
<th>Theta1</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>Target Beta</th>
<th>Actual Beta</th>
<th>Theta</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.21695</td>
<td>10</td>
<td>1.000</td>
<td>2.0</td>
<td>1.0</td>
<td>0.01000</td>
<td>0.01000</td>
<td>0.78305</td>
<td>0.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.45485</td>
<td>20</td>
<td>1.000</td>
<td>2.0</td>
<td>1.0</td>
<td>0.01000</td>
<td>0.01000</td>
<td>0.54515</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.67159</td>
<td>30</td>
<td>1.000</td>
<td>2.0</td>
<td>1.0</td>
<td>0.01000</td>
<td>0.01000</td>
<td>0.32841</td>
<td>1.1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.80628</td>
<td>40</td>
<td>1.000</td>
<td>2.0</td>
<td>1.0</td>
<td>0.01000</td>
<td>0.01000</td>
<td>0.19372</td>
<td>1.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.46940</td>
<td>10</td>
<td>1.000</td>
<td>2.0</td>
<td>1.0</td>
<td>0.05000</td>
<td>0.05000</td>
<td>0.53060</td>
<td>1.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.71828</td>
<td>20</td>
<td>1.000</td>
<td>2.0</td>
<td>1.0</td>
<td>0.05000</td>
<td>0.05000</td>
<td>0.28172</td>
<td>1.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.86665</td>
<td>30</td>
<td>1.000</td>
<td>2.0</td>
<td>1.0</td>
<td>0.05000</td>
<td>0.05000</td>
<td>0.13335</td>
<td>1.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.93730</td>
<td>40</td>
<td>1.000</td>
<td>2.0</td>
<td>1.0</td>
<td>0.05000</td>
<td>0.05000</td>
<td>0.06270</td>
<td>1.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Report Definitions

Power is the probability of rejecting a false null hypothesis.
N is the size of the sample drawn from the population.
Alpha is the probability of rejecting a true null hypothesis.
Beta is the probability of accepting a false null hypothesis.
Theta0 is the Mean Life under the null hypothesis.
Theta1 is the Mean Life under the alternative hypothesis.
t0 is the test duration time. It provides the scale for Theta0 and Theta1.
r is the number of failures.

Summary Statements

A sample size of 10 achieves 22% power to detect the difference between the null hypothesis mean lifetime of 2.0 and the alternative hypothesis mean lifetime of 1.0 at a 0.01000 significance level (alpha) using a one-sided test based on the elapsed time. Failing items are not replaced with new items. The study is terminated when it has run for 1.000 time units.

This report shows the power for each of the scenarios. The critical value, Theta C, is also provided.
Example 2 – Validation using Epstein

Epstein (1960), page 438, presents a table giving values of $r$ necessary to meet risk criteria for various values of alpha, beta, theta0, and theta1 for the fixed failures case. Specifically, when theta0 = 5, theta1 = 2, beta = 0.05, and alpha = 0.01, 0.05, and 0.10, he finds $r = 21, 14, and 11$. We will now duplicate these results.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for One Exponential Mean procedure window by expanding Means, then One Mean, then clicking on Test (Inequality), and then clicking on Tests for One Exponential Mean. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>$r$</td>
</tr>
<tr>
<td>Power</td>
<td>0.95</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.01 0.05 0.10</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Theta0 (Baseline Mean Life)</td>
<td>5</td>
</tr>
<tr>
<td>Theta1 (Alternative Mean Life)</td>
<td>2</td>
</tr>
<tr>
<td>Replacement Method</td>
<td>Without Replacement</td>
</tr>
<tr>
<td>Termination Criterion</td>
<td>Fixed Failures, Fixed E(t0)</td>
</tr>
<tr>
<td>$t_0$ (Test Duration Time)</td>
<td>1</td>
</tr>
<tr>
<td>$E(t_0)$ based on Theta1</td>
<td>Unchecked</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>$Ha: \text{Theta0} &gt; \text{Theta1}$</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

**Numeric Results**

<table>
<thead>
<tr>
<th>Power</th>
<th>$r_0 / N$</th>
<th>$E(t_0)$</th>
<th>Theta0</th>
<th>Theta1</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>Target Beta</th>
<th>Actual Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.95841</td>
<td>21/115</td>
<td>1.000</td>
<td>5.0</td>
<td>2.0</td>
<td>0.01000</td>
<td>0.01000</td>
<td>0.05000</td>
<td>0.04159</td>
</tr>
<tr>
<td>0.95956</td>
<td>14/77</td>
<td>1.000</td>
<td>5.0</td>
<td>2.0</td>
<td>0.05000</td>
<td>0.05000</td>
<td>0.05000</td>
<td>0.04044</td>
</tr>
<tr>
<td>0.96221</td>
<td>11/60</td>
<td>1.000</td>
<td>5.0</td>
<td>2.0</td>
<td>0.10000</td>
<td>0.10000</td>
<td>0.05000</td>
<td>0.03779</td>
</tr>
</tbody>
</table>

*PASS* has calculated 21, 14, and 11 for $r$ as in Epstein.

We should note that occasionally our results differ from those of Epstein. We have checked a few of these carefully by hand, and, in every case, we have found our results to be correct.
Chapter 410

Inequality Tests for One Mean (Simulation)

Introduction

This procedure allows you to study the power and sample size of several statistical tests of the hypothesis that the population mean is equal to a specific value versus the alternative that it is greater than, less than, or not equal to that value. The one-sample t-test is commonly used in this situation, but other tests have been developed for situations where the data are not normally distributed. These additional tests include the Wilcoxon signed-rank test, the sign test, and the computer-intensive bootstrap test. When the population follows the exponential distribution, a test based on this distribution should be used.

The t-test assumes that the data are normally distributed. When this assumption does not hold, the t-test is still used hoping that its robustness will produce accurate results. This procedure allows you to study the accuracy of various tests using simulation techniques. A wide variety of distributions can be simulated to allow you to assess the impact of various forms of non-normality on each test’s accuracy.

The details of the power analysis of the t-test using analytic techniques are presented in the PASS chapter entitled “Inequality Tests for One Mean” and will not be duplicated here. This chapter will be confined to power analysis using computer simulation.

Technical Details

Computer simulation allows one to estimate the power and significance level that is actually achieved by a test procedure in situations that are not mathematically tractable. Computer simulation was once limited to mainframe computers. Currently, due to increased computer speeds, simulation studies can be completed on desktop and laptop computers in a reasonable period of time.

The steps to a simulation study are as follows.

1. Specify the method by which the test is to be carried out. This includes specifying how the test statistic is calculated and how the significance level is specified.
2. Generate a random sample, \( X_1, X_2, \ldots, X_n \), from the distribution specified by the alternative hypothesis. Calculate the test statistic from the simulated data and determine if the null hypothesis is accepted or rejected. Each of these samples is used to calculate the power of the test.

3. Generate a random sample, \( Y_1, Y_2, \ldots, Y_n \), from the distribution specified by the null hypothesis. Calculate the test statistic from the simulated data and determine if the null hypothesis is accepted or rejected. Each of these samples is used to calculate the significance level of the test.

4. Repeat steps 2 and 3 several thousand times, tabulating the number of times the simulated data lead to a rejection of the null hypothesis. The power is the proportion of simulation samples in step 2 that lead to rejection. The significance level is the proportion of simulated samples in step 3 that lead to rejection.

---

### Data Distributions

A wide variety of distributions may be studied. These distributions can vary in skewness, elongation, or other features such as bimodality. A detailed discussion of the distributions that may be used in the simulation is provided in the chapter ‘Data Simulator’.

### Test Statistics

This section describes the test statistics that are available in this procedure.

#### One-Sample t-Test

The one-sample t-test assumes that the data are a simple random sample from a population of normally-distributed values that all have the same mean and variance. This assumption implies that the data are continuous and their distribution is symmetric. The calculation of the t-test proceeds as follows.

\[
t_{n-1} = \frac{\bar{X} - M_0}{s_{\bar{X}}}
\]

where

\[
\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n},
\]

\[
s_{\bar{X}} = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}},
\]

and \( M_0 \) is the value of the mean hypothesized by the null hypothesis.

The significance of the test statistic is determined by computing the p-value. If this p-value is less than a specified level (often 0.05), the null hypothesis is rejected. Otherwise, no conclusion can be reached.
Wilcoxon Signed-Rank Test

The Wilcoxon signed-rank test is a popular, nonparametric substitute for the t-test. This test assumes that the data follow a symmetric distribution. The test is computed using the following steps.

1. Subtract the hypothesized mean, $M_0$, from each data value. Rank the values according to their absolute values.
2. Compute the sum of the positive ranks, $S_p$, and the sum of the negative ranks, $S_n$. The test statistic, $W$, is the minimum of $S_p$ and $S_n$.
3. Compute the mean and standard deviation of $W$ using the formulas

$$\mu_w = \frac{n(n+1)}{4}$$

and

$$s_w = \sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{\sum t_i^3 - \sum t_i}{48}}$$

respectively, where $t_i$ represents the number of times the $i^{th}$ value occurs.
4. Compute the $z$ value using

$$z_w = \frac{W - \mu_w}{s_w}$$

For cases when $n$ is less than 38, the significance level is found from a table of exact probabilities for the Wilcoxon test. When $n$ is greater than or equal to 38, the significance of the test statistic is determined by comparing the $z$ value to a normal probability table. If this p-value is less than a specified level (often 0.05), the null hypothesis is rejected. Otherwise, no conclusion can be reached.

Sign Test

The sign test is popular because it is simple to compute. This test assumes that the data all follow the same distribution. The test is computed using the following steps.

1. Count the number of values strictly greater than $M_0$. Call this value $X$.
2. Count the number of values strictly less than $M_0$. Call this value $Y$.
3. Set $m = X + Y$.
4. Under the null hypothesis, $X$ is distributed as a binomial random variable with a proportion of 0.5 and sample size of $m$.

The significance of $X$ is calculated using binomial probabilities.

Bootstrap Test

The one-sample bootstrap procedure for testing whether the mean is equal to a specific value is given in Efron & Tibshirani (1993), pages 224-227. The bootstrap procedure is as follows.

1. Compute the mean of the sample. Call it $\bar{X}$.
2. Compute the t-value using the standard t-test. The formula for this computation is

$$t_X = \frac{\bar{X} - M_0}{s_{\bar{X}}}$$
where $M_0$ is the hypothesized mean.

3. Draw a random, with-replacement sample of size $n$ from the original $X$ values. Call this sample $Y_1, Y_2, \ldots, Y_n$.

4. Compute the t-value of this bootstrap sample using the formula

$$t = \frac{\bar{Y} - \bar{X}}{s_T}$$

5. For a two-tailed test, if $|t_Y| > |t_x|$ then add one to a counter variable, $A$.

6. Repeat steps 3 – 5 $B$ times. $B$ may be anywhere from 100 to 10,000.

7. Compute the $p$-value of the bootstrap test as $(A + 1) / (B + 1)$

8. Steps 1 – 7 complete one simulation iteration. Repeat these steps $M$ times, where $M$ is the number of simulations. The power and significance level are equal to the percent of the time the $p$-value is less than the nominal alpha of the test in their respective simulations.

Note that the bootstrap test is a time-consuming test to analyze, especially if you set $B$ to a value much larger than 100.

**Exponential Test**

The exponential distribution is a highly skewed distribution, so it is very different from the normal distribution. Thus, the t-test does not work well with exponential data.

There is an exact test for the mean of a sample drawn from the exponential distribution. It is well known that a simple function of the mean of exponential data follows the chi-square distribution. This relationship is given in Epstein (1960) as

$$\frac{2n\bar{X}}{M_0} \sim \chi^2_{2n}$$

This expression can be used to test hypotheses about the value of the mean, $M_0$.

**Standard Deviations**

Care must be used when either the null or alternative distribution is not normal. In these cases, the standard deviation is usually not specified directly. For example, you might use a gamma distribution with a shape parameter of 1.5 and a mean of 4 as the null distribution and a gamma distribution with the same shape parameter and a mean of 5 as the alternative distribution. This allows you to compare the two means. However, note that although the shape parameters are constant, the standard deviations are not. In cases such as this, the null and alternatives not only have different means, but different standard deviations!
Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains most of the parameters and options that you will be concerned with.

Solve For

Find (Solve For)

This option specifies the parameter to be calculated using the values of the other parameters. Under most conditions, you would select either Power or N.

Select Power when you want to estimate the power for a specific scenario.

Select N when you want to determine the sample size needed to achieve a given power and alpha error level. This option can be very computationally intensive, and may take considerable time to complete.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.
Sample Size

N (Sample Size)
This option specifies one or more values of the sample size, the number of individuals in the study. This value must be an integer greater than one. Note that you may enter a list of values using the syntax 50 100 150 200 250 or 50 to 250 by 50.

Test

Test Type
Specify which test statistic (t-test, Wilcoxon test, sign test, bootstrap test, or exponential test) is to be simulated. Although the t-test is the most commonly used test statistic, it is based on assumptions that may not be viable in many situations. For your data, you may find that one of the other tests is more accurate (actual alpha = target alpha) and more precise (higher power).

Note that the bootstrap test is computationally intensive, so it can be very slow to evaluate.

Alternative Hypothesis
This option specifies the alternative hypothesis, H1. This implicitly specifies the direction of the hypothesis test. The null hypothesis is always H0: Mean = M0.

Note that the alternative hypothesis enters into power calculations by specifying the rejection region of the hypothesis test. Its accuracy is critical.

Possible selections are:

- **Mean <> M0**
  This is the most common selection. It yields a *two-tailed test*. Use this option when you are testing whether the mean is different from a specified value, M0, but you do not want to specify beforehand whether it is smaller or larger. Most scientific journals require two-tailed tests.

- **Mean < M0**
  This option yields a *one-tailed test*. Use it when you want to test whether the true mean is less than M0.

- **Mean > M0**
  This option yields a *one-tailed test*. Use it when you want to test whether the true mean is greater than M0.

Simulations

Simulations
This option specifies the number of iterations, M, used in the simulation. Larger numbers of iterations result in longer running time and more accurate results.

The precision of the simulated power estimates can be determined by recognizing that they follow the binomial distribution. Thus, confidence intervals may be constructed for power estimates. The following table gives an estimate of the precision that is achieved for various simulation sizes when the power is either 0.50 or 0.95. The table values are interpreted as follows: a 95%
confidence interval of the true power is given by the power reported by the simulation plus and minus the ‘Precision’ amount given in the table.

<table>
<thead>
<tr>
<th>Simulation Size</th>
<th>Precision when Power = 0.50</th>
<th>Precision when Power = 0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>M</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.100</td>
<td>0.044</td>
</tr>
<tr>
<td>500</td>
<td>0.045</td>
<td>0.019</td>
</tr>
<tr>
<td>1000</td>
<td>0.032</td>
<td>0.014</td>
</tr>
<tr>
<td>2000</td>
<td>0.022</td>
<td>0.010</td>
</tr>
<tr>
<td>5000</td>
<td>0.014</td>
<td>0.006</td>
</tr>
<tr>
<td>10000</td>
<td>0.010</td>
<td>0.004</td>
</tr>
<tr>
<td>50000</td>
<td>0.004</td>
<td>0.002</td>
</tr>
<tr>
<td>100000</td>
<td>0.003</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Notice that a simulation size of 1000 gives a precision of plus or minus 0.014 when the true power is 0.95. Also note that as the simulation size is increased beyond 5000, there is only a small amount of additional precision achieved.

**Effect Size**

**Distribution Assuming H0 (Null Hypothesis)**

This option specifies the mean and distribution under the null hypothesis, H0. Usually, the mean is specified by entering ‘M0’ for the mean parameter in the distribution expression and then entering values for the M0 parameter described below. All of the distributions are parameterized so that the mean is entered first. For example, if you wanted to test whether the mean of a normal distributed variable is five, you could enter N(5, S) or N(M0, S) here.

The parameters of each distribution are specified using numbers or letters. If letters are used, their values are specified in the boxes below. The value ‘M0’ is reserved for the value of the mean under the null hypothesis.

Following is a list of the distributions that are available and the syntax used to specify them:

- Beta=A(M0,A,B,Minimum)
- Binomial=B(M0,N)
- Cauchy=C(M0,Scale)
- Constant=K(Value)
- Exponential=E(M0)
- F=F(M0,DF1)
- Gamma=G(M0,A)
- Multinomial=M(P1,P2,P3,…,Pk)
- Normal=N(M0,SD)
- Poisson=P(M0)
- Student's T=T(M0,D)
- Tukey's Lambda=L(M0,S,Skewness,Elongation)
- Uniform=U(M0,Minimum)
- Weibull=W(M0,B)

Details of writing mixture distributions, combined distributions, and compound distributions are found in the chapter on Data Simulation and will not be repeated here.
Finding the Value of the Mean under H0
The distributions have been parameterized in terms of their means, since this is the parameter being tested. The mean of a distribution created as a linear combination of other distributions is found by applying the linear combination to the individual means. However, the mean of a distribution created by multiplying or dividing other distributions is not necessarily equal to applying the same function to the individual means. For example, the mean of \(4N(4, 5) + 2N(5, 6)\) is \(4*4 + 2*5 = 26\), but the mean of \(4N(4, 5) * 2N(5, 6)\) is not exactly \(4*4*2*5 = 160\) (although it is close).

Specifying the Mean for Paired (Matched) Data
Depending on the formula that is entered, the mean is not necessarily the value of M0. For example, a common use of the one-group t-test is to test whether the mean of a set of differences is zero. Differences may be specified (ignoring the correlation between paired observations) as the difference between two normal distributions. This would be specified as \(N(M0, S) - N(M0, S)\). The mean of the resulting distribution is \(M0 – M0 = 0\) (not \(M0\)).

Distribution Assuming H1 (Alternative Hypothesis)
This option specifies the mean and distribution under the alternative hypothesis, H1. That is, this is the actual (true) value of the mean at which the power is computed. Usually, the mean is specified by entering ‘M1’ for the mean parameter in the distribution expression and then entering values for the M1 parameter below. All of the distributions are parameterized so that the mean is entered first.

The parameters of each distribution are specified using numbers or letters. If letters are used, their values are specified in the boxes below. The value ‘M1’ is reserved for the value of the mean under the alternative hypothesis.

Following is a list of the distributions that are available and the syntax used to specify them:

- Beta=A(M1,A,B,Minimum)
- Binomial=B(M1,N)
- Cauchy=C(M1,Scale)
- Constant=K(Value)
- Exponential=E(M1)
- F=F(M1,DF1)
- Gamma=G(M1,A)
- Multinomial=M(P1,P2,P3,…,Pk)
- Normal=N(M1,SD)
- Poisson=P(M1)
- Student's T=T(M1,D)
- Tukey's Lambda=L(M1,S,Skewness,Elongation)
- Uniform=U(M1,Minimum)
- Weibull=W(M1,B)

Details of writing mixture distributions, combined distributions, and compound distributions are found in the chapter on Data Simulation and will not be repeated here.

Finding the Value of the Mean under H1
The distributions have been parameterized in terms of their means, since this is the parameter being tested. The mean of a distribution created as a linear combination of other distributions is found by applying the linear combination to the individual means. However, the mean of a
distribution created by multiplying or dividing other distributions is not necessarily equal to applying the same function to the individual means. For example, the mean of $4N(4, 5) + 2N(5, 6)$ is $4*4 + 2*5 = 26$, but the mean of $4N(4, 5) * 2N(5, 6)$ is not exactly $4*4*2*5 = 160$ (although it is close).

**Specifying the Mean for Paired (Matched) Data**

Depending on the formula that is entered, the mean is not necessarily the value of M1. For example, a common use of the one-group t-test is to test whether the mean of a set of differences is zero. Differences may be specified (ignoring the correlation between paired observations) as the difference between two normal distributions. This would be specified as $N(M1, S) - N(M0, S)$. The mean of the resulting distribution is $M1 - M0$.

---

**Effect Size – Distribution Parameters**

* M0 (Mean|H0)

These values are substituted for the M0 in the distribution specifications given above. M0 is intended to be the value of the mean hypothesized by the null hypothesis, H0.

You can enter a list of values using syntax such as `0 1 2 3` or `0 to 3 by 1`.

Note that whether M0 is the mean of the simulated distribution depends on the formula you have entered. For example, $N(M0, S)$ has a mean of M0, but $N(M0, S) - N(M0, S)$ has a mean of zero.

* M1 (Mean|H1)

These values are substituted for the M1 in the distribution specifications given above. M1 is intended to be the value of the mean hypothesized by the alternative hypothesis, H1.

You can enter a list of values using syntax such as `0 1 2 3` or `0 to 3 by 1`.

Note that whether M1 is the mean of the simulated distribution depends on the formula you have entered. For example, $N(M1, S)$ has a mean of M1, but $N(M1, S) - N(M0, S)$ has a mean of M1 - M0.

* Parameter Values (S, A, B, C)

Enter the numeric value(s) of parameter listed above. These values are substituted for the corresponding letter in the distribution specifications for H0 and H1.

You can enter a list of values using syntax such as `0 1 2 3` or `0 to 3 by 1`.

You can also change the letter that is used as the name of this parameter.

---

**Iterations Tab**

The Iterations tab contains limits on the number of iterations and various options about individual tests.

* **Maximum Iterations**

**Maximum Iterations Before Search Termination**

Specify the maximum number of iterations before the search for the sample size, N, is aborted. When the maximum number of iterations is reached without convergence, the sample size is not reported. We recommend a value of at least 500.
Bootstrap Iterations

Bootstrap Iterations
Specify the number of iterations used in the bootstrap hypothesis test. This value is only used if the bootstrap test is displayed on the reports. The running time of the procedure depends heavily on the number of iterations specified here.

Recommendations by authors of books discussing the bootstrap range from 100 to 10,000. If you enter a large (greater than 500) value, the procedure may take several hours to run.

Example 1 – Power at Various Sample Sizes

A researcher is planning an experiment to test whether the mean response level to a certain drug is significantly different from zero. The researcher wants to use a t-test with an alpha level of 0.05. He wants to compute the power at various sample sizes from 5 to 40, assuming the true mean is one. He assumes that the data are normally distributed with a standard deviation of 2. Since this is an exploratory analysis, he sets the number of simulation iterations to 1000.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for One Mean (Simulation) procedure window by expanding Means, then One Mean, then clicking on Test (Inequality), and then clicking on Tests for One Mean (Simulation). You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>5 to 40 by 5</td>
</tr>
<tr>
<td>Distribution</td>
<td>H0 (Null Hypothesis)</td>
</tr>
<tr>
<td>Distribution</td>
<td>H1 (Alt Hypothesis)</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H1)</td>
</tr>
<tr>
<td>S</td>
<td>2</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Mean ≠ M0</td>
</tr>
<tr>
<td>Test Type</td>
<td>T-Test</td>
</tr>
<tr>
<td>Simulations</td>
<td>1000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Reports Tab</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Show Numeric Report</td>
<td>Checked</td>
</tr>
<tr>
<td>Show Inc's &amp; 95% C.I.'s</td>
<td>Checked</td>
</tr>
<tr>
<td>Show Definitions</td>
<td>Checked</td>
</tr>
<tr>
<td>Show Plots</td>
<td>Checked</td>
</tr>
<tr>
<td>Show Summary Statements</td>
<td>1</td>
</tr>
</tbody>
</table>
Annotated Output

Click the Run button to perform the calculations and generate the following output.

**Numeric Results**

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>Mean0</th>
<th>Mean1</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>Beta</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.138</td>
<td>5</td>
<td>0.0</td>
<td>1.0</td>
<td>0.050</td>
<td>0.050</td>
<td>0.862</td>
<td>0.0</td>
<td>1.0</td>
<td>2.0</td>
</tr>
<tr>
<td>(0.021)</td>
<td>[0.117</td>
<td>0.159]</td>
<td></td>
<td></td>
<td>(0.014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.293</td>
<td>10</td>
<td>0.0</td>
<td>1.0</td>
<td>0.050</td>
<td>0.058</td>
<td>0.563</td>
<td>0.0</td>
<td>1.0</td>
<td>2.0</td>
</tr>
<tr>
<td>(0.028)</td>
<td>[0.265</td>
<td>0.321]</td>
<td></td>
<td></td>
<td>(0.014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.437</td>
<td>15</td>
<td>0.0</td>
<td>1.0</td>
<td>0.050</td>
<td>0.058</td>
<td>0.563</td>
<td>0.0</td>
<td>1.0</td>
<td>2.0</td>
</tr>
<tr>
<td>(0.031)</td>
<td>[0.406</td>
<td>0.468]</td>
<td></td>
<td></td>
<td>(0.014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.582</td>
<td>20</td>
<td>0.0</td>
<td>1.0</td>
<td>0.050</td>
<td>0.058</td>
<td>0.418</td>
<td>0.0</td>
<td>1.0</td>
<td>2.0</td>
</tr>
<tr>
<td>(0.031)</td>
<td>[0.551</td>
<td>0.613]</td>
<td></td>
<td></td>
<td>(0.014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.643</td>
<td>25</td>
<td>0.0</td>
<td>1.0</td>
<td>0.050</td>
<td>0.048</td>
<td>0.357</td>
<td>0.0</td>
<td>1.0</td>
<td>2.0</td>
</tr>
<tr>
<td>(0.030)</td>
<td>[0.613</td>
<td>0.763]</td>
<td></td>
<td></td>
<td>(0.013)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.772</td>
<td>30</td>
<td>0.0</td>
<td>1.0</td>
<td>0.050</td>
<td>0.042</td>
<td>0.228</td>
<td>0.0</td>
<td>1.0</td>
<td>2.0</td>
</tr>
<tr>
<td>(0.026)</td>
<td>[0.746</td>
<td>0.798]</td>
<td></td>
<td></td>
<td>(0.012)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.806</td>
<td>35</td>
<td>0.0</td>
<td>1.0</td>
<td>0.050</td>
<td>0.054</td>
<td>0.194</td>
<td>0.0</td>
<td>1.0</td>
<td>2.0</td>
</tr>
<tr>
<td>(0.025)</td>
<td>[0.781</td>
<td>0.831]</td>
<td></td>
<td></td>
<td>(0.014)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.872</td>
<td>40</td>
<td>0.0</td>
<td>1.0</td>
<td>0.050</td>
<td>0.044</td>
<td>0.128</td>
<td>0.0</td>
<td>1.0</td>
<td>2.0</td>
</tr>
<tr>
<td>(0.021)</td>
<td>[0.851</td>
<td>0.893]</td>
<td></td>
<td></td>
<td>(0.013)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:
Second Row: (Power Inc.) [95% LCL and UCL Power] (Alpha Inc.) [95% LCL and UCL Alpha]
Number of Monte Carlo Samples: 1000. Simulation Run Time: 17.81 seconds.

Report Definitions
Power is the probability of rejecting a false null hypothesis.
N is the size of the sample drawn from the population.
Mean0 is the value of the mean assuming the null hypothesis. This is the value being tested.
Mean1 is the actual value of the mean. The procedure tests whether Mean0 = Mean1.
Target Alpha is the probability of rejecting a true null hypothesis. It is set by the user.
Actual Alpha is the alpha level that was actually achieved by the experiment.
Beta is the probability of accepting a false null hypothesis.

This report shows the estimated power for each scenario. The first row shows the parameter settings and the estimated power and significance level (Actual Alpha). Note that because these are results of a simulation study, the computed power and alpha will vary from run to run. Thus, another report obtained using the same input parameters will be slightly different than the one above.

The second row shows two 95% confidence intervals in brackets: the first for the power and the second for the significance level. Half the width of each confidence interval is given in parentheses as a fundamental measure of the accuracy of the simulation. As the number of simulations is increased, the width of the confidence interval will decrease.
Inequality Tests for One Mean (Simulation)

Plots Section

This plot shows the relationship between sample size and power.

Example 2 – Finding the Sample Size for Skewed Data

In studying deaths from SIDS (Sudden Infant Death Syndrome), one hypothesis put forward is that infants dying of SIDS weigh less than normal at birth. Suppose the average birth weight of infants is 3300 grams with a standard deviation of 663 grams. The researchers decide to examine the effect of a skewed distribution on the test used by adding skewness to the simulated data using Tukey’s Lambda distribution with a skewness factor of 0.5.

Using the Data Simulator program, the researchers found that the actual standard deviation using the above parameters was almost 800. This occurs because adding skewness changes the standard deviation. They found that setting the standard deviation in Tukey’s Lambda distribution to 563 resulted in a standard deviation in the data of about 663.

A histogram of 10,000 pseudo-random values from this distribution appears as follows.
The researchers want to determine how large a sample of SIDS infants will be needed to detect a drop in average weight of 25%? Note that applying this percentage to the average weight of 3300 yields 2475. Use an alpha of 0.05 and 80% power.

Although a one-sided hypothesis might be considered, sample size estimates will assume a two-sided alternative to keep the research design in line with other studies. To decrease the running time of this example, the number of simulation iterations is set to 1000. In practice, you would probably use a value of about 5000.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for One Mean (Simulation) procedure window by expanding Means, then One Mean, then clicking on Test (Inequality), and then clicking on Tests for One Mean (Simulation). You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N</td>
</tr>
<tr>
<td>Power</td>
<td>0.80</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Distribution</td>
<td>H0 (Null Hypothesis)</td>
</tr>
<tr>
<td>Distribution</td>
<td>H1 (Alt Hypothesis)</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H1)</td>
</tr>
<tr>
<td>S</td>
<td>563</td>
</tr>
<tr>
<td>G</td>
<td>0.5 (Note that parameter A was changed to G.)</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Mean ≠ M0</td>
</tr>
<tr>
<td>Test Type</td>
<td>T-Test</td>
</tr>
<tr>
<td>Simulations</td>
<td>1000</td>
</tr>
</tbody>
</table>

Reports Tab

Show Numeric Report...Checked
Show Inc's & 95% C.I.'s...Checked

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results of Search for N

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>H0 Mean0</th>
<th>H1 Mean1</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>Beta</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.817</td>
<td>6</td>
<td>2475.0</td>
<td>3300.0</td>
<td>0.050</td>
<td>0.073</td>
<td>0.183</td>
<td>2475.0</td>
<td>3300.0</td>
<td>563.0</td>
</tr>
<tr>
<td>(0.024)</td>
<td></td>
<td>[0.793]</td>
<td>[0.841]</td>
<td>(0.016)</td>
<td>(0.016)</td>
<td>[0.057]</td>
<td>[0.089]</td>
<td>[0.989]</td>
<td></td>
</tr>
</tbody>
</table>

The required sample size was 6. Notice how wide the confidence interval of power is. We re-ran this simulation several times and obtained sample sizes of 5, 6, and 7. Note that the actual alpha
value is between 0.057 and 0.089, which is definitely greater than 0.05. This shows one of the problems of using the t-test with a skewed distribution.

To be more accurate and yet avoid the long running time of the search for N, a reasonable strategy would be to run simulations to obtain the powers using N’s from 4 to 10. The result of this study is displayed next.

### Numeric Results of Power Search for Various N

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>H0 Mean0</th>
<th>H1 Mean1</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>Beta</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.414</td>
<td>4</td>
<td>2475.0</td>
<td>3300.0</td>
<td>0.050</td>
<td>0.093</td>
<td>0.586</td>
<td>2475.0</td>
<td>3300.0</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.014)</td>
<td>(0.008)</td>
<td>(0.085)</td>
<td>(0.008)</td>
<td></td>
<td>(0.101)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.645</td>
<td>5</td>
<td>2475.0</td>
<td>3300.0</td>
<td>0.050</td>
<td>0.084</td>
<td>0.355</td>
<td>2475.0</td>
<td>3300.0</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.013)</td>
<td>(0.006)</td>
<td>(0.076)</td>
<td>(0.007)</td>
<td></td>
<td>(0.091)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.811</td>
<td>6</td>
<td>2475.0</td>
<td>3300.0</td>
<td>0.050</td>
<td>0.088</td>
<td>0.189</td>
<td>2475.0</td>
<td>3300.0</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.011)</td>
<td>(0.008)</td>
<td>(0.081)</td>
<td>(0.008)</td>
<td></td>
<td>(0.096)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.912</td>
<td>7</td>
<td>2475.0</td>
<td>3300.0</td>
<td>0.050</td>
<td>0.089</td>
<td>0.088</td>
<td>2475.0</td>
<td>3300.0</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.008)</td>
<td>(0.008)</td>
<td>(0.081)</td>
<td>(0.008)</td>
<td></td>
<td>(0.097)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.960</td>
<td>8</td>
<td>2475.0</td>
<td>3300.0</td>
<td>0.050</td>
<td>0.077</td>
<td>0.040</td>
<td>2475.0</td>
<td>3300.0</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.005)</td>
<td>(0.007)</td>
<td>(0.069)</td>
<td>(0.007)</td>
<td></td>
<td>(0.084)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.983</td>
<td>9</td>
<td>2475.0</td>
<td>3300.0</td>
<td>0.050</td>
<td>0.082</td>
<td>0.017</td>
<td>2475.0</td>
<td>3300.0</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.004)</td>
<td>(0.008)</td>
<td>(0.074)</td>
<td>(0.008)</td>
<td></td>
<td>(0.089)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.994</td>
<td>10</td>
<td>2475.0</td>
<td>3300.0</td>
<td>0.050</td>
<td>0.079</td>
<td>0.006</td>
<td>2475.0</td>
<td>3300.0</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.002)</td>
<td>(0.007)</td>
<td>(0.071)</td>
<td>(0.007)</td>
<td></td>
<td>(0.086)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The sample size of 6 appears to meet the design parameters the best. The actual significance level still appears to be greater than 0.05. The researchers decide that they must use a smaller value of Alpha so that the actual alpha is about 0.05. After some experimentation, they find that setting Alpha to 0.025 results in the desired power and significance level.

### Numeric Results with Alpha = 0.025

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>H0 Mean0</th>
<th>H1 Mean1</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>Beta</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.593</td>
<td>6</td>
<td>2475.0</td>
<td>3300.0</td>
<td>0.025</td>
<td>0.057</td>
<td>0.407</td>
<td>2475.0</td>
<td>3300.0</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.014)</td>
<td>(0.006)</td>
<td>(0.051)</td>
<td>(0.006)</td>
<td></td>
<td>(0.064)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.754</td>
<td>7</td>
<td>2475.0</td>
<td>3300.0</td>
<td>0.025</td>
<td>0.058</td>
<td>0.246</td>
<td>2475.0</td>
<td>3300.0</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.012)</td>
<td>(0.006)</td>
<td>(0.051)</td>
<td>(0.006)</td>
<td></td>
<td>(0.064)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.862</td>
<td>8</td>
<td>2475.0</td>
<td>3300.0</td>
<td>0.025</td>
<td>0.049</td>
<td>0.138</td>
<td>2475.0</td>
<td>3300.0</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.010)</td>
<td>(0.006)</td>
<td>(0.043)</td>
<td>(0.006)</td>
<td></td>
<td>(0.055)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.929</td>
<td>9</td>
<td>2475.0</td>
<td>3300.0</td>
<td>0.025</td>
<td>0.044</td>
<td>0.071</td>
<td>2475.0</td>
<td>3300.0</td>
<td>0.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>(0.007)</td>
<td>(0.006)</td>
<td>(0.039)</td>
<td>(0.006)</td>
<td></td>
<td>(0.050)</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It appears that a sample size of 8 with a Target Alpha of 0.025 will result in an experimental design with the characteristics the researchers wanted.

Notice that when working with non-normal distributions, you must change both N and the Target Alpha to achieve the design you want!
Example 3 – Comparative results with Skewed Data

Continuing with Example 2, the researchers want to study the characteristics of various test statistics as the amount of skewness is increased. To do this, they let the skewness parameter of Tukey’s Lambda distribution vary between 0 and 1. The researchers realize that the standard deviation will change as the skewness parameter is increased, but they decide to ignore this complication.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for One Mean (Simulation) procedure window by expanding Means, then One Mean, then clicking on Test (Inequality), and then clicking on Tests for One Mean (Simulation). You may then make the appropriate entries as listed below, or open Example 3 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>6</td>
</tr>
<tr>
<td>Distribution</td>
<td>H0 (Null Hypothesis)</td>
</tr>
<tr>
<td>Distribution</td>
<td>H1 (Alt Hypothesis)</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H1)</td>
</tr>
<tr>
<td>S</td>
<td>563</td>
</tr>
<tr>
<td>G</td>
<td>0.0 0.2 0.4 0.6 0.8 1.0</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Mean ≠ M0</td>
</tr>
<tr>
<td>Test Type</td>
<td>T-Test</td>
</tr>
<tr>
<td>Simulations</td>
<td>1000</td>
</tr>
</tbody>
</table>

Report Tab

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Show Comparative Reports</td>
<td>Checked</td>
</tr>
<tr>
<td>Show Comparative Plots</td>
<td>Checked</td>
</tr>
<tr>
<td>Include T-Test Results</td>
<td>Checked</td>
</tr>
<tr>
<td>Include Wilcoxon &amp; Sign Test</td>
<td>Checked</td>
</tr>
</tbody>
</table>
Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power Comparison for Testing One Mean = Mean0. Hypotheses: H0: Mean1=Mean0; H1: Mean1&lt;&gt;Mean0</th>
</tr>
</thead>
<tbody>
<tr>
<td>H0 Distribution: Tukey(M0 S G 0)</td>
</tr>
<tr>
<td>H1 Distribution: Tukey(M1 S G 0)</td>
</tr>
<tr>
<td>N</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Alpha Comparison for Testing One Mean = Mean0. Hypotheses: H0: Mean1=Mean0; H1: Mean1&lt;&gt;Mean0</th>
</tr>
</thead>
<tbody>
<tr>
<td>N</td>
</tr>
<tr>
<td>--------</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>6</td>
</tr>
<tr>
<td>6</td>
</tr>
</tbody>
</table>


Several interesting trends become apparent from this study. First, for a sample size of 6, the power of the Wilcoxon test and the sign test are the same (this is not the case for larger sample sizes). The alpha of the t-test decreases as the amount of skewness increases. The alpha of the Wilcoxon and sign tests does not increase as rapidly for the non-parametric tests.

Example 4 – Validation using Zar

Zar (1984), pages 111-112, presents an example in which Mean0 = 0.0, Mean1 = 1.0, S = 1.25, alpha = 0.05, and N = 12. Zar obtains an approximate power of 0.72. We will validate this procedure by running this example. To make certain that the results are very accurate, the number of simulations will be set to 10,000.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for One Mean (Simulation) procedure window by expanding Means, then One Mean, then clicking on Test (Inequality), and then clicking on Tests for One Mean (Simulation). You may then make the appropriate entries as listed below, or open Example 4 by going to the File menu and choosing Open Example Template.
Inequality Tests for One Mean (Simulation)  410-17

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>12</td>
</tr>
<tr>
<td>Distribution</td>
<td>H0 (Null Hypothesis)</td>
</tr>
<tr>
<td>Distribution</td>
<td>H1 (Alt Hypothesis)</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H1)</td>
</tr>
<tr>
<td>S</td>
<td>1.25</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Mean ≠ M0</td>
</tr>
<tr>
<td>Test Type</td>
<td>T-Test</td>
</tr>
<tr>
<td>Simulations</td>
<td>10000</td>
</tr>
</tbody>
</table>

| Reports Tab                 |                             |
| Show Numeric Report         | Checked                     |
| Show Inc's & 95% C.I.'s     | Checked                     |

**Output**

Click the Run button to perform the calculations and generate the following output.

**Numeric Results**

This simulation obtained a power of 0.717 which rounds to the 0.72 computed by Zar. Note that another repetition of this same analysis will probably be slightly different since a different set of random numbers will be used.
Example 5 – Validation using Machin

Machin, et. al. (1997), page 37, present an example in which Mean0 = 0.0, Mean1 = 0.2, S = 1.0, alpha = 0.05, and beta = 0.20. They obtain a sample size of 199. Because of the long running time, we will set the number of simulations at only 200. In practice you would usually set this to a value greater than 1000.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for One Mean (Simulation) procedure window by expanding Means, then One Mean, then clicking on Test (Inequality), and then clicking on Tests for One Mean (Simulation). You may then make the appropriate entries as listed below, or open Example 5 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N</td>
</tr>
<tr>
<td>Power</td>
<td>0.80</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Distribution</td>
<td>H0 (Null Hypothesis)</td>
</tr>
<tr>
<td>Distribution</td>
<td>H1 (Alt Hypothesis)</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H1)</td>
</tr>
<tr>
<td>S</td>
<td>1</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Mean ≠ M0</td>
</tr>
<tr>
<td>Test Type</td>
<td>T-Test</td>
</tr>
<tr>
<td>Simulations</td>
<td>200</td>
</tr>
</tbody>
</table>

Reports Tab

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Show Numeric Report</td>
<td>Checked</td>
</tr>
<tr>
<td>Show Inc’s &amp; 95% C.I.’s</td>
<td>Checked</td>
</tr>
</tbody>
</table>
Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>Mean0</th>
<th>Mean1</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>Beta</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.785</td>
<td>211</td>
<td>0.0</td>
<td>0.2</td>
<td>0.050</td>
<td>0.045</td>
<td>0.215</td>
<td>0.0</td>
<td>0.2</td>
<td>1.0</td>
</tr>
<tr>
<td>(0.057)</td>
<td>[0.728</td>
<td>0.842)</td>
<td></td>
<td>(0.029)</td>
<td>[0.016</td>
<td>0.074)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:
Second Row: (Power Inc.) [95% LCL and UCL Power] (Alpha Inc.) [95% LCL and UCL Alpha]
Number of Monte Carlo Samples: 200. Simulation Run Time: 39.83 seconds.

Note that using a simulation size of only 200, the estimated sample size of 211 is still close to the exact value of 199. We ran this simulation with 10000 simulations and obtained a sample size of 199.

Example 6 – Power of the Wilcoxon Test

The Wilcoxon nonparametric test was designed for data that do not follow the normal distribution but are symmetric. This type of data often occurs when differences between two non-normal variables are taken, as in a study that analyzes differences in pre- and post-test scores.

For this example, suppose the pre-test and the post-test scores are exponentially distributed. Here are examples of exponentially-distributed data with means of 4 and 2, respectively.

It has been shown that the differences between two identically-distributed variables are symmetric. The histogram below on the left shows differences in the null case in which the difference is between two exponential variables both with a mean of 4. The histogram below on the right shows differences in the alternative case in which the difference is between an exponential variable with a mean of 4 and an exponential variable with a mean of 2. Careful inspection shows that the second histogram is skewed to the right and the mean difference is about 2, not 0.
The researchers want to study the power of the two-sided Wilcoxon test when sample sizes of 10, 20, 30, and 40 are used, and testing is done at the 5% significance level.

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for One Mean (Simulation) procedure window by expanding Means, then One Mean, then clicking on Test (Inequality), and then clicking on Tests for One Mean (Simulation). You may then make the appropriate entries as listed below, or open Example 6 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Tab</strong></td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>10 20 30 40 50</td>
</tr>
<tr>
<td>Distribution</td>
<td>H0 (Null Hypothesis)</td>
</tr>
<tr>
<td>Distribution</td>
<td>H1 (Alt Hypothesis)</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H1)</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Mean ≠ M0</td>
</tr>
<tr>
<td>Test Type</td>
<td>Wilcoxon</td>
</tr>
<tr>
<td>Simulations</td>
<td>5000</td>
</tr>
</tbody>
</table>

| **Reports Tab**          |                                            |
| Show Numeric Report      | Checked                                    |
| Show Inc’s & 95% C.I.’s  | Checked                                    |
Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>Mean0</th>
<th>Mean1</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>Beta</th>
<th>M0</th>
<th>M1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.204</td>
<td>10</td>
<td>0.0</td>
<td>2.0</td>
<td>0.050</td>
<td>0.038</td>
<td>0.796</td>
<td>4.0</td>
<td>2.0</td>
</tr>
<tr>
<td>(0.011)</td>
<td></td>
<td>[0.193]</td>
<td>[0.216]</td>
<td></td>
<td>(0.005)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.480</td>
<td>20</td>
<td>0.0</td>
<td>2.0</td>
<td>0.050</td>
<td>0.051</td>
<td>0.520</td>
<td>4.0</td>
<td>2.0</td>
</tr>
<tr>
<td>(0.014)</td>
<td></td>
<td>[0.466]</td>
<td>[0.494]</td>
<td></td>
<td>(0.006)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.647</td>
<td>30</td>
<td>0.0</td>
<td>2.0</td>
<td>0.050</td>
<td>0.050</td>
<td>0.353</td>
<td>4.0</td>
<td>2.0</td>
</tr>
<tr>
<td>(0.013)</td>
<td></td>
<td>[0.634]</td>
<td>[0.660]</td>
<td></td>
<td>(0.006)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.789</td>
<td>40</td>
<td>0.0</td>
<td>2.0</td>
<td>0.050</td>
<td>0.047</td>
<td>0.211</td>
<td>4.0</td>
<td>2.0</td>
</tr>
<tr>
<td>(0.011)</td>
<td></td>
<td>[0.778]</td>
<td>[0.800]</td>
<td></td>
<td>(0.006)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.863</td>
<td>50</td>
<td>0.0</td>
<td>2.0</td>
<td>0.050</td>
<td>0.049</td>
<td>0.137</td>
<td>4.0</td>
<td>2.0</td>
</tr>
<tr>
<td>(0.010)</td>
<td></td>
<td>[0.853]</td>
<td>[0.872]</td>
<td></td>
<td>(0.006)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:
Second Row: (Power Inc.) [95% LCL and UCL Power]    (Alpha Inc.) [95% LCL and UCL Alpha]
Number of Monte Carlo Samples: 5000.   Simulation Run Time: 79.70 seconds.

Reasonable power is achieved for N = 50.

Example 7 – Likert-Scale Data

Likert-scale data occurs commonly in survey research. A Likert Scale is discrete, ordinal data. It usually occurs when a survey poses a question and the respondent must pick among strongly agree, agree, undecided, disagree, or strongly disagree. The responses are usually coded as 1, 2, 3, 4, and 5.

Likert data can be analyzed in a number of ways. Perhaps the most common is to use a t-test or a Wilcoxon test. (Using the Wilcoxon test is invalid in this case because the data are seldom distributed symmetrically.)

In this example, a questionnaire is planned on which Likert-scale questions will be asked. The researchers want to study the power and actual significance levels of various sample sizes. They decide to look at what happens as the proportion of strongly agree responses is increased beyond a perfectly uniform response pattern. They want to compute the power when the strongly agree response is twice as likely, four times as likely, and eight times as likely. The sample size is 20, alpha is 0.05, and the test is two-sided.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for One Mean (Simulation) procedure window by expanding Means, then One Mean, then clicking on Test (Inequality), and then clicking on
Tests for One Mean (Simulation). You may then make the appropriate entries as listed below, or open Example 7 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>20</td>
</tr>
<tr>
<td>Distribution</td>
<td>H0 (Null Hypothesis)</td>
</tr>
<tr>
<td>Distribution</td>
<td>H1 (Alt Hypothesis)</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H1)</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Mean ≠ M0</td>
</tr>
<tr>
<td>Test Type</td>
<td>T-Test</td>
</tr>
<tr>
<td>Simulations</td>
<td>5000</td>
</tr>
</tbody>
</table>

Reports Tab
Show Numeric Report Checked
Show Inc's & 95% C.I.'s Checked

Output
Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>H0 Mean0</th>
<th>H1 Mean1</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>Beta</th>
<th>M0</th>
<th>M1</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.167</td>
<td>20</td>
<td>3.0</td>
<td>2.7</td>
<td>0.050</td>
<td>0.050</td>
<td>0.833</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>0.558</td>
<td>20</td>
<td>3.0</td>
<td>2.3</td>
<td>0.050</td>
<td>0.052</td>
<td>0.442</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>0.910</td>
<td>20</td>
<td>3.0</td>
<td>1.8</td>
<td>0.050</td>
<td>0.055</td>
<td>0.090</td>
<td>1</td>
<td>8</td>
</tr>
</tbody>
</table>

Notes:
Second Row: (Power Inc.) [95% LCL and UCL Power] (Alpha Inc.) [95% LCL and UCL Alpha]
Number of Monte Carlo Samples: 5000. Simulation Run Time: 12.53 seconds.

Note that M0 and M1 are no longer the H0 and H1 means. Now, they represent the relative weighting given to the strongly agree response. Under H0, the mean is 3.0. As M1 is increased, the mean under H1 changes from 2.7 to 2.3 to 1.8. We note that the actual significance level, alpha, remains close to the target value of 0.05.
Example 8 – Computing the Power after Completing an Experiment

A group of researchers has completed an experiment designed to determine if a particular hormone increases weight gain in rats. The researchers inject 20 rats of the same age with the hormone and measure their weight gain after 1 month. The investigators use the two-sided bootstrap test with alpha = 0.05 and 100 bootstrap samples to determine if the average weight gained by these rats (171 grams) is significantly greater than the known average weight gained by rats of the same age over the same period of time (155 grams). Unfortunately, the results indicate that there is no significant difference between the two means. Therefore, the researchers decide to compute the power achieved by this test for alternative means ranging from 160 to 190 grams. They decide to use 1000 simulations for the study. For comparative purposes, they also decide to look at the power achieved by the bootstrap test in comparison to various other applicable tests. Suppose that they know that the standard deviation for weight gain is 33 grams.

Note that the researchers compute the power for a range of practically significant alternatives. The range chosen should represent likely values based on historical evidence.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for One Mean (Simulation) procedure window by expanding Means, then One Mean, then clicking on Test (Inequality), and then clicking on Tests for One Mean (Simulation). You may then make the appropriate entries as listed below, or open Example 8 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>20</td>
</tr>
<tr>
<td>Distribution</td>
<td>H0 (Null Hypothesis)</td>
</tr>
<tr>
<td>Distribution</td>
<td>H1 (Alt Hypothesis)</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H1)</td>
</tr>
<tr>
<td>S</td>
<td>33</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Mean ≠ M0</td>
</tr>
<tr>
<td>Test Type</td>
<td>Bootstrap</td>
</tr>
<tr>
<td>Simulations</td>
<td>1000</td>
</tr>
<tr>
<td>Reports Tab</td>
<td></td>
</tr>
<tr>
<td>Show Numeric Report</td>
<td>Checked</td>
</tr>
<tr>
<td>Show Inc's &amp; 95% C.I.'s</td>
<td>Checked</td>
</tr>
<tr>
<td>Show Comparative Reports</td>
<td>Checked</td>
</tr>
<tr>
<td>Show Plots</td>
<td>Checked</td>
</tr>
<tr>
<td>Show Comparative Plots</td>
<td>Checked</td>
</tr>
<tr>
<td>Include T-Test Results</td>
<td>Checked</td>
</tr>
</tbody>
</table>
410-24 Inequality Tests for One Mean (Simulation)

Reports Tab (continued)
Include Wilcoxon & Sign Test .................. Checked
Include Bootstrap Test Results ............... Checked

Iterations Tab
Bootstrap Iterations .............................. 100

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results for Power of Bootstrap

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>H0 Mean0</th>
<th>H1 Mean1</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>Beta</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.108</td>
<td>20</td>
<td>155.0</td>
<td>160.0</td>
<td>0.050</td>
<td>0.045</td>
<td>0.892</td>
<td>155.0</td>
<td>160.0</td>
<td>33.0</td>
</tr>
<tr>
<td>(0.019)</td>
<td>[0.989 0.127]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.453</td>
<td>20</td>
<td>155.0</td>
<td>170.0</td>
<td>0.050</td>
<td>0.044</td>
<td>0.547</td>
<td>155.0</td>
<td>170.0</td>
<td>33.0</td>
</tr>
<tr>
<td>(0.031)</td>
<td>[0.422 0.484]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.872</td>
<td>20</td>
<td>155.0</td>
<td>180.0</td>
<td>0.050</td>
<td>0.044</td>
<td>0.128</td>
<td>155.0</td>
<td>180.0</td>
<td>33.0</td>
</tr>
<tr>
<td>(0.021)</td>
<td>[0.851 0.893]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.994</td>
<td>20</td>
<td>155.0</td>
<td>190.0</td>
<td>0.050</td>
<td>0.042</td>
<td>0.006</td>
<td>155.0</td>
<td>190.0</td>
<td>33.0</td>
</tr>
<tr>
<td>(0.005)</td>
<td>[0.989 0.999]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:
Second Row: (Power Inc.) [95% LCL and UCL Power] (Alpha Inc.) [95% LCL and UCL Alpha]
Number of Monte Carlo Samples: 1000. Simulation Run Time: 2.99 minutes.

Reasonable power is achieved by this test for alternative means larger than 180. The accuracy of these results, of course, depends on the assumption that the data are normally distributed.
Comparative Results for Power of Various Tests

<table>
<thead>
<tr>
<th>N</th>
<th>H0 Mean (Mean0)</th>
<th>H1 Mean (Mean1)</th>
<th>Target Alpha</th>
<th>T-Test Power</th>
<th>Wilcoxon Power</th>
<th>Sign Power</th>
<th>Bootstrap Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>155.0</td>
<td>160.0</td>
<td>0.050</td>
<td>0.105</td>
<td>0.097</td>
<td>0.078</td>
<td>0.108</td>
</tr>
<tr>
<td>20</td>
<td>155.0</td>
<td>170.0</td>
<td>0.050</td>
<td>0.472</td>
<td>0.439</td>
<td>0.312</td>
<td>0.453</td>
</tr>
<tr>
<td>20</td>
<td>155.0</td>
<td>180.0</td>
<td>0.050</td>
<td>0.903</td>
<td>0.882</td>
<td>0.697</td>
<td>0.872</td>
</tr>
<tr>
<td>20</td>
<td>155.0</td>
<td>190.0</td>
<td>0.050</td>
<td>0.997</td>
<td>0.991</td>
<td>0.935</td>
<td>0.994</td>
</tr>
</tbody>
</table>

Number of Monte Carlo Iterations: 1000. Simulation Run Time: 2.99 minutes.

It is apparent from these results that the bootstrap performs about as well as the t-test and nonparametric tests for this design.
Example 9 – Comparison of Tests for Exponential Data

A researcher is designing an experiment. She believes that the data will follow an exponential distribution. Consequently, she does not believe that the t-test will be useful for her situation. She would like to compare several possible tests to determine which would be best for analyzing exponential data. She is interested in determining the power when the alternative mean is twice the null mean, which is 10. She wants to find the power achieved for sample sizes ranging from 20 to 60 with alpha = 0.05.

The number of simulations will be set at 1000 to expedite the analysis. Greater accuracy could be achieved by setting this number higher. This example will still take a few minutes to run because the bootstrap is included in the report.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for One Mean (Simulation) procedure window by expanding Means, then One Mean, then clicking on Test (Inequality), and then clicking on Tests for One Mean (Simulation). You may then make the appropriate entries as listed below, or open Example 9 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Tab</strong></td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>20 to 60 by 20</td>
</tr>
<tr>
<td>Distribution</td>
<td>H0 (Null Hypothesis)</td>
</tr>
<tr>
<td>Distribution</td>
<td>H1 (Alt Hypothesis)</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H1)</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Mean ≠ M0</td>
</tr>
<tr>
<td>Test Type</td>
<td>T-Test</td>
</tr>
<tr>
<td>Simulations</td>
<td>1000</td>
</tr>
<tr>
<td><strong>Reports Tab</strong></td>
<td></td>
</tr>
<tr>
<td>Show Comparative Reports</td>
<td>Checked</td>
</tr>
<tr>
<td>Show Comparative Plots</td>
<td>Checked</td>
</tr>
<tr>
<td>Include T-Test Results</td>
<td>Checked</td>
</tr>
<tr>
<td>Include Wilcoxon &amp; Sign Test</td>
<td>Checked</td>
</tr>
<tr>
<td>Include Bootstrap Test Results</td>
<td>Checked</td>
</tr>
<tr>
<td>Include Exponential Test Results</td>
<td>Checked</td>
</tr>
<tr>
<td><strong>Iterations Tab</strong></td>
<td></td>
</tr>
<tr>
<td>Bootstrap Iterations</td>
<td>100</td>
</tr>
</tbody>
</table>
Output

Click the Run button to perform the calculations and generate the following output.

**Numeric Results and Plots**

<table>
<thead>
<tr>
<th>N</th>
<th>H0 Mean (Mean0)</th>
<th>H1 Mean (Mean1)</th>
<th>Target Alpha</th>
<th>T-Test Power</th>
<th>Wilcoxon Power</th>
<th>Sign Power</th>
<th>Bootstrap Power</th>
<th>Expo Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>10.0</td>
<td>20.0</td>
<td>0.050</td>
<td>0.626</td>
<td>0.445</td>
<td>0.125</td>
<td>0.427</td>
<td>0.886</td>
</tr>
<tr>
<td>40</td>
<td>10.0</td>
<td>20.0</td>
<td>0.050</td>
<td>0.964</td>
<td>0.790</td>
<td>0.254</td>
<td>0.875</td>
<td>0.989</td>
</tr>
<tr>
<td>60</td>
<td>10.0</td>
<td>20.0</td>
<td>0.050</td>
<td>0.996</td>
<td>0.898</td>
<td>0.271</td>
<td>0.983</td>
<td>0.999</td>
</tr>
</tbody>
</table>

Number of Monte Carlo Iterations: 1000.  Simulation Run Time: 2.43 minutes.

<table>
<thead>
<tr>
<th>N</th>
<th>H0 Mean (Mean0)</th>
<th>H1 Mean (Mean1)</th>
<th>Target Alpha</th>
<th>T-Test Alpha</th>
<th>Wilcoxon Alpha</th>
<th>Sign Alpha</th>
<th>Bootstrap Alpha</th>
<th>Expo Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>20</td>
<td>10.0</td>
<td>20.0</td>
<td>0.050</td>
<td>0.094</td>
<td>0.130</td>
<td>0.202</td>
<td>0.074</td>
<td>0.048</td>
</tr>
<tr>
<td>40</td>
<td>10.0</td>
<td>20.0</td>
<td>0.050</td>
<td>0.057</td>
<td>0.172</td>
<td>0.342</td>
<td>0.046</td>
<td>0.044</td>
</tr>
<tr>
<td>60</td>
<td>10.0</td>
<td>20.0</td>
<td>0.050</td>
<td>0.060</td>
<td>0.268</td>
<td>0.489</td>
<td>0.049</td>
<td>0.049</td>
</tr>
</tbody>
</table>

Number of Monte Carlo Iterations: 1000.  Simulation Run Time: 2.43 minutes.

As would be expected for exponential data, the exponential test performs the best. The bootstrap test performs nearly as well for larger sample sizes. The other tests fail to achieve the target alpha level. Note that these simulation results will vary from run to run because the samples generated are random. The researcher must now decide which test to use based on her level of confidence in the data being truly exponentially distributed and the size of a sample she can afford to take.
Chapter 412

Inequality Tests for One Poisson Mean

Introduction

The Poisson probability law gives the probability distribution of the number of events occurring in a specified interval of time or space. The Poisson distribution is often used to fit count data, such as the number of defects on an item, the number of accidents at an intersection during a year, the number of calls to a call center during an hour, or the number of meteors seen in the evening sky during an hour.

The Poisson distribution is characterized by a single parameter, \( \lambda \), which is the mean number of occurrences during the interval. This procedure calculates the power or sample size for testing whether \( \lambda \) is less than or greater than a specified value. This test is usually called the test of the Poisson mean.

The test is described in Ostle (1988) and the power calculation is given in Guenther (1977).

Test Procedure

Assume that the mean is \( \lambda_0 \). To test \( H_0 : \lambda \leq \lambda_0 \) vs. \( H_a : \lambda > \lambda_0 \), you would take the following steps.

1. Find the critical value. Choose the critical value \( X^* \) so that the probability of rejecting \( H_0 \) when it is true is equal to \( \alpha \). This is done by solving the following inequality for \( X^* \).

\[
\sum_{x=X^*}^{\infty} e^{-\lambda_0} \frac{(n\lambda_0)^x}{x!} \leq \alpha.
\]

Note that because \( X \) is an integer, equality will seldom occur. Therefore, the minimum value of \( X^* \) is found for which the inequality holds.

2. Select a sample of \( n \) items compute the total number of events \( X = \sum_{i=1}^{n} x_i \). If \( X > X^* \) reject \( H_0 \) in favor of \( H_a \).
The test in the other direction \((H_0: \lambda \geq \lambda_0 \text{ vs. } H_a: \lambda < \lambda_0)\) is computed similarly.

### Assumptions

The assumptions of the one-sample Poisson test are:

1. The data are counts (discrete) that follow the Poisson distribution.
2. The sample is a simple random sample from its population. Each individual in the population has an equal probability of being selected in the sample.

### Limitations

There are few limitations when using these tests. As long as the assumption that the mean occurrence rate is constant is met, the test is valid.

### Technical Details

#### Computing Power

The power is computing for a specific alternative value \(\lambda_1\) using the following formula.

\[
\text{Power} = 1 - \beta = \sum_{x=X}^{\infty} e^{-n\lambda_1} \frac{(n\lambda_1)^x}{x!}
\]

#### Computing Sample Size

Following Guenther (1977), the sample size, \(n\), is found by increasing the value of \(d\) in the following expression until the left-hand end-point is less than the right-hand end-point and the interval contains at least one integer.

\[
\frac{X_{\chi^2,1-\beta}^2}{2\lambda_1} \leq n \leq \frac{X_{\chi^2,\alpha}^2}{2\lambda_0}, \quad d = 1, 2, 3, \ldots
\]

Here \(X_{\chi^2,\nu}^2\) is a percentage point of the chi-square distribution with \(\nu\) degrees of freedom.

### Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

### Data Tab

The Data tab contains most of the parameters and options that you will be concerned with.
Solve For

Find (Solve For)
This option specifies the parameter to be calculated from the values of the other parameters. Under most conditions, you would select either Power and Beta or n.

Select n when you want to determine the sample size needed to achieve a given power and alpha error level.

Select Power and Beta when you want to calculate the power of an experiment that has already been run.

Error Rates

Power or Beta
This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

Values must be between zero and one. Historically, the value of 0.80 (beta = 0.20) was used for power. Now, 0.90 (beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

Alpha (Significance Level)
This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

Values must be between zero and one. For one-sided tests such as this, the value of 0.025 is recommended for alpha. You may enter a range of values such as 0.025 0.05 0.10 or 0.025 to 0.05 by 0.005.

Sample Size

n (Sample Size)
This option specifies one or more values of the sample size, the number of individuals in the study. This value must be an integer greater than one. Note that you may enter a list of values using the syntax 50,100,150,200,250 or 50 to 250 by 50.

Effect Size – Means

λ₀ (Null or Baseline Mean)
This option specifies one or more values of the mean occurrence rate corresponding to the null hypothesis. This value must be greater than zero.

λ₁ (Alternative Mean)
This option specifies one or more values of the mean occurrence rate corresponding to the alternative hypothesis. This value must be greater than zero.
Test

Alternative Hypothesis
This option specifies the alternative hypothesis. This implicitly specifies the direction of the hypothesis test. The null hypothesis is always $H_0: \lambda_1 = \lambda_0$.

Possible selections for the alternative hypothesis are:

1. $H_a: \lambda_1 \leq \lambda_0$. This option yields a one-tailed t test.

2. $H_a: \lambda_1 \geq \lambda_0$. This option yields a one-tailed t test.

Example 1 – Power after a Study
This example demonstrates how to calculate the power for specific values of the other parameters. Suppose that accidents have occurred at an intersection at an average rate of 1 per month for the last several years. Recently, a distraction has been constructed near the intersection that appears to have increased the accident rate. Suppose the sample sizes are 12 and 24 months and alpha is 0.025. What is the power to detect alternatives of 1.1, 1.5, 2.0, and 2.5?

Setup
This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for One Poisson Mean procedure window by expanding Means, then One Mean, then clicking on Test (Inequality), and then clicking on Tests for One Poisson Mean. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.025</td>
</tr>
<tr>
<td>n (Sample Size)</td>
<td>12 24</td>
</tr>
<tr>
<td>$\lambda_0$ (Null or Baseline)</td>
<td>1.0</td>
</tr>
<tr>
<td>$\lambda_1$ (Alternative)</td>
<td>1.1 1.4 1.8 2.2 2.5</td>
</tr>
<tr>
<td>Ha (Alternative Hypothesis)</td>
<td>Ha: $\lambda_0 &lt; \lambda_1$</td>
</tr>
</tbody>
</table>
Annotated Output

Click the Run button to perform the calculations and generate the following output.

**Numeric Results**

<table>
<thead>
<tr>
<th>Power</th>
<th>n</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>$\lambda_0$</th>
<th>$\lambda_1$</th>
<th>Diff ($\lambda_0$-$\lambda_1$)</th>
<th>Effect Size</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0484</td>
<td>12</td>
<td>0.0250</td>
<td>0.0213</td>
<td>1.00</td>
<td>1.10</td>
<td>-0.10</td>
<td>0.0953</td>
<td>0.9516</td>
</tr>
<tr>
<td>0.0623</td>
<td>24</td>
<td>0.0250</td>
<td>0.0206</td>
<td>1.00</td>
<td>1.10</td>
<td>-0.10</td>
<td>0.0953</td>
<td>0.9377</td>
</tr>
<tr>
<td>0.2476</td>
<td>12</td>
<td>0.0250</td>
<td>0.0213</td>
<td>1.00</td>
<td>1.40</td>
<td>0.40</td>
<td>0.3381</td>
<td>0.7524</td>
</tr>
<tr>
<td>0.4273</td>
<td>24</td>
<td>0.0250</td>
<td>0.0206</td>
<td>1.00</td>
<td>1.40</td>
<td>0.40</td>
<td>0.3381</td>
<td>0.5727</td>
</tr>
<tr>
<td>0.6638</td>
<td>12</td>
<td>0.0250</td>
<td>0.0213</td>
<td>1.00</td>
<td>1.80</td>
<td>-0.80</td>
<td>0.5963</td>
<td>0.3362</td>
</tr>
<tr>
<td>0.9108</td>
<td>24</td>
<td>0.0250</td>
<td>0.0206</td>
<td>1.00</td>
<td>1.80</td>
<td>-0.80</td>
<td>0.5963</td>
<td>0.0892</td>
</tr>
<tr>
<td>0.9154</td>
<td>12</td>
<td>0.0250</td>
<td>0.0213</td>
<td>1.00</td>
<td>2.20</td>
<td>-1.20</td>
<td>0.8090</td>
<td>0.0846</td>
</tr>
<tr>
<td>0.9962</td>
<td>24</td>
<td>0.0250</td>
<td>0.0206</td>
<td>1.00</td>
<td>2.20</td>
<td>-1.20</td>
<td>0.8090</td>
<td>0.0038</td>
</tr>
<tr>
<td>0.9781</td>
<td>12</td>
<td>0.0250</td>
<td>0.0213</td>
<td>1.00</td>
<td>2.50</td>
<td>-1.50</td>
<td>0.9487</td>
<td>0.0219</td>
</tr>
<tr>
<td>0.9998</td>
<td>24</td>
<td>0.0250</td>
<td>0.0206</td>
<td>1.00</td>
<td>2.50</td>
<td>-1.50</td>
<td>0.9487</td>
<td>0.0002</td>
</tr>
</tbody>
</table>

**References**


**Report Definitions**

- **Power** is the probability of rejecting a false null hypothesis. It should be close to one.
- **n** is the size of the sample drawn from the population. To conserve resources, it should be small.
- **Alpha** is the probability of rejecting a true null hypothesis. It should be small.
- **Diff** is the value of $\lambda_0$ - $\lambda_1$, the difference being tested.
- $\lambda_0$ is the value of the population mean under the null hypothesis.
- $\lambda_1$ is the value of the population mean under the alternative hypothesis.
- **Effect Size** is the value of $(\lambda_0 - \lambda_1) / \sqrt{\lambda_1}$.
- **Beta** is the probability of accepting a false null hypothesis. It should be small.

**Summary Statements**

A sample size of 12 achieves 5% power to detect a difference of -0.10 between the null hypothesis mean of 1.00 and the alternative hypothesis mean of 1.10 and with a significance level (alpha) of 0.0250 using a one-sided one-sample Poisson test.

This report shows the values of each of the parameters, one scenario per row. The values of power and beta were calculated from the other parameters.

Note that the actual power achieved is greater than the target power. Similarly, the actual alpha is less than the target alpha. These differences occur because only integer values of the count variable occur.
Plots Section

This plot shows the relationship between sample size and power for various values of the alternative mean and the sample size.
Inequality Tests for One Poisson Mean  412-7

Example 2 – Finding the Sample Size

This example will extend Example 1 to the case in which we want to find the necessary sample size to achieve at least 90% power. This is done as follows.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for One Poisson Mean procedure window by expanding Means, then One Mean, then clicking on Test (Inequality), and then clicking on Tests for One Poisson Mean. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>n (sample size)</td>
</tr>
<tr>
<td>Power</td>
<td>0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.025</td>
</tr>
<tr>
<td>n (Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>λ0 (Null or Baseline)</td>
<td>1.0</td>
</tr>
<tr>
<td>λ1 (Alternative)</td>
<td>1.1 1.4 1.8 2.2 2.5</td>
</tr>
<tr>
<td>Ha (Alternative Hypothesis)</td>
<td>Ha: λ0 &lt; λ1</td>
</tr>
</tbody>
</table>

Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>λ0</th>
<th>λ1</th>
<th>Diff (λ0-λ1)</th>
<th>Effect Size</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9002</td>
<td>0.0250</td>
<td>0.0250</td>
<td>1.00</td>
<td>1.10</td>
<td>-0.10</td>
<td>0.0953</td>
<td>0.0998</td>
</tr>
<tr>
<td>0.9009</td>
<td>0.0250</td>
<td>0.0220</td>
<td>1.00</td>
<td>1.40</td>
<td>-0.40</td>
<td>0.3381</td>
<td>0.0991</td>
</tr>
<tr>
<td>0.9154</td>
<td>0.0250</td>
<td>0.0206</td>
<td>1.00</td>
<td>1.80</td>
<td>-0.80</td>
<td>0.5963</td>
<td>0.0892</td>
</tr>
<tr>
<td>0.9366</td>
<td>0.0250</td>
<td>0.0213</td>
<td>1.00</td>
<td>2.20</td>
<td>-1.20</td>
<td>0.8090</td>
<td>0.0846</td>
</tr>
<tr>
<td>0.9366</td>
<td>0.0250</td>
<td>0.0220</td>
<td>1.00</td>
<td>2.50</td>
<td>-1.50</td>
<td>0.9487</td>
<td>0.0634</td>
</tr>
</tbody>
</table>

This report shows the sample sizes that are necessary to achieve the required power.
Example 3 – Finding the Minimum Detectable Difference

Continuing with the previous example, suppose only 10 months of data are available. What is the minimum detectable difference that can be detected by this design?

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for One Poisson Mean procedure window by expanding Means, then One Mean, then clicking on Test (Inequality), and then clicking on Tests for One Poisson Mean. You may then make the appropriate entries as listed below, or open Example 3 by going to the File menu and choosing Open Example Template.

Option | Value
--- | ---
Data Tab | 
Find (Solve For) | $\lambda_1$
Power | 0.90
Alpha | 0.025
n (Sample Size) | 10
$\lambda_0$ (Null or Baseline) | 1.0
$\lambda_1$ (Alternative) | Ignored since this is the Find setting
Ha (Alternative Hypothesis) | $H_a: \lambda_0 < \lambda_1$

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>n</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>$\lambda_0$</th>
<th>$\lambda_1$</th>
<th>Diff ((\lambda_0 - \lambda_1))</th>
<th>Effect Size</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9000</td>
<td>10</td>
<td>0.0250</td>
<td>0.0143</td>
<td>1.00</td>
<td>2.36</td>
<td>-1.36</td>
<td>0.8856</td>
<td>0.1000</td>
</tr>
</tbody>
</table>

This report shows that the minimum detectable difference is $2.36 - 1.00 = 1.36$. 
Example 4 – Validation using Guenther

Guenther (1977) page 27 gives an example in which \( \lambda_0 = 0.05, \lambda_1 = .2, \alpha = 0.05, \beta = 0.10, \) and \( n = 47. \) We will now run this example.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for One Poisson Mean procedure window by expanding Means, then One Mean, then clicking on Test (Inequality), and then clicking on Tests for One Poisson Mean. You may then make the appropriate entries as listed below, or open Example 4 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab Find (Solve For)</td>
<td>( n ) (sample size)</td>
</tr>
<tr>
<td>Power</td>
<td>0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.025</td>
</tr>
<tr>
<td>( n ) (Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>( \lambda_0 ) (Null or Baseline)</td>
<td>0.05</td>
</tr>
<tr>
<td>( \lambda_1 ) (Alternative)</td>
<td>0.20</td>
</tr>
<tr>
<td>Ha (Alternative Hypothesis)</td>
<td>Ha: ( \lambda_0 &lt; \lambda_1 )</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

**Numeric Results**

<table>
<thead>
<tr>
<th>Power</th>
<th>( n )</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>( \lambda_0 )</th>
<th>( \lambda_1 )</th>
<th>Diff (( \lambda_0 - \lambda_1 ))</th>
<th>Effect Size</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9065</td>
<td>47</td>
<td>0.0500</td>
<td>0.0327</td>
<td>0.05</td>
<td>0.20</td>
<td>-0.15</td>
<td>0.3354</td>
<td>0.0935</td>
</tr>
</tbody>
</table>

Note that the value of \( n \) is indeed 47.
412-10 Inequality Tests for One Poisson Mean
Chapter 413

Tests for One Mean with Non-Zero Null Hypothesis

Introduction

This module computes power and sample size for tests in one-sample designs with a superiority margin in which the outcome is distributed as a normal random variable. This includes the analysis of the differences between paired values.

The details of sample size calculation for the one-sample design are presented in the Inequality Tests for One Mean chapter and they will not be duplicated here. This chapter only discusses those changes necessary for non-zero null tests. Sample size formulas for non-inferiority and non-zero null hypothesis tests of a single mean are presented in Chow et al. (2003) page 50.

The one-sample t-test is used to test whether a population mean is different from a specific value. When the data are differences between paired values, this test is known as the paired t-test. This module also calculates the power of the non-parametric analog of the t-test, the Wilcoxon test.

Paired Designs

Paired data may occur because two measurements are made on the same subject or because measurements are made on two subjects that have been matched according to other variables. Hypothesis tests on paired data can be analyzed by considering the difference between the paired items as the response. The distribution of differences is usually symmetric. In fact, the distribution must be symmetric if the individual distributions of the two items are identical. Hence, the paired t-test and the Wilcoxon signed-rank test are appropriate for paired data even when the distributions of the individual items are not normal.

In paired designs, the variable of interest is the difference between two individual measurements. Although the non-inferiority hypothesis refers to the difference between two individual means, the actual values of those means are not needed. All that is needed is their difference.
The Statistical Hypotheses

Both non-inferiority and superiority tests are examples of directional (one-sided) tests and their power and sample size could be calculated using the One-Sample T-Test procedure. However, at the urging of our users, we have developed this module which provides the input and output options that are convenient for non-zero null hypothesis tests. This section will review the specifics of non-zero null testing.

Remember that in the usual t-test setting, the null (H0) and alternative (H1) hypotheses for one-sided tests are defined as

\[ H_0: \mu_X \leq A \text{ versus } H_1: \mu_X > A \]

Rejecting H0 implies that the mean is larger than the value A. This test is called an upper-tail test because H0 is rejected in samples in which the sample mean is larger than A.

Following is an example of a lower-tail test.

\[ H_0: \mu_X \geq A \text{ versus } H_1: \mu_X < A \]

Non-zero null tests are special cases of the above directional tests. It will be convenient to adopt the following specialize notation for the discussion of these tests.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PASS Input/Output</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu_T )</td>
<td>Not used</td>
<td>Population mean. If the data are paired differences, this is the mean of those differences.</td>
</tr>
<tr>
<td>( \mu_R )</td>
<td>Not used</td>
<td>Reference value. Usually, this is the mean of a reference population. If the data are paired differences, this is the hypothesized value of the mean difference.</td>
</tr>
<tr>
<td>( M_s )</td>
<td>SM</td>
<td>Margin of superiority. This is a tolerance value that defines the magnitude of difference that is required for practical importance. This may be thought of as the smallest difference from the reference value that is considered to be different.</td>
</tr>
<tr>
<td>( \delta )</td>
<td>D</td>
<td>True difference. This is the value of ( \mu_T - \mu_R ), the difference between the mean and the reference value, at which the power is calculated.</td>
</tr>
</tbody>
</table>

Note that the actual values of \( \mu_T \) and \( \mu_R \) are not needed. Only their difference is needed for power and sample size calculations.
Non-Zero Null Tests

A non-zero null test tests that the mean is better than that of the baseline (reference) population by more than a small superiority margin. The actual direction of the hypothesis depends on the whether higher values of the response are good or bad.

Case 1: High Values Good

In this case, higher values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the mean is greater than the reference value by at least the margin of superiority. The value of $\delta$ must be greater than $|M_s|$. Equivalent sets of the null and alternative hypotheses are

$$H_0 : \mu_T \leq \mu_R + |M_s| \quad \text{versus} \quad H_1 : \mu_T > \mu_R + |M_s|$$

$$H_0 : \mu_T - \mu_R \leq |M_s| \quad \text{versus} \quad H_1 : \mu_T - \mu_R > |M_s|$$

$$H_0 : \delta \leq |M_s| \quad \text{versus} \quad H_1 : \delta > |M_s|$$

Case 2: High Values Bad

In this case, lower values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the mean is less than the reference value by at least the margin of superiority. The value of $\delta$ must be less than $-|M_s|$. Equivalent sets of the null and alternative hypotheses are

$$H_0 : \mu_T \geq \mu_R - |M_s| \quad \text{versus} \quad H_1 : \mu_T < \mu_R - |M_s|$$

$$H_0 : \mu_T - \mu_R \geq -|M_s| \quad \text{versus} \quad H_1 : \mu_T - \mu_R < -|M_s|$$

$$H_0 : \delta \geq -|M_s| \quad \text{versus} \quad H_1 : \delta < -|M_s|$$

Example

A non-zero null test example will set the stage for the discussion of the terminology that follows. Suppose that a test is to be conducted to determine if a new cancer treatment substantially improves the mean bone density. The adjusted mean bone density (AMBD) in the population of interest is 0.002300 gm/cm with a standard deviation of 0.000300 gm/cm. Clinicians decide that if the treatment increases AMBD by more than 5% (0.000115 gm/cm), it provides a significant health benefit.

The hypothesis of interest is whether the AMBD in the treated group is greater than $0.002300 + 0.000115 = 0.002415$. The statistical test will be set up so that if the null hypothesis that the AMBD is greater than or equal to 0.002415 is rejected, the conclusion will be that the new treatment is superior, at least in terms of AMBD. The value 0.000115 gm/cm is called the margin of superiority.

Test Statistics

This section describes the test statistics that are available in this procedure.
413-4 Tests for One Mean with Non-Zero Null Hypothesis

One-Sample T-Test

The one-sample t-test assumes that the data are a simple random sample from a population of normally-distributed values that all have the same mean and variance. This assumption implies that the data are continuous and their distribution is symmetric. The calculation of the t-test proceeds as follow

\[ t_{n-1} = \frac{\bar{X} - D_0}{s_X/N} \]

where

\[ \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}, \]
\[ s_X = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}}, \]

and \( D_0 \) is the value of the mean hypothesized by the null hypothesis.

The significance of the test statistic is determined by computing the p-value. If this p-value is less than a specified level (usually 0.05), the hypothesis is rejected. Otherwise, no conclusion can be reached.

Wilcoxon Signed-Rank Test

The Wilcoxon signed-rank test is a popular, nonparametric substitute for the t-test. It assumes that the data follow a symmetric distribution. The test is computed using the following steps.

1. Subtract the hypothesized mean, \( D_0 \), from each data value. Rank the values according to their absolute values.
2. Compute the sum of the positive ranks \( S_p \) and the sum of the negative ranks \( S_n \). The test statistic, \( W \), is the minimum of \( S_p \) and \( S_n \).
3. Compute the mean and standard deviation of \( W \) using the formulas

\[ \mu_{w} = \frac{n(n+1)}{4} \]
\[ \sigma_{w} = \sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{\sum t^2 - \sum t}{48}} \]

where \( t \) represents the number of times the \( i \)th value occurs.
4. Compute the \( z \) value using

\[ z_{w} = \frac{W - \mu_{w}}{\sigma_{w}} \]

The significance of the test statistic is determined by computing the p-value using the standard normal distribution. If this p-value is less than a specified level (usually 0.05), the null hypothesis is rejected in favor of the alternative hypothesis. Otherwise, no conclusion can be reached.
Computing the Power

The power is calculated as follows for a directional alternative (one-tailed test) in which \( D1 > D0 \). \( D1 \) is the value of the mean at which the power is computed.

1. Find \( t_\alpha \) such that \( 1 - T_{n-1}(t_\alpha) = \alpha \), where \( T_{n-1}(t_\alpha) \) is the area to the left of \( x \) under a central-t curve with \( n - 1 \) degrees of freedom.

2. Calculate \( x = D0 + t_\alpha \frac{\sigma}{\sqrt{n}}. \)

3. Calculate the noncentrality parameter \( \lambda = \frac{D1 - D0}{\frac{\sigma}{\sqrt{n}}} \).

4. Calculate \( t' = \frac{x - D1}{\frac{\sigma}{\sqrt{n}}} + \lambda. \)

5. Calculate the power = \( 1 - T'_{n-1,\lambda}(t') \), where \( T'_{n-1,\lambda}(x) \) is the area to the left of \( x \) under a noncentral-t curve with degrees of freedom \( n - 1 \) and noncentrality parameter \( \lambda. \).

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains most of the parameters and options that will be of interest.

Solve For

Find (Solve For)

This option specifies the parameter to be calculated from the values of the other parameters. Under most conditions, you would select either Power and Beta or \( N \).

Select \( N \) when you want to determine the sample size needed to achieve a given power and alpha error level.

Select Power and Beta when you want to calculate the power.
Error Rates

Power or Beta
This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used. A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

Alpha (Significance Level)
This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation. You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.

Sample Size

N (Sample Size)
This option specifies one or more values of the sample size, the number of individuals in the study. This value must be an integer greater than one. You may enter a list of values using the syntax 50 100 150 200 250 or 50 to 250 by 50.

Effect Size – Mean Difference

SM (Superiority Margin)
This is the magnitude of the margin of superiority. It must be entered as a positive number. When higher means are better, this value is the distance above the reference value that is required to be considered superior. When higher means are worse, this value is the distance below the reference value that is required to be considered superior.

D (True Difference)
This is the actual difference between the mean and the reference value. When higher means are better, this value should be greater than SM. When higher means are worse, this value should be negative and greater in magnitude than SM.

Effect Size – Standard Deviation

Standard Deviation
This option specifies one or more values of the standard deviation. This must be a positive value. PASS includes a special module for estimating the standard deviation. This module may be
Tests for One Mean with Non-Zero Null Hypothesis

Test

Higher Means Are
This option defines whether higher values of the response variable are to be considered better or worse. The choice here determines the direction of the test.

If Higher Means Are Better the null hypothesis is \( \text{Diff} \leq \text{SM} \) and the alternative hypothesis is \( \text{Diff} > \text{SM} \). If Higher Means Are Worse the null hypothesis is \( \text{Diff} \geq -\text{SM} \) and the alternative hypothesis is \( \text{Diff} < -\text{SM} \).

Nonparametric Adjustment
This option makes appropriate sample size adjustments for the Wilcoxon test. Results by Al-Sunduqchi and Guenther (1990) indicate that power calculations for the Wilcoxon test may be made using the standard \( t \) test formulations with a simple adjustment to the sample size. The size of the adjustment depends upon the actual distribution of the data. They give sample size adjustment factors for four distributions. These are 1 for the uniform distribution, \( \frac{2}{3} \) for the double exponential distribution, \( \frac{9}{\pi^2} \) for the logistic distribution, and \( \frac{\pi}{3} \) for the normal distribution.

The options are as follows:

- **Ignore**
  Do not make a Wilcoxon adjustment. This indicates that you want to analyze a \( t \) test, not the Wilcoxon test.

- **Uniform**
  Make the Wilcoxon sample size adjustment assuming the uniform distribution. Since the factor is one, this option performs the same as Ignore. It is included for completeness.

- **Double Exponential**
  Make the Wilcoxon sample size adjustment assuming that the data actually follow the double exponential distribution.

- **Logistic**
  Make the Wilcoxon sample size adjustment assuming that the data actually follow the logistic distribution.

- **Normal**
  Make the Wilcoxon sample size adjustment assuming that the data actually follow the normal distribution.

Population Size
This is the number of subjects in the population. Usually, you assume that samples are drawn from a very large (infinite) population. Occasionally, however, situations arise in which the population of interest is of limited size. In these cases, appropriate adjustments must be made.

When a finite population size is specified, the standard deviation is reduced according to the formula:

\[
\sigma_{\text{p}} = \frac{\sigma}{\sqrt{N}}
\]
\[ \sigma_1^2 = \left(1 - \frac{n}{N}\right)\sigma^2 \]

where \( n \) is the sample size, \( N \) is the population size, \( \sigma \) is the original standard deviation, and \( \sigma_1 \) is the new standard deviation.

The quantity \( n/N \) is often called the sampling fraction. The quantity \( 1 - \frac{n}{N} \) is called the finite population correction factor.

**Example 1 – Power Analysis**

Suppose that a test is to be conducted to determine if a new cancer treatment improves the mean bone density. The adjusted mean bone density (AMBD) in the population of interest is 0.002300 gm/cm with a standard deviation of 0.000300 gm/cm. Clinicians decide that if the treatment increases AMBD by more than 5% (0.000115 gm/cm), it generates a significant health benefit. They also want to consider what would happen if the margin of superiority is set to 2.5% (0.0000575 gm/cm).

The analysis will be a non-zero null test using the t-test at the 0.025 significance level. Power is to be calculated assuming that the new treatment has 7.5% improvement on AMBD. Several sample sizes between 20 and 300 will be analyzed. The researchers want to achieve a power of at least 90%. All numbers have been multiplied by 10000 to make the reports and plots easier to read.

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for One Mean with Non-Zero Null procedure window by expanding Means, then One Mean, then clicking on Test (Non-Zero Null), and then clicking on Tests for One Mean with Non-Zero Null. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Power</td>
<td></td>
</tr>
<tr>
<td>Alpha</td>
<td>0.025</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>20 40 60 80 100 150 200 300</td>
</tr>
<tr>
<td>SM (Superiority Margin)</td>
<td>0.575 1.15</td>
</tr>
<tr>
<td>D (True Difference)</td>
<td>1.725</td>
</tr>
<tr>
<td>S (Standard Deviation)</td>
<td>3</td>
</tr>
<tr>
<td>Higher Means Are</td>
<td>Better</td>
</tr>
<tr>
<td>Nonparametric Adjustment</td>
<td>Ignore</td>
</tr>
<tr>
<td>Population Size</td>
<td>Infinite</td>
</tr>
</tbody>
</table>
Tests for One Mean with Non-Zero Null Hypothesis

Annotated Output
Click the Run button to perform the calculations and generate the following output.

Numeric Results for Superiority Test (H0: Diff <= SM; H1: Diff > SM)
Higher Means are Better
Test Statistic: T-Test

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>Superiority Margin (SM)</th>
<th>Actual Difference (D)</th>
<th>Significance Level (Alpha)</th>
<th>Standard Deviation (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.36990</td>
<td>20</td>
<td>0.575</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.63010</td>
</tr>
<tr>
<td>0.65705</td>
<td>40</td>
<td>0.575</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.34295</td>
</tr>
<tr>
<td>0.83164</td>
<td>60</td>
<td>0.575</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.16836</td>
</tr>
<tr>
<td>0.92317</td>
<td>80</td>
<td>0.575</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.07683</td>
</tr>
<tr>
<td>0.96682</td>
<td>100</td>
<td>0.575</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.03318</td>
</tr>
<tr>
<td>0.99658</td>
<td>150</td>
<td>0.575</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.00342</td>
</tr>
<tr>
<td>0.99970</td>
<td>200</td>
<td>0.575</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.00030</td>
</tr>
<tr>
<td>1.00000</td>
<td>300</td>
<td>0.575</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.00000</td>
</tr>
<tr>
<td>0.12601</td>
<td>20</td>
<td>1.150</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.87399</td>
</tr>
<tr>
<td>0.21844</td>
<td>40</td>
<td>1.150</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.78156</td>
</tr>
<tr>
<td>0.30873</td>
<td>60</td>
<td>1.150</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.69127</td>
</tr>
<tr>
<td>0.39493</td>
<td>80</td>
<td>1.150</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.60507</td>
</tr>
<tr>
<td>0.47532</td>
<td>100</td>
<td>1.150</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.52468</td>
</tr>
<tr>
<td>0.64517</td>
<td>150</td>
<td>1.150</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.35483</td>
</tr>
<tr>
<td>0.76959</td>
<td>200</td>
<td>1.150</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.23041</td>
</tr>
<tr>
<td>0.91135</td>
<td>300</td>
<td>1.150</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.08865</td>
</tr>
</tbody>
</table>

Report Definitions
H0 (null hypothesis) is Diff <= SM, where Diff = Mean - Reference Value for the one-sample case, and Diff = Mean of Differences - Reference Value for the paired case.
H1 (alternative hypothesis) is Diff > SM.
Power is the probability of rejecting H0 when it is false.
N is the sample size, the number of subjects in the study.
SM is the magnitude of the margin of superiority. Since higher means are better, this value is positive and is the distance above the reference value that is required to be considered superior.
D is the mean difference (treatment - reference value) at which the power is computed.
Alpha is the probability of rejecting H0 when it is true, which is the probability of a false positive.
Beta is the probability of accepting H0 when it is false, which is the probability of a false negative.
S is the standard deviation of the response. It measures the variability in the population.

Summary Statements
A sample size of 20 achieves 37% power to detect superiority using a one-sided t-test when the margin of superiority is 0.575 and the true difference between the mean and the reference value is 1.725. The data are drawn from a single population with a standard deviation of 3.000. The significance level (alpha) of the test is 0.02500.

The above report shows that for SM = 1.15, the sample size necessary to obtain 90% power is just under 300. However, if SM = 0.575, the required sample size is only about 75.
Example 2 – Finding the Sample Size

Continuing with Example 1, the researchers want to know the exact sample size for each value of SM to achieve 90% power.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for One Mean with Non-Zero Null procedure window by expanding Means, then One Mean, then clicking on Test (Non-Zero Null), and then clicking on Tests for One Mean with Non-Zero Null. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N (Sample Size)</td>
</tr>
<tr>
<td>Power</td>
<td>0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.025</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>SM (Superiority Margin)</td>
<td>0.575 1.15</td>
</tr>
<tr>
<td>D (True Difference)</td>
<td>1.725</td>
</tr>
<tr>
<td>S (Standard Deviation)</td>
<td>3</td>
</tr>
<tr>
<td>Higher Means Are</td>
<td>Better</td>
</tr>
<tr>
<td>Nonparametric Adjustment</td>
<td>Ignore</td>
</tr>
<tr>
<td>Population Size</td>
<td>Infinite</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>Superiority Margin</th>
<th>Actual Difference</th>
<th>Significance Level</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.90215</td>
<td>74</td>
<td>0.575</td>
<td>1.725</td>
<td>0.02500</td>
<td>3.000</td>
</tr>
<tr>
<td>0.90005</td>
<td>288</td>
<td>1.150</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.09995</td>
</tr>
</tbody>
</table>

This report shows the exact sample size requirement for each value of SM.

Example 3 – Validation

This procedure uses the same mechanics as the Non-Inferiority Tests for One Mean procedure. We refer the user to Example 3 of Chapter 415 for the validation.
Chapter 415

Non-Inferiority Tests for One Mean

Introduction

This module computes power and sample size for non-inferiority tests in one-sample designs in which the outcome is distributed as a normal random variable. This includes the analysis of the differences between paired values.

The details of sample size calculation for the one-sample design are presented in the Inequality Tests for One Mean chapter and they will not be duplicated here. This chapter only discusses those changes necessary for non-inferiority tests. Sample size formulas for non-inferiority tests of a single mean are presented in Chow et al. (2003) page 50.

The one-sample t-test is used to test whether a population mean is different from a specific value. When the data are differences between paired values, this test is known as the paired t-test. This module also calculates the power of the nonparametric analog of the t-test, the Wilcoxon test.

Paired Designs

Paired data may occur because two measurements are made on the same subject or because measurements are made on two subjects that have been matched according to other variables. Hypothesis tests on paired data can be analyzed by considering the difference between the paired items as the response. The distribution of differences is usually symmetric. In fact, the distribution must be symmetric if the individual distributions of the two items are identical. Hence, the paired t-test and the Wilcoxon signed-rank test are appropriate for paired data even when the distributions of the individual items are not normal.

In paired designs, the variable of interest is the difference between two individual measurements. Although the non-inferiority hypothesis refers to the difference between two individual means, the actual values of those means are not needed. All that is needed is their difference.
The Statistical Hypotheses

Both non-inferiority and superiority tests are examples of directional (one-sided) tests and their power and sample size could be calculated using the One-Sample T-Test procedure. However, at the urging of our users, we have developed this module which provides the input and output options that are convenient for non-inferiority tests. This section will review the specifics of non-inferiority and superiority testing.

Remember that in the usual t-test setting, the null (H0) and alternative (H1) hypotheses for one-sided tests are defined as

\[ H_0: \mu_X \leq A \text{ versus } H_1: \mu_X > A \]

Rejecting H0 implies that the mean is larger than the value A. This test is called an upper-tail test because H0 is rejected in samples in which the sample mean is larger than A.

Following is an example of a lower-tail test.

\[ H_0: \mu_X \geq A \text{ versus } H_1: \mu_X < A \]

Non-inferiority tests are special cases of the above directional tests. It will be convenient to adopt the following specialize notation for the discussion of these tests.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PASS Input/Output</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu_T )</td>
<td>Not used</td>
<td>Population mean. If the data are paired differences, this is the mean of those differences.</td>
</tr>
<tr>
<td>( \mu_R )</td>
<td>Not used</td>
<td>Reference value. Usually, this is the mean of a reference population. If the data are paired differences, this is the hypothesized value of the mean difference.</td>
</tr>
<tr>
<td>( M_{NI} )</td>
<td>NIM</td>
<td>Margin of non-inferiority. This is a tolerance value that defines the magnitude of difference that is not of practical importance. This may be thought of as the largest difference from the reference value that is considered to be trivial. The absolute value symbols are used to emphasize that this is a magnitude. The sign is determined by the specific design.</td>
</tr>
<tr>
<td>( \delta )</td>
<td>D</td>
<td>True difference. This is the value of ( \mu_T - \mu_R ), the difference between the mean and the reference value, at which the power is calculated.</td>
</tr>
</tbody>
</table>

Note that the actual values of \( \mu_T \) and \( \mu_R \) are not needed. Only their difference is needed for power and sample size calculations.
Non-Inferiority Tests

A non-inferiority test tests that the mean is not worse than that of the baseline (reference) population by more than a small non-inferiority margin. The actual direction of the hypothesis depends on whether higher values of the response are good or bad.

Case 1: High Values Good

In this case, higher values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the mean of the treatment group is no less than a small amount below the reference value. The value of $\delta$ is often set to zero. Equivalent sets of the null and alternative hypotheses are

$$H_0: \mu_T \leq \mu_R - |M_{NI}| \quad \text{versus} \quad H_1: \mu_T > \mu_R - |M_{NI}|$$

$$H_0: \mu_T - \mu_R \leq -|M_{NI}| \quad \text{versus} \quad H_1: \mu_T - \mu_R > -|M_{NI}|$$

$$H_0: \delta \leq -|M_{NI}| \quad \text{versus} \quad H_1: \delta > -|M_{NI}|$$

Case 2: High Values Bad

In this case, lower values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the mean of the treatment group is no more than a small amount above the reference value. The value of $\delta$ is often set to zero. Equivalent sets of the null and alternative hypotheses are

$$H_0: \mu_T \geq \mu_R + |M_{NI}| \quad \text{versus} \quad H_1: \mu_T < \mu_R + |M_{NI}|$$

$$H_0: \mu_T - \mu_R \geq |M_{NI}| \quad \text{versus} \quad H_1: \mu_T - \mu_R < |M_{NI}|$$

$$H_0: \delta \geq |M_{NI}| \quad \text{versus} \quad H_1: \delta < |M_{NI}|$$

Example

A non-inferiority test example will set the stage for the discussion of the terminology that follows. Suppose that a test is to be conducted to determine if a new cancer treatment adversely affects the mean bone density. The adjusted mean bone density (AMBD) in the population of interest is 0.002300 gm/cm with a standard deviation of 0.000300 gm/cm. Clinicians decide that if the treatment reduces AMBD by more than 5% ($0.000115$ gm/cm), it poses a significant health threat.

The hypothesis of interest is whether the AMBD in the treated group is greater than $0.002300 - 0.000115 = 0.002185$. The statistical test will be set up so that if the null hypothesis that the AMBD is less than or equal to 0.002185 is rejected, the conclusion will be that the new treatment is non-inferior, at least in terms of AMBD. The value 0.000115 gm/cm is called the margin of non-inferiority.

Test Statistics

This section describes the test statistics that are available in this procedure.
One-Sample T-Test

The one-sample t-test assumes that the data are a simple random sample from a population of normally-distributed values that all have the same mean and variance. This assumption implies that the data are continuous and their distribution is symmetric. The calculation of the t-test proceeds as follow

\[ t_{n-1} = \frac{\bar{X} - D_0}{s_{\bar{X}}} \]

where

\[ \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}, \]

\[ s_{\bar{X}} = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}}, \]

and \( D_0 \) is the value of the mean hypothesized by the null hypothesis.

The significance of the test statistic is determined by computing the p-value. If this p-value is less than a specified level (usually 0.05), the hypothesis is rejected. Otherwise, no conclusion can be reached.

Wilcoxon Signed-Rank Test

The Wilcoxon signed-rank test is a popular, nonparametric substitute for the t-test. It assumes that the data follow a symmetric distribution. The test is computed using the following steps.

1. Subtract the hypothesized mean, \( D_0 \), from each data value. Rank the values according to their absolute values.
2. Compute the sum of the positive ranks \( S_P \) and the sum of the negative ranks \( S_n \). The test statistic, \( W \), is the minimum of \( S_P \) and \( S_n \).
3. Compute the mean and standard deviation of \( W \) using the formulas

\[ \mu_{w_s} = \frac{n(n+1)}{4} \quad \text{and} \quad \sigma^2_{w_s} = \sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{\sum t^2 - \sum t}{48}} \]

where \( t \) represents the number of times the \( i \)th value occurs.
4. Compute the \( z \) value using

\[ z_{w_s} = \frac{W - \mu_{w_s}}{\sigma_{w_s}} \]

The significance of the test statistic is determined by computing the p-value using the standard normal distribution. If this p-value is less than a specified level (usually 0.05), the null hypothesis is rejected in favor of the alternative hypothesis. Otherwise, no conclusion can be reached.
Computing the Power

The power is calculated as follows for a directional alternative (one-tailed test) in which $D1 > D0$. $D1$ is the value of the mean at which the power is computed.

1. Find $t_\alpha$ such that $1 - T_{n-1}(t_\alpha) = \alpha$, where $T_{n-1}(t_\alpha)$ is the area to the left of $x$ under a central-t curve with $n – 1$ degrees of freedom.

2. Calculate $x_\alpha = D0 + t_\alpha \frac{\sigma}{\sqrt{n}}$.

3. Calculate the noncentrality parameter $\lambda = \frac{D1 - D0}{\sigma \sqrt{n}}$.

4. Calculate $t_\alpha = \frac{x_\alpha - D1}{\sigma \sqrt{n}} + \lambda$.

5. Calculate the power $= 1 - T'_{n-1,\lambda}(t_\alpha)$, where $T'_{n-1,\lambda}(x)$ is the area to the left of $x$ under a noncentral-t curve with degrees of freedom $n – 1$ and noncentrality parameter $\lambda$.

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains most of the parameters and options that will be of interest.

Solve For

Find (Solve For)

This option specifies the parameter to be calculated from the values of the other parameters. Under most conditions, you would select either Power and Beta or $N$.

Select $N$ when you want to determine the sample size needed to achieve a given power and alpha error level.

Select Power and Beta when you want to calculate the power.
Non-Inferiority Tests for One Mean

**Error Rates**

**Power or Beta**
This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as **0.8 to 0.95 by 0.05** may be entered.

**Alpha (Significance Level)**
This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as **0.01 0.05 0.10** or **0.01 to 0.10 by 0.01**.

---

**Sample Size**

**N (Sample Size)**
This option specifies one or more values of the sample size, the number of individuals in the study. This value must be an integer greater than one. You may enter a list of values using the syntax **50 100 150 200 250** or **50 to 250 by 50**.

**Effect Size – Mean Difference**

**NIM (Non-Inferiority Margin)**
This is the magnitude of the margin of non-inferiority. It must be entered as a positive number.

When higher means are better, this value is the distance below the reference value that is still considered non-inferior. When higher means are worse, this value is the distance above the reference value that is still considered non-inferior.

**D (True Value)**
This is the actual difference between the mean and the reference value.

For non-inferiority tests, this value is often set to zero. When this value is non-zero, care should be taken that this value is consistent with whether higher means are better or worse.

**Effect Size – Standard Deviation**

**Standard Deviation**
This option specifies one or more values of the standard deviation. This must be a positive value. **PASS** includes a special module for estimating the standard deviation. This module may be
loaded by pressing the $SD$ button. Refer to the Standard Deviation Estimator chapter for further details.

**Test**

**Higher Means Are**
This option defines whether higher values of the response variable are to be considered better or worse.

The choice here determines the direction of the non-inferiority test.

If Higher Means Are Better the null hypothesis is $\text{Diff} \leq -\text{NIM}$ and the alternative hypothesis is $\text{Diff} > -\text{NIM}$. If Higher Means Are Worse the null hypothesis is $\text{Diff} \geq \text{NIM}$ and the alternative hypothesis is $\text{Diff} < \text{NIM}$.

**Nonparametric Adjustment**
This option makes appropriate sample size adjustments for the Wilcoxon test. Results by Al-Sunduqchi and Guenther (1990) indicate that power calculations for the Wilcoxon test may be made using the standard $t$ test formulations with a simple adjustment to the sample size. The size of the adjustment depends upon the actual distribution of the data. They give sample size adjustment factors for four distributions. These are 1 for the uniform distribution, $2/3$ for the double exponential distribution, $9/\pi^2$ for the logistic distribution, and $\pi/3$ for the normal distribution.

The options are as follows:

- **Ignore**
  Do not make a Wilcoxon adjustment. This indicates that you want to analyze a $t$ test, not the Wilcoxon test.

- **Uniform**
  Make the Wilcoxon sample size adjustment assuming the uniform distribution. Since the factor is one, this option performs the same as Ignore. It is included for completeness.

- **Double Exponential**
  Make the Wilcoxon sample size adjustment assuming that the data actually follow the double exponential distribution.

- **Logistic**
  Make the Wilcoxon sample size adjustment assuming that the data actually follow the logistic distribution.

- **Normal**
  Make the Wilcoxon sample size adjustment assuming that the data actually follow the normal distribution.

**Population Size**
This is the number of subjects in the population. Usually, you assume that samples are drawn from a very large (infinite) population. Occasionally, however, situations arise in which the population of interest is of limited size. In these cases, appropriate adjustments must be made.
When a finite population size is specified, the standard deviation is reduced according to the formula

\[ \sigma_1^2 = \left(1 - \frac{n}{N}\right)\sigma^2 \]

where \( n \) is the sample size, \( N \) is the population size, \( \sigma \) is the original standard deviation, and \( \sigma_1 \) is the new standard deviation.

The quantity \( n/N \) is often called the sampling fraction. The quantity \( 1 - \frac{n}{N} \) is called the finite population correction factor.

### Example 1 – Power Analysis

Suppose that a test is to be conducted to determine if a new cancer treatment adversely affects the mean bone density. The adjusted mean bone density (AMBD) in the population of interest is 0.002300 gm/cm with a standard deviation of 0.000300 gm/cm. Clinicians decide that if the treatment reduces AMBD by more than 5% (0.000115 gm/cm), it poses a significant health threat. They also want to consider what would happen if the margin of non-inferiority is set to 2.5% (0.0000575 gm/cm).

Following accepted procedure, the analysis will be a non-inferiority test using the t-test at the 0.025 significance level. Power is to be calculated assuming that the new treatment has no effect on AMBD. Several sample sizes between 20 and 300 will be analyzed. The researchers want to achieve a power of at least 90%. All numbers have been multiplied by 10000 to make the reports and plots easier to read.

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Non-Inferiority Tests for One Mean procedure window by expanding Means, then One Mean, then clicking on Non-Inferiority, and then clicking on Non-Inferiority Tests for One Mean. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td></td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.025</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>20 40 60 80 100 150 200 300</td>
</tr>
<tr>
<td>NIM (Non-Inferiority Margin)</td>
<td>0.575 1.15</td>
</tr>
<tr>
<td>D (True Difference)</td>
<td>0</td>
</tr>
<tr>
<td>S (Standard Deviation)</td>
<td>3</td>
</tr>
<tr>
<td>Higher Means Are</td>
<td>Better</td>
</tr>
<tr>
<td>Nonparametric Adjustment</td>
<td>Ignore</td>
</tr>
<tr>
<td>Population Size</td>
<td>Infinite</td>
</tr>
</tbody>
</table>
Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results for Non-Inferiority Test (H0: Diff <= -NIM; H1: Diff > -NIM)
Higher Means are Better
Test Statistic: T-Test

<table>
<thead>
<tr>
<th>Power</th>
<th>Non-Inferiority Margin</th>
<th>Actual Difference</th>
<th>Significance Level</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.12601</td>
<td>20 -0.575</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.87399</td>
</tr>
<tr>
<td>0.21844</td>
<td>40 -0.575</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.78156</td>
</tr>
<tr>
<td>0.30873</td>
<td>60 -0.575</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.69127</td>
</tr>
<tr>
<td>0.39493</td>
<td>80 -0.575</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.60507</td>
</tr>
<tr>
<td>0.47532</td>
<td>100 -0.575</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.52468</td>
</tr>
<tr>
<td>0.64517</td>
<td>150 -0.575</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.35483</td>
</tr>
<tr>
<td>0.76959</td>
<td>200 -0.575</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.23041</td>
</tr>
<tr>
<td>0.91262</td>
<td>300 -0.575</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.08738</td>
</tr>
<tr>
<td>0.36990</td>
<td>20 -1.150</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.63010</td>
</tr>
<tr>
<td>0.65705</td>
<td>40 -1.150</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.34295</td>
</tr>
<tr>
<td>0.83164</td>
<td>60 -1.150</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.16836</td>
</tr>
<tr>
<td>0.92317</td>
<td>80 -1.150</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.07683</td>
</tr>
<tr>
<td>0.96682</td>
<td>100 -1.150</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.03318</td>
</tr>
<tr>
<td>0.99658</td>
<td>150 -1.150</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.00342</td>
</tr>
<tr>
<td>0.99970</td>
<td>200 -1.150</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.00000</td>
</tr>
<tr>
<td>1.00000</td>
<td>300 -1.150</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.00000</td>
</tr>
</tbody>
</table>

Report Definitions

- Power is the probability of rejecting H0 when it is false.
- N is the sample size, the number of subjects (or pairs) in the study.
- -NIM is the magnitude and direction of the margin of non-inferiority. Since higher means are better, this value is negative and is the distance below the reference value that is still considered non-inferior.
- D is the mean difference (treatment - reference value) at which the power is computed.
- Alpha is the probability of rejecting H0 when it is true, which is the probability of a false positive.
- Beta is the probability of accepting H0 when it is false, which is the probability of a false negative.
- S is the standard deviation of the response. It measures the variability in the population.

Summary Statements

A sample size of 20 achieves 13% power to detect non-inferiority using a one-sided t-test when the margin of equivalence is -0.575 and the true difference between the mean and the reference value is 0.000. The data are drawn from a single population with a standard deviation of 3.000. The significance level (alpha) of the test is 0.02500.

The above report shows that for NIM = 1.15, the sample size necessary to obtain 90% power is just under 80. However, if NIM = 0.575, the required sample size is about 300.
**Example 2 – Finding the Sample Size**

Continuing with Example 1, the researchers want to know the exact sample size for each value of NIM.

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Non-Inferiority Tests for One Mean procedure window by expanding Means, then One Mean, then clicking on Non-Inferiority, and then clicking on Non-Inferiority Tests for One Mean. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N</td>
</tr>
<tr>
<td>Power</td>
<td>0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.025</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>NIM (Non-Inferiority Margin)</td>
<td>0.575 1.15</td>
</tr>
<tr>
<td>D (True Difference)</td>
<td>0</td>
</tr>
<tr>
<td>S (Standard Deviation)</td>
<td>3</td>
</tr>
<tr>
<td>Higher Means Are</td>
<td>Better</td>
</tr>
<tr>
<td>Nonparametric Adjustment</td>
<td>Ignore</td>
</tr>
<tr>
<td>Population Size</td>
<td>Infinite</td>
</tr>
</tbody>
</table>

**Output**

Click the Run button to perform the calculations and generate the following output.

**Numeric Results**

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>Non-Inferiority Margin (-NIM)</th>
<th>Actual Difference (D)</th>
<th>Significance Level (Alpha)</th>
<th>Beta</th>
<th>Standard Deviation (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.90005</td>
<td>288</td>
<td>-0.575</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.09995</td>
<td>3.000</td>
</tr>
<tr>
<td>0.90215</td>
<td>74</td>
<td>-1.150</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.09785</td>
<td>3.000</td>
</tr>
</tbody>
</table>

This report shows the exact sample size requirement for each value of NIM.
Example 3 – Validation using Chow

Chow, Shao, Wang (2003) pages 54-55 has an example of a sample size calculation for a non-inferiority trial. Their example obtains a sample size of 8 when \(D = 0.5\), NIM = 0.5, \(S = 1\), Alpha = 0.05, and Beta = 0.20.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Non-Inferiority Tests for One Mean procedure window by expanding Means, then One Mean, then clicking on Non-Inferiority, and then clicking on Non-Inferiority Tests for One Mean. You may then make the appropriate entries as listed below, or open Example 3 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>(N)</td>
</tr>
<tr>
<td>Power</td>
<td>0.80</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>(N) (Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>NIM (Non-Inferiority Margin)</td>
<td>0.5</td>
</tr>
<tr>
<td>(D) (True Difference)</td>
<td>0.5</td>
</tr>
<tr>
<td>(S) (Standard Deviation)</td>
<td>1</td>
</tr>
<tr>
<td>Higher Means Are</td>
<td>Better</td>
</tr>
<tr>
<td>Nonparametric Adjustment</td>
<td>Ignore</td>
</tr>
<tr>
<td>Population Size</td>
<td>Infinite</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>Non-Inferiority Margin (-NIM)</th>
<th>Actual Difference (D)</th>
<th>Significance Level (Alpha)</th>
<th>Beta</th>
<th>Standard Deviation (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.81502</td>
<td>8</td>
<td>-0.500</td>
<td>0.500</td>
<td>0.05000</td>
<td>0.18498</td>
<td>1.000</td>
</tr>
</tbody>
</table>

PASS has also obtained a sample size of 8.
Example 4 – Validation of a Cross-Over Design given in Julious

Julious (2004) page 1953 gives an example of a sample size calculation for a cross-over design. His example obtains a sample size of 87 when $D = 0$, $\text{NIM} = 10$, $S = 28.28427$, $\alpha = 0.025$, and $\beta = 0.10$. When $D$ is changed to 2, the resulting sample size is 61.

Note that in Julius’s example, the population standard deviation is given as 20. Assuming that the correlation between items in a pair is 0, the standard deviation of the difference is calculated to be $S = \sqrt{20^2 + 20^2 - 0(20)(20)} = 28.28471$. Actually, the value of $S$ probably should be less because the correlation is usually greater than 0 (at least 0.2).

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Non-Inferiority Tests for One Mean procedure window by expanding Means, then One Mean, then clicking on Non-Inferiority, and then clicking on Non-Inferiority Tests for One Mean. You may then make the appropriate entries as listed below, or open Example 4 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>$N$</td>
</tr>
<tr>
<td>Power</td>
<td>0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.025</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>NIM (Non-Inferiority Margin)</td>
<td>10</td>
</tr>
<tr>
<td>D (True Difference)</td>
<td>2</td>
</tr>
<tr>
<td>S (Standard Deviation)</td>
<td>28.28427</td>
</tr>
<tr>
<td>Higher Means Are</td>
<td>Better</td>
</tr>
<tr>
<td>Nonparametric Adjustment</td>
<td>Ignore</td>
</tr>
<tr>
<td>Population Size</td>
<td>Infinite</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

<table>
<thead>
<tr>
<th>Numeric Results for Non-Inferiority Test (H0: Diff &lt;= -NIM; H1: Diff &gt; -NIM)</th>
<th>Higher Means are Better</th>
<th>Test Statistic: T-Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power</td>
<td>Non-Inferiority Margin (NIM)</td>
<td>Actual Difference (D)</td>
</tr>
<tr>
<td>-------</td>
<td>-----------------------------</td>
<td>-----------------------</td>
</tr>
<tr>
<td>0.90332</td>
<td>-10.000</td>
<td>0.000</td>
</tr>
<tr>
<td>0.90323</td>
<td>-10.000</td>
<td>2.000</td>
</tr>
</tbody>
</table>

PASS has also obtained sample sizes of 87 and 61.
Example 5 – Validation of a Cross-Over Design given in Chow, Shao, and Wang

Chow, Shao, and Wang (2004) page 67 give an example of a sample size calculation for a cross-over design. Their example calculates sample sizes of 13 and 14 (13 by formula and 14 from their table) in each sequence (26 or 28 total) when $D = -0.1$, NIM = 0.2, $S = 0.2$, $\alpha = 0.05$, and $\beta = 0.20$.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Non-Inferiority Tests for One Mean procedure window by expanding Means, then One Mean, then clicking on Non-Inferiority, and then clicking on Non-Inferiority Tests for One Mean. You may then make the appropriate entries as listed below, or open Example 5 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>$N$</td>
</tr>
<tr>
<td>Power</td>
<td>0.80</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>NIM (Non-Inferiority Margin)</td>
<td>0.2</td>
</tr>
<tr>
<td>D (True Difference)</td>
<td>-0.1</td>
</tr>
<tr>
<td>S (Standard Deviation)</td>
<td>0.2</td>
</tr>
<tr>
<td>Higher Means Are</td>
<td>Better</td>
</tr>
<tr>
<td>Nonparametric Adjustment</td>
<td>Ignore</td>
</tr>
<tr>
<td>Population Size</td>
<td>Infinite</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>Non-Inferiority Margin (-NIM)</th>
<th>Actual Difference (D)</th>
<th>Significance Level (Alpha)</th>
<th>Standard Deviation (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.81183</td>
<td>27</td>
<td>-0.200</td>
<td>-0.100</td>
<td>0.05000</td>
<td>0.18817</td>
</tr>
</tbody>
</table>

PASS obtained a sample size of 27 which is between the values of 26 and 28 that were obtained by Chow et al.
415-14 Non-Inferiority Tests for One Mean
Chapter 420

Confidence Intervals for One Mean

Introduction

This routine calculates the sample size necessary to achieve a specified distance from the mean to
the confidence limit(s) at a stated confidence level for a confidence interval about the mean when
the underlying data distribution is normal.

Caution: This procedure assumes that the standard deviation of the future sample will be the same
as the standard deviation that is specified. If the standard deviation to be used in the procedure is
estimated from a previous sample or represents the population standard deviation, the Confidence
Intervals for One Mean with Tolerance Probability procedure should be considered. That
procedure controls the probability that the distance from the mean to the confidence limits will be
less than or equal to the value specified.

Technical Details

For a single mean from a normal distribution with known variance, a two-sided, 100(1 – \(\alpha\))% confidence interval is calculated by

\[
\bar{X} \pm \frac{z_{1-\alpha/2} \sigma}{\sqrt{n}}
\]

A one-sided 100(1 – \(\alpha\))% upper confidence limit is calculated by

\[
\bar{X} + \frac{z_{1-\alpha} \sigma}{\sqrt{n}}
\]

Similarly, the one-sided 100(1 – \(\alpha\))% lower confidence limit is

\[
\bar{X} - \frac{z_{1-\alpha} \sigma}{\sqrt{n}}
\]
For a single mean from a normal distribution with unknown variance, a two-sided, 100(1 – \( \alpha \))% confidence interval is calculated by

\[
\bar{X} \pm \frac{t_{1-\alpha/2,n-1} \hat{\sigma}}{\sqrt{n}}
\]

A one-sided 100(1 – \( \alpha \))% upper confidence limit is calculated by

\[
\bar{X} + \frac{t_{1-\alpha,n-1} \hat{\sigma}}{\sqrt{n}}
\]

Similarly, the one-sided 100(1 – \( \alpha \))% lower confidence limit is

\[
\bar{X} - \frac{t_{1-\alpha,n-1} \hat{\sigma}}{\sqrt{n}}
\]

Each confidence interval is calculated using an estimate of the mean plus and/or minus a quantity that represents the distance from the mean to the edge of the interval. For two-sided confidence intervals, this distance is sometimes called the precision, margin of error, or half-width. We will label this distance, \( D \).

The basic equation for determining sample size when \( D \) has been specified is

\[
D = \frac{z_{1-\alpha/2} \sigma}{\sqrt{n}}
\]

when the standard deviation is known, and

\[
D = \frac{t_{1-\alpha/2,n-1} \hat{\sigma}}{\sqrt{n}}
\]

when the standard deviation is unknown. These equations can be solved for any of the unknown quantities in terms of the others. The value \( \alpha/2 \) is replaced by \( \alpha \) when a one-sided interval is used.

Finite Population Size

The above calculations assume that samples are being drawn from a large (infinite) population. When the population is of finite size (\( N \)), an adjustment must be made. The adjustment reduces the standard deviation as follows:

\[
\sigma_{\text{finite}} = \sigma \sqrt{\left(1 - \frac{n}{N}\right)}
\]

This new standard deviation replaces the regular standard deviation in the above formulas.

Confidence Level

The confidence level, 1 – \( \alpha \), has the following interpretation. If thousands of samples of \( n \) items are drawn from a population using simple random sampling and a confidence interval is calculated for each sample, the proportion of those intervals that will include the true population mean is 1 - \( \alpha \).
Confidence Intervals for One Mean  420-3

Procedure Options
This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab
The Data tab contains most of the parameters and options that you will be concerned with.

Solve For
Find (Solve For)
This option specifies the parameter to be solved for from the other parameters.

Confidence
Confidence Level
The confidence level, $1 - \alpha$, has the following interpretation. If thousands of samples of $n$ items are drawn from a population using simple random sampling and a confidence interval is calculated for each sample, the proportion of those intervals that will include the true population mean is $1 - \alpha$.

Often, the values 0.95 or 0.99 are used. You can enter single values or a range of values such as 0.90, 0.95 or 0.90 to 0.99 by 0.01.

Sample Size
N (Sample Size)
Enter one or more values for the sample size. This is the number of individuals selected at random from the population to be in the study.

You can enter a single value or a range of values.

Precision
Distance from Mean to Limit(s)
This is the distance from the confidence limit(s) to the mean. For two-sided intervals, it is also known as the precision, half-width, or margin of error.

You can enter a single value or a list of values. The value(s) must be greater than zero.

Standard Deviation
S (Standard Deviation)
Enter a value (or range of values) for the standard deviation. Roughly speaking, this value estimates the average absolute difference between each individual and every other individual.

You can use the results of a pilot study, a previous study, or a ball park estimate based on the range (e.g., Range/4) to estimate this parameter.
Know Standard Deviation
Check this box when you want to base your results on the normal distribution. When the box is not checked, calculations are based on the t-distribution. The difference between the two distributions is negligible when the sample sizes are large (>50).

One-Sided or Two-Sided Interval
Interval Type
Specify whether the interval to be used will be a one-sided or a two-sided confidence interval.

Population
Population Size
This is the number of individuals in the population. Usually, you assume that samples are drawn from a very large (infinite) population. Occasionally, however, situations arise in which the population of interest is of limited size. In these cases, appropriate adjustments must be made. This option sets the population size.

Iterations Tab
This tab sets an option used in the iterative procedures.

Maximum Iterations
Maximum Iterations Before Search Termination
Specify the maximum number of iterations allowed before the search for the criterion of interest is aborted. When the maximum number of iterations is reached without convergence, the criterion is left blank. A value of 500 is recommended.
Example 1 – Calculating Sample Size

Suppose a study is planned in which the researcher wishes to construct a two-sided 95% confidence interval for the mean such that the width of the interval is no wider than 14 units. The confidence level is set at 0.95, but 0.99 is included for comparative purposes. The standard deviation estimate, based on the range of data values, is 28. Instead of examining only the interval half-width of 7, a series of half-widths from 5 to 9 will also be considered.

The goal is to determine the necessary sample size.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Confidence Intervals for One Mean procedure window by expanding Means, then One Mean, then clicking on Confidence Interval, and then clicking on Confidence Intervals for One Mean. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N (Sample Size)</td>
</tr>
<tr>
<td>Confidence Level</td>
<td>0.95 0.99</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Distance from Mean to Limit(s)</td>
<td>5 to 9 by 1</td>
</tr>
<tr>
<td>S (Standard Deviation)</td>
<td>28</td>
</tr>
<tr>
<td>Interval Type</td>
<td>Two-Sided</td>
</tr>
<tr>
<td>Population Size</td>
<td>Infinite</td>
</tr>
</tbody>
</table>

Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>Sample Size</th>
<th>Target Distance from Mean to Limits</th>
<th>Actual Distance from Mean to Limits</th>
<th>Standard Deviation ($)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.95</td>
<td>123</td>
<td>5.000</td>
<td>4.998</td>
<td>28.000</td>
</tr>
<tr>
<td>0.99</td>
<td>212</td>
<td>5.000</td>
<td>4.999</td>
<td>28.000</td>
</tr>
<tr>
<td>0.95</td>
<td>87</td>
<td>6.000</td>
<td>5.986</td>
<td>28.000</td>
</tr>
<tr>
<td>0.99</td>
<td>149</td>
<td>6.000</td>
<td>5.986</td>
<td>28.000</td>
</tr>
<tr>
<td>0.95</td>
<td>64</td>
<td>7.000</td>
<td>6.994</td>
<td>28.000</td>
</tr>
<tr>
<td>0.99</td>
<td>110</td>
<td>7.000</td>
<td>6.999</td>
<td>28.000</td>
</tr>
<tr>
<td>0.95</td>
<td>50</td>
<td>8.000</td>
<td>7.958</td>
<td>28.000</td>
</tr>
<tr>
<td>0.99</td>
<td>86</td>
<td>8.000</td>
<td>7.956</td>
<td>28.000</td>
</tr>
<tr>
<td>0.95</td>
<td>40</td>
<td>9.000</td>
<td>8.955</td>
<td>28.000</td>
</tr>
<tr>
<td>0.99</td>
<td>69</td>
<td>9.000</td>
<td>8.933</td>
<td>28.000</td>
</tr>
</tbody>
</table>

References

420-6  Confidence Intervals for One Mean

### Report Definitions
Confidence level is the proportion of confidence intervals (constructed with this same confidence level, sample size, etc.) that would contain the population mean.

N is the size of the sample drawn from the population.

Distance from Mean to Limit is the distance from the confidence limit(s) to the mean. For two-sided intervals, it is also know as the precision, half-width, or margin of error.

Target Distance from Mean to Limit is the value of the distance that is entered into the procedure.

Actual Distance from Mean to Limit is the value of the distance that is obtained from the procedure.

The standard deviation of the population measures the variability in the population.

### Summary Statements
A sample size of 123 produces a two-sided 95% confidence interval with a distance from the mean to the limits that is equal to 4.998 when the estimated standard deviation is 28.000.

This report shows the calculated sample size for each of the scenarios.

### Plots Section

This plot shows the sample size versus the distance from the mean to the limits (precision) for the two confidence levels.
Example 2 – Validation using Moore and McCabe

Moore and McCabe (1999) page 443 give an example of a sample size calculation for a confidence interval on the mean when the confidence coefficient is 95%, the standard deviation is known to be 3, and the margin of error is 2. The necessary sample size is 9.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Confidence Intervals for One Mean procedure window by expanding Means, then One Mean, then clicking on Confidence Interval, and then clicking on Confidence Intervals for One Mean. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N (Sample Size)</td>
</tr>
<tr>
<td>Confidence Level</td>
<td>0.95</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Distance from Mean to Limit(s)</td>
<td>2</td>
</tr>
<tr>
<td>S (Standard Deviation)</td>
<td>3</td>
</tr>
<tr>
<td>Known Standard Deviation</td>
<td>Checked</td>
</tr>
<tr>
<td>Interval Type</td>
<td>Two-Sided</td>
</tr>
<tr>
<td>Population Size</td>
<td>Infinite</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>Sample Size (N)</th>
<th>Target Distance from Mean to Limits</th>
<th>Actual Distance from Mean to Limits</th>
<th>Standard Deviation (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.95</td>
<td>9</td>
<td>2.000</td>
<td>1.960</td>
<td>3.000</td>
</tr>
</tbody>
</table>

PASS also calculated the necessary sample size to be 9.
Example 3 – Validation using Ostle and Malone

Ostle and Malone (1988) page 536 give an example of a sample size calculation for a confidence interval on the mean when the confidence coefficient is 95%, the standard deviation is known to be 7, and the margin of error is 5. The necessary sample size is 8.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Confidence Intervals for One Mean procedure window by expanding Means, then One Mean, then clicking on Confidence Interval, and then clicking on Confidence Intervals for One Mean. You may then make the appropriate entries as listed below, or open Example 3 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N (Sample Size)</td>
</tr>
<tr>
<td>Confidence Level</td>
<td>0.95</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Distance from Mean to Limit(s)</td>
<td>5</td>
</tr>
<tr>
<td>S (Standard Deviation)</td>
<td>7</td>
</tr>
<tr>
<td>Known Standard Deviation</td>
<td>Checked</td>
</tr>
<tr>
<td>Interval Type</td>
<td>Two-Sided</td>
</tr>
<tr>
<td>Population Size</td>
<td>Infinite</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>Sample Size (N)</th>
<th>Target Distance from Mean to Limits</th>
<th>Actual Distance from Mean to Limits</th>
<th>Standard Deviation (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.95</td>
<td>8</td>
<td>5.000</td>
<td>4.851</td>
<td>7.000</td>
</tr>
</tbody>
</table>

PASS also calculated the necessary sample size to be 8.
Chapter 421

Confidence Intervals for One Mean with Tolerance Probability

Introduction
This procedure calculates the sample size necessary to achieve a specified distance from the mean to the confidence limit(s) with a given tolerance probability at a stated confidence level for a confidence interval about a single mean when the underlying data distribution is normal.

Technical Details
For a single mean from a normal distribution with unknown variance, a two-sided, 100(1 – \(\alpha\))% confidence interval is calculated by

\[
\bar{X} \pm \frac{t_{1-a/2, n-1} \hat{\sigma}}{\sqrt{n}}
\]

A one-sided 100(1 – \(\alpha\))% upper confidence limit is calculated by

\[
\bar{X} + \frac{t_{1-a, n-1} \hat{\sigma}}{\sqrt{n}}
\]

Similarly, the one-sided 100(1 – \(\alpha\))% lower confidence limit is

\[
\bar{X} - \frac{t_{1-a, n-1} \hat{\sigma}}{\sqrt{n}}
\]

Each confidence interval is calculated using an estimate of the mean plus and/or minus a quantity that represents the distance from the mean to the edge of the interval. For two-sided confidence intervals, this distance is sometimes called the precision, margin of error, or half-width. We will label this distance, \(D\).
The basic equation for determining sample size when $D$ has been specified is

$$D = \frac{t_{1-\alpha/2,n-1} \hat{\sigma}}{\sqrt{n}}$$

Solving for $n$, we obtain

$$n = \left( \frac{t_{1-\alpha/2,n-1} \hat{\sigma}}{D} \right)^2$$

This equation can be solved for any of the unknown quantities in terms of the others. The value $\alpha/2$ is replaced by $\alpha$ when a one-sided interval is used.

There is an additional subtlety that arises when the standard deviation is to be chosen for estimating sample size. The sample sizes determined from the formula above produce confidence intervals with the specified widths only when the future sample has a sample standard deviation that is no greater than the value specified.

As an example, suppose that 15 individuals are sampled in a pilot study, and a standard deviation estimate of 3.5 is obtained from the sample. The purpose of a later study is to estimate the mean within 10 units. Suppose further that the sample size needed is calculated to be 57 using the formula above with 3.5 as the estimate for the standard deviation. The sample of size 57 is then obtained from the population, but the standard deviation of the 57 individuals turns out to be 3.9 rather than 3.5. The confidence interval is computed and the distance from the mean to the confidence limits is greater than 10 units.

This example illustrates the need for an adjustment to adjust the sample size such that the distance from the mean to the confidence limits will be below the specified value with known probability.

Such an adjustment for situations where a previous sample is used to estimate the standard deviation is derived by Harris, Horvitz, and Mood (1948) and discussed in Zar (1984) and Hahn and Meeker (1991). The adjustment is

$$n = \left( \frac{t_{1-\alpha/2,n-1} \hat{\sigma}}{D} \right)^2 F_{1-\gamma;n-1,m-1}$$

where $1 - \gamma$ is the probability that the distance from the mean to the confidence limit(s) will be below the specified value, and $m$ is the sample size in the previous sample that was used to estimate the standard deviation.

The corresponding adjustment when no previous sample is available is discussed in Kupper and Hafner (1989) and Hahn and Meeker (1991). The adjustment in this case is

$$n = \left( \frac{t_{1-\alpha/2,n-1} \hat{\sigma}}{D} \right)^2 \left( \frac{\chi^2_{1-\gamma,n-1}}{n-1} \right)$$

where, again, $1 - \gamma$ is the probability that the distance from the mean to the confidence limit(s) will be below the specified value.

Each of these adjustments accounts for the variability in a future estimate of the standard deviation. In the first adjustment formula (Harris, Horvitz, and Mood, 1948), the distribution of the standard deviation is based on the estimate from a previous sample. In the second adjustment formula, the distribution of the standard deviation is based on a specified value that is assumed to be the population standard deviation.
Finite Population Size

The above calculations assume that samples are being drawn from a large (infinite) population. When the population is of finite size ($N$), an adjustment must be made. The adjustment reduces the standard deviation as follows:

$$\sigma_{finite} = \sigma \sqrt{1 - \frac{n}{N}}$$

This new standard deviation replaces the regular standard deviation in the above formulas.

Confidence Level

The confidence level, $1 - \alpha$, has the following interpretation. If thousands of samples of $n$ items are drawn from a population using simple random sampling and a confidence interval is calculated for each sample, the proportion of those intervals that will include the true population mean is $1 - \alpha$.

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains most of the parameters and options that you will be concerned with.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for from the other parameters.

Confidence and Tolerance

Confidence Level ($1 - \alpha$)

The confidence level, $1 - \alpha$, has the following interpretation. If thousands of samples of $n$ items are drawn from a population using simple random sampling and a confidence interval is calculated for each sample, the proportion of those intervals that will include the true population mean is $1 - \alpha$.

Often, the values 0.95 or 0.99 are used. You can enter single values or a range of values such as 0.90, 0.95 or 0.90 to 0.99 by 0.01.

Tolerance Probability

This is the probability that a future interval with sample size $N$ and the specified confidence level will have a distance from the mean to the limit(s) that is less than or equal to the distance specified.

If a tolerance probability is not used, as in the 'Confidence Intervals for One Mean' procedure, the sample size is calculated for the expected distance from the mean to the limit(s), which assumes that the future standard deviation will also be the one specified.
Using a tolerance probability implies that the standard deviation of the future sample will not be known in advance, and therefore, an adjustment is made to the sample size formula to account for the variability in the standard deviation. Use of a tolerance probability is similar to using an upper bound for the standard deviation in the 'Confidence Intervals for One Mean' procedure.

Values between 0 and 1 can be entered. The choice of the tolerance probability depends upon how important it is that the distance from the interval limit(s) to the mean is at most the value specified. You can enter a range of values such as 0.70 0.80 0.90 or 0.70 to 0.95 by 0.05.

**Sample Size**

**N (Sample Size)**
Enter one or more values for the sample size. This is the number of individuals selected at random from the population to be in the study.

You can enter a single value or a range of values.

**Precision**

**Distance from Mean to Limit(s)**

This is the distance from the confidence limit(s) to the mean. For two-sided intervals, it is also known as the precision, half-width, or margin of error.

You can enter a single value or a list of values. The value(s) must be greater than zero.

**Standard Deviation**

**Standard Deviation Source**

This procedure permits two sources for estimates of the standard deviation:

- **S is a Population Standard Deviation**
  
  This option should be selected if there is no previous sample that can be used to obtain an estimate of the standard deviation. In this case, the algorithm assumes that future sample obtained will be from a population with standard deviation S.

- **S from a Previous Sample**
  
  This option should be selected if the estimate of the standard deviation is obtained from a previous random sample from the same distribution as the one to be sampled. The sample size of the previous sample must also be entered under 'Sample Size of Previous Sample'.

**Standard Deviation – S is a Population Standard Deviation**

**S (Standard Deviation)**
Enter an estimate of the standard deviation (must be positive). In this case, the algorithm assumes that future samples obtained will be from a population with standard deviation S.

One common method for estimating the standard deviation is the range divided by 4, 5, or 6.

You can enter a range of values such as 1 2 3 or 1 to 10 by 1.
Press the Standard Deviation Estimator button to load the Standard Deviation Estimator window.

**Standard Deviation – S from a Previous Sample**

**S (SD Estimated from a Previous Sample)**
Enter an estimate of the standard deviation from a previous (or pilot) study. This value must be positive.

A range of values may be entered.

Press the Standard Deviation Estimator button to load the Standard Deviation Estimator window.

**Sample Size of Previous Sample**
Enter the sample size that was used to estimate the standard deviation entered in S (SD Estimated from a Previous Sample).

This value is entered only when 'Standard Deviation Source:' is set to 'S from a Previous Sample'.

**One-Sided or Two-Sided Interval**

**Interval Type**
Specify whether the interval to be used will be a one-sided or a two-sided confidence interval.

**Population**

**Population Size**
This is the number of individuals in the population. Usually, you assume that samples are drawn from a very large (infinite) population. Occasionally, however, situations arise in which the population of interest is of limited size. In these cases, appropriate adjustments must be made.

This option sets the population size.

**Iterations Tab**

This tab sets an option used in the iterative procedures.

**Maximum Iterations**

**Maximum Iterations Before Search Termination**
Specify the maximum number of iterations allowed before the search for the criterion of interest is aborted. When the maximum number of iterations is reached without convergence, the criterion is left blank. A value of 500 is recommended.
Example 1 – Calculating Sample Size

A researcher would like to estimate the mean weight of a population with 95% confidence. It is very important that the mean weight is estimated within 15 grams. Data available from a previous study are used to provide an estimate of the standard deviation. The estimate of the standard deviation is 45.1 grams, from a sample of size 14.

The goal is to determine the sample size necessary to obtain a two-sided confidence interval such that the mean weight is estimated within 15 grams. Tolerance probabilities of 0.70 to 0.95 will be examined.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Confidence Intervals for One Mean with Tolerance Probability procedure window by expanding Means, then One Mean, then clicking on Confidence Interval, and then clicking on Confidence Intervals for One Mean with Tolerance Probability. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N (Sample Size)</td>
</tr>
<tr>
<td>Confidence Level</td>
<td>0.95</td>
</tr>
<tr>
<td>Tolerance Probability</td>
<td>0.70 to 0.95 by 0.05</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Distance from Mean to Limit(s)</td>
<td>15</td>
</tr>
<tr>
<td>Standard Deviation Source</td>
<td>S from a Previous Sample</td>
</tr>
<tr>
<td>S</td>
<td>45.1</td>
</tr>
<tr>
<td>Sample Size of Previous Sample</td>
<td>14</td>
</tr>
<tr>
<td>Interval Type</td>
<td>Two-Sided</td>
</tr>
<tr>
<td>Population Size</td>
<td>Infinite</td>
</tr>
</tbody>
</table>

Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>Sample Size (N)</th>
<th>Target Distance from Mean to Limits</th>
<th>Actual Distance from Mean to Limits</th>
<th>Standard Deviation (S)</th>
<th>Tolerance Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.950</td>
<td>49</td>
<td>15,000</td>
<td>14.899</td>
<td>45.100</td>
<td>0.700</td>
</tr>
<tr>
<td>0.950</td>
<td>52</td>
<td>15,000</td>
<td>14.959</td>
<td>45.100</td>
<td>0.750</td>
</tr>
<tr>
<td>0.950</td>
<td>56</td>
<td>15,000</td>
<td>14.969</td>
<td>45.100</td>
<td>0.800</td>
</tr>
<tr>
<td>0.950</td>
<td>61</td>
<td>15,000</td>
<td>14.997</td>
<td>45.100</td>
<td>0.850</td>
</tr>
<tr>
<td>0.950</td>
<td>69</td>
<td>15,000</td>
<td>14.925</td>
<td>45.100</td>
<td>0.900</td>
</tr>
<tr>
<td>0.950</td>
<td>82</td>
<td>15,000</td>
<td>14.943</td>
<td>45.100</td>
<td>0.950</td>
</tr>
</tbody>
</table>

Sample size for estimate of S from previous sample = 14.
Confidence Intervals for One Mean with Tolerance Probability 421-7

References
Harris, M., Horvitz, D. J., and Mood, A. M. 1948. ‘On the Determination of Sample Sizes in Designing

Report Definitions
Confidence level is the proportion of confidence intervals (constructed with this same confidence level,
sample size, etc.) that would contain the population mean.
N is the size of the sample drawn from the population.
Distance from Mean to Limit is the distance from the confidence limit(s) to the mean. For two-sided intervals,
it is also known as the precision, half-width, or margin of error.
Target Distance from Mean to Limit is the value of the distance that is entered into the procedure.
Actual Distance from Mean to Limit is the value of the distance that is obtained from the procedure.
The standard deviation of the population measures the variability in the population.
Tolerance Probability is the probability that a future interval with sample size N and corresponding
confidence level will have a distance from the mean to the limit(s) that is less than or equal to the
specified distance.

Summary Statements
The probability is 0.700 that a sample size of 49 will produce a two-sided 95% confidence
interval with a distance from the mean to the limits that is less than or equal to 14.899 if
the population standard deviation is estimated to be 45.100 by a previous sample of size 14.

This report shows the calculated sample size for each of the scenarios.

Plots Section

This plot shows the sample size versus the tolerance probability.
**Example 2 – Validation using Hahn and Meeker**

Hahn and Meeker (1991) page 139 give an example of a sample size calculation for a two-sided confidence interval on the mean when the confidence level is 95%, the population standard deviation is assumed to be 2500, the distance from the mean to the limit is 1500, and the tolerance probability is 0.90. The necessary sample size is 19.

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Confidence Intervals for One Mean with Tolerance Probability** procedure window by expanding **Means**, then **One Mean**, then clicking on **Confidence Interval**, and then clicking on **Confidence Intervals for One Mean with Tolerance Probability**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Tab</strong></td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N (Sample Size)</td>
</tr>
<tr>
<td>Confidence Level</td>
<td>0.95</td>
</tr>
<tr>
<td>Tolerance Probability</td>
<td>0.90</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Distance from Mean to Limit(s)</td>
<td>1500</td>
</tr>
<tr>
<td>Standard Deviation Source</td>
<td>S is a Population Standard Deviation</td>
</tr>
<tr>
<td>S</td>
<td>2500</td>
</tr>
<tr>
<td>Interval Type</td>
<td>Two-Sided</td>
</tr>
<tr>
<td>Population Size</td>
<td>Infinite</td>
</tr>
</tbody>
</table>

**Output**

Click the Run button to perform the calculations and generate the following output.

**Numeric Results**

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>Sample Size (N)</th>
<th>Target Distance from Mean to Limits</th>
<th>Actual Distance from Mean to Limits</th>
<th>Standard Deviation (S)</th>
<th>Tolerance Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.95</td>
<td>19</td>
<td>1500.000</td>
<td>1447.889</td>
<td>2500.000</td>
<td>0.90</td>
</tr>
</tbody>
</table>

**PASS** also calculated the necessary sample size to be 19.
Example 3 – Validation using Zar

Zar (1984) pages 109-110 give an example of a sample size calculation for a two-sided confidence interval on the mean when the confidence level is 95%, the standard deviation is estimated to be 4.247211 by a previous sample of size 25, the distance from the mean to the limit is 1.5, and the tolerance probability is 0.90. The necessary sample size is 53.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Confidence Intervals for One Mean with Tolerance Probability procedure window by expanding Means, then One Mean, then clicking on Confidence Interval, and then clicking on Confidence Intervals for One Mean with Tolerance Probability. You may then make the appropriate entries as listed below, or open Example 3 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N (Sample Size)</td>
</tr>
<tr>
<td>Confidence Level</td>
<td>0.95</td>
</tr>
<tr>
<td>Tolerance Probability</td>
<td>0.90</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Distance from Mean to Limit(s)</td>
<td>1.5</td>
</tr>
<tr>
<td>Standard Deviation Source</td>
<td>S from a Previous Sample</td>
</tr>
<tr>
<td>S</td>
<td>4.247211</td>
</tr>
<tr>
<td>Sample Size of Previous Sample</td>
<td>25</td>
</tr>
<tr>
<td>Interval Type</td>
<td>Two-Sided</td>
</tr>
<tr>
<td>Population Size</td>
<td>Infinite</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Numeric Results for Two-Sided Confidence Intervals</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidence Level</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>0.95</td>
</tr>
</tbody>
</table>

PASS also calculated the necessary sample size to be 53.
Example 4 – Validation using Harris, Horvitz, and Mood

Harris, Horvitz, and Mood (1948) pages 392-393 give an example of a sample size calculation for a two-sided confidence interval on the mean when the confidence level is 99%, the standard deviation is estimated to be 3 by a previous sample of size 9, the distance from the mean to the limit is 2, and the tolerance probability is 0.95. The necessary sample size is 49.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Confidence Intervals for One Mean with Tolerance Probability procedure window by expanding Means, then One Mean, then clicking on Confidence Interval, and then clicking on Confidence Intervals for One Mean with Tolerance Probability. You may then make the appropriate entries as listed below, or open Example 4 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N (Sample Size)</td>
</tr>
<tr>
<td>Confidence Level</td>
<td>0.99</td>
</tr>
<tr>
<td>Tolerance Probability</td>
<td>0.95</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Distance from Mean to Limit(s)</td>
<td>2</td>
</tr>
<tr>
<td>Standard Deviation Source</td>
<td>S from a Previous Sample</td>
</tr>
<tr>
<td>S</td>
<td>3</td>
</tr>
<tr>
<td>Sample Size of Previous Sample</td>
<td>9</td>
</tr>
<tr>
<td>Interval Type</td>
<td>Two-Sided</td>
</tr>
<tr>
<td>Population Size</td>
<td>Infinite</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>Sample Size (N)</th>
<th>Target Distance from Mean to Limits</th>
<th>Actual Distance from Mean to Limits</th>
<th>Standard Deviation (S)</th>
<th>Tolerance Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.99</td>
<td>49</td>
<td>2.000</td>
<td>1.999</td>
<td>3.000</td>
<td>0.95</td>
</tr>
</tbody>
</table>

PASS also calculated the necessary sample size to be 49.
Chapter 430

Inequality Tests for Two Means using Differences (Two-Sample T-Test)

Introduction

A common research task is to compare the means of two populations (groups) by taking independent samples from each. This is sometimes referred to as a parallel-groups design. This design is used in situations such as the comparison of the income level of two regions, the nitrogen content of two lakes, or the effectiveness of two drugs.

The mean represents the center of the population. If the means are different, then the populations are different. Other parameters of the two populations (such as the variance) can also be considered, but the mean is usually the starting point.

If assumptions about the other features of the two populations are met (such as that they are normally distributed and their variances are equal), the two-sample $t$ test can be used to compare the means of random samples drawn from these two populations. If the normality assumption is violated but the distributions are still symmetric, the nonparametric Mann-Whitney $U$ test may be used instead.

Test Procedure

Let the means of populations one and two be $\mu_1$ and $\mu_2$. Let $H_0$, the null hypothesis, represent the hypothesis that the two means are equal. That is, $H_0: \mu_1 - \mu_2 = 0$.

The formal steps in conducting a two-sample $t$ test and analyzing its power are as follows:
1. **Find the critical value.** Assume that the true difference between the means ( \( \mu_1 - \mu_2 \) ) is zero. Choose a value \( T_\alpha \) so that the probability of rejecting \( H_0 \) when \( H_0 \) is true is equal to a specified value, \( \alpha \). Using the \( t \) distribution, select \( T_\alpha \) so that \( \Pr(T > T_\alpha) = \alpha \).

Again, select \( T_\alpha \) so that if the means of the two populations are equal, \( t \) statistics calculated from two samples drawn from those populations will only exceed \( T_\alpha \) exactly 100\( \alpha \)% of the time.

![Figure 1 - Find the Critical Value](image)

2. **Conduct the experiment.** Select two samples of \( N_1 \) and \( N_2 \) items from the populations and compute the \( t \) value. Call this number \( T_S \).

3. **Look for statistical significance.** If \( T_S > T_\alpha \) reject the null hypothesis that \( \mu_1 - \mu_2 = 0 \) in favor of an alternative hypothesis that \( \mu_1 - \mu_2 = d > 0 \), where \( \mu_1 > \mu_2 \).

4. **Compute the power.** Now suppose that you want to compute the power of this test. First, you must specify an alternative value, \( d \), for the difference between the two means so that \( \mu_1 = \mu_2 + d \). You now consider a new probability distribution centered at \( d \) which is called the noncentral-\( t \) distribution. It appears as a bell-shaped curve as shown below.

The power is the probability of rejecting \( H_0 \) when the true difference is \( d \). Since we reject \( H_0 \) when our computed \( T_S \) value is greater than \( T_\alpha \), the power is the area under the noncentral-\( t \) curve to the right of \( T_\alpha \). The area to the left of \( T_\alpha \) represents the probability of a type-II error, or beta, since when the computed \( T_S \) value is less than \( T_\alpha \), we do not reject the false \( H_0 \).
Notice that in order to compute the power of the test, we must specify the true values of the means. Since we do not know these values, we compute the power at several possible values of $d$. This lets us understand what the power might have been.

Note that we can set the value of alpha (probability of a type-I error). However, we cannot set the value of beta (probability of a type-II error). Beta is computed based on a hypothesized value of $d$. We do not know what the value $d$ really is. So we can compute beta for a variety of $d$ values, but unless we know the true values of the population means, we do not know the true value of $d$, and hence, we do not know the true value of beta. This is why so much attention is paid to alpha, but so little attention is paid to beta.

**Assumptions**

The following assumptions are made when using the two-sample $t$ test or the Mann-Whitney $U$ test. One of the reasons for the popularity of the $t$ test is its robustness in the face of assumption violation. However, if an assumption is not met even approximately, the significance levels and the power of the $t$ test are unknown. Unfortunately, in practice it often happens that several assumptions are not met. This makes matters even worse! Hence, you should take the appropriate steps to check the assumptions before you make important decisions based on these tests.

**Two-Sample T Test Assumptions**

The assumptions of the two-sample $t$ test are:

1. The data are continuous (not discrete).
2. The data follow the normal probability distribution.
3. The variances of the two populations are equal. (If not, the Aspin-Welch Unequal-Variance test is used.)
4. The two samples are independent. There is no relationship between the individuals in one sample as compared to the other (as there is in the paired $t$ test).
5. Both samples are simple random samples from their respective populations. Each individual in the population has an equal probability of being selected in the sample.
Mann-Whitney U Test Assumptions
The assumptions of the Mann-Whitney U test for difference in means are:

1. The variable of interest is continuous (not discrete). The measurement scale is at least ordinal.
2. The probability distributions of the two populations are identical, except for location. That is, the variances are equal.
3. The two samples are independent.
4. Both samples are simple random samples from their respective populations. Each individual in the population has an equal probability of being selected in the sample.

Limitations
There are few limitations when using these tests. Sample sizes may range from a few to several hundred. If your data are discrete with at least five unique values, you can often ignore the continuous variable assumption. Perhaps the greatest restriction is that your data come from a random sample of the population. If you do not have a random sample, your significance levels will probably be incorrect.

Technical Details
There are four separate situations each requiring different formulas. Let the means of the two populations be represented by $\mu_1$ and $\mu_2$. The difference between these means will be represented by $d$. Let the standard deviations of the two populations be represented as $\sigma_1$ and $\sigma_2$.

Case 1 – Standard Deviations Known and Equal
When $\sigma_1 = \sigma_2 = \sigma$ and are known, the power of the t test is calculated as follows for a directional alternative (one-tailed test) in which $d > 0$.

1. Find $z_\alpha$ such that $1 - \Phi(z_\alpha) = \alpha$, where $\Phi(x)$ is the area under the standardized normal curve to the left of $x$.
2. Calculate: $\sigma_\tau = \sigma \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}$.
3. Calculate: $z_p = \frac{z_\alpha \sigma_\tau - d}{\sigma_\tau}$.
4. Calculate: $\text{Power} = 1 - \Phi(z_p)$. 

Case 2 – Standard Deviations Known and Unequal

When $\sigma_1 \neq \sigma_2$ and are known, the power is calculated as follows for a directional alternative (one-tailed test) in which $d > 0$.

1. Find $z_\alpha$ such that $1 - \Phi(z_\alpha) = \alpha$, where $\Phi(x)$ is the area under the standardized normal curve to the left of $x$.

2. Calculate: $\sigma_x = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$.

3. Calculate: $z_p = \frac{z_\alpha \sigma_x - d}{\sigma_x}$.

4. Calculate: Power = $1 - \Phi(z_p)$.

Case 3 – Standard Deviations Unknown and Equal

When $\sigma_1 = \sigma_2 = \sigma$ and are unknown, the power of the $t$ test is calculated as follows for a directional alternative (one-tailed test) in which $d > 0$.

1. Find $t_\alpha$ such that $1 - T_{df}(t_\alpha) = \alpha$, where $T_{df}(t_\alpha)$ is the area under a central-$t$ curve to the left of $x$ and $df = N_1 + N_2 - 2$.

2. Calculate: $\sigma_x = \sigma \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}$.

3. Calculate the noncentrality parameter: $\lambda = \frac{d}{\sigma_x}$.

4. Calculate: $t_p = \frac{t_\alpha \sigma_x - d}{\sigma_x} + \lambda$.

5. Calculate: Power = $1 - T_{df,\lambda'}(t_p)$, where $T_{df,\lambda'}(x)$ is the area under a noncentral-$t$ curve with degrees of freedom $df$ and noncentrality parameter $\lambda$ to the left of $x$. 
Case 4 – Standard Deviations Unknown and Unequal

When $\sigma_1 \neq \sigma_2$ and are unknown, the power is calculated as follows for a directional alternative (one-tailed test) in which $d > 0$. Note that in this case, an approximate $t$ test is used.

1. Calculate: $\sigma_\tau = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$.

2. Calculate: $f = \frac{\sigma_\tau^4}{\frac{\sigma_1^4}{N_1(N_1+1)} + \frac{\sigma_2^4}{N_2(N_2+1)}} - 2$.

which is the adjusted degrees of freedom. Often, this is rounded to the next highest integer.

3. Find $t_\alpha$ such that $1 - T_f(t_\alpha) = \alpha$, where $T_f(t_\alpha)$ is the area to the left of $x$ under a central-$t$ curve with $f$ degrees of freedom.

4. Calculate: $\lambda = \frac{d}{\sigma_\tau}$, the noncentrality parameter.

5. Calculate: $t_p = \frac{t_\alpha \sigma_\tau - d}{\sigma_\tau} + \lambda$.

6. Calculate: Power = $1 - T'_{f, \lambda}(t_p)$, where $T'_{f, \lambda}(x)$ is the area to the left of $x$ under a noncentral-$t$ curve with degrees of freedom $f$ and noncentrality parameter $\lambda$.

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains most of the parameters and options that you will be concerned with.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are Mean1, Mean2, Sigma1, Sigma2, Alpha, Power and Beta, N1, and N2. In most situations, you will select either Power and Beta or N1.

Select N1 when you want to calculate the sample size needed to achieve a given power and alpha level.

Select Power and Beta when you want to calculate the power of an experiment.
Error Rates

Power or Beta
This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of equal means when in fact the means are different.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

Alpha (Significance Level)
This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when you reject the null hypothesis of equal means when in fact the means are equal.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.

<table>
<thead>
<tr>
<th>Alpha</th>
<th>Approximate Odds of Rejecting a true null hypothesis</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.01</td>
<td>1 in 100</td>
</tr>
<tr>
<td>0.02</td>
<td>1 in 50</td>
</tr>
<tr>
<td>0.03</td>
<td>1 in 33</td>
</tr>
<tr>
<td>0.04</td>
<td>1 in 25</td>
</tr>
<tr>
<td>0.05</td>
<td>1 in 20</td>
</tr>
<tr>
<td>0.06</td>
<td>1 in 17</td>
</tr>
<tr>
<td>0.07</td>
<td>1 in 14</td>
</tr>
<tr>
<td>0.08</td>
<td>1 in 12</td>
</tr>
<tr>
<td>0.09</td>
<td>1 in 11</td>
</tr>
<tr>
<td>0.10</td>
<td>1 in 10</td>
</tr>
<tr>
<td>0.15</td>
<td>1 in 7</td>
</tr>
<tr>
<td>0.20</td>
<td>1 in 5</td>
</tr>
<tr>
<td>0.25</td>
<td>1 in 4</td>
</tr>
<tr>
<td>0.33</td>
<td>1 in 3</td>
</tr>
<tr>
<td>0.50</td>
<td>1 in 2</td>
</tr>
</tbody>
</table>

Sample Size

N1 (Sample Size Group 1)
Enter a value (or range of values) for the sample size of this group. Note that these values are ignored when you are solving for N1. You may enter a range of values such as 10 to 100 by 10.

N2 (Sample Size Group 2)
Enter a value (or range of values) for the sample size of group 2 or enter Use R to base N2 on the value of N1. You may enter a range of values such as 10 to 100 by 10.
Use R

When Use R is entered here, N2 is calculated using the formula

\[ N2 = \lceil R(N1) \rceil \]

where \( R \) is the Sample Allocation Ratio and the operator \( \lceil \cdot \rceil \) is the first integer greater than or equal to \( Y \). For example, if you want \( N1 = N2 \), select Use R and set \( R = 1 \).

**R (Sample Allocation Ratio)**

Enter a value (or range of values) for \( R \), the allocation ratio between samples. This value is only used when \( N2 \) is set to Use R.

When used, \( N2 \) is calculated from \( N1 \) using the formula: \( N2 = \lceil R(N1) \rceil \) where \( \lceil Y \rceil \) is the next integer greater than or equal to \( Y \). Note that setting \( R = 1.0 \) forces \( N2 = N1 \).

**Effect Size – Means**

**Mean1 (Mean of Group 1)**

This option specifies the mean of the first group. Under the null hypothesis of no difference between groups, the means of both groups are assumed to be equal. Hence, under the null hypothesis, this is also the mean of the second group.

**Mean2 (Mean of Group 2)**

This option specifies the mean of the second group in the alternative hypothesis. The difference between this value and the value of Mean1 represents the amount that is tested by the \( t \) test.

**Effect Size – Standard Deviations**

**S1 and S2 (Standard Deviations)**

These options specify the values of the standard deviations for each group. When the \( S2 \) is set to \( S1 \), only \( S1 \) needs to be specified. The value of \( S1 \) will be copied into \( S2 \).

When these values are not known, you must supply estimates of them. Press the SD button to display the Standard Deviation Estimator window. This procedure will help you find appropriate values for the standard deviation.

**Known Standard Deviation**

This option specifies whether the standard deviations (sigmas) are known or unknown. In almost all experimental situations, sigma is not known. However, since great calculation efficiencies are obtained if we can assume that sigma is known, and since this option has only a small impact on the final result, we usually leave it checked until we are ready for the final results.

When this box is checked, the program makes its calculations assuming that the standard deviations are known. This results in the use of the normal distribution in all probability calculations. Calculations using this option will be much faster than for the unknown sigma case. The results for either case will be close when the sample size is over 30.

When this box is not checked, the program assumes that sigma is not known and will be estimated from the data. This results in probability calculations using the noncentral-\( t \) distribution. This distribution requires a lot more calculations than does the normal distribution.
The calculation speed comes into play whenever the Find option is set to something besides Beta. In these cases, the program uses a special searching algorithm which requires many iterations. You will note a real difference in calculation speed depending on whether this option is checked.

A reasonable strategy would be to leave this option checked while you are experimenting with the parameters and then leave it unchecked when you are ready for your final results.

**Test**

**Alternative Hypothesis**

This option specifies the alternative hypothesis. This implicitly specifies the direction of the hypothesis test. The null hypothesis is always $H_0: \text{Mean}_1 = \text{Mean}_2$.

Note that the alternative hypothesis enters into power calculations by specifying the rejection region of the hypothesis test. Its accuracy is critical.

Possible selections are:

- **Ha: Mean1 \neq Mean2**
  
  This is the most common selection. It yields the *two-tailed* $t$ test. Use this option when you are testing whether the means are different, but you do not want to specify beforehand which mean is larger.

- **Ha: Mean1 < Mean2**

  This option yields a *one-tailed* $t$ test. Use it when you are only interested in the case in which $\text{Mean}_2$ is greater than $\text{Mean}_1$.

- **Ha: Mean1 > Mean2**

  This option yields a *one-tailed* $t$ test. Use it when you are only interested in the case in which $\text{Mean}_2$ is less than $\text{Mean}_1$.

**Nonparametric Adjustment (Mann-Whitney Test)**

This option lets you make sample size adjustments appropriate for when you are using the Mann-Whitney test rather than the $t$ test. Results by Al-Sunduqchi and Guenther (1990) indicate that power calculations for the Mann-Whitney test may be made using the standard $t$ test formulations with a simple adjustment to the sample sizes, $N_1$ and $N_2$. The size of the adjustment depends on the actual distribution of the data. They give sample size adjustment factors for four distributions. These are $1$ for uniform, $2/3$ for double exponential, $9 / \pi^2$ for logistic, and $\pi / 3$ for normal.

The options are as follows:

- **Ignore**
  
  Do not make a Mann-Whitney adjustment. This indicates that you want to analyze a $t$ test, not the Mann-Whitney test.

- **Uniform**

  Make the Mann-Whitney sample size adjustment assuming the uniform distribution. Since the factor is one, this option performs the same as Ignore. It is included for completeness.
• **Double Exponential**
  Make the Mann-Whitney sample size adjustment assuming the double exponential distribution.

• **Logistic**
  Make the Mann-Whitney sample size adjustment assuming the logistic distribution.

• **Normal**
  Make the Mann-Whitney sample size adjustment assuming the normal distribution.

---

**Example 1 – Power after a Study**

This example will cover the situation in which you are calculating the power of a $t$ test after the data have been collected.

A clinical trial was run to compare the effectiveness of two drugs. The ten responses in each group are shown below.

<table>
<thead>
<tr>
<th>Drug A</th>
<th>Drug B</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>15</td>
</tr>
<tr>
<td>20</td>
<td>17</td>
</tr>
<tr>
<td>25</td>
<td>17</td>
</tr>
<tr>
<td>20</td>
<td>19</td>
</tr>
<tr>
<td>23</td>
<td>22</td>
</tr>
<tr>
<td>20</td>
<td>12</td>
</tr>
<tr>
<td>13</td>
<td>16</td>
</tr>
<tr>
<td>18</td>
<td>21</td>
</tr>
<tr>
<td>25</td>
<td>20</td>
</tr>
<tr>
<td>24</td>
<td>19</td>
</tr>
</tbody>
</table>

These data were run through the **NCSS** statistical program with the following results.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Count</th>
<th>Mean</th>
<th>Standard Deviation</th>
<th>Standard Error</th>
<th>95% LCL of Mean</th>
<th>95% UCL of Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Drug A</td>
<td>10</td>
<td>20.9</td>
<td>3.665151</td>
<td>1.159023</td>
<td>18.27811</td>
<td>23.52189</td>
</tr>
<tr>
<td>Drug B</td>
<td>10</td>
<td>17.8</td>
<td>3.011091</td>
<td>0.9521905</td>
<td>15.646</td>
<td>19.954</td>
</tr>
<tr>
<td>Alternative Hypothesis (Drug A)-(Drug B)&lt;0</td>
<td>T Value</td>
<td>Prob Level</td>
<td>Decision (5%)</td>
<td>Accept Ho</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2.0667</td>
<td>0.053460</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notice that the probability level of 0.05346 is not significant. When a test is not significant, its power should be evaluated. The researchers decide to calculate the power using the sample values as estimates for the population values for various sample sizes and for alphas of 0.01 and 0.05.
Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means (Two-Sample T-Test) [Differences] procedure window by expanding Means, then Two Independent Means, then clicking on Test (Inequality), and then clicking on Tests for Two Means (Two-Sample T-Test) [Differences]. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.01 0.05</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>5 10 15 20 25 30 50</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Mean1 (Mean of Group 1)</td>
<td>20.9</td>
</tr>
<tr>
<td>Mean2 (Mean of Group 2)</td>
<td>17.8</td>
</tr>
<tr>
<td>S1 (Standard Deviation Group 1)</td>
<td>3.67</td>
</tr>
<tr>
<td>S2 (Standard Deviation Group 2)</td>
<td>3.01</td>
</tr>
<tr>
<td>Known Standard Deviation</td>
<td>Not checked</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Ha: Mean1 ≠ Mean2</td>
</tr>
<tr>
<td>Nonparametric Adjustment</td>
<td>Ignore</td>
</tr>
</tbody>
</table>

Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N1</th>
<th>N2</th>
<th>Ratio</th>
<th>Alpha</th>
<th>Beta</th>
<th>Mean1</th>
<th>Mean2</th>
<th>S1</th>
<th>S2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.08825</td>
<td>5</td>
<td>5</td>
<td>1.00</td>
<td>0.01000</td>
<td>0.91175</td>
<td>20.900</td>
<td>17.800</td>
<td>3.670</td>
<td>3.010</td>
</tr>
<tr>
<td>0.24642</td>
<td>10</td>
<td>10</td>
<td>1.00</td>
<td>0.01000</td>
<td>0.75358</td>
<td>20.900</td>
<td>17.800</td>
<td>3.670</td>
<td>3.010</td>
</tr>
<tr>
<td>0.42417</td>
<td>15</td>
<td>15</td>
<td>1.00</td>
<td>0.01000</td>
<td>0.57583</td>
<td>20.900</td>
<td>17.800</td>
<td>3.670</td>
<td>3.010</td>
</tr>
<tr>
<td>0.58661</td>
<td>20</td>
<td>20</td>
<td>1.00</td>
<td>0.01000</td>
<td>0.41339</td>
<td>20.900</td>
<td>17.800</td>
<td>3.670</td>
<td>3.010</td>
</tr>
<tr>
<td>0.71790</td>
<td>25</td>
<td>25</td>
<td>1.00</td>
<td>0.01000</td>
<td>0.28210</td>
<td>20.900</td>
<td>17.800</td>
<td>3.670</td>
<td>3.010</td>
</tr>
<tr>
<td>0.81541</td>
<td>30</td>
<td>30</td>
<td>1.00</td>
<td>0.01000</td>
<td>0.18459</td>
<td>20.900</td>
<td>17.800</td>
<td>3.670</td>
<td>3.010</td>
</tr>
<tr>
<td>0.97513</td>
<td>50</td>
<td>50</td>
<td>1.00</td>
<td>0.01000</td>
<td>0.02487</td>
<td>20.900</td>
<td>17.800</td>
<td>3.670</td>
<td>3.010</td>
</tr>
<tr>
<td>0.26033</td>
<td>5</td>
<td>5</td>
<td>1.00</td>
<td>0.05000</td>
<td>0.73967</td>
<td>20.900</td>
<td>17.800</td>
<td>3.670</td>
<td>3.010</td>
</tr>
<tr>
<td>0.50069</td>
<td>10</td>
<td>10</td>
<td>1.00</td>
<td>0.05000</td>
<td>0.49931</td>
<td>20.900</td>
<td>17.800</td>
<td>3.670</td>
<td>3.010</td>
</tr>
<tr>
<td>0.68601</td>
<td>15</td>
<td>15</td>
<td>1.00</td>
<td>0.05000</td>
<td>0.31399</td>
<td>20.900</td>
<td>17.800</td>
<td>3.670</td>
<td>3.010</td>
</tr>
<tr>
<td>0.81252</td>
<td>20</td>
<td>20</td>
<td>1.00</td>
<td>0.05000</td>
<td>0.18748</td>
<td>20.900</td>
<td>17.800</td>
<td>3.670</td>
<td>3.010</td>
</tr>
<tr>
<td>0.89246</td>
<td>25</td>
<td>25</td>
<td>1.00</td>
<td>0.05000</td>
<td>0.10754</td>
<td>20.900</td>
<td>17.800</td>
<td>3.670</td>
<td>3.010</td>
</tr>
<tr>
<td>0.94028</td>
<td>30</td>
<td>30</td>
<td>1.00</td>
<td>0.05000</td>
<td>0.05972</td>
<td>20.900</td>
<td>17.800</td>
<td>3.670</td>
<td>3.010</td>
</tr>
<tr>
<td>0.99550</td>
<td>50</td>
<td>50</td>
<td>1.00</td>
<td>0.05000</td>
<td>0.00450</td>
<td>20.900</td>
<td>17.800</td>
<td>3.670</td>
<td>3.010</td>
</tr>
</tbody>
</table>
430-12 Inequality Tests for Two Means using Differences (Two-Sample T-Test)

Report Definitions
Power is the probability of rejecting a false null hypothesis. Power should be close to one.
N1 and N2 are the number of items sampled from each population. To conserve resources, they should be small.
Alpha is the probability of rejecting a true null hypothesis. It should be small.
Beta is the probability of accepting a false null hypothesis. It should be small.
Mean1 is the mean of populations 1 and 2 under the null hypothesis of equality.
Mean2 is the mean of population 2 under the alternative hypothesis. The mean of population 1 is unchanged.
S1 and S2 are the population standard deviations. They represent the variability in the populations.

Summary Statements
Group sample sizes of 5 and 5 achieve 9% power to detect a difference of 3.1 between the null hypothesis that both group means are 20.9 and the alternative hypothesis that the mean of group 2 is 17.8 with estimated group standard deviations of 3.7 and 3.0 and with a significance level (alpha) of 0.01000 using a two-sided two-sample t-test.

This report shows the values of each of the parameters, one scenario per row. At alpha = 0.05 and N1 = 10, the power was only 0.50. The researchers only had a 50-50 chance of rejecting the null hypothesis in this case.

Plots Section

This plot shows the relationship between alpha and power in this example. Notice that the range of power values over the range of alpha values. Clearly, the sample size should have been doubled to twenty per group in order to achieve a power greater than 0.80.
Example 2 – Finding the Sample Size Necessary to Reject

Continuing with the last example, determine the sample size that the researchers would have needed for the null hypothesis to be rejected at the alpha = 0.01 and 0.05 levels, all other parameters remaining unchanged. They decided to use a beta error level of 0.20.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means (Two-Sample T-Test) [Differences] procedure window by expanding Means, then Two Independent Means, then clicking on Test (Inequality), and then clicking on Tests for Two Means (Two-Sample T-Test) [Differences]. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N1</td>
</tr>
<tr>
<td>Power</td>
<td>0.80</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.01 0.05</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>Ignored</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Mean1 (Mean of Group 1)</td>
<td>20.9</td>
</tr>
<tr>
<td>Mean2 (Mean of Group 2)</td>
<td>17.8</td>
</tr>
<tr>
<td>S1 (Standard Deviation Group 1)</td>
<td>3.67</td>
</tr>
<tr>
<td>S2 (Standard Deviation Group 2)</td>
<td>3.01</td>
</tr>
<tr>
<td>Known Standard Deviation</td>
<td>Not checked</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Ha: Mean1 ≠ Mean2</td>
</tr>
<tr>
<td>Nonparametric Adjustment</td>
<td>Ignore</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N1</th>
<th>N2</th>
<th>Ratio</th>
<th>Alpha</th>
<th>Beta</th>
<th>Mean1</th>
<th>Mean2</th>
<th>S1</th>
<th>S2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.81541</td>
<td>30</td>
<td>30</td>
<td>1.00</td>
<td>0.0100</td>
<td>0.18459</td>
<td>20.900</td>
<td>17.800</td>
<td>3.670</td>
<td>3.010</td>
</tr>
<tr>
<td>0.81252</td>
<td>20</td>
<td>20</td>
<td>1.00</td>
<td>0.0500</td>
<td>0.18748</td>
<td>20.900</td>
<td>17.800</td>
<td>3.670</td>
<td>3.010</td>
</tr>
</tbody>
</table>

We note that the required sample size is 20 when alpha is 0.05 and 30 when alpha is 0.01. Note that although the power was set at 0.80, the actual power achieved was 0.81. This is due to the fact that sample sizes must be integers, so specified power levels are not met exactly.
Example 3 – Minimum Detectable Difference

The minimum detectable difference is the difference between the two means that would be significant if all other parameters are kept at their experimental values. The minimum detectable difference is found by setting Mean1 to zero and solving for Mean2.

Continuing with the previous example, what is the minimum detectable difference when \( N1 = \ N2 = 10 \), \( \alpha = 0.05 \), \( \beta = 0.20 \), \( S1 = 3.67 \), and \( S2 = 3.01 \).

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means (Two-Sample T-Test) [Differences] procedure window by expanding Means, then Two Independent Means, then clicking on Test (Inequality), and then clicking on Tests for Two Means (Two-Sample T-Test) [Differences]. You may then make the appropriate entries as listed below, or open Example 3 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Mean2 (Search&gt;Mean1)</td>
</tr>
<tr>
<td>Power</td>
<td>0.80</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>10</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Mean1 (Mean of Group 1)</td>
<td>0</td>
</tr>
<tr>
<td>Mean2 (Mean of Group 2)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>S1 (Standard Deviation Group 1)</td>
<td>3.67</td>
</tr>
<tr>
<td>S2 (Standard Deviation Group 2)</td>
<td>3.01</td>
</tr>
<tr>
<td>Known Standard Deviation</td>
<td>Not checked</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Ha: Mean1 ≠ Mean2</td>
</tr>
<tr>
<td>Nonparametric Adjustment</td>
<td>Ignore</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N1</th>
<th>N2</th>
<th>Ratio</th>
<th>Alpha</th>
<th>Beta</th>
<th>Mean1</th>
<th>Mean2</th>
<th>S1</th>
<th>S2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.80000</td>
<td>10</td>
<td>10</td>
<td>1.00</td>
<td>0.05000</td>
<td>0.20000</td>
<td>0.000</td>
<td>4.431</td>
<td>3.670</td>
<td>3.010</td>
</tr>
</tbody>
</table>
The minimum detectable difference for this experiment is 4.431 minutes. If the true population means were this far apart, at a significance level of 0.05 and the power would be 0.80. Hence, the researchers should not have proceeded with the experiment if they thought the true difference was less than 4.431.

Example 4 – Finding the Sample Size

This example will show how the sample size for a new study is determined. A researcher decides to use a parallel-group design to study the impact of a new exercise program on body weight. Participants will be divided into two groups: those using and those not using the exercise program. Each participant’s weight loss (or gain) will be measured after three months. How many participants are needed to achieve 90% power at significance levels of 0.01 and 0.05?

Past experiments of this type have had standard deviations in the range of 10 to 15 pounds. The researcher wants to detect a difference of 15 pounds or more.

Although a drop in the mean is hypothesized, two-sided testing will be used because this is the standard method used and the researcher plans on publishing the results.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means (Two-Sample T-Test) [Differences] procedure window by expanding Means, then Two Independent Means, then clicking on Test (Inequality), and then clicking on Tests for Two Means (Two-Sample T-Test) [Differences]. You may then make the appropriate entries as listed below, or open Example 4 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Find (Solve For)</td>
<td>N1</td>
</tr>
<tr>
<td>Power</td>
<td>0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.01 0.05</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Mean1 (Mean of Group 1)</td>
<td>0</td>
</tr>
<tr>
<td>Mean2 (Mean of Group 2)</td>
<td>15</td>
</tr>
<tr>
<td>S1 (Standard Deviation Group 1)</td>
<td>10 12.5 15</td>
</tr>
<tr>
<td>S2 (Standard Deviation Group 2)</td>
<td>S1</td>
</tr>
<tr>
<td>Known Standard Deviation</td>
<td>Not checked</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Ha: Mean1 ≠ Mean2</td>
</tr>
<tr>
<td>Nonparametric Adjustment</td>
<td>Ignore</td>
</tr>
</tbody>
</table>
Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N1</th>
<th>N2</th>
<th>Ratio</th>
<th>Alpha</th>
<th>Beta</th>
<th>Mean1</th>
<th>Mean2</th>
<th>S1</th>
<th>S2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.90052</td>
<td>15</td>
<td>15</td>
<td>1.00</td>
<td>0.01000</td>
<td>0.09948</td>
<td>0.000</td>
<td>15.000</td>
<td>10.000</td>
<td>10.000</td>
</tr>
<tr>
<td>0.91690</td>
<td>11</td>
<td>11</td>
<td>1.00</td>
<td>0.05000</td>
<td>0.08310</td>
<td>0.000</td>
<td>15.000</td>
<td>10.000</td>
<td>10.000</td>
</tr>
<tr>
<td>0.90961</td>
<td>23</td>
<td>23</td>
<td>1.00</td>
<td>0.01000</td>
<td>0.09039</td>
<td>0.000</td>
<td>15.000</td>
<td>12.500</td>
<td>12.500</td>
</tr>
<tr>
<td>0.90719</td>
<td>16</td>
<td>16</td>
<td>1.00</td>
<td>0.05000</td>
<td>0.08281</td>
<td>0.000</td>
<td>15.000</td>
<td>12.500</td>
<td>12.500</td>
</tr>
<tr>
<td>0.90596</td>
<td>32</td>
<td>32</td>
<td>1.00</td>
<td>0.01000</td>
<td>0.08404</td>
<td>0.000</td>
<td>15.000</td>
<td>15.000</td>
<td>15.000</td>
</tr>
<tr>
<td>0.91250</td>
<td>23</td>
<td>23</td>
<td>1.00</td>
<td>0.05000</td>
<td>0.08750</td>
<td>0.000</td>
<td>15.000</td>
<td>15.000</td>
<td>15.000</td>
</tr>
</tbody>
</table>

After looking at these reports, the researcher decides to enroll 20 subjects per group and test the hypothesis at the 0.05 significance level. He chooses 20 because it is a little larger than the 16 that are required when the standard deviation is 12.5.

Example 5 – Mann-Whitney Test

The Mann-Whitney test is a popular nonparametric analog of the two-sample t test. It is recommended when the distribution of the data is not normal. A study by Al-Sunduqchi (1990) showed that sample size and power calculations for the Mann-Whitney test can be made using the standard t test results with an adjustment to the sample size.

Suppose that the researcher in Example 4 wants to compare sample size requirements of the t test with those of the Mann-Whitney test. To do this, he would use the same values, only this time the Nonparametric Adjustment would be set to a specific distribution. In this example, the double exponential is selected since it requires the largest adjustment of the distributions listed and the actual distribution is not known.
Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means (Two-Sample T-Test) [Differences] procedure window by expanding Means, then Two Independent Means, then clicking on Test (Inequality), and then clicking on Tests for Two Means (Two-Sample T-Test) [Differences]. You may then make the appropriate entries as listed below, or open Example 5 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N1</td>
</tr>
<tr>
<td>Power</td>
<td>0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.01 0.05</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Mean1 (Mean of Group 1)</td>
<td>0</td>
</tr>
<tr>
<td>Mean2 (Mean of Group 2)</td>
<td>15</td>
</tr>
<tr>
<td>S1 (Standard Deviation Group 1)</td>
<td>10 12.5 15</td>
</tr>
<tr>
<td>S2 (Standard Deviation Group 2)</td>
<td>S1</td>
</tr>
<tr>
<td>Known Standard Deviation</td>
<td>Not checked</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Ha: Mean1 ≠ Mean2</td>
</tr>
<tr>
<td>Nonparametric Adjustment</td>
<td>Double Exponential</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

| Numeric Results for Mann-Whitney Test (Double Exponenton Distribution) |
|-----------------------------|-----------------------------|
| Null Hypothesis: Mean1=Mean2. Alternative Hypothesis: Mean1<->Mean2     |
| The standard deviations were assumed to be unknown and equal.            |

<table>
<thead>
<tr>
<th>Power</th>
<th>N1</th>
<th>N2</th>
<th>Ratio</th>
<th>Alpha</th>
<th>Beta</th>
<th>Mean1</th>
<th>Mean2</th>
<th>S1</th>
<th>S2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.90052</td>
<td>10</td>
<td>10</td>
<td>1.000</td>
<td>0.01000</td>
<td>0.09948</td>
<td>0.000</td>
<td>15.000</td>
<td>10.000</td>
<td>10.000</td>
</tr>
<tr>
<td>0.91690</td>
<td>8</td>
<td>8</td>
<td>1.000</td>
<td>0.05000</td>
<td>0.08310</td>
<td>0.000</td>
<td>15.000</td>
<td>10.000</td>
<td>10.000</td>
</tr>
<tr>
<td>0.90961</td>
<td>16</td>
<td>16</td>
<td>1.000</td>
<td>0.01000</td>
<td>0.09039</td>
<td>0.000</td>
<td>15.000</td>
<td>12.500</td>
<td>12.500</td>
</tr>
<tr>
<td>0.90719</td>
<td>11</td>
<td>11</td>
<td>1.000</td>
<td>0.05000</td>
<td>0.09281</td>
<td>0.000</td>
<td>15.000</td>
<td>12.500</td>
<td>12.500</td>
</tr>
<tr>
<td>0.90596</td>
<td>22</td>
<td>22</td>
<td>1.000</td>
<td>0.01000</td>
<td>0.09404</td>
<td>0.000</td>
<td>15.000</td>
<td>15.000</td>
<td>15.000</td>
</tr>
<tr>
<td>0.91250</td>
<td>16</td>
<td>16</td>
<td>1.000</td>
<td>0.05000</td>
<td>0.08750</td>
<td>0.000</td>
<td>15.000</td>
<td>15.000</td>
<td>15.000</td>
</tr>
</tbody>
</table>

Comparing the sample sizes found here with those of the corresponding \( t \) test found in the last example at the 0.05 significance level, note that there is a reduction in the maximum sample size from 23 to 16. That is, if the Mann-Whitney test is used instead of the \( t \) test when the actual distribution follows the double exponential distribution, the sample size necessary to achieve 90% power at the 0.05 significance level is reduced from 23 to 16 per group.
Example 6 – Validation of Sample Size using Machin et al.

Machin et al. (1997) page 35 present an example in which the mean difference is 5, the common standard deviation is 10, the power is 90%, and the significance level is 0.05. They calculate the per group sample size as 86.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means (Two-Sample T-Test) [Differences] procedure window by expanding Means, then Two Independent Means, then clicking on Test (Inequality), and then clicking on Tests for Two Means (Two-Sample T-Test) [Differences]. You may then make the appropriate entries as listed below, or open Example 6 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N1</td>
</tr>
<tr>
<td>Power</td>
<td>0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Mean1 (Mean of Group 1)</td>
<td>0</td>
</tr>
<tr>
<td>Mean2 (Mean of Group 2)</td>
<td>5</td>
</tr>
<tr>
<td>S1 (Standard Deviation Group 1)</td>
<td>10</td>
</tr>
<tr>
<td>S2 (Standard Deviation Group 2)</td>
<td>S1</td>
</tr>
<tr>
<td>Known Standard Deviation</td>
<td>Not checked</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Ha: Mean1 ≠ Mean2</td>
</tr>
<tr>
<td>Nonparametric Adjustment</td>
<td>Ignore</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N1</th>
<th>N2</th>
<th>Allocation Ratio</th>
<th>Alpha</th>
<th>Beta</th>
<th>Mean1</th>
<th>Mean2</th>
<th>S1</th>
<th>S2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.90323</td>
<td>86</td>
<td>86</td>
<td>1.00</td>
<td>0.05000</td>
<td>0.09677</td>
<td>0.000</td>
<td>5.000</td>
<td>10.000</td>
<td>10.000</td>
</tr>
</tbody>
</table>

Note that the sample size of 86 per group matches Machin’s result exactly.
Example 7 – Validation using Zar

Zar (1984) page 136 give an example in which the mean difference is 1, the common standard deviation is 0.7206, the sample sizes are 15 in each group, and the significance level is 0.05. They calculate the power as 0.96.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means (Two-Sample T-Test) [Differences] procedure window by expanding Means, then Two Independent Means, then clicking on Test (Inequality), and then clicking on Tests for Two Means (Two-Sample T-Test) [Differences]. You may then make the appropriate entries as listed below, or open Example 7 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>15</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Mean1 (Mean of Group 1)</td>
<td>0</td>
</tr>
<tr>
<td>Mean2 (Mean of Group 2)</td>
<td>1</td>
</tr>
<tr>
<td>S1 (Standard Deviation Group 1)</td>
<td>0.7206</td>
</tr>
<tr>
<td>S2 (Standard Deviation Group 2)</td>
<td>S1</td>
</tr>
<tr>
<td>Known Standard Deviation</td>
<td>Not checked</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Ha: Mean1 ≠ Mean2</td>
</tr>
<tr>
<td>Nonparametric Adjustment</td>
<td>Ignore</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N1</th>
<th>N2</th>
<th>Allocation Ratio</th>
<th>Alpha</th>
<th>Beta</th>
<th>Mean1</th>
<th>Mean2</th>
<th>S1</th>
<th>S2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9561</td>
<td>15</td>
<td>15</td>
<td>1.00</td>
<td>0.05</td>
<td>0.04</td>
<td>1.00</td>
<td>1.00</td>
<td>0.72</td>
<td>0.72</td>
</tr>
</tbody>
</table>

Note that the power of 0.9561 matches Zar’s result of 0.96 to the two decimal places given.
430-20 Inequality Tests for Two Means using Differences (Two-Sample T-Test)
Chapter 431

Inequality Tests for Two Means in a Repeated Measures Design

Introduction

This module calculates the power for testing the time-averaged difference (TAD) between two means in a repeated measures design. A repeated measures design is one in which subjects are observed repeatedly over time. Measurements may be taken at pre-determined intervals (e.g. weekly or at specified time points following the administration of a particular treatment), or at random times so there are variable intervals between repeated measurements.

Time-averaged difference analysis is often used when the outcome to be measured varies with time. For example, suppose that you want to compare two treatment groups based on the means of a certain outcome such as blood pressure. It is known that a person's blood pressure depends on several instantaneous factors such as amount of sleep, excitement level, mood, exercise, etc. If only a single measurement is taken from each patient then the comparison of mean values from the two groups may be invalid because of the large degree of variation in blood pressure levels among patients. The precision of the experiment is increased by taking multiple measurements from each individual and comparing the time-averaged difference between the two groups. Care must be taken in the analysis because of the correlation that is introduced when several measurements are taken from the same individual. The covariance structure may take on several forms depending on the nature of the experiment and the subjects involved. This procedure allows you to calculate sample sizes using four different covariance patterns: Compound Symmetry, AR(1), Banded(1), and Simple.

This procedure can be used to calculate sample size and power for tests of pairwise contrasts in a mixed models analysis of repeated measures data. Mixed models analysis of repeated measures data is also employed to provide more flexibility in covariance specification and a greater degree of robustness in the presence of missing data, provided that the data can be assumed to be missing at random.
Technical Details

Theory and Notation

For a study with \(n_1\) subjects in group 1 and \(n_2\) subjects in group 2 (for a total of \(N\) subjects), each measured \(m\) times, the time-averaged difference \((d)\) of a continuous response between two groups can be estimated using the following model:

\[
y_{ij} = \beta_0 + \beta_1 x_i + \epsilon_{ij}, \quad i = 1, \ldots, N; \ j = 1, \ldots, m,\]

where

- \(y_{ij}\) is the \(j\)th response from subject \(i\),
- \(\beta_0\) is the model intercept,
- \(\beta_1\) is the treatment effect or the time-averaged difference between groups 1 and 2 (i.e. \(\beta_1 = d\)),
- \(x_i\) is a binary group assignment variable, which is equal to 1 if the \(i\)th subject is in group 1 and equal to 0 if the \(i\)th subject is in group 2, and
- \(\epsilon_{ij}\) is the normal, random error associated with the observation \(y_{ij}\).

Accounting for the relationship between repeated measurements, the model presented above can be written in matrix form as

\[
y_i = X_i \beta + \epsilon_i,\]

where

- \(y_i = (y_{i1} \ y_{i2} \ \cdots \ y_{im})\) is an \(m \times 1\) vector of responses from subject \(i\),
- \(X_i = \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & 1 \end{pmatrix}_{m \times 2}\) if the \(i\)th subject is in group 1,
- \(X_i = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ \vdots & \vdots \\ 1 & 0 \end{pmatrix}_{m \times 2}\) if the \(i\)th subject is in group 2,
- \(\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}\) is the vector of model parameters, and
- \(\epsilon_i \sim N(0, \sigma^2 R_i)\) is the vector of correlated random errors for the observations from subject \(i\),
Inequality Tests for Two Means in a Repeated Measures Design  431-3

where

\[ \text{var}(y_{ij}) = \sigma^2 \] is the residual variance for a single observation, and \( R \) is the \( m \times m \) common correlation matrix for all subjects. The contents of \( R \) depend on the assumed within-subject correlation structure.

We can stack the data in a single vector and matrix form as follows:

\[
y = (y_1', y_2', \ldots, y_N')'
\]

\[
X = (X_1', X_2', \ldots, X_N')'
\]

\[
\varepsilon = (\varepsilon_1', \varepsilon_2', \ldots, \varepsilon_N')'
\]

and the model for the \( N \) equations can be compressed into one as

\[
y = X'\beta + \varepsilon,
\]

with

\[
V = \text{var}(y)
\]

\[
= \sigma^2 \left(\begin{array}{ccc}
R_1 & 0 & 0 \\
0 & \ddots & 0 \\
0 & 0 & R_N
\end{array}\right)
\]

\[
= \sigma^2 R
\]

as the covariance (or variance - covariance) matrix.

**Covariance Pattern**

In a repeated measures design with \( N \) subjects, each measured \( m \) times, observations from a single subject may be correlated, and a pattern for their covariance must be specified. In this case, \( V \) will have a block-diagonal form:

\[
V = \begin{pmatrix}
V_1 & 0 & 0 & \cdots & 0 \\
0 & V_2 & 0 & \cdots & 0 \\
0 & 0 & V_3 & \cdots & 0 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
0 & 0 & 0 & \cdots & V_N
\end{pmatrix},
\]

where \( V_i \) are \( m \times m \) covariance matrices corresponding to the \( i^{th} \) subject. The 0's represent \( m \times m \) matrices of zeros giving zero covariances for observations on different subjects. This routine allows the specification of four different covariance matrix types: Compound Symmetry, AR(1), Banded(1), and Simple.
Compound Symmetry
A compound symmetry covariance model assumes that all covariances are equal, and all variances on the diagonal are equal. That is

$$V_i = \sigma^2 \begin{pmatrix} 1 & \rho & \rho & \cdots & \rho \\ \rho & 1 & \rho & \cdots & \rho \\ \rho & \rho & 1 & \cdots & \rho \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho & \rho & \rho & \cdots & 1 \end{pmatrix}$$

where $\sigma^2$ is the residual variance and $\rho$ is the correlation between observations on the same subject.

AR(1)
An AR(1) (autoregressive order 1) covariance model assumes that all variances on the diagonal are equal and that covariances $t$ time periods apart are equal to $\sigma^2 \rho^t$. That is

$$V_i = \sigma^2 \begin{pmatrix} 1 & \rho & \rho^2 & \rho^3 & \cdots & \rho^{m-1} \\ \rho & 1 & \rho^2 & \rho^3 & \cdots & \rho^{m-2} \\ \rho^2 & \rho & 1 & \rho & \cdots & \rho^{m-3} \\ \rho^3 & \rho^2 & \rho & 1 & \cdots & \rho^{m-4} \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ \rho^{m-1} & \rho^{m-2} & \rho^{m-3} & \rho^{m-4} & \cdots & 1 \end{pmatrix}$$

where $\sigma^2$ is the residual variance and $\rho$ is the correlation between observations on the same subject.

Banded(1)
A Banded(1) (banded order 1) covariance model assumes that all variances on the diagonal are equal, covariances for observations one time period apart are equal to $\sigma^2 \rho$, and covariances for measurements greater than one time period apart are equal to zero. That is

$$V_i = \sigma^2 \begin{pmatrix} 1 & \rho & 0 & 0 & \cdots & 0 \\ \rho & 1 & \rho & 0 & \cdots & 0 \\ 0 & \rho & 1 & \rho & \cdots & 0 \\ 0 & 0 & \rho & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 \end{pmatrix}$$

where $\sigma^2$ is the residual variance and $\rho$ is the correlation between observations on the same subject.
Simple

A simple covariance model assumes that all variances on the diagonal are equal and that all covariances are equal to zero. That is

\[ V_i = \sigma^2 \begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix}_{m \times m} \]

where \( \sigma^2 \) is the residual variance.

Model Estimation

With \( \hat{V} = \hat{\sigma}^2 \hat{R} \), then estimates of the regression coefficients from the above regression model are given as

\[ \hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix} = (X' \hat{V}^{-1} X)^{-1} X' \hat{V}^{-1} y \]

and the variance of \( \hat{\beta} \) is

\[ \text{var}(\hat{\beta}) = \begin{pmatrix} \text{var}(\hat{\beta}_0) & \text{cov}(\hat{\beta}_0, \hat{\beta}_1) \\ \text{cov}(\hat{\beta}_0, \hat{\beta}_1) & \text{var}(\hat{\beta}_1) \end{pmatrix} = (X' V^{-1} X)^{-1} \]

which is estimated by substituting \( \hat{V} \) for \( V \).

Hypothesis Test

A two-sided test that the time-averaged difference between the two groups is equal to zero is equivalent to the test of \( H_0 : \beta_1 = 0 \) vs. \( H_1 : \beta_1 \neq 0 \). Similarly, the upper and lower one-sided tests are \( H_0 : \beta_1 \leq 0 \) vs. \( H_1 : \beta_1 > 0 \) and \( H_0 : \beta_1 \geq 0 \) vs. \( H_1 : \beta_1 < 0 \), respectively. The test can be carried out using the test statistic

\[ z = \frac{\hat{\beta}_1}{\sqrt{\text{var}(\hat{\beta}_1)}} \rightarrow N(0,1). \]

If the standard deviation is unknown and estimated, a \( t \) test should be used. In practice, this test is often carried out by calculating the average response for each individual and then using a two-sample \( t \) test. If the data are balanced, the test can also be carried out in NCSS using Repeated Measures GLM and specifying a comparison such as "Each with First". In the case where the data
are not balanced, the test could be carried out using SAS® PROC MIXED or SAS® PROC GLM. In both cases a REPEATED statement should be used, along with a statement such as

\text{ESTIMATE 'A-B' treat 1 -1; or LSMEANS treat/ PDIFF;}

\section*{Power Calculations}

Sample sizes for repeated measures studies are often calculated as if a simple trial with no repeated measures was planned, which results in a higher calculated sample size than would be found if the correlation between repeated measures were taken into consideration. With an idea of the correct covariance structure, and an estimate of the within-patient correlation, you can get a better estimate of the power and sample size necessary to achieve your objectives. If you have no indication of the correct covariance structure for the experiment, then the compound symmetry (program default) is likely to be adequate. If you have no previous estimate of the within-patient correlation, then Brown and Prescott (2006) suggest using a conservative prediction of the correlation, i.e. a higher correlation than anticipated.

For a two-sided test where it is assumed that \( d > 0 \) (without loss of generality),

\[
\text{Power} = 1 - \beta = \Pr(\text{rejecting } H_0 | H_1) \\
= \Pr\left( \frac{\hat{\beta}_1}{\sqrt{\text{var}(\hat{\beta}_1)}} > z_{1-\alpha/2} | H_1 \right) \\
\approx \Pr\left( \frac{\hat{\beta}_1}{\sqrt{\text{var}(\hat{\beta}_1)}} > z_{1-\alpha/2} | H_1 \right) \text{ since it is assumed that } d > 0 \\
= \Pr\left( \frac{\hat{\beta}_1 - d}{\sqrt{\text{var}(\hat{\beta}_1)}} > z_{1-\alpha/2} - \frac{d}{\sqrt{\text{var}(\hat{\beta}_1)}} | H_1 \right) \\
= 1 - \Phi \left( z_{1-\alpha/2} - \frac{d}{\sqrt{\text{var}(\hat{\beta}_1)}} \right),
\]

where \( \Phi() \) is the standard normal density function, and \( \alpha \) and \( \beta \) are the probabilities of type I and type II error, respectively. For a one-sided test, \( \alpha \) is used in place of \( \alpha/2 \).

Since a \( t \) test is usually used to test for a group difference in a case such as this, we should note here that the power calculation using the standard normal distribution represents an approximation to the actual power achieved by the \( t \) test. We feel that it might be more appropriate to use the non-central \( t \) distribution; however, since the calculation is based on numerous assumptions about the covariance structure that influence the results, it seems unnecessary to worry about the small gain in precision that may occur by using the non-central \( t \) distribution. For this reason, along with the fact that this is the published method, we have elected to follow the methods of Brown and Prescott (2006), Liu and Wu (2005), Diggle et al. (1994) and use the standard normal distribution in power and sample size calculations.
Calculating Power for Testing Pairwise Contrasts of Fixed Effects in Mixed Models

Mixed Model Theory and Notation

A linear mixed model incorporates both fixed and random effects. Fixed effects are those effects in the model whose values are assumed constant, or unchanging. Random effects are those effects in the model that are assumed to have arisen from a distribution, resulting in another source of random variation other than residual variation. Brown and Prescott (2006) demonstrates how this methodology may be used to calculate the sample size and power for testing pairwise contrasts of fixed effects in a mixed models analysis of repeated measures data. For an experiment with $N$ subjects, $p$ fixed effect parameters, and $q$ random effect parameters, the general mixed model can be expressed using matrix notation as

$$y_i = X_i \beta + Z_i u_i + \varepsilon_i, \quad i = 1, \ldots, N$$

where

$y_i$ is an $n_i \times 1$ vector of responses for subject $i$,

$X_i$ is an $n_i \times p$, full-rank design matrix of fixed effects for subject $i$,

$\beta$ is an $p \times 1$ vector of fixed effects parameters,

$Z_i$ is an $n_i \times q$ design matrix of the random effects for subject $i$,

$u_i$ is a $q \times 1$ vector of random effects for subject $i$ which has means of zero and scaled covariance matrix $G_i$,

$\varepsilon_i$ is an $n_i \times 1$ vector of errors for subject $i$ with zero mean and scaled covariance $\Sigma_i$.

The covariance of $y_i$, $\text{var}(y_i) = V_i$, can be written as

$$V_i = \text{var}(X_i \beta + Z_i u_i + \varepsilon_i)$$

$$= Z_i \text{var}(u_i) Z_i' + \text{var}(\varepsilon_i)$$

$$= Z_i G_i Z_i' + \Sigma_i.$$ 

We can stack the data in a single vector and matrix form as follows:

$$y = (y_1, y_2, \ldots, y_N)'$$

$$X = (X_1, X_2, \ldots, X_N)'$$

$$Z = \begin{pmatrix}
Z_1 & 0 & 0 \\
0 & \ddots & 0 \\
0 & 0 & Z_N
\end{pmatrix}$$

$$u = (u_1, u_2, \ldots, u_N)'$$

$$\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N)'$$

and the mixed model for the $N$ equations can be compressed into one as

$$y = X\beta + Zu + \varepsilon,$$
Inequality Tests for Two Means in a Repeated Measures Design

with

$$V = \begin{pmatrix} V_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & V_N \end{pmatrix} = ZGZ' + \Sigma$$

where

$$\Sigma = \begin{pmatrix} \Sigma_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \Sigma_N \end{pmatrix}$$

is the covariance (or variance-covariance) matrix.

Mixed Model Estimation

Estimates of the variance components $G$ and $\Sigma$ are found using maximum likelihood (ML) or restricted/residual maximum likelihood (REML) methods. From these estimates, $\hat{G}$ and $\hat{\Sigma}$, an estimate of $V$ is obtained as $\hat{V} = ZGZ' + \hat{\Sigma}$. The fixed effects are then estimated as

$$\hat{\beta} = (X'\hat{V}^{-1}X)^{-1}X'\hat{V}^{-1}y$$

with the variance of $\hat{\beta}$ estimated as

$$\text{var}(\hat{\beta}) = (X'\hat{V}^{-1}X)^{-1}.$$ 

If $ZGZ = 0$ and $\Sigma = \sigma^2R$, then these estimation equations are identical to the TAD estimation equations presented earlier, except for the fact that $\hat{\beta}$ may contain more than two parameters, i.e. a parameter for each fixed effect being modeled. In the TAD model presented above, $\beta_i$ represents the difference between two treatment means, $d$. In the mixed model formulation presented here, $\beta_1, \beta_2, \ldots$ represent individual treatment effects. If there are no random effects, then we can use this routine for TAD to calculate the approximate power for testing pairwise contrasts of fixed effects in mixed models designs.

Brown and Prescott (2006) presents an example on page 228 of an experiment for which the power for testing pairwise contrasts can be calculated using this procedure. To determine the relative efficacy of three treatments in controlling hypertension, patients are assigned to one of the three treatments and blood pressure is measured at four follow-up visits. The study aims to determine the differences in average blood pressure among the three treatments.

Testing Fixed Effects

Significance tests for fixed effects can be done using tests based on the $t$ distribution. We can define tests of fixed and random effects as contrasts

$$C = L'\hat{\beta} = 0,$$

respectively. For example, in a trial containing three treatments A, B, and C, a pairwise comparison of treatments A and C is given by the contrast

$$C_{AC} = L'\hat{\beta} = (0 1 0 -1)\hat{\beta} = \hat{\beta}_A - \hat{\beta}_C,$$
where the first term in $\beta$ is the intercept term, and the other three terms are the treatment effects. For a single comparison, the test statistic is given by

$$t_{df} = \frac{L'\hat{\beta} - \hat{\beta}_h}{\text{SE}(L'\hat{\beta})} = \frac{\hat{\beta}_j - \hat{\beta}_h}{\text{SE}(\hat{\beta}_j - \hat{\beta}_h)},$$

where $df$ is the degrees of freedom, usually determined using the Satterthwaite approximation, and $\hat{\beta}_j$ and $\hat{\beta}_h$ ($j \neq h$) are estimated treatment effects.

Contrasts such as this can be tested in SAS® using the `ESTIMATE` statement or by including the `PDIFF` option in an `LSMEANS` statement. For example, if the variable designating three treatments, A, B, and C, were called "treat", then I could use the following statements in `PROC MIXED` to test for a difference between A and C

```sas
ESTIMATE 'A-C' treat 1 0 -1;
```

or

```sas
LSMEANS treat/ PDIFF;
```

The latter statement would produce tests of all pairwise comparisons of the levels of the treatment variable. The former would only test the difference between groups A and C. Of course, these comparison statements must be used in conjunction with appropriate model and class statements (see pages 233-237 of Brown and Prescott (2006) for an example analyzed using SAS® `PROC MIXED`).

Estimates of the correlation ($\rho$) and the standard deviation ($\sigma$) for use in power calculations can be found using SAS® `PROC MIXED`. For a model fit using compound symmetry, $\sigma^2$ and $\rho$ can be estimated as the sum of the variance parameters, and the compound symmetry variance parameter divided by the sum of the variance parameters, respectively. For AR(1), Banded(1), and Simple covariance models, $\sigma^2$ and $\rho$ can be estimated as the residual variance, and the correlation between adjacent measurements, respectively. Alternatively, the `R` and `RCORR` options may be used within the `REPEATED` statement to display the covariance and correlation matrices, from which the parameter estimates can be determined.
Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains most of the parameters and options that you will be concerned with.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for. When you choose to solve for \( n \), the program searches for the lowest sample size that meets the alpha and beta criterion you have specified for each of the terms. The "solve for" parameter is displayed on the vertical axis of the plot.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of equal means when in fact the means are different.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when you reject the null hypothesis of equal means when in fact the means are equal.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.

Sample Size

N1 (Sample Size Group 1)

Enter a value (or range of values) for the sample size of this group. Note that these values are ignored when you are solving for \( N1 \). You may enter a range of values such as 10 to 100 by 10.
N2 (Sample Size Group 2)
Enter a value (or range of values) for the sample size of group 2 or enter Use R to base N2 on the value of N1. You may enter a range of values such as 10 to 100 by 10.

- Use R
  When Use R is entered here, N2 is calculated using the formula
  \[ N2 = \lceil R(N1) \rceil \]
  where R is the Sample Allocation Ratio and the operator \( \lceil \cdot \rceil \) is the first integer greater than or equal to \( Y \). For example, if you want \( N1 = N2 \), select Use R and set \( R = 1 \).

R (Sample Allocation Ratio)
Enter a value (or range of values) for R, the allocation ratio between samples. This value is only used when N2 is set to Use R.

When used, N2 is calculated from N1 using the formula: \( N2 = \lceil R(N1) \rceil \) where \( \lceil \cdot \rceil \) is the next integer greater than or equal to \( Y \). Note that setting \( R = 1.0 \) forces \( N2 = N1 \).

Effect Size – Difference to Detect

D1 (Difference|H1)
Enter a value for the treatment difference to be detected. This difference represents a contrast of interest between two treatments in the study. You may enter a single value or a range of values such as 5 10 20 or 5 to 25 by 5. The items in the list may be separated with commas or blanks.

Effect Size – Repeated Measurements

M (Number of Time Points)
Enter a value for the number of time points (repeated measurements) at which each subject will be observed. You may enter a single value or a range of values such as 3 5 7 or 2 to 8 by 1. The items in the list may be separated with commas or blanks.

Effect Size – Covariance Structure

Covariance Type
Select the within-subject covariance structure that will be used in the mixed models analysis. The options are:

- Compound Symmetry
  All variances on the diagonal of the within-subject variance-covariance matrix are equal to \( \sigma^2 \), and all covariances are equal to \( \rho \sigma^2 \).

- AR(1)
  All variances on the diagonal of the within-subject variance-covariance matrix are equal to \( \sigma^2 \), and the covariance between observations \( t \) time periods apart is \( \rho^t \sigma^2 \).
• Banded(1)
  All variances on the diagonal of the within-subject variance-covariance matrix are equal to \( \sigma^2 \), and the covariance between observations one time period apart is \( \rho \sigma^2 \). Covariances between observations more than one time period apart are equal to zero.

• Simple
  All variances are equal to \( \sigma^2 \), and all covariances are equal to zero.

**Sigma (Std Dev of a Single Observation)**
Enter a value for the standard deviation (the square root of the residual variance). This standard deviation is assumed to be equal for the two groups. This parameter is equal to the square root of the sum of the variance parameters when compound symmetry is fit in a mixed models analysis of repeated measures data. This is equal to the square root of the residual variance parameter when an AR(1), Banded(1), or Simple model is fit in a mixed models analysis. You may enter a single value or a range of values such as 5 10 20 or 5 to 25 by 5. The items in the list may be separated with commas or blanks.

**Rho (Autocorrelation)**
Enter a value for the correlation between observations on the same subject. When no previous estimate of the within-patient correlation is available, you should use a conservative prediction of the correlation, i.e. a correlation that is higher than anticipated. You may enter a single value or a range of values such as 0.5 0.6 0.7 or 0.4 to 0.9 by 0.1. The items in the list may be separated with commas or blanks.

**Test**

**Alternative Hypothesis**
This option specifies the alternative hypothesis. This implicitly specifies the direction of the hypothesis test. The null hypothesis is always \( H_0 : d = 0 \).

Note that the alternative hypothesis enters into power calculations by specifying the rejection region of the hypothesis test. Its accuracy is critical.

Possible selections are:

• **One-Sided**
  This option yields a *one-tailed* test. Use it for testing the alternative hypotheses \( H_1 : d > 0 \) or \( H_1 : d < 0 \).

• **Two-Sided**
  This is the most common selection. It yields the *two-tailed* test. Use this option when you are testing whether the means are different, but you do not want to specify beforehand which mean is larger.
Example 1 – Determining Power

Researchers are planning a study of the impact of a new drug on heart rate. They want to evaluate the time-averaged difference in heart rate between subjects who take the new drug, and subjects who take the standard drug. Their experimental protocol calls for a baseline heart rate measurement, followed by administration of a certain level of the drug, followed by three additional measurements 30 minutes apart. They want to be able to detect a 10% difference in heart rate between the two treatments.

Similar studies have found an average heart rate of 93 for individuals taking the standard drug, a standard deviation of 9, and an autocorrelation between adjacent measurements on the same individual of 0.7. The researchers assume that first-order autocorrelation adequately represents the autocorrelation pattern. From a heart rate of 93, a 10% reduction gives 83.7, for a difference of 9.3. The test will be conducted at the 0.05 significance level.

What power does the study achieve over a range of possible sample sizes?

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means in a Repeated Measures Design procedure window by expanding Means, then Two Independent Means, then clicking on Repeated Measures, and then clicking on Tests for Two Means in a Repeated Measures Design. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Find</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Power</td>
<td>0.05</td>
</tr>
<tr>
<td>Alpha</td>
<td>4 to 20 by 2</td>
</tr>
<tr>
<td>N1</td>
<td>1.0</td>
</tr>
<tr>
<td>N2</td>
<td>9.3</td>
</tr>
<tr>
<td>R</td>
<td>4</td>
</tr>
<tr>
<td>Covariance Type</td>
<td>AR(1)</td>
</tr>
<tr>
<td>Sigma</td>
<td>0.7</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Two-Sided</td>
</tr>
</tbody>
</table>
Annoted Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>Group 1 Sample Size (N1)</th>
<th>Group 2 Sample Size (N2)</th>
<th>Sample Allocation Ratio (R)</th>
<th>Sample Time Points (M)</th>
<th>Difference to be Detected (D1)</th>
<th>Standard Deviation (Sigma)</th>
<th>Auto-corr. (Rho)</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.42660</td>
<td>4</td>
<td>4</td>
<td>1.000</td>
<td>4</td>
<td>9.300</td>
<td>9.000</td>
<td>0.700</td>
<td>0.050</td>
<td>0.57340</td>
</tr>
<tr>
<td>0.58468</td>
<td>6</td>
<td>6</td>
<td>1.000</td>
<td>4</td>
<td>9.300</td>
<td>9.000</td>
<td>0.700</td>
<td>0.050</td>
<td>0.41532</td>
</tr>
<tr>
<td>0.70890</td>
<td>8</td>
<td>8</td>
<td>1.000</td>
<td>4</td>
<td>9.300</td>
<td>9.000</td>
<td>0.700</td>
<td>0.050</td>
<td>0.29110</td>
</tr>
<tr>
<td>0.80135</td>
<td>10</td>
<td>10</td>
<td>1.000</td>
<td>4</td>
<td>9.300</td>
<td>9.000</td>
<td>0.700</td>
<td>0.050</td>
<td>0.19865</td>
</tr>
<tr>
<td>0.86742</td>
<td>12</td>
<td>12</td>
<td>1.000</td>
<td>4</td>
<td>9.300</td>
<td>9.000</td>
<td>0.700</td>
<td>0.050</td>
<td>0.13258</td>
</tr>
<tr>
<td>0.91318</td>
<td>14</td>
<td>14</td>
<td>1.000</td>
<td>4</td>
<td>9.300</td>
<td>9.000</td>
<td>0.700</td>
<td>0.050</td>
<td>0.08692</td>
</tr>
<tr>
<td>0.94407</td>
<td>16</td>
<td>16</td>
<td>1.000</td>
<td>4</td>
<td>9.300</td>
<td>9.000</td>
<td>0.700</td>
<td>0.050</td>
<td>0.05593</td>
</tr>
<tr>
<td>0.96448</td>
<td>18</td>
<td>18</td>
<td>1.000</td>
<td>4</td>
<td>9.300</td>
<td>9.000</td>
<td>0.700</td>
<td>0.050</td>
<td>0.03552</td>
</tr>
<tr>
<td>0.97773</td>
<td>20</td>
<td>20</td>
<td>1.000</td>
<td>4</td>
<td>9.300</td>
<td>9.000</td>
<td>0.700</td>
<td>0.050</td>
<td>0.02227</td>
</tr>
</tbody>
</table>

References


Report Definitions

Power is the probability of rejecting a false null hypothesis. It should be close to one. N1 & N2 are the number of subjects in groups 1 and 2, respectively. R is the ratio of the number of subjects in group 2 to the number in group 1 (R = N2/N1). M is the number of time points (repeated measurements) at which each subject is observed. D1 is the difference between the means of groups 1 and 2 under the alternative hypothesis. Sigma is the standard deviation of a single observation. It is the same for both groups. Rho is the correlation between observations on the same subject. Alpha is the probability of rejecting a true null hypothesis. It should be small. Beta is the probability of accepting a false null hypothesis. It should be small.

Summary Statements

Group sample sizes of 4 and 4 achieve 43% power to detect a difference of 9.300 in a design with 4 repeated measurements having a AR(1) covariance structure when the standard deviation is 9.000, the correlation between observations on the same subject is 0.700, and the alpha level is 0.050.

This report gives the power for each value of the other parameters.

Power

This is the computed power for detecting the time-averaged difference between the two group means.

Group 1 Sample Size (N1)

The value of N1 is the number of subjects in group 1.

Group 2 Sample Size (N2)

The value of N2 is the number of subjects in group 2.
**Sample Allocation Ratio (R)**
This is the ratio of the number of subjects in group 2 to the number in group 1 ($R = N_2/N_1$).

**Time Points (M)**
This is the number of repeated measurements taken.

**Difference to be Detected (D1)**
This is the treatment difference that is to be detected.

**Standard Deviation (Sigma)**
This is the value of $\sigma$, the standard deviation or the square root of the residual variance.

**Autocorr. (Rho)**
This is the correlation between observations from the same subject.

**Alpha**
Alpha is the significance level of the test.

**Beta**
Beta is the probability of failing to reject the null hypothesis when the alternative hypothesis is true.

---

**Plots Section**

The chart shows the relationship between power and $N_1$ when the other parameters in the design are held constant.
Example 2 – Finding the Sample Size

Continuing with Example 1, the researchers want to determine the exact sample size necessary to achieve at least 80% power.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means in a Repeated Measures Design procedure window by expanding Means, then Two Independent Means, then clicking on Repeated Measures, and then clicking on Tests for Two Means in a Repeated Measures Design. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find</td>
<td>N1 (Group 1 Sample Size)</td>
</tr>
<tr>
<td>Power</td>
<td>0.8</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N1</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>N2</td>
<td>Use R</td>
</tr>
<tr>
<td>R</td>
<td>1.0</td>
</tr>
<tr>
<td>D1</td>
<td>9.3</td>
</tr>
<tr>
<td>M</td>
<td>4</td>
</tr>
<tr>
<td>Covariance Type</td>
<td>AR(1)</td>
</tr>
<tr>
<td>Sigma</td>
<td>9</td>
</tr>
<tr>
<td>Rho</td>
<td>0.7</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Two-Sided</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

## Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>Group 1 Sample Size (N1)</th>
<th>Group 2 Sample Size (N2)</th>
<th>Sample Allocation Ratio (R)</th>
<th>Time Points (M)</th>
<th>Difference to be Detected (D1)</th>
<th>Standard Deviation (Sigma)</th>
<th>Auto-corr. (Rho)</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.80135</td>
<td>10</td>
<td>10</td>
<td>1.000</td>
<td>4</td>
<td>9.300</td>
<td>9.000</td>
<td>0.700</td>
<td>0.050</td>
<td>0.19865</td>
</tr>
</tbody>
</table>

A group sample size of 10 is required to achieve at least 80% power.
Example 3 – Varying the Difference between the Means

Continuing with Examples 1 and 2, the researchers want to evaluate the impact on power of varying the size of the difference between the means for a range of sample sizes from 2 to 8 per group. In the output to follow, we only display the plots. You may want to display the numeric reports as well, but we do not here in order to save space.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means in a Repeated Measures Design procedure window by expanding Means, then Two Independent Means, then clicking on Repeated Measures, and then clicking on Tests for Two Means in a Repeated Measures Design. You may then make the appropriate entries as listed below, or open Example 3 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Find</td>
<td>.Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N1</td>
<td>4 to 8 by 1</td>
</tr>
<tr>
<td>N2</td>
<td>Use R</td>
</tr>
<tr>
<td>R</td>
<td>1.0</td>
</tr>
<tr>
<td>D1</td>
<td>4 to 11 by 1</td>
</tr>
<tr>
<td>M</td>
<td>4</td>
</tr>
<tr>
<td>Covariance Type</td>
<td>AR(1)</td>
</tr>
<tr>
<td>Sigma</td>
<td>9</td>
</tr>
<tr>
<td>Rho</td>
<td>0.7</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Two-Sided</td>
</tr>
</tbody>
</table>
Output
Click the Run button to perform the calculations and generate the following output.

Plots Section

This chart shows how the power depends on the difference to be detected, \( d \), as well as the group sample size, \( n_1 \).

Example 4 – Impact of the Number of Repeated Measurements

Continuing with Examples 1 - 3, the researchers want to study the impact on the sample size if they changing the number of measurements made on each individual. Their experimental protocol calls for four measurements that are 30 minutes apart. They want to see the impact of taking twice that many measurements.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means in a Repeated Measures Design procedure window by expanding Means, then Two Independent Means, then clicking on Repeated Measures, and then clicking on Tests for Two Means in a Repeated Measures Design. You may then make the appropriate entries as listed below, or open Example 4 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find</td>
<td>( N_1 ) (Group 1 Sample Size)</td>
</tr>
<tr>
<td>Power</td>
<td>0.8</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>( N_1 )</td>
<td>Ignored since this is the Find setting</td>
</tr>
</tbody>
</table>
Inequality Tests for Two Means in a Repeated Measures Design 431-19

Data Tab (continued)

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Use R</td>
<td></td>
</tr>
<tr>
<td>R</td>
<td>1.0</td>
</tr>
<tr>
<td>D1</td>
<td>9.3</td>
</tr>
<tr>
<td>M</td>
<td>4.8</td>
</tr>
<tr>
<td>Covariance Type</td>
<td>AR(1)</td>
</tr>
<tr>
<td>Sigma</td>
<td>9.0</td>
</tr>
<tr>
<td>Rho</td>
<td>0.7</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Two-Sided</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Group 1 Sample Size</th>
<th>Group 2 Sample Size</th>
<th>Sample Allocation Ratio</th>
<th>Time Points</th>
<th>Difference to be Detected</th>
<th>Standard Deviation</th>
<th>Auto-cor. (Rho)</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.80135</td>
<td>10</td>
<td>10</td>
<td>1.000</td>
<td>4</td>
<td>9.300</td>
<td>9.000</td>
<td>0.700</td>
<td>0.19865</td>
</tr>
<tr>
<td>0.84737</td>
<td>8</td>
<td>8</td>
<td>1.000</td>
<td>8</td>
<td>9.300</td>
<td>9.000</td>
<td>0.700</td>
<td>0.15283</td>
</tr>
</tbody>
</table>

Doubling the number of repeated measurements per individual decreases the group sample size by 2. This reduction in sample size may not justify the additional four measurements on each subject.

Example 5 – Validation using Diggle et al.

Diggle et al. (1994) page 31 presents an example of calculating the sample size for a TAD study. They calculate the group sample sizes for the cases where \( \frac{d}{\sigma} \) ranges from 0.2 to 0.5, \( \rho \) ranges from 0.2 to 0.8, \( \alpha = 0.05 \), \( M = 3 \), and power = 0.8. Note that Diggle et al. (1994) uses a one-sided test.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means in a Repeated Measures Design procedure window by expanding Means, then Two Independent Means, then clicking on Repeated Measures, and then clicking on Tests for Two Means in a Repeated Measures Design. You may then make the appropriate entries as listed below, or open Example 5 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab Find</td>
<td>N1 (Group 1 Sample Size)</td>
</tr>
<tr>
<td>Power</td>
<td>0.8</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N1</td>
<td>Ignored since this is the Find setting</td>
</tr>
</tbody>
</table>
431-20  Inequality Tests for Two Means in a Repeated Measures Design

Data Tab (continued)
N2 ........................................................... Use R
R ............................................................ 1.0
D1 ........................................................... 0.2 to 0.5 by 0.1
M ........................................................... 3
Covariance Type ........................................ Compound Symmetry
Sigma ....................................................... 1
Rho .......................................................... 0.2 0.5 0.8
Alternative Hypothesis ............................ One-Sided

Output
Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>Group 1 Sample Size (N1)</th>
<th>Group 2 Sample Size (N2)</th>
<th>Sample Allocation Ratio (R)</th>
<th>Time Points (M)</th>
<th>Difference to be Detected (D1)</th>
<th>Standard Deviation (Sigma)</th>
<th>Auto-corr. (Rho)</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.80178</td>
<td>145</td>
<td>145</td>
<td>1.000</td>
<td>3</td>
<td>0.200</td>
<td>1.000</td>
<td>0.200</td>
<td>0.050</td>
<td>0.19822</td>
</tr>
<tr>
<td>0.80154</td>
<td>207</td>
<td>207</td>
<td>1.000</td>
<td>3</td>
<td>0.200</td>
<td>1.000</td>
<td>0.500</td>
<td>0.050</td>
<td>0.19846</td>
</tr>
<tr>
<td>0.80012</td>
<td>288</td>
<td>288</td>
<td>1.000</td>
<td>3</td>
<td>0.200</td>
<td>1.000</td>
<td>0.800</td>
<td>0.050</td>
<td>0.19988</td>
</tr>
<tr>
<td>0.80475</td>
<td>65</td>
<td>65</td>
<td>1.000</td>
<td>3</td>
<td>0.300</td>
<td>1.000</td>
<td>0.200</td>
<td>0.050</td>
<td>0.19625</td>
</tr>
<tr>
<td>0.80154</td>
<td>92</td>
<td>92</td>
<td>1.000</td>
<td>3</td>
<td>0.300</td>
<td>1.000</td>
<td>0.500</td>
<td>0.050</td>
<td>0.19846</td>
</tr>
<tr>
<td>0.80270</td>
<td>120</td>
<td>120</td>
<td>1.000</td>
<td>3</td>
<td>0.300</td>
<td>1.000</td>
<td>0.800</td>
<td>0.050</td>
<td>0.19730</td>
</tr>
<tr>
<td>0.80885</td>
<td>37</td>
<td>37</td>
<td>1.000</td>
<td>3</td>
<td>0.400</td>
<td>1.000</td>
<td>0.200</td>
<td>0.050</td>
<td>0.19115</td>
</tr>
<tr>
<td>0.80321</td>
<td>52</td>
<td>52</td>
<td>1.000</td>
<td>3</td>
<td>0.400</td>
<td>1.000</td>
<td>0.500</td>
<td>0.050</td>
<td>0.19679</td>
</tr>
<tr>
<td>0.80012</td>
<td>67</td>
<td>67</td>
<td>1.000</td>
<td>3</td>
<td>0.400</td>
<td>1.000</td>
<td>0.800</td>
<td>0.050</td>
<td>0.19988</td>
</tr>
<tr>
<td>0.81343</td>
<td>24</td>
<td>24</td>
<td>1.000</td>
<td>3</td>
<td>0.500</td>
<td>1.000</td>
<td>0.200</td>
<td>0.050</td>
<td>0.19657</td>
</tr>
<tr>
<td>0.80028</td>
<td>33</td>
<td>33</td>
<td>1.000</td>
<td>3</td>
<td>0.500</td>
<td>1.000</td>
<td>0.800</td>
<td>0.050</td>
<td>0.19972</td>
</tr>
<tr>
<td>0.80109</td>
<td>43</td>
<td>43</td>
<td>1.000</td>
<td>3</td>
<td>0.500</td>
<td>1.000</td>
<td>0.800</td>
<td>0.050</td>
<td>0.19891</td>
</tr>
</tbody>
</table>

The sample sizes calculated by PASS match the results of Diggle et al. (1994) very closely, with slight differences due to rounding. If you calculate the sample sizes by hand, using the formula given in Diggle et al. (1994), page 31, your answers will match those of PASS.


Brown and Prescott (2006) pages 268 and 269 presents an example of calculating the sample size for pairwise contrasts in a hypertension trial to be analyzed using mixed models. The analysis of repeated DBP measurements from four post-treatment visits using a compound symmetry covariance pattern resulted in the following covariance matrix for each subject:

\[
V = \begin{pmatrix}
  1 & 0.53 & 0.53 & 0.53 \\
  0.53 & 1 & 0.53 & 0.53 \\
  0.53 & 0.53 & 1 & 0.53 \\
  0.53 & 0.53 & 0.53 & 1 \\
\end{pmatrix}
\]

\[V_i = 76\]

From this matrix they determine that \( \rho = 0.53 \) and \( \sigma^2 = 76 \) (\( \sigma = 8.718 \)).
The trial followed several hundred patients given one of three treatments. Brown and Prescott calculate the group sample size to be 31 for a future study involving four post-treatment visits to detect a difference in DBP of 5 mmHg at the 5% significance level with 80% power.

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means in a Repeated Measures Design procedure window by expanding Means, then Two Independent Means, then clicking on Repeated Measures, and then clicking on Tests for Two Means in a Repeated Measures Design. You may then make the appropriate entries as listed below, or open Example 6 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find</td>
<td>N1 (Group 1 Sample Size)</td>
</tr>
<tr>
<td>Power</td>
<td>0.8</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N1</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>N2</td>
<td>Use R</td>
</tr>
<tr>
<td>R</td>
<td>1.0</td>
</tr>
<tr>
<td>D1</td>
<td>5</td>
</tr>
<tr>
<td>M</td>
<td>4</td>
</tr>
<tr>
<td>Covariance Type</td>
<td>Compound Symmetry</td>
</tr>
<tr>
<td>Sigma</td>
<td>8.718</td>
</tr>
<tr>
<td>Rho</td>
<td>0.53</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Two-Sided</td>
</tr>
</tbody>
</table>

**Output**

Click the Run button to perform the calculations and generate the following output.

**Numeric Results for M = 4**

<table>
<thead>
<tr>
<th>Power</th>
<th>Group 1 Sample Size (N1)</th>
<th>Group 2 Sample Size (N2)</th>
<th>Sample Allocation Ratio (R)</th>
<th>Time Points (M)</th>
<th>Difference to be Detected (D1)</th>
<th>Standard Deviation (Sigma)</th>
<th>Auto-cor. (Rho)</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.80125</td>
<td>31</td>
<td>31</td>
<td>1.000</td>
<td>4</td>
<td>5.000</td>
<td>8.718</td>
<td>0.530</td>
<td>0.050</td>
<td>0.19875</td>
</tr>
</tbody>
</table>

The sample size of 31 calculated by **PASS** matches the results of Brown and Prescott (2006) exactly.

Brown and Prescott further calculate the sample size for the case where no account is taken of repeated measurements and the case of 10 repeated measurements. If we change the number of repeated measurements to 1 and 10, we get the following output (Example6b template):
431-22 Inequality Tests for Two Means in a Repeated Measures Design

Numeric Results for M = 1, 10

<table>
<thead>
<tr>
<th>Power</th>
<th>Group 1 Sample Size (N1)</th>
<th>Group 2 Sample Size (N2)</th>
<th>Sample Allocation Ratio (R)</th>
<th>Time Points (M)</th>
<th>Difference to be Detected (D1)</th>
<th>Standard Deviation (Sigma)</th>
<th>Auto-cor. (Rho)</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.80226</td>
<td>48</td>
<td>48</td>
<td>1.000</td>
<td>1</td>
<td>5.000</td>
<td>8.718</td>
<td>0.530</td>
<td>0.050</td>
<td>0.19774</td>
</tr>
<tr>
<td>0.80651</td>
<td>28</td>
<td>28</td>
<td>1.000</td>
<td>10</td>
<td>5.000</td>
<td>8.718</td>
<td>0.530</td>
<td>0.050</td>
<td>0.19349</td>
</tr>
</tbody>
</table>

In both cases, the results of P\textit{ASS} match those of Brown and Prescott (2006) exactly.
**Chapter 435**

**Inequality Tests for Two Exponential Means**

**Introduction**

This program module designs studies for testing hypotheses about the means of two exponential distributions. Such a test is used when you want to make a comparison between two groups that both follow the exponential distribution. The responses from the samples are assumed to be continuous, positive numbers such as lifetime.

We adopt the basic methodology outlined in the books by Bain and Engelhardt (1991) and Desu and Raghavarao (1990).

**Technical Details**

The test procedure described here makes the assumption that lifetimes in each group follow an exponential distribution. The densities of the two exponential distributions are written as

\[
 f_i(t) = \frac{1}{\theta_i} \exp\left(-\frac{t}{\theta_i}\right), \quad i = 1, 2
\]

The parameters \( \theta_i \) are interpreted as the average failure times, the mean time to failure (MTTF), or the mean time between failures (MTBF) of the two groups. The reliability, or the probability that a unit continues running beyond time \( t \), is

\[
 R_i(t) = e^{-\frac{t}{\theta_i}}
\]
Hypothesis Test

The relevant statistical hypothesis is $H_0: \frac{\theta_1}{\theta_2} = 1$ versus one of the following alternatives:

$H_A: \frac{\theta_1}{\theta_2} = \rho > 1$, $H_A: \frac{\theta_1}{\theta_2} = \rho < 1$, or $H_A: \frac{\theta_1}{\theta_2} = \rho \neq 1$. The test procedure is to reject the null hypothesis $H_0$ if the ratio of the observed mean lifetimes $\hat{\rho} = \hat{\theta}_1 / \hat{\theta}_2$ is too large or too small.

The samples of size $n_i$ are assumed to be drawn without replacement. The experiment is run until all items fail.

If the experiment is curtailed before all $n_1 + n_2$ items fail, the sample size results are based on the number of failures $r_1 + r_2$, not the total number of samples $n_1 + n_2$.

The mean lifetimes are estimated as follows

$$\hat{\theta}_i = \frac{\sum_{j} t_{ij}}{r_i}, \quad i = 1, 2$$

where $t_{ij}$ is the time that the $j$th item in the $i$th group is tested, whether measured until failure or until the study is completed.

Power and sample size calculations are based on the fact that the estimated lifetime ratio is proportional to the $F$ distribution. That is,

$$\frac{\hat{\theta}_1}{\hat{\theta}_2} \sim \frac{\theta_1}{\theta_2} F_{r_1, r_2}$$

which, under the null hypothesis of equality, becomes

$$\frac{\hat{\theta}_1}{\hat{\theta}_2} \sim F_{r_1, r_2}$$

Note that only the actual numbers of failures are used in these distributions. Hence, we assume that the experiment is run until all items fail so that $r_i = n_i$. That is, the sample sizes are the number of failures, not the number of items. Enough units must be sampled to ensure that the stated number of failures occur.
Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains most of the parameters and options that you will be concerned with.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for from the other parameters. Under most situations, you will select either *Power and Beta* for a power analysis or *N* for sample size determination.

Select *N* when you want to calculate the sample size needed to achieve a given power and alpha level. Select *Power and Beta* when you want to calculate the power of an experiment.

Error Rates

Power or Beta (Beta is Consumer’s Risk)

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta (consumer’s risk) is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of equal thetas when in fact they are different.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as *0.8 to 0.95 by 0.05* may be entered.

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when you reject the null hypothesis of equal thetas when in fact they are equal.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as *0.01 0.05 0.10* or *0.01 to 0.10 by 0.01*. 
Sample Size

N1 (Sample Size Group 1)
Enter a value (or range of values) for the sample size of this group. Note that these values are ignored when you are solving for \( N1 \). You may enter a range of values such as 10 to 100 by 10.

N2 (Sample Size Group 2)
Enter a value (or range of values) for the sample size of group 2 or enter Use R to base \( N2 \) on the value of \( N1 \). You may enter a range of values such as 10 to 100 by 10.

- **Use R**
  
  When Use R is entered here, \( N2 \) is calculated using the formula
  \[
  N2 = \lceil R(N1) \rceil
  \]
  
  where \( R \) is the Sample Allocation Ratio and the operator \( \lceil Y \rceil \) is the first integer greater than or equal to \( Y \). For example, if you want \( N1 = N2 \), select Use R and set \( R = 1 \).

R (Sample Allocation Ratio)
Enter a value (or range of values) for \( R \), the allocation ratio between samples. This value is only used when \( N2 \) is set to Use R.

When used, \( N2 \) is calculated from \( N1 \) using the formula: \( N2 = \lceil R(N1) \rceil \) where \( \lceil Y \rceil \) is the next integer greater than or equal to \( Y \). Note that setting \( R = 1.0 \) forces \( N2 = N1 \).

Effect Size

Theta1 (Group 1 Mean Life)
Enter one or more values for the mean life of group 1 under the alternative hypothesis. This value is usually scaled in terms of elapsed time such as hours, days, or years. Of course, all time values must be on the same time scale.

Note that the value of theta may be calculated from the estimated probability of failure using the relationship

\[
P(Failure) = 1 - e^{-\text{time}/\theta}
\]

so that

\[
\theta = \frac{-\text{time}}{\ln(1 - P(Failure))}
\]

Any positive values are valid. You may enter a range of values such as ‘10 20 30’ or ‘100 to 1000 by 100.’

Note that only the ratio of theta1 and theta2 is used in the calculations.
Theta2 (Group 2 Mean Life)
Enter one or more values for the mean life of group 2 under the alternative hypothesis. This value is usually scaled in terms of elapsed time such as hours, days, or years. Of course, all time values must be on the same time scale.

Note that the value of theta may be calculated from the estimated probability of failure using the relationship

\[ P(\text{Failure}) = 1 - e^{-\text{time}/\theta} \]

so that

\[ \theta = \frac{-\text{time}}{\ln(1 - P(\text{Failure}))} \]

Any positive values are valid. You may enter a range of values such as ‘10 20 30’ or ‘100 to 1000 by 100.’

Note that only the ratio of theta1 and theta2 is used in the calculations.

---

Test

Alternative Hypothesis
Specify the alternative hypothesis of the test. Since the null hypothesis is equality (a difference between theta1 and theta2 of zero), the alternative is all that needs to be specified.

Note that the alternative hypothesis should match the values of Theta1 and Theta2. That is, if you select Ha: Theta1 > Theta, then the value of Theta1 should be greater than the value of Theta2.

---

Example 1 – Power for Several Sample Sizes

This example will calculate power for several sample sizes of a study designed to compare the average failure time of (supposedly) identical components manufactured by two companies. Management wants the study to be large enough to detect a ratio of mean lifetimes of 1.3 at the 0.05 significance level. The analysts decide to look at sample sizes between 5 and 500.

---

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Exponential Means procedure window by expanding Means, then Two Independent Means, then clicking on Test (Inequality), and then clicking on Tests for Two Exponential Means. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>5 20 50 100 200 300 400 500</td>
</tr>
</tbody>
</table>
435-6 Inequality Tests for Two Exponential Means

Data Tab (continued)
N2 (Sample Size Group 2) ...................... Use R
R (Sample Allocation Ratio) .................... 1.0
Theta1 (Group 1 Mean Life) .................... 1.3
Theta2 (Group 2 Mean Life) .................... 1.0
Alternative Hypothesis ............................ Ha: Theta1 ≠ Theta2

Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N1</th>
<th>N2</th>
<th>Ratio</th>
<th>Alpha</th>
<th>Beta</th>
<th>Theta1</th>
<th>Theta2</th>
<th>Theta1/Theta2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.06652</td>
<td>5</td>
<td>5</td>
<td>1.00000</td>
<td>0.05000</td>
<td>0.93348</td>
<td>1.3</td>
<td>1.0</td>
<td>1.30000</td>
</tr>
<tr>
<td>0.12839</td>
<td>20</td>
<td>20</td>
<td>1.00000</td>
<td>0.05000</td>
<td>0.87161</td>
<td>1.3</td>
<td>1.0</td>
<td>1.30000</td>
</tr>
<tr>
<td>0.25602</td>
<td>50</td>
<td>50</td>
<td>1.00000</td>
<td>0.05000</td>
<td>0.74398</td>
<td>1.3</td>
<td>1.0</td>
<td>1.30000</td>
</tr>
<tr>
<td>0.45619</td>
<td>100</td>
<td>100</td>
<td>1.00000</td>
<td>0.05000</td>
<td>0.54381</td>
<td>1.3</td>
<td>1.0</td>
<td>1.30000</td>
</tr>
<tr>
<td>0.74551</td>
<td>200</td>
<td>200</td>
<td>1.00000</td>
<td>0.05000</td>
<td>0.25449</td>
<td>1.3</td>
<td>1.0</td>
<td>1.30000</td>
</tr>
<tr>
<td>0.89447</td>
<td>300</td>
<td>300</td>
<td>1.00000</td>
<td>0.05000</td>
<td>0.10553</td>
<td>1.3</td>
<td>1.0</td>
<td>1.30000</td>
</tr>
<tr>
<td>0.95976</td>
<td>400</td>
<td>400</td>
<td>1.00000</td>
<td>0.05000</td>
<td>0.04024</td>
<td>1.3</td>
<td>1.0</td>
<td>1.30000</td>
</tr>
<tr>
<td>0.98559</td>
<td>500</td>
<td>500</td>
<td>1.00000</td>
<td>0.05000</td>
<td>0.01441</td>
<td>1.3</td>
<td>1.0</td>
<td>1.30000</td>
</tr>
</tbody>
</table>

Report Definitions
Power is the probability of rejecting a false null hypothesis.
N1 is the number of failures needed in Group 1.
N2 is the number of failures needed in Group 2.
Alpha is the probability of rejecting a true null hypothesis.
Beta is the probability of accepting a false null hypothesis.
Theta1 is the Mean Life in Group 1
Theta2 is the Mean Life in Group 2.

Summary Statements
Samples of size 5 and 5 achieve 7% power to detect a difference between the mean lifetime in group 1 of 1.3 and the mean lifetime in group 2 of 1.0 at a 0.05000 significance level (alpha) using a two-sided hypothesis based on the F distribution.

This report shows the power for each of the scenarios.

Plots Section

![Power vs N1](image)
Example 2 – Validation using Manual Calculations

We could not find published results that could be used to validate this procedure. Instead, we will compare the results to those computed using our probability distribution calculator.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Exponential Means procedure window by expanding Means, then Two Independent Means, then clicking on Test (Inequality), and then clicking on Tests for Two Exponential Means. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>20</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Theta1 (Group 1 Mean Life)</td>
<td>1.3</td>
</tr>
<tr>
<td>Theta2 (Group 2 Mean Life)</td>
<td>1.0</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Ha: Theta1 &gt; Theta2</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

### Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N1</th>
<th>N2</th>
<th>Ratio</th>
<th>Alpha</th>
<th>Beta</th>
<th>Theta1</th>
<th>Theta2</th>
<th>Theta1/Theta2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.20369</td>
<td>20</td>
<td>20</td>
<td>1.00000</td>
<td>0.05000</td>
<td>0.79631</td>
<td>1.3</td>
<td>1.0</td>
<td>1.30000</td>
</tr>
</tbody>
</table>

We will now check these results using manual calculations. First, we find critical value

\[ F_{0.95, 40, 40} = 1.6927972097 \]

using the probability calculator. Now, to calculate the power, we find the inverse \( F \) of

\[ 1.6927972097/1.3 = 1.302152 \]

to be 0.79631, which matches the reported value of Beta.
Inequality Tests for Two Exponential Means
Chapter 437

Inequality Tests for Two Poisson Means

Introduction

The Poisson probability law gives the probability distribution of the number of events occurring in a specified interval of time or space. The Poisson distribution is often used to fit count data, such as the number of defects on an item, the number of accidents at an intersection during a year, the number of calls to a call center during an hour, or the number of meteors seen in the evening sky during an hour.

The Poisson distribution is characterized by a single parameter which is the mean number of occurrences during the specified interval.

The procedure documented in this chapter calculates the power or sample size for testing whether the ratio of two Poisson means is different from a specified value (usually one). The test procedure is described in Gu et al. (2008).

Test Procedure

Assume that all subjects in each group are observed for a fixed time period and the number of events (outcomes or defects) is recorded. The following table presents the various terms that are used.

<table>
<thead>
<tr>
<th>Group</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fixed time interval</td>
<td>$t_1$</td>
<td>$t_2$</td>
</tr>
<tr>
<td>Sample Size</td>
<td>$N_1$</td>
<td>$N_2$</td>
</tr>
<tr>
<td>Number of events</td>
<td>$X_1$</td>
<td>$X_2$</td>
</tr>
<tr>
<td>Individual event rates</td>
<td>$\lambda_1$</td>
<td>$\lambda_2 = \rho \lambda_1$</td>
</tr>
<tr>
<td>Distribution of X</td>
<td>Poisson($\lambda_1 t_1$)</td>
<td>Poisson($\lambda_2 t_2$)</td>
</tr>
</tbody>
</table>
Inequality Tests for Two Poisson Means

Gu (2008) considered several test statistics that can be used to test one-sided hypotheses about the ratio.

\[ H_0 : \frac{\lambda_2}{\lambda_1} = \rho \text{ versus } H_a : \frac{\lambda_2}{\lambda_1} > \rho. \]

Two test statistics are available in this case. The first is based on unconstrained maximum likelihood estimates

\[
W_1 = \frac{X_2 - X_1 \left( \frac{\rho}{d} \right)}{\sqrt{X_2 + X_1 \left( \frac{\rho}{d} \right)^2}}
\]

The second test is based on constrained maximum likelihood estimates

\[
W_2 = \frac{X_2 - X_1 \left( \frac{\rho}{d} \right)}{\sqrt{(X_2 + X_1) \left( \frac{\rho}{d} \right)}}
\]

An equivalent pair of test statistics are available if logarithms are used. The statistical hypothesis is

\[ H_0 : \ln\left(\frac{\lambda_2}{\lambda_1}\right) - \ln(\rho) = 0 \text{ versus } H_a : \ln\left(\frac{\lambda_2}{\lambda_1}\right) - \ln(\rho) > 0 \]

Two test statistics are available in this case as well. The first is based on unconstrained maximum likelihood estimates

\[
W_3 = \frac{\ln\left(\frac{X_2}{X_1}\right) - \ln\left(\frac{\rho}{d}\right)}{\sqrt{\frac{1}{X_2} + \frac{1}{X_1}}}
\]

The second test is based on constrained maximum likelihood estimates

\[
W_4 = \frac{\ln\left(\frac{X_2}{X_1}\right) - \ln\left(\frac{\rho}{d}\right)}{\sqrt{\frac{2 + \frac{d}{\rho} + \frac{\rho}{d}}{X_1 + X_2}}}
\]

After extensive simulation, they recommend the following extension of the variance-stabilized test proposed by Huffman (1984) for the case when \( \rho/d > 1 \).
Inequality Tests for Two Poisson Means 437-3

\[
W_5 = \frac{2\left(\sqrt{X_2 + k} - \sqrt{\frac{\rho}{d}(X_1 + k)}\right)}{\sqrt{1 + \frac{\rho}{d}}}
\]

where \( d = t_1N_1 / t_2N_2 \) and \( k = 3/8 \).

Gu et al. (2008) show that all of these test statistics are approximately distributed as a standard normal and thus use the normal distribution as the basis of significance testing and power analysis.

**Assumptions**

The assumptions of the two-sample Poisson test are:

1. The data in each group are counts (discrete) that follow the Poisson distribution.
2. Each sample is a simple random sample from its population. Unlike most designs, in this design the sample size involves a fixed time parameter. That is, instead of specifying the number of people in a study, the number of man-hours is specified. Hence, a sample size of 10 hours could be achieved by ten people being observed for one hour or two people being observed for five hours.

   Also, since time is continuous, the sample size can be fractional.

**Technical Details**

**Computing Power**

The power analysis for testing the hypothesis

\[ H_0 : \lambda_2 / \lambda_1 = \rho_0 \text{ versus } H_a : \lambda_2 / \lambda_1 = \rho_a > \rho_0 \]

using the test statistics defined above is completed as follows.

1. **Find the critical value.** Choose the critical value \( z_{1-\alpha} \) using the standard normal distribution so that the probability of rejecting \( H_0 \) when it is true is \( \alpha \).
2. **Compute the power.** Compute the power for each test as follows.

For $W_1$, $W_3$, and $W_4$, the power is given by

$$Power(W_i) = 1 - \Phi\left(\frac{z_{1-\alpha} \sigma_i - \mu_i}{\sigma_i}\right)$$

where

$$\Phi(z) = \int_{-\infty}^{z} \text{Normal}(0,1)$$

$$\mu_i = \left(\frac{\rho_a}{d} - \frac{\rho_0}{d}\right) t_i \lambda_i, \quad \sigma_i^2 = \left(\frac{d \rho_a + \rho_0^2}{d^2}\right) t_i \lambda_i$$

$$\mu_3 = \ln\left(\frac{\rho_a}{\rho_0}\right), \quad \sigma_3^2 = \frac{d + \rho_a}{t_i \lambda_i \rho_a}$$

$$\mu_4 = \ln\left(\frac{\rho_a}{\rho_0}\right), \quad \sigma_4^2 = \left(\frac{2 + \frac{d}{\rho_0}}{1 + \frac{\rho_a}{d}}\right) t_i \lambda_i$$

For $W_2$ the power is computed using

$$Power(W_2) = 1 - \Phi\left(\frac{E z_{1-\alpha} - F}{G}\right)$$

where

$$E = \sqrt{\left(\frac{\rho_0}{\rho_a}\right)^2 + \frac{\rho_0^2}{\rho_a d}}$$

$$F = \left(1 - \frac{\rho_0}{\rho_a}\right) \sqrt{\frac{\lambda_i t_i \lambda_i}{d}}$$

$$G = \sqrt{\frac{\rho_0}{\rho_a} \left(1 + \frac{\rho_0^2}{d \rho_a}\right)}$$

For $W_5$ the power is computed using

$$Power(W_5) = 1 - \Phi\left(\frac{A \sqrt{B} - z_{1-\alpha} C}{D}\right)$$

where

$$A = 2 \left(1 - \sqrt{\frac{\rho_0}{\rho_a}}\right), \quad B = \lambda_i t_i N_1 + k$$
Inequality Tests for Two Poisson Means

\[ C = \sqrt{\frac{\rho_0 + d}{\rho_a}}, \quad D = \sqrt{\frac{\rho_a + d}{\rho_a}} \]

**Computing Sample Size**

The sample size is found using the formula

\[ N_1 = \frac{\left( z_{1-\alpha} C + z_{\text{Power}} D \right)^2}{A} \cdot \frac{1}{\lambda_i f_i} - k \]

**Procedure Options**

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

**Data Tab**

The Data tab contains most of the parameters and options that you will be concerned with.

**Solve For**

**Find (Solve For)**

This option specifies the parameter to be calculated from the values of the other parameters. Under most conditions, you would select either *Power and Beta* or *N1*.

Select *N1* when you want to determine the sample size needed to achieve a given power and alpha error level.

Select *Power and Beta* when you want to calculate the power of an experiment that has already been run.

**Error Rates**

**Power or Beta**

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

Values must be between zero and one. Historically, the value of 0.80 (beta = 0.20) was used for power. Now, 0.90 (beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.
Inequality Tests for Two Poisson Means

**Alpha (Significance Level)**
This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

Values must be between zero and one. For one-sided tests such as this, the value of 0.025 is recommended for alpha. You may enter a range of values such as 0.025 0.05 0.10 or 0.025 to 0.05 by 0.005.

**Sample Size**

**t1 (Observation Time Group 1)**
This is the fixed observation time for group 1. Each subject is observed for this amount of time. If the observation times are variable, this is the average observed time per subject.

**t2 (Observation Time Group 2)**
This is the fixed observation time for group 2. Each subject is observed for this amount of time. If the observation times are variable, this is the average observed time per subject.

**N1 (Sample Size 1)**
This option specifies one or more values of the sample size in group 1. This value must be positive and greater than zero. Note that you may enter a list of values using the syntax 50,100,150,200,250 or 50 to 250 by 50.

Note that the effective sample size for group 1 is t1(N1). For example, the sample size might be 1400 man hours.

**N2 (Sample Size 2)**
Enter a value (or range of values) for the sample size of group 2. Enter Use R if you want N2 to be calculated using the formula: N2 = R(N1) where R is the Sample Allocation Ratio.

For example, if you want N1 = N2, select Use R here and set R = 1.

**R (Sample Allocation Ratio)**
Enter a value (or range of values) for R, the allocation ratio between samples. That is, R = N2/N1. This value is only used when N2 is set to Use R. When used, N2 is calculated from N1 using the formula: N2 = R(N1). Note that setting R = 1.0 forces N2 = N1.

**Effect Size – Means**

**λ1 (Group 1 Event Rate)**
This option specifies one or more values of the mean occurrence rate in group 1 assuming the null hypothesis. This value must be greater than zero. This mean is compared to λ2 by the statistical test.

The difference in the ratios of λ2 and λ1 under the null and alternative hypotheses is the amount that this design can detect. You can enter a range of values such as “1 1.2 1.4” or “1 to 5 by 1.”
$\rho_0 \ (\lambda_2 / \lambda_1, \text{Assuming H0})$

This is the value of the ratio of the two mean event rates assumed by the null hypothesis, $H_0$. Usually, $\rho_0 = 1.0$ which implies that the two rates are equal. However, you may test other values of $\rho_0$ as well.

Strictly speaking, any positive number is valid. However, usually 1.0 is used.

Warning: you cannot use the same value for both $\rho_0$ and $\rho_a$.

$\rho_a \ (\lambda_2 / \lambda_1, \text{Assuming Ha})$

This is the value of the ratio of the two mean event rates assumed by the alternative hypothesis, $H_a$. The magnitude of the difference between this value and $\rho_0$ is the amount that the design can detect.

Warning: you cannot use the same value for both $\rho_0$ and $\rho_a$.

**Test**

**Alternative Hypothesis**

This option specifies the alternative hypothesis. This implicitly specifies the direction of the hypothesis test. The null hypothesis is $H_0 : \rho_a = \rho_0$. Possible selections for the alternative hypothesis are:

1. $H_a : \rho_0 < \rho_a$. This option yields a one-tailed t test.

2. $H_a : \rho_0 > \rho_a$. This option yields a one-tailed t test.

**Example 1 – Finding the Sample Size**

We will use the example of Gu (2008) in which epidemiologist wish to examine the relationship of post-menopausal hormone use and coronary heart disease (CHD). The incidence rate for those not using the hormone is $0.0005 \ (\lambda_1 = 0.0005)$. How large of a sample is needed to detect a change in the incidence ratio from $\rho_0 = 1$ to $\rho_a = 2, 3, 4, 5, \text{ or } 6$. Assume that 90% power is required and $\alpha = 0.05$. Assume that each subject will be observed for two years and that the design calls for an equal number of subjects in both groups.

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Poisson Means [Ratios] procedure window by expanding Means, then Two Independent Means, then clicking on Test (Inequality), and then clicking on Tests for Two Poisson Means [Ratios]. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.
437-8 Inequality Tests for Two Poisson Means

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For) N1</td>
<td></td>
</tr>
<tr>
<td>Power</td>
<td>0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>t1 (Observation Time Group 1)</td>
<td>2</td>
</tr>
<tr>
<td>t2 (Observation Time Group 2)</td>
<td>2</td>
</tr>
<tr>
<td>N1 (Sample Size 1)</td>
<td>Ignored since this is the search parameter.</td>
</tr>
<tr>
<td>N2 (Sample Size 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1</td>
</tr>
<tr>
<td>λ1 (Group 1 Event Rate)</td>
<td>0.0005</td>
</tr>
<tr>
<td>ρ0 (λ2 / λ1 Assuming H0)</td>
<td>1</td>
</tr>
<tr>
<td>ρa (λ2 / λ1 Assuming Ha)</td>
<td>2 3 4 5 6</td>
</tr>
<tr>
<td>Ha (Alternative Hypothesis)</td>
<td>Ha: ρ0 &lt; ρa</td>
</tr>
</tbody>
</table>

Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Numeric Results for Two-Sample Poisson Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null Hypothesis: λ2 / λ1 = ρ0</td>
</tr>
<tr>
<td>Alternative Hypothesis: λ2 / λ1 = ρa &gt; ρ0</td>
</tr>
<tr>
<td>Fixed Observation Time Length: Group 1 = 2, Group 2 = 2</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Power</th>
<th>N1</th>
<th>N2</th>
<th>Mean Rate 1</th>
<th>Mean Rate 2</th>
<th>Ratio</th>
<th>Ratio</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9000</td>
<td>29736.2</td>
<td>29736.2</td>
<td>0.0005</td>
<td>0.0010</td>
<td>1.0000</td>
<td>2.0000</td>
<td>0.0500</td>
<td>0.1000</td>
</tr>
<tr>
<td>0.9000</td>
<td>10776.9</td>
<td>10776.9</td>
<td>0.0005</td>
<td>0.0015</td>
<td>1.0000</td>
<td>3.0000</td>
<td>0.0500</td>
<td>0.1000</td>
</tr>
<tr>
<td>0.9000</td>
<td>6363.7</td>
<td>6363.7</td>
<td>0.0005</td>
<td>0.0020</td>
<td>1.0000</td>
<td>4.0000</td>
<td>0.0500</td>
<td>0.1000</td>
</tr>
<tr>
<td>0.9000</td>
<td>4512.5</td>
<td>4512.5</td>
<td>0.0005</td>
<td>0.0025</td>
<td>1.0000</td>
<td>5.0000</td>
<td>0.0500</td>
<td>0.1000</td>
</tr>
<tr>
<td>0.9000</td>
<td>3513.9</td>
<td>3513.9</td>
<td>0.0005</td>
<td>0.0030</td>
<td>1.0000</td>
<td>6.0000</td>
<td>0.0500</td>
<td>0.1000</td>
</tr>
</tbody>
</table>

References


Report Definitions

Power is the probability of rejecting a false null hypothesis. It should be close to one.
N1 is the number of subjects in the first group.
N2 is the number of subjects in the second group.
λ1 is the mean response rate for a group 1 subject during the fixed observation time.
λ2 / ρa is the mean response rate for a group 2 subject during the fixed observation time assuming ρa.
ρ0 is the ratio λ2 / λ1 assuming the null hypothesis.
ρa is the ratio λ2 / λ1 assuming the alternative hypothesis.
Alpha is the probability of rejecting a true null hypothesis. It should be small.
Beta is the probability of accepting a false null hypothesis. It should be small.

Summary Statements

Samples of 29736.2 subjects in group 1 observed for 2 time periods and 29736.2 subjects in group 2 observed for 2 time periods achieve 90% power to detect a change in the mean response ratio (λ2 / λ1) from 1.0000 to 2.0000 when λ1 = 0.0005 and the significance level (alpha) is 0.0500 using a one-sided test.

This report shows the values of each of the parameters, one scenario per row. The values of power and beta were calculated from the other parameters.
**Power**

Power is the probability of rejecting a false null hypothesis.

**N1**

N1 is the number of subjects in the first group. In the case of Poisson random variables, N1 does not have to be a whole number. A fractional subject simply refers to a subject that is observed for a fraction of the time period. Also, the effective sample size is the number of subjects multiplied by the fixed time period. For example, the sample size in the first row is $29736.2 \times 2 = 59472.4$ person-years.

**N2**

N2 is the number of subjects in the second group.

**λ1**

$\lambda_1$ is the mean response rate for a group 1 subject during the fixed observation time.

**λ2 | ρa**

$\lambda_2 | \rho_a$ is the mean response rate for a group 2 subject during the fixed observation time assuming $\rho_a$.

**ρ0**

$\rho_0$ is the ratio $\lambda_2 / \lambda_1$ assuming the null hypothesis.

**ρa**

$\rho_a$ is the ratio $\lambda_2 / \lambda_1$ assuming the alternative hypothesis.

**Alpha**

Alpha is the probability of rejecting a true null hypothesis. It should be small.
### Example 2 – Validation using Gu

Gu et al. (2008) present an example that we will use to validate this procedure. Using the scenario cited in Example 1 above, they give a sample size calculation on page 295. In this example, $\lambda_1 = 0.0005$, $\rho_0 = 1$, $\rho_a = 4$, $t_1 = t_2 = 2$, $\alpha = 0.05$, $R = 0.5$, and power $= 0.9$. In their Table 6, they list the sample size for $p^{(e)}_3$ in this scenario as 8627. However, this number is inaccurate because of the two-decimal place rounding that was done during their calculation. In a private communication, they agreed that the more accurate number is 8589.

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Poisson Means [Ratios]** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Test (Inequality)**, and then clicking on **Tests for Two Poisson Means [Ratios]**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.
Inequality Tests for Two Poisson Means

### Data Tab

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N1</td>
</tr>
<tr>
<td>Power</td>
<td>0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>t1 (Observation Time Group 1)</td>
<td>2</td>
</tr>
<tr>
<td>t2 (Observation Time Group 2)</td>
<td>2</td>
</tr>
<tr>
<td>N1 (Sample Size 1)</td>
<td>Ignored since this is the search parameter.</td>
</tr>
<tr>
<td>N2 (Sample Size 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>0.5</td>
</tr>
<tr>
<td>λ1 (Group 1 Event Rate)</td>
<td>0.0005</td>
</tr>
<tr>
<td>ρ0 (λ2 / λ1 Assuming H0)</td>
<td>1</td>
</tr>
<tr>
<td>ρa (λ2 / λ1 Assuming Ha)</td>
<td>4</td>
</tr>
<tr>
<td>Ha (Alternative Hypothesis)</td>
<td>Ha: ρ0 &lt; ρa</td>
</tr>
</tbody>
</table>

### Annotated Output

Click the Run button to perform the calculations and generate the following output.

### Numeric Results

<table>
<thead>
<tr>
<th>Numeric Results for Two-Sample Poisson Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Null Hypothesis: λ2 / λ1 = ρ0</td>
</tr>
<tr>
<td>Alternative Hypothesis: λ2 / λ1 = ρa &gt; ρ0</td>
</tr>
<tr>
<td>Fixed Observation Time Length: Group 1 = 2, Group 2 = 2</td>
</tr>
<tr>
<td>Power</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>0.9000</td>
</tr>
</tbody>
</table>
437-12  Inequality Tests for Two Poisson Means
Chapter 440

Inequality Tests for Two Means (Simulation)

Introduction

This procedure allows you to study the power and sample size of several statistical tests of the null hypothesis that the difference between two means is equal to a specific value versus the alternative hypothesis that it is greater than, less than, or not-equal to that value. Because the mean represents the center of the population, if the means are different, the populations are different. Other attributes of the two populations (such as the shape and spread) might also be compared, but this module focuses on comparisons of the means only.

Measurements are made on individuals that have been randomly assigned to, or randomly chosen from, one of two groups. This is sometimes referred to as a parallel-groups design. This design is used in situations such as the comparison of the income level of two regions, the nitrogen content of two lakes, or the effectiveness of two drugs.

The two-sample t-test is commonly used in this situation. When the variances of the two groups are unequal, Welch’s t-test is often used. When the data are not normally distributed, the Mann-Whitney (Wilcoxon signed-ranks) U test may be used.

The details of the power analysis of the two-sample t-test using analytic techniques are presented in another PASS chapter and they won’t be duplicated here. This chapter will only consider power analysis using computer simulation.

Technical Details

Computer simulation allows us to estimate the power and significance level that is actually achieved by a test procedure in situations that are not mathematically tractable. Computer simulation was once limited to mainframe computers. But, in recent years, as computer speeds have increased, simulation studies can be completed on desktop and laptop computers in a reasonable period of time.

The steps to a simulation study are

1. Specify how the test is carried out. This includes indicating how the test statistic is calculated and how the significance level is specified.
2. Generate random samples from the distributions specified by the alternative hypothesis. Calculate the test statistics from the simulated data and determine if the null hypothesis is accepted or rejected. Tabulate the number of rejections and use this to calculate the test’s power.

3. Generate random samples from the distributions specified by the null hypothesis. Calculate each test statistic from the simulated data and determine if the null hypothesis is accepted or rejected. Tabulate the number of rejections and use this to calculate the test’s significance level.

4. Repeat steps 2 and 3 several thousand times, tabulating the number of times the simulated data leads to a rejection of the null hypothesis. The power is the proportion of simulated samples in step 2 that lead to rejection. The significance level is the proportion of simulated samples in step 3 that lead to rejection.

Generating Random Distributions

Two methods are available in PASS to simulate random samples. The first method generates the random variates directly, one value at a time. The second method generates a large pool (over 10,000) of random values and then draw the random numbers from this pool. This second method can cut the running time of the simulation by 70%.

As mentioned above, the second method begins by generating a large pool of random numbers from the specified distributions. Each of these pools is evaluated to determine if its mean is within a small relative tolerance (0.0001) of the target mean. If the actual mean is not within the tolerance of the target mean, individual members of the population are replaced with new random numbers if the new random number moves the mean towards its target. Only a few hundred such swaps are required to bring the actual mean to within tolerance of the target mean. This population is then sampled with replacement using the uniform distribution. We have found that this method works well as long as the size of the pool is the maximum of twice the number of simulated samples desired and 10,000.

Test Statistics

This section describes the test statistics that are available in this procedure.

Two-Sample T-Test

The two-sample t-test assumes that the data are a simple random sample from a population of normally-distributed values that all have the same mean and variance. This assumption implies that the data are continuous and their distribution is symmetric. The calculation of the t statistic is as follows

\[ t_{df} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{X}_1 - \bar{X}_2}} \]
where
\[
\bar{X}_k = \frac{\sum_{i=1}^{N_k} X_{ki}}{N_k}
\]
\[
s_{\bar{X}_1, \bar{X}_2} = \sqrt{\frac{\sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2 + \sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2}{N_1 + N_2 - 2} \left( \frac{1}{N_1} + \frac{1}{N_2} \right)}
\]
\[
df = N_1 + N_2 - 2
\]
The significance of the test statistic is determined by computing the p-value based on the t distribution with degrees of freedom df. If this p-value is less than a specified level (often 0.05), the null hypothesis is rejected. Otherwise, no conclusion can be reached.

**Welch’s T-Test**

Welch (1938) proposed the following test for use when the two variances cannot be assumed equal.

\[
t_f^* = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{s_{\bar{X}_1, \bar{X}_2}}
\]

where
\[
s_{\bar{X}_1, \bar{X}_2}^* = \sqrt{\left( \frac{\sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2}{N_1(N_1 - 1)} + \frac{\sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2}{N_2(N_2 - 1)} \right)}
\]
\[
f = \frac{\left( \frac{s_1^2}{N_1} + \frac{s_2^2}{N_2} \right)^2}{\frac{s_1^4}{N_1(N_1 - 1)} + \frac{s_2^4}{N_2(N_2 - 1)}}
\]
\[
s_1 = \sqrt{\frac{\sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2}{N_1 - 1}}, \quad s_2 = \sqrt{\frac{\sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2}{N_2 - 1}}
\]

**Trimmed T-Test assuming Equal Variances**

The notion of trimming off a small proportion of possibly outlying observations and using the remaining data to form a t-test was first proposed for one sample by Tukey and McLaughlin (1963). Dixon and Tukey (1968) consider a slight modification of this one sample test, called
Winsorization, which replaces the trimmed data with the nearest remaining value. The two-sample trimmed t-test was proposed by Yuen and Dixon (1973).

Assume that the data values have been sorted from lowest to highest. The trimmed mean is defined as

\[
\bar{X}_{tg} = \frac{\sum_{k=g+1}^{N-g} X_k}{h}
\]

where \( h = N - 2g \) and \( g = \lceil N(G/100) \rceil \). Here we use \( \lceil Z \rceil \) to mean the largest integer smaller than \( Z \) with the modification that if \( G \) is non-zero, the value of \( \lceil N(G/100) \rceil \) is at least one. \( G \) is the percent trimming and should usually be less than 25\%, often between 5\% and 10\%. Thus, the \( g \) smallest and \( g \) largest observation are omitted in the calculation.

To calculate the modified t-test, calculate the Winsorized mean and the Winsorized sum of squared deviations as follows.

\[
\bar{X}_{wg} = \frac{g(X_{g+1} + X_{N-g}) + \sum_{k=g+1}^{N-g} X_k}{N}
\]

\[
SSD_{wg} = \frac{g(X_{g+1} - \bar{X}_{wg})^2 + g(X_{N-g} - \bar{X}_{wg})^2 + \sum_{k=g+1}^{N-g} (X_k - \bar{X}_{wg})^2}{N}
\]

Using the above definitions, the two-sample trimmed t-test is given by

\[
T_{tg} = \frac{(\bar{X}_{1tg} - \bar{X}_{2tg}) - (\mu_1 - \mu_2)}{\sqrt{SSD_{1wg} + SSD_{2wg}} \left( \frac{1}{h_1} + \frac{1}{h_2} \right)}
\]

The distribution of this \( t \) statistic is approximately that of a \( t \) distribution with degrees of freedom equal to \( h_1 + h_2 - 2 \). This approximation is often reasonably accurate if both sample sizes are greater than 6.

**Trimmed T-Test assuming Unequal Variances**

Yuen (1974) combines trimming (see above) with Welch’s (1938) test. The resulting trimmed Welch test is resistant to outliers and seems to alleviate some of the problems that occur because of skewness in the underlying distributions. Extending the results from above, the trimmed version of Welch’s t-test is given by

\[
T^{*}_{tg} = \frac{(\bar{X}_{1tg} - \bar{X}_{2tg}) - (\mu_1 - \mu_2)}{\sqrt{SSD_{1wg} + SSD_{2wg}}} \left( \frac{1}{h_1(h_1 - 1)} + \frac{1}{h_2(h_2 - 1)} \right)
\]
with degrees of freedom $f$ given by

$$\frac{1}{f} = \frac{c^2}{h_1 - 1} + \frac{1 - c^2}{h_2 - 1}$$

where

$$c = \frac{SSD_{1wg}}{h_1(h_1 - 1)} + \frac{SSD_{2wg}}{h_2(h_2 - 1)}$$

**Mann-Whitney U Test**

This test is the nonparametric substitute for the equal-variance t-test. Two key assumptions for this test are that the distributions are at least ordinal and that they are identical under $H_0$. This implies that ties (repeated values) are not acceptable. When ties are present, the approximation provided can be used, but know that the theoretic results no longer hold.

The Mann-Whitney test statistic is defined as follows in Gibbons (1985).

$$z = \frac{W_1 - \frac{N_1(N_1 + N_2 + 1)}{2} + C}{s_w}$$

where

$$W_1 = \sum_{k=1}^{N_1} \text{Rank}(X_{1k})$$

The ranks are determined after combining the two samples. The standard deviation is calculated as

$$s_w = \sqrt{\frac{N_1N_2(N_1 + N_2 + 1)}{12} - \frac{N_1N_2}{12(N_1 + N_2)(N_1 + N_2 - 1)} \sum_{i=1}^{t_1} (t_1^2 - t_1)$$

where $t_1$ is the number of observations tied at value one, $t_2$ is the number of observations tied at some value two, and so forth.

The correction factor, $C$, is 0.5 if the rest of the numerator of $z$ is negative or -0.5 otherwise. The value of $z$ is then compared to the standard normal distribution.

**Standard Deviations**

Care must be used when either the null or alternative distribution is not normal. In these cases, the standard deviation is usually not specified directly. For example, you might use a gamma distribution with a shape parameter of 1.5 and a mean of 4 as the null distribution and a gamma distribution with the same shape parameter and a mean of 5 as the alternative distribution. This allows you to compare the two means. However, although the shape parameters are constant, the standard deviations, which are based on both the shape parameter and the mean, are not. Thus the distributions not only have different means, but different standard deviations!
Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains most of the parameters and options that you will be concerned with.

Solve For

Find (Solve For)

This option specifies whether you want to find Power or N1 from the simulation. Select Power when you want to estimate the power of a certain scenario. Select N1 when you want to determine the sample size needed to achieve a given power and alpha error level. Finding N1 is very computationally intensive, and so it may take a long time to complete.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of equal means when in fact the means are different.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when you reject the null hypothesis of equal means when in fact the means are equal.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.

Sample Size

N1 (Sample Size Group 1)

Enter a value (or range of values) for the sample size of group 1. Note that these values are ignored when you are solving for N1. You may enter a range of values such as 10 to 100 by 10.
N2 (Sample Size Group 2)
Enter a value (or range of values) for the sample size of group 2 or enter Use R to base N2 on the value of N1. You may enter a range of values such as 10 to 100 by 10.

- **Use R**
  When Use R is entered here, N2 is calculated using the formula
  \[ N2 = \lfloor R (N1) \rfloor \]
  where R is the Sample Allocation Ratio and the operator \( \lfloor \) is the first integer greater than or equal to \( Y \). For example, if you want \( N1 = N2 \), select Use R and set \( R = 1 \).

R (Sample Allocation Ratio)
Enter a value (or range of values) for R, the allocation ratio between samples. This value is only used when \( N2 \) is set to Use R.

When used, \( N2 \) is calculated from \( N1 \) using the formula: \( N2 = \lfloor R (N1) \rfloor \) where \( \lfloor Y \rfloor \) is the next integer greater than or equal to \( Y \). Note that setting \( R = 1.0 \) forces \( N2 = N1 \).

Test

**Test Type**
Specify which test statistic is to be used in the simulation. Although the t-test is the most commonly used test statistic, it is based on assumptions that may not be viable in many situations. For your data, you may find that one of the other tests is more accurate (actual alpha = target alpha) and more precise (better power).

**Alternative Hypothesis**
This option specifies the alternative hypothesis, H1. This implicitly specifies the direction of the hypothesis test. The null hypothesis is always \( H0: \text{Diff} = \text{Diff0} \).

Note that the alternative hypothesis enters into power calculations by specifying the rejection region of the hypothesis test. Its accuracy is critical.

Possible selections are:

- **Difference <> Diff0**
  This is the most common selection. It yields a two-tailed test. Use this option when you are testing whether the mean is different from a specified value Diff0, but you do not want to specify beforehand whether it is smaller or larger. Most scientific journals require two-tailed tests.

- **Difference < Diff0**
  This option yields a one-tailed test. Use it when you want to test whether the true mean is less than Diff0.

- **Difference > Diff0**
  This option yields a one-tailed test. Use it when you want to test whether the true mean is greater than Diff0. Note that this option could be used for a non-inferiority test.
Simulations

This option specifies the number of iterations, $M$, used in the simulation. As the number of iterations is increased, the accuracy and running time of the simulation will be increased also.

The precision of the simulated power estimates are calculated from the binomial distribution. Thus, confidence intervals may be constructed for various power values. The following table gives an estimate of the precision that is achieved for various simulation sizes when the power is either 0.50 or 0.95. The table values are interpreted as follows: a 95% confidence interval of the true power is given by the power reported by the simulation plus and minus the ‘Precision’ amount given in the table.

<table>
<thead>
<tr>
<th>Simulation Size</th>
<th>Precision when Power = 0.50</th>
<th>Precision when Power = 0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td></td>
<td></td>
</tr>
<tr>
<td>100</td>
<td>0.100</td>
<td>0.044</td>
</tr>
<tr>
<td>500</td>
<td>0.045</td>
<td>0.019</td>
</tr>
<tr>
<td>1000</td>
<td>0.032</td>
<td>0.014</td>
</tr>
<tr>
<td>2000</td>
<td>0.022</td>
<td>0.010</td>
</tr>
<tr>
<td>5000</td>
<td>0.014</td>
<td>0.006</td>
</tr>
<tr>
<td>10000</td>
<td>0.010</td>
<td>0.004</td>
</tr>
<tr>
<td>50000</td>
<td>0.004</td>
<td>0.002</td>
</tr>
<tr>
<td>100000</td>
<td>0.003</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Notice that a simulation size of 1000 gives a precision of plus or minus 0.01 when the true power is 0.95. Also note that as the simulation size is increased beyond 5000, there is only a small amount of additional accuracy achieved.

Effect Size

Group 1 (and 2) Distribution|H0

These options specify the distributions of the two groups under the null hypothesis, H0. The difference between the means of these two distributions is the difference that is tested, $Diff0$.

Usually, these two distributions will be identical and $Diff0 = 0$. However, if you are planning a non-inferiority test, the means will be different.

All of the distributions are parameterized so that the mean is entered first. For example, if you wanted to specify that the mean of a normally-distributed variable is to be five, you could enter $N(5, S)$ or $N(M0, S)$ here and $M0 = 5$ later.

The parameters of each distribution are specified using numbers or letters. If letters are used, their values are specified in the boxes below. The value $M0$ is reserved for the value of the mean under the null hypothesis.

Following is a list of the distributions that are available and the syntax used to specify them. Note that, except for the multinomial, the distributions are parameterized so that the mean is entered first.

- Beta=\(A(M0,A,B,Minimum)\)
- Binomial=\(B(M0,N)\)
- Cauchy=\(C(M0,Scale)\)
- Constant=\(K(Value)\)
Exponential=E(M0)
F=F(M0,DF1)
Gamma=G(M0,A)
Multinomial=M(P1,P2,…,Pk)
Normal=N(M0,SD)
Poisson=P(M0)
Student’s T=T(M0,D)
Tukey’s Lambda=L(M0,S,Skewness,Elongation)
Uniform=U(M0,Minimum)
Weibull=W(M0,B)

Details of writing mixture distributions, combined distributions, and compound distributions are found in the chapter on Data Simulation and are not repeated here.

**Finding the Value of the Mean of a Specified Distribution**

Except for the multinomial distribution, the distributions have been parameterized in terms of their means, since this is the parameter being tested. The mean of a distribution created as a linear combination of other distributions is found by applying the linear combination to the individual means. However, the mean of a distribution created by multiplying or dividing other distributions is not necessarily equal to applying the same function to the individual means. For example, the mean of $4N(4, 5) + 2N(5, 6)$ is $4*4 + 2*5 = 26$, but the mean of $4N(4, 5) * 2N(5, 6)$ is not exactly $4*4*2*5 = 160$ (although it is close).

**Group 1 (and 2) Distribution|H1**

These options specify the distributions of the two groups under the alternative hypothesis, H1. The difference between the means of these two distributions is the difference that is assumed to be the true value of the difference. That is, this is the difference at which the power is computed.

Usually, the mean difference is specified by entering $M0$ for the mean parameter in the distribution expression for group 1 and $M1$ for the mean parameter in the distribution expression for group 2. The mean difference under H1 then becomes the value of $M0 – M1$.

The parameters of each distribution are specified using numbers or letters. If letters are used, their values are specified in the boxes below. The value $M1$ is reserved for the value of the mean of group 2 under the alternative hypothesis.

Following is a list of the distributions that are available and the syntax used to specify them. Note that, except for the multinomial, the distributions are parameterized so that the mean, $M1$, is entered first.

Beta=A(M1,A,B,Minimum)
Binomial=B(M1,N)
Cauchy=C(M1,Scale)
Constant=K(Value)
Exponential=E(M1)
F=F(M1,DF1)
Gamma=G(M1,A)
Multinomial=M(P1,P2,…,Pk)
Normal=N(M1,SD)
Poisson=P(M1)
Student’s T=T(M1,D)
Tukey’s Lambda=L(M1,S,Skewness,Elongation)
Uniform=U(M1,Minimum)
Weibull=W(M1,B)

Details of writing mixture distributions, combined distributions, and compound distributions are found in the chapter on Data Simulation and will not be repeated here.

**Effect Size – Distribution Parameters**

**M0 (Mean|H0)**
These values are substituted for $M0$ in the distribution specifications given above. $M0$ is intended to be the value of the mean hypothesized by the null hypothesis, H0.

You can enter a list of values using the syntax 0 1 2 3 or 0 to 3 by 1.

**M1 (Mean|H1)**
These values are substituted for $M1$ in the distribution specifications given above. Although it can be used wherever you want, $M1$ is intended to be the value of the mean hypothesized by the alternative hypothesis, H1.

You can enter a list of values using the syntax 0 1 2 3 or 0 to 3 by 1.

**Parameter Values (S, A, B)**
Enter the numeric value(s) of the parameters listed above. These values are substituted for the corresponding letter in all four distribution specifications.

You can enter a list of values for each letter using the syntax 0 1 2 3 or 0 to 3 by 1.

You can also change the letter that is used as the name of this parameter using the pull-down menu to the side.

**Reports Tab**
The Reports tab contains settings about the format of the output.

**Select Output – Numeric Reports**

**Show Numeric Reports & Plots**
These options let you specify whether you want to generate the standard reports and plots.

**Show Inc’s & 95% C.I.**
Checking this option causes an additional line to be printed showing a 95% confidence interval for both the power and actual alpha and half the width of the confidence interval (the increment).

**Select Output – Plots**

**Show Comparative Reports & Plots**
These options let you specify whether you want to generate reports and plots that compare the test statistics that are available.
Comparative Report/Plot Options

Include T-Test Results – Include Mann-Whitney-Test Results
These options let you specify whether to include each test statistic in the comparative reports. These options are only used if comparative reports and/or plots are generated.

Options Tab
The Options tab contains limits on the number of iterations and various options about individual tests.

Maximum Iterations

Maximum Iterations Before Search Termination
Specify the maximum number of iterations before the search for the sample size, N1, is aborted. When the maximum number of iterations is reached without convergence, the sample size is left blank. We recommend a value of at least 500.

Random Numbers

Random Number Pool Size
This is the size of the pool of random values from which the random samples will be drawn. Pools should be at least the maximum of 10,000 and twice the number of simulations. You can enter Automatic and an appropriate value will be calculated.

If you do not want to draw numbers from a pool, enter 0 here.

Trimmed T-Test

Percent Trimmed at Each End
Specify the percent of each end of the sorted data that is to be trimmed (constant G above) when using the trimmed means procedures. This percentage is applied to the sample size to determine how many of the lowest and highest data values are to be trimmed by the procedure. For example, if the sample size (N1) is 27 and you specify 10 here, then \([27\times10/100] = 2\) observations will be trimmed at the bottom and the top. For any percentage, at least one observation is trimmed from each end of the sorted dataset.

The range of possible values is 0 to 25.
Example 1 – Power at Various Sample Sizes

Researchers are planning a parallel-group experiment to test whether the difference in response to a certain drug is zero. The researchers will use a two-sided t-test with an alpha level of 0.05. They want to compare the power at sample sizes of 50, 100, and 200 when the shift in the means is 0.6 from drug 1 to drug 2. They assume that the data are normally distributed with a standard deviation of 2. Since this is an exploratory analysis, they set the number of simulation iterations to 2000.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means (Simulation) procedure window by expanding Means, then Two Independent Means, then clicking on Test (Inequality), and then clicking on Tests for Two Means (Simulation). You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>50 100 200</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Group 1 Dist’n</td>
<td>H0</td>
</tr>
<tr>
<td>Group 2 Dist’n</td>
<td>H0</td>
</tr>
<tr>
<td>Group 1 Dist’n</td>
<td>H1</td>
</tr>
<tr>
<td>Group 2 Dist’n</td>
<td>H1</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H1)</td>
</tr>
<tr>
<td>S</td>
<td>2</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Diff#Diff0</td>
</tr>
<tr>
<td>Simulations</td>
<td>2000</td>
</tr>
<tr>
<td>Test Type</td>
<td>T-Test</td>
</tr>
</tbody>
</table>
Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results and Plots

<table>
<thead>
<tr>
<th>Power</th>
<th>N1/N2</th>
<th>Diff0</th>
<th>H0 H1</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>Beta</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.324</td>
<td>50/50</td>
<td>0.0</td>
<td>0.0</td>
<td>-0.6</td>
<td>0.050</td>
<td>0.056</td>
<td>0.0</td>
<td>0.6</td>
<td>2.0</td>
</tr>
<tr>
<td>(0.021)</td>
<td>[0.303, 0.345]</td>
<td></td>
<td></td>
<td>(0.010)</td>
<td>[0.045, 0.066]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.563</td>
<td>100/100</td>
<td>0.0</td>
<td>0.0</td>
<td>-0.6</td>
<td>0.050</td>
<td>0.047</td>
<td>0.0</td>
<td>0.6</td>
<td>2.0</td>
</tr>
<tr>
<td>(0.022)</td>
<td>[0.541, 0.585]</td>
<td></td>
<td></td>
<td>(0.009)</td>
<td>[0.038, 0.056]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.855</td>
<td>200/200</td>
<td>0.0</td>
<td>0.0</td>
<td>-0.6</td>
<td>0.050</td>
<td>0.045</td>
<td>0.0</td>
<td>0.6</td>
<td>2.0</td>
</tr>
<tr>
<td>(0.015)</td>
<td>[0.840, 0.870]</td>
<td></td>
<td></td>
<td>(0.009)</td>
<td>[0.035, 0.054]</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:
- Pool Size: 10000.
- Run Time: 34.78 seconds.

Report Definitions
- Power is the probability of rejecting a false null hypothesis.
- N1 is the size of the sample drawn from population 1.
- N2 is the size of the sample drawn from population 2.
- Diff0 is the mean difference between (Grp1 - Grp2) assuming the null hypothesis, H0.
- Diff1 is the mean difference between (Grp1 - Grp2) assuming the alternative hypothesis, H1.
- Target Alpha is the probability of rejecting a true null hypothesis. It is set by the user.
- Actual Alpha is the alpha level that was actually achieved by the experiment.
- Beta is the probability of accepting a false null hypothesis.

Second Row: (Power Prec.) [95% LCL and UCL Power] (Alpha Prec.) [95% LCL and UCL Alpha]

Summary Statements
- Group sample sizes of 50 and 50 achieve 32% power to detect a difference of -0.6 between the null hypothesis mean difference of 0.0 and the actual mean difference of -0.6 at the 0.050 significance level (alpha) using a two-sided T-Test. These results are based on 2000 Monte Carlo samples from the null distributions: Normal(M0 S) and Normal(M0 S), and the alternative distributions: Normal(M0 S) and Normal(M1 S).

Chart Section

![Power vs N1](chart.png)
This report shows the estimated power for each scenario. The first row shows the parameter settings and the estimated power and significance level (Actual Alpha).

The second row shows two 95% confidence intervals in brackets: the first for the power and the second for the significance level. Half the width of each confidence interval is given in parentheses as a fundamental measure of the accuracy of the simulation. As the number of simulations is increased, the width of the confidence intervals will decrease.

---

**Example 2 – Finding the Sample Size**

Continuing with Example1, the researchers want to determine how large a sample is needed to obtain a power of 0.90.

---

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Means (Simulation)** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Test (Inequality)**, and then clicking on **Tests for Two Means (Simulation)**. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Tab</strong></td>
<td><strong>Value</strong></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N</td>
</tr>
<tr>
<td>Power</td>
<td>0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Group 1 Dist’n</td>
<td>H0</td>
</tr>
<tr>
<td>Group 2 Dist’n</td>
<td>H0</td>
</tr>
<tr>
<td>Group 1 Dist’n</td>
<td>H1</td>
</tr>
<tr>
<td>Group 2 Dist’n</td>
<td>H1</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H1)</td>
</tr>
<tr>
<td>S</td>
<td>2</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Diff#Diff0</td>
</tr>
<tr>
<td>Simulations</td>
<td>2000</td>
</tr>
<tr>
<td>Test Type</td>
<td>T-Test</td>
</tr>
</tbody>
</table>
Output

Click the Run button to perform the calculations and generate the following output.

**Numeric Results of Search for N**

<table>
<thead>
<tr>
<th>Power</th>
<th>N1/N2</th>
<th>H0 Diff0</th>
<th>H1 Diff1</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>Beta</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.904</td>
<td>231/231</td>
<td>0.0</td>
<td>-0.6</td>
<td>0.050</td>
<td>0.053</td>
<td>0.097</td>
<td>0.0</td>
<td>0.6</td>
<td>2.0</td>
</tr>
<tr>
<td>(0.013)</td>
<td>[0.891, 0.916]</td>
<td>(0.010)</td>
<td>[0.043, 0.063]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:

The required sample size was 231 which achieved a power of 0.904. To check the accuracy of this simulation, we ran this scenario through the analytic procedure in PASS which gave the sample size as 234 per group. The simulation answer of 231 was reasonably close.

**Example 3 – Comparative Results**

Continuing with Example 2, the researchers want to study the characteristics of alternative test statistics. They want to compare the results of all test statistics for N1 = 50, 100, and 200.

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means (Simulation) procedure window by expanding Means, then Two Independent Means, then clicking on Test (Inequality), and then clicking on Tests for Two Means (Simulation). You may then make the appropriate entries as listed below, or open Example 3 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>50 100 200</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Group 1 Dist’n</td>
<td>H0</td>
</tr>
<tr>
<td>Group 2 Dist’n</td>
<td>H0</td>
</tr>
<tr>
<td>Group 1 Dist’n</td>
<td>H1</td>
</tr>
<tr>
<td>Group 2 Dist’n</td>
<td>H1</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H1)</td>
</tr>
<tr>
<td>S</td>
<td>2</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Diff#Diff0</td>
</tr>
<tr>
<td>Simulations</td>
<td>2000</td>
</tr>
<tr>
<td>Test Type</td>
<td>T-Test</td>
</tr>
</tbody>
</table>
Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>N1/N2</th>
<th>H0 Diff (Diff0)</th>
<th>H1 Diff (Diff1)</th>
<th>Target Alpha</th>
<th>T-Test Power</th>
<th>Welch Power</th>
<th>Trim. T-Test Power</th>
<th>Trim. Welch Power</th>
<th>Mann Whit'y Power</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>50/50</td>
<td>0.0</td>
<td>-0.6</td>
<td>0.050</td>
<td>0.304</td>
<td>0.303</td>
<td>0.283</td>
<td>0.283</td>
<td>0.288</td>
<td>0.0</td>
<td>0.6</td>
<td>2.0</td>
</tr>
<tr>
<td>100/100</td>
<td>0.0</td>
<td>-0.6</td>
<td>0.050</td>
<td>0.577</td>
<td>0.577</td>
<td>0.538</td>
<td>0.538</td>
<td>0.544</td>
<td>0.0</td>
<td>0.6</td>
<td>2.0</td>
</tr>
<tr>
<td>200/200</td>
<td>0.0</td>
<td>-0.6</td>
<td>0.050</td>
<td>0.859</td>
<td>0.859</td>
<td>0.848</td>
<td>0.848</td>
<td>0.850</td>
<td>0.0</td>
<td>0.6</td>
<td>2.0</td>
</tr>
</tbody>
</table>


These results show that for data that fit the assumptions of the t-test, all five test statistics have accurate alpha values and reasonably close power values. It is interesting to note that the powers of the trimmed procedures, when N1 = 50, are only 7% less than that of the t-test, even though about 20% of the data were trimmed.

Example 4 – Validation using Zar

Zar (1984) page 136 give an example in which the mean difference is 1, the common standard deviation is 0.7206, the sample sizes are 15 in each group, and the significance level is 0.05. They calculate the power to be 0.96.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means (Simulation) procedure window by expanding Means, then Two Independent Means, then clicking on Test (Inequality), and then clicking on Tests for Two Means (Simulation). You may then make the appropriate entries as listed below, or open Example 4 by going to the File menu and choosing Open Example Template.
Inequality Tests for Two Means (Simulation) 440-17

**Option** | **Value**
--- | ---
Data Tab | 
Find (Solve For) | Power
Power | Ignored since this is the Find setting
Alpha | 0.05
N1 (Sample Size Group 1) | 15
N2 (Sample Size Group 2) | Use R
R (Allocation Ratio) | 1.0
Group 1 Dist'n | H0 | N(M0 S)
Group 2 Dist'n | H0 | N(M0 S)
Group 1 Dist'n | H1 | N(M0 S)
Group 2 Dist'n | H1 | N(M1 S)
M0 (Mean|H0) | 0
M1 (Mean|H1) | 1
S | 0.7206
Alternative Hypothesis | Diff\#Diff0
Simulations | 10000
Test Type | T-Test

**Output**
Click the Run button to perform the calculations and generate the following output.

**Numeric Results**

<table>
<thead>
<tr>
<th>Power</th>
<th>N1/N2</th>
<th>H0 Diff0</th>
<th>H1 Diff1</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>Beta</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.956</td>
<td>15/15</td>
<td>0.000</td>
<td>-1.000</td>
<td>0.050</td>
<td>0.045</td>
<td>0.044</td>
<td>0.0</td>
<td>1.0</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Notes:

The power matches the exact value of 0.96.

**Example 5 – Non-Inferiority Test**

A non-inferiority test is used to show that a new treatment is not significantly worse than the standard (or reference) treatment. The maximum deviation that is ‘not significantly worse’ is called the margin of equivalence.

Suppose that the mean diastolic BP of subjects on a certain drug is 96mmHg. If the mean diastolic BP of a new drug is not more than 100mmHg, the drug will be considered non-inferior to the standard drug. The standard deviation among these subjects is 6 mmHg.

The developers of this new drug must design an experiment to test the hypothesis that the mean difference between the two mean BP’s is less than 4. The statistical hypothesis to be tested is
Notice that when the null hypothesis is rejected, the conclusion is that the average difference is less than 4. Following proper procedure, they use a significance level of 0.025 for this one-sided test to keep it comparable to the usual value of 0.05 for a two-sided test. They decide to find the sample size at which the power is 0.90 when the two means are actually equal.

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means (Simulation) procedure window by expanding Means, then Two Independent Means, then clicking on Test (Inequality), and then clicking on Tests for Two Means (Simulation). You may then make the appropriate entries as listed below, or open Example 5 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N1</td>
</tr>
<tr>
<td>Power</td>
<td>0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.025</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Group 1 Dist’n</td>
<td>H0</td>
</tr>
<tr>
<td>Group 2 Dist’n</td>
<td>H0</td>
</tr>
<tr>
<td>Group 1 Dist’n</td>
<td>H1</td>
</tr>
<tr>
<td>Group 2 Dist’n</td>
<td>H1</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H1)</td>
</tr>
<tr>
<td>S</td>
<td>6</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Diff&lt;Diff0</td>
</tr>
<tr>
<td>Simulations</td>
<td>2000</td>
</tr>
<tr>
<td>Test Type</td>
<td>T-Test</td>
</tr>
</tbody>
</table>

**Output**

Click the Run button to perform the calculations and generate the following output.

**Numeric Results**

<table>
<thead>
<tr>
<th>Power</th>
<th>N1/N2</th>
<th>H0 Diff0</th>
<th>H1 Diff1</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>Beta</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.918</td>
<td>49/49</td>
<td>4.0</td>
<td>0.0</td>
<td>0.025</td>
<td>0.024</td>
<td>0.083</td>
<td>96.0</td>
<td>100.0</td>
<td>6.0</td>
</tr>
</tbody>
</table>

We see that 49 subjects are required to achieve the desired experimental design.
Example 6 – Selecting a Test Statistic when the Data Contain Outliers

The two-sample t-test is known to be robust to the violation of some assumptions, but it is susceptible to inaccuracy because the data contain outliers. This example will investigate the impact of outliers on the power and precision of the five test statistics available in PASS.

A mixture of two normal distributions will be used to randomly generate outliers. The mixture will draw 95% of the data from a normal distribution with a mean of 0 and a standard deviation of 1. The other 5% of the data will come from a normal distribution with a mean of 0 and a standard deviation that ranges from 1 to 10.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means (Simulation) procedure window by expanding Means, then Two Independent Means, then clicking on Test (Inequality), and then clicking on Tests for Two Means (Simulation). You may then make the appropriate entries as listed below, or open Example 6 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>20</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Group 1 Dist’n</td>
<td>H0</td>
</tr>
<tr>
<td>Group 2 Dist’n</td>
<td>H0</td>
</tr>
<tr>
<td>Group 1 Dist’n</td>
<td>H1</td>
</tr>
<tr>
<td>Group 2 Dist’n</td>
<td>H1</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H1)</td>
</tr>
<tr>
<td>S</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>1 5 10</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Diff≠Diff0</td>
</tr>
<tr>
<td>Simulations</td>
<td>2000</td>
</tr>
<tr>
<td>Test Type</td>
<td>T-Test</td>
</tr>
</tbody>
</table>

Reports Tab

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Show Comparative Reports</td>
<td>Checked</td>
</tr>
<tr>
<td>Show Comparative Plots</td>
<td>Checked</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.
### Numeric Results

#### Power Comparison for Testing Mean Difference = Diff0

<table>
<thead>
<tr>
<th>N1/N2</th>
<th>H0 Diff (Diff0)</th>
<th>H1 Diff (Diff1)</th>
<th>Target Alpha</th>
<th>T-Test Power</th>
<th>Welch Power</th>
<th>Trim. T-Test Power</th>
<th>Trim. Welch Power</th>
<th>Mann Whit'Y Power</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>20/20</td>
<td>0.0</td>
<td>-1.0</td>
<td>0.050</td>
<td>0.865</td>
<td>0.864</td>
<td>0.835</td>
<td>0.835</td>
<td>0.841</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>20/20</td>
<td>0.0</td>
<td>-1.0</td>
<td>0.050</td>
<td>0.638</td>
<td>0.637</td>
<td>0.789</td>
<td>0.787</td>
<td>0.781</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>5.0</td>
</tr>
<tr>
<td>20/20</td>
<td>0.0</td>
<td>-1.0</td>
<td>0.050</td>
<td>0.469</td>
<td>0.463</td>
<td>0.778</td>
<td>0.775</td>
<td>0.776</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>10.0</td>
</tr>
</tbody>
</table>


#### Alpha Comparison for Testing Mean Difference = Diff0

<table>
<thead>
<tr>
<th>N1/N2</th>
<th>H0 Diff (Diff0)</th>
<th>H1 Diff (Diff1)</th>
<th>Target Alpha</th>
<th>T-Test Alpha</th>
<th>Welch Alpha</th>
<th>Trim. T-Test Alpha</th>
<th>Trim. Welch Alpha</th>
<th>Mann Whit'Y Alpha</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>20/20</td>
<td>0.0</td>
<td>-1.0</td>
<td>0.050</td>
<td>0.046</td>
<td>0.046</td>
<td>0.045</td>
<td>0.044</td>
<td>0.047</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>20/20</td>
<td>0.0</td>
<td>-1.0</td>
<td>0.050</td>
<td>0.037</td>
<td>0.034</td>
<td>0.054</td>
<td>0.052</td>
<td>0.061</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>10.0</td>
</tr>
</tbody>
</table>


The first line gives the results for the standard case in which the two standard deviations (S and A) are equal. Note that in this case, the power of the t-test is a little higher than for the other tests. As the amount of contamination is increased (A equal 5 and then 10), the power of the trimmed tests and the Mann Whitney test remain high, but the power of the t-test falls from 86% to 47%. Also, the value of alpha remains constant for the trimmed and nonparametric tests, but the alpha of the t-test becomes very conservative.

The conclusion this simulation is that if there is a possibility of outliers, you should use either the nonparametric test or the trimmed test.

### Example 7 – Selecting a Test Statistic when the Data are Skewed

The two-sample t-test is known to be robust to the violation of some assumptions, but it is susceptible to inaccuracy when the underlying distributions are skewed. This example will investigate the impact of skewness on the power and precision of the five test statistics available in **PASS**.

Tukey’s lambda distribution will be used because it allows the amount of skewness to be gradually increased.

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Tests for Two Means (Simulation)** procedure window by expanding Means, then Two Independent Means, then clicking on Test (Inequality), and then clicking on Tests for Two Means (Simulation). You may then make the appropriate entries as listed below, or open Example 7 by going to the File menu and choosing Open Example Template.
Inequality Tests for Two Means (Simulation)  440-21

Option  Value

Data Tab

Find (Solve For) .................................. Power

Power ...................................................... Ignored since this is the Find setting

Alpha .................................................. 0.05

N1 (Sample Size Group 1) ................. 20

N2 (Sample Size Group 2) ................. Use R

R (Allocation Ratio) ......................... 1.0

Group 1 Dist’n | H0 .................. L(M0 S G 0)

Group 2 Dist’n | H0 .................. L(M0 S G 0)

Group 1 Dist’n | H1 .................. L(M0 S G 0)

Group 2 Dist’n | H1 .................. L(M1 S G 0)

M0 (Mean|H0) .................. 0

M1 (Mean|H1) .................. 1

S .............................................................. 1

G .............................................................. 0 0.5 0.9

Alternative Hypothesis ................. Diff # Diff0

Simulations ......................... 2000

Test Type ......................... T-Test

Reports Tab

Show Comparative Reports .................... Checked

Show Comparative Plots ...................... Checked

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

Power Comparison for Testing Mean Difference = Diff0.  Hypotheses: H0: Diff1=Diff0; H1: Diff1<>Diff0

<table>
<thead>
<tr>
<th>N1/N2</th>
<th>Diff (Diff0)</th>
<th>H0 Diff</th>
<th>Target</th>
<th>T-Test Power</th>
<th>Welch Power</th>
<th>Trim. T-Test Power</th>
<th>Welch Power</th>
<th>Trim. Mann Power</th>
<th>Welch Power</th>
<th>Whit'y Power</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>20/20</td>
<td>0.0</td>
<td>-1.0</td>
<td>0.050</td>
<td>0.869</td>
<td>0.867</td>
<td>0.833</td>
<td>0.833</td>
<td>0.933</td>
<td>0.938</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>20/20</td>
<td>0.0</td>
<td>-1.0</td>
<td>0.050</td>
<td>0.867</td>
<td>0.966</td>
<td>0.960</td>
<td>0.993</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>0.5</td>
<td>0.5</td>
<td>0.9</td>
<td></td>
</tr>
</tbody>
</table>


Alpha Comparison for Testing Mean Difference = Diff0.  Hypotheses: H0: Diff1=Diff0; H1: Diff1<>Diff0

<table>
<thead>
<tr>
<th>N1/N2</th>
<th>Diff (Diff0)</th>
<th>H0 Diff</th>
<th>Target</th>
<th>T-Test Alpha</th>
<th>Welch Alpha</th>
<th>Trim. T-Test Alpha</th>
<th>Welch Alpha</th>
<th>Trim. Mann Alpha</th>
<th>Welch Alpha</th>
<th>Whit'y Alpha</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>20/20</td>
<td>0.0</td>
<td>-1.0</td>
<td>0.050</td>
<td>0.051</td>
<td>0.051</td>
<td>0.043</td>
<td>0.043</td>
<td>0.948</td>
<td>0.948</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>20/20</td>
<td>0.0</td>
<td>-1.0</td>
<td>0.050</td>
<td>0.038</td>
<td>0.038</td>
<td>0.043</td>
<td>0.043</td>
<td>0.944</td>
<td>0.944</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>0.5</td>
<td></td>
</tr>
<tr>
<td>20/20</td>
<td>0.0</td>
<td>-1.0</td>
<td>0.050</td>
<td>0.049</td>
<td>0.051</td>
<td>0.047</td>
<td>0.054</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>0.9</td>
<td>0.9</td>
<td>0.9</td>
<td></td>
</tr>
</tbody>
</table>


The first line gives the results for the standard case in which there is no skewness (G = 0). Note that in this case, the power of the t-test is a little higher than that of the other tests. As the amount of skewness is increased (G equal 0.5 and then 0.9), the power of the trimmed tests and the Mann
Whitney test increases, but the power of the t-test remains about the same. Also, the value of alpha remains constant for all tests.

The conclusion of this simulation is that if there is skewness, you will gain power by using the nonparametric or trimmed test.
Chapter 445

Inequality Tests for Two Means using Ratios (Two-Sample T-Test)

Introduction

This procedure calculates power and sample size for t-tests from a parallel-groups design in which the logarithm of the outcome is a continuous normal random variable. This routine deals with the case in which the statistical hypotheses are expressed in terms of mean ratios instead of mean differences.

The details of testing two treatments using data from a two-group design are given in another chapter, and they will not be repeated here. If the logarithms of the responses can be assumed to follow a normal distribution, hypotheses stated in terms of the ratio can be transformed into hypotheses about the difference. The details of this analysis are given in Julious (2004). They will only be summarized here.

Testing Using Ratios

It will be convenient to adopt the following specialized notation for the discussion of these tests.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PASS Input/Output</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_T$</td>
<td>Not used</td>
<td>Treatment mean. This is the treatment (group 2) mean.</td>
</tr>
<tr>
<td>$\mu_R$</td>
<td>Not used</td>
<td>Reference mean. This is the reference (group 1) mean.</td>
</tr>
<tr>
<td>$\phi$</td>
<td>R1</td>
<td>True ratio. This is the value of $\phi = \frac{\mu_T}{\mu_R}$ at which the power is calculated.</td>
</tr>
</tbody>
</table>

Note that the actual values of $\mu_T$ and $\mu_R$ are not needed. Only the ratio of these values is needed for power and sample size calculations.
In the two-sided case, the null hypothesis is

\[ H_0: \phi = \phi_0 \]

and the alternative hypothesis is

\[ H_1: \phi \neq \phi_0 \]

**Log-Transformation**

In many cases, hypotheses stated in terms of ratios are more convenient than hypotheses stated in terms of differences. This is because ratios can be interpreted as percentages, but differences must be interpreted as actual amounts in their original scale. Hence, it has become a common practice to take the following steps in hypothesis testing.

1. State the statistical hypotheses in terms of the ratio of the means.
2. Transform this into hypotheses about a difference by taking logarithms.
3. Analyze the logged data—that is, do the analysis in terms of the difference.
4. Draw the conclusion in terms of the ratio.

The details of step 2 are as follows for the null hypothesis.

\[ \phi = \phi_0 \]

\[ \Rightarrow \phi = \frac{\mu_T}{\mu_R} \]

\[ \Rightarrow \ln(\phi) \neq \{\ln(\mu_T) - \ln(\mu_R)\} \]

Thus, a hypothesis about the ratio of the means on the original scale can be translated into a hypothesis about the difference of two means on the logged scale.

**Coefficient of Variation**

The coefficient of variation (COV) is the ratio of the standard deviation to the mean. This parameter can be used to represent the variation in the data because of a unique relationship that it has in the case of log-normal data.

Suppose the variable \( X \) is the logarithm of the original variable \( Y \). That is, \( X = \ln(Y) \) and \( Y = \exp(X) \). Label the mean and variance of \( X \) as \( \mu_X \) and \( \sigma_X^2 \), respectively. Similarly, label the mean and variance of \( Y \) as \( \mu_Y \) and \( \sigma_Y^2 \), respectively. If \( X \) is normally distributed, then \( Y \) is log-normally distributed. Julious (2004) presents the following well-known relationships between these two variables

\[ \mu_Y = e^{\mu_X + \frac{\sigma_X^2}{2}} \]

\[ \sigma_Y^2 = \mu_Y^2 \left(e^{\sigma_X^2} - 1\right) \]
From this relationship, the coefficient of variation of \( Y \) can be found to be

\[
COV_y = \sqrt{\frac{\mu_y^2 (e^{\sigma_y^2} - 1)}{\mu_y}}
\]

\[
= \sqrt{e^{\sigma_y^2} - 1}
\]

Solving this relationship for \( \sigma_x^2 \), the standard deviation of \( X \) can be stated in terms of the coefficient of variation of \( Y \). This equation is

\[
\sigma_x = \sqrt{\ln(COV_y^2 + 1)}
\]

Similarly, the mean of \( X \) is

\[
\mu_x = \frac{\mu_y}{\ln(COV_y^2 + 1)}
\]

One final note: for parallel-group designs, \( \sigma_x^2 \) equals \( \sigma_d^2 \), the average variance used in the t-test of the logged data.

Thus, the hypotheses can be stated in the original (\( Y \)) scale and then the power can be analyzed in the transformed (\( X \)) scale.

### Power Calculation

As is shown above, the hypotheses can be stated in the original (\( Y \)) scale using ratios or the logged (\( X \)) scale using differences. In either case, the power and sample size calculations are made using the formulas for testing the difference in two means. These formulas are presented in another chapter and are not duplicated here.

### Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

### Data Tab

The Data tab contains the parameters associated with this test such as the means, sample sizes, alpha, and power.

### Solve For

Find (Solve For)

This option specifies the parameter to be solved for from the other parameters. In most situations, you will select either Power and Beta for a power analysis or \( N1 \) for sample size determination.
Error Rates

Power or Beta
This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of equal means when in fact the means are different.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

Alpha (Significance Level)
This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when you reject the null hypothesis of equal means when in fact the means are equal.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.

Sample Size

N1 (Sample Size)
Enter a value (or range of values) for the sample size of group 1 (the reference group). Note that these values are ignored when you are solving for N1. You may enter a range of values such as 10 to 100 by 10.

N2 (Sample Size Group 2)
Enter a value (or range of values) for the sample size of group 2 (the treatment group) or enter Use R to base N2 on the value of N1. You may enter a range of values such as 10 to 100 by 10.

- Use R
  When Use R is entered here, N2 is calculated using the formula
  \[ N2 = \lceil R(N1) \rceil \]
  where \( R \) is the Sample Allocation Ratio and the operator \( \lceil Y \rceil \) is the first integer greater than or equal to \( Y \). For example, if you want \( N1 = N2 \), select Use R and set \( R = 1 \).

R (Sample Allocation Ratio)
Enter a value (or range of values) for \( R \), the allocation ratio between samples. This value is only used when \( N2 \) is set to Use R.

When used, N2 is calculated from N1 using the formula: \( N2 = \lceil R(N1) \rceil \) where \( \lceil Y \rceil \) is the next integer greater than or equal to \( Y \). Note that setting \( R = 1.0 \) forces \( N2 = N1 \).
Effect Size – Ratios

R0 (Ratio Under H0)
This is the value of the ratio of the two means assumed by the null hypothesis, H0. Usually, R0 = 1.0 which implies that the two means are equal. However, you may test other values of R0 as well. Strictly speaking, any positive number is valid, but values near to, or equal to, 1.0 are usually used.

Warning: you cannot use the same value for both R0 and R1.

R1 (True Ratio)
This is the value of the ratio of the two means at which the power is to be calculated. Often, a range of values will be tried. For example, you might try the four values:

1.05 1.10 1.15 1.20

Strictly speaking, any positive number is valid. However, numbers between 0.50 and 2.00 are usually used.

Warning: you cannot use the same value for both R0 and R1.

Effect Size – Coefficient of Variation

COV (Coefficient of Variation)
The coefficient of variation is used to specify the variability (standard deviation). It is important to realize that this is the COV defined on the original (not log) scale. This value must be determined from past experience or from a pilot study. See the discussion above for more details on the definition of the coefficient of variation.

Test

Alternative Hypothesis
This option specifies the alternative hypothesis. This implicitly specifies the direction of the hypothesis test. Possible selections are:

- **H1: R1 <> R0**
  This is the most common selection. It yields the two-tailed t-test. Use this option when you are testing whether the means are different, but you do not want to specify beforehand which mean is larger.

- **H1: R1 < R0**
  This option yields a one-tailed t-test. Use it when you are only interested in the case in which Mean1 is greater than Mean2.

- **H1: R1 > R0**
  This option yields a one-tailed t-test. Use it when you are only interested in the case in which Mean1 is less than Mean2.
Example 1 – Finding Power

A company has developed a generic drug for treating rheumatism and wants to show that it is better than the standard drug. From previous studies, responses for either treatment are known to follow a lognormal distribution. A parallel-group design will be used and the logged data will be analyzed with a one-sided, two-sample t-test.

Past experience leads the researchers to set the COV to 1.20. The significance level is 0.025. The power will be computed for R1 equal 1.10 and 1.20. Sample sizes between 100 and 900 will be examined in the analysis.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means (Two-Sample T-Test) [Ratios] procedure window by expanding Means, then Two Independent Means, then clicking on Test (Inequality), and then clicking on Tests for Two Means (Two-Sample T-Test) [Ratios]. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.025</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>100 to 900 by 200</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>R0 (Ratio under H0)</td>
<td>1.0</td>
</tr>
<tr>
<td>R1 (True Ratio)</td>
<td>1.1 1.2</td>
</tr>
<tr>
<td>COV (Coefficient of Variation)</td>
<td>1.2</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>R1&gt;R0 (One-Sided)</td>
</tr>
</tbody>
</table>
Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>Group Sample Sizes</th>
<th>Mean Ratio Under H0 (R0)</th>
<th>Mean Ratio Under H1 (R1)</th>
<th>Effect Size (ES)</th>
<th>Coefficient of Variation (COV)</th>
<th>Significance Level (Alpha)</th>
<th>Significance Level (Beta)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1057</td>
<td>100/100</td>
<td>1.00</td>
<td>1.100</td>
<td>0.1009</td>
<td>1.200</td>
<td>0.0250</td>
<td>0.8943</td>
</tr>
<tr>
<td>0.2339</td>
<td>300/300</td>
<td>1.00</td>
<td>1.100</td>
<td>0.1009</td>
<td>1.200</td>
<td>0.0250</td>
<td>0.7661</td>
</tr>
<tr>
<td>0.3581</td>
<td>500/500</td>
<td>1.00</td>
<td>1.100</td>
<td>0.1009</td>
<td>1.200</td>
<td>0.0250</td>
<td>0.6419</td>
</tr>
<tr>
<td>0.4715</td>
<td>700/700</td>
<td>1.00</td>
<td>1.100</td>
<td>0.1009</td>
<td>1.200</td>
<td>0.0250</td>
<td>0.5285</td>
</tr>
<tr>
<td>0.5718</td>
<td>900/900</td>
<td>1.00</td>
<td>1.100</td>
<td>0.1009</td>
<td>1.200</td>
<td>0.0250</td>
<td>0.4282</td>
</tr>
<tr>
<td>0.2737</td>
<td>100/100</td>
<td>1.00</td>
<td>1.200</td>
<td>0.1930</td>
<td>1.200</td>
<td>0.0250</td>
<td>0.7263</td>
</tr>
<tr>
<td>0.6556</td>
<td>300/300</td>
<td>1.00</td>
<td>1.200</td>
<td>0.1930</td>
<td>1.200</td>
<td>0.0250</td>
<td>0.3444</td>
</tr>
<tr>
<td>0.8625</td>
<td>500/500</td>
<td>1.00</td>
<td>1.200</td>
<td>0.1930</td>
<td>1.200</td>
<td>0.0250</td>
<td>0.1375</td>
</tr>
<tr>
<td>0.9506</td>
<td>700/700</td>
<td>1.00</td>
<td>1.200</td>
<td>0.1930</td>
<td>1.200</td>
<td>0.0250</td>
<td>0.0494</td>
</tr>
<tr>
<td>0.9836</td>
<td>900/900</td>
<td>1.00</td>
<td>1.200</td>
<td>0.1930</td>
<td>1.200</td>
<td>0.0250</td>
<td>0.0164</td>
</tr>
</tbody>
</table>

Report Definitions

Power is the probability of rejecting a false null hypothesis. Power should be close to one.
N1 and N2 are the number of items sampled from each population.
Alpha is the probability of rejecting a true null hypothesis.
Beta is the probability of accepting a false null hypothesis.
R0 is the ratio of the means (Mean2/Mean1) under the null hypothesis, H0.
R1 is the ratio of the means (Mean2/Mean1) at which the power is calculated.
COV is the coefficient of variation on the original scale. The value of sigma is calculated from this.
ES is the effect size which is |ln(R0)-ln(R1)| / (sigma).

Summary Statements

A one-sided, two-sample t-test with group sample sizes of 100 and 100 achieves 11% power to detect a ratio of 1.100 when the ratio under the null hypothesis is 1.000. The coefficient of variation on the original scale is 1.200. The significance level (alpha) is 0.0250.

This report shows the power for the indicated scenarios.

Plots Section

This plot shows the power versus the sample size.
Example 2 – Validation

We will validate this procedure by showing that it gives the identical results to the regular test on differences—a procedure that has been validated. We will use the same settings as those given in Example 1. Since the output for this example is shown above, only the output from the procedure that uses differences is shown below.

To run the power analysis of a *t*-test on differences, we need the values of Mean2 (which correspond to R1) and S1. The value of Mean1 will be zero.

\[
S1 = \sqrt{\ln(COV^2 + 1)}
\]

\[
= \sqrt{\ln(1.2^2 + 1)}
\]

\[
= 0.944456
\]

\[
Mean2 = \ln(R1)
\]

\[
= \ln(1.10) = \ln(1.20)
\]

\[
= 0.095310 = 0.182322
\]

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means (Two-Sample T-Test) [Differences] procedure window by expanding Means, then Two Independent Means, then clicking on Test (Inequality), and then clicking on Tests for Two Means (Two-Sample T-Test) [Differences]. You may then make the appropriate entries as listed below, or open Example 1b by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.025</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>100 to 900 by 200</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Mean1 (Mean of Group 1)</td>
<td>0</td>
</tr>
<tr>
<td>Mean2 (Mean of Group 2)</td>
<td>0.095310 0.182322</td>
</tr>
<tr>
<td>S1 (Standard Deviation Group 1)</td>
<td>0.944456</td>
</tr>
<tr>
<td>S1 (Standard Deviation Group 2)</td>
<td>S1</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Ha: Mean1&lt;Mean2</td>
</tr>
<tr>
<td>Nonparametric Adjustment</td>
<td>Ignore</td>
</tr>
</tbody>
</table>
**Output**

Click the Run button to perform the calculations and generate the following output.

**Numeric Results**

<table>
<thead>
<tr>
<th>Power</th>
<th>N1</th>
<th>N2</th>
<th>Ratio</th>
<th>Alpha</th>
<th>Beta</th>
<th>Mean1</th>
<th>Mean2</th>
<th>S1</th>
<th>S2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1057</td>
<td>100</td>
<td>100</td>
<td>1.00</td>
<td>0.0250</td>
<td>0.8943</td>
<td>0.0000</td>
<td>0.0953</td>
<td>0.9445</td>
<td>0.9445</td>
</tr>
<tr>
<td>0.2339</td>
<td>300</td>
<td>300</td>
<td>1.00</td>
<td>0.0250</td>
<td>0.7649</td>
<td>0.0000</td>
<td>0.0953</td>
<td>0.9445</td>
<td>0.9445</td>
</tr>
<tr>
<td>0.3581</td>
<td>500</td>
<td>500</td>
<td>1.00</td>
<td>0.0250</td>
<td>0.6419</td>
<td>0.0000</td>
<td>0.0953</td>
<td>0.9445</td>
<td>0.9445</td>
</tr>
<tr>
<td>0.4715</td>
<td>700</td>
<td>700</td>
<td>1.00</td>
<td>0.0250</td>
<td>0.5285</td>
<td>0.0000</td>
<td>0.0953</td>
<td>0.9445</td>
<td>0.9445</td>
</tr>
<tr>
<td>0.5718</td>
<td>900</td>
<td>900</td>
<td>1.00</td>
<td>0.0250</td>
<td>0.4282</td>
<td>0.0000</td>
<td>0.0953</td>
<td>0.9445</td>
<td>0.9445</td>
</tr>
<tr>
<td>0.2737</td>
<td>100</td>
<td>100</td>
<td>1.00</td>
<td>0.0250</td>
<td>0.7263</td>
<td>0.0000</td>
<td>0.1823</td>
<td>0.9445</td>
<td>0.9445</td>
</tr>
<tr>
<td>0.6556</td>
<td>300</td>
<td>300</td>
<td>1.00</td>
<td>0.0250</td>
<td>0.3429</td>
<td>0.0000</td>
<td>0.1823</td>
<td>0.9445</td>
<td>0.9445</td>
</tr>
<tr>
<td>0.8625</td>
<td>500</td>
<td>500</td>
<td>1.00</td>
<td>0.0250</td>
<td>0.1375</td>
<td>0.0000</td>
<td>0.1823</td>
<td>0.9445</td>
<td>0.9445</td>
</tr>
<tr>
<td>0.9506</td>
<td>700</td>
<td>700</td>
<td>1.00</td>
<td>0.0250</td>
<td>0.0494</td>
<td>0.0000</td>
<td>0.1823</td>
<td>0.9445</td>
<td>0.9445</td>
</tr>
<tr>
<td>0.9836</td>
<td>900</td>
<td>900</td>
<td>1.00</td>
<td>0.0250</td>
<td>0.0164</td>
<td>0.0000</td>
<td>0.1823</td>
<td>0.9445</td>
<td>0.9445</td>
</tr>
</tbody>
</table>

You can compare these power values with those shown above in Example 1 to validate the procedure. You will find that the power values are identical.
Chapter 448

Tests for Two Means with Non-Zero Null using Differences

Introduction

This procedure computes power and sample size for non-zero null tests in two-sample designs in which the outcome is a continuous normal random variable. Measurements are made on individuals that have been randomly assigned to one of two groups. This is sometimes referred to as a parallel-groups design. This design is used in situations such as the comparison of the income level of two regions, the nitrogen content of two lakes, or the effectiveness of two drugs.

The two-sample t-test is commonly used with this situation. When the variances of the two groups are unequal, Welch’s t-test may be used. When the data are not normally distributed, the Mann-Whitney (Wilcoxon signed-ranks) U test may be used.

The details of sample size calculation for the two-sample design are presented in the Two-Sample T-Test chapter and they will not be duplicated here. This chapter only discusses those changes necessary for non-inferiority and superiority tests. Sample size formulas for non-inferiority and superiority tests of two means are presented in Chow et al. (2003) pages 57-59.

The Statistical Hypotheses

Both non-inferiority and superiority tests are examples of directional (one-sided) tests and their power and sample size could be calculated using the Two-Sample T-Test procedure. However, at the urging of our users, we have developed this module, which provides the input and output in formats that are convenient for these types of tests. This section will review the specifics of non-inferiority and superiority testing.
Remember that in the usual t-test setting, the null (H0) and alternative (H1) hypotheses for one-sided tests are defined as

\[ H_0: \mu_1 - \mu_2 \leq D \quad \text{versus} \quad H_1: \mu_1 - \mu_2 > D \]

Rejecting this test implies that the mean difference is larger than the value \( D \). This test is called an upper-tailed test because it is rejected in samples in which the difference between the sample means is larger than \( D \).

Following is an example of a lower-tailed test.

\[ H_0: \mu_1 - \mu_2 \geq D \quad \text{versus} \quad H_1: \mu_1 - \mu_2 < D \]

Non-inferiority and non-zero null tests are special cases of the above directional tests. It will be convenient to adopt the following specialized notation for the discussion of these tests.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PASS Input/Output</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu_1 )</td>
<td>Not used</td>
<td>Mean of population 1. Population 1 is assumed to consist of those who have received the new treatment.</td>
</tr>
<tr>
<td>( \mu_2 )</td>
<td>Not used</td>
<td>Mean of population 2. Population 2 is assumed to consist of those who have received the reference treatment.</td>
</tr>
<tr>
<td>( M_s )</td>
<td>SM</td>
<td>Margin of superiority. This is a tolerance value that defines the magnitude of difference that is required for practical importance. This may be thought of as the smallest difference from the reference that is considered to be different.</td>
</tr>
<tr>
<td>( \delta )</td>
<td>D</td>
<td>True difference. This is the value of ( \mu_1 - \mu_2 ), the difference between the means. This is the value at which the power is calculated.</td>
</tr>
</tbody>
</table>

Note that the actual values of \( \mu_1 \) and \( \mu_2 \) are not needed. Only their difference is needed for power and sample size calculations.

**Non-Zero Null Tests**

A non-zero null test tests that the treatment mean is better than the reference mean by more than the superiority margin. The actual direction of the hypothesis depends on the response variable being studied.
Case 1: High Values Good
In this case, higher values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is greater than the reference mean by at least the margin of superiority. The value of $\delta$ must be greater than $|M_S|$. The following are equivalent sets of hypotheses.

\[
\begin{align*}
H_0 : \mu_1 &\leq \mu_2 + |M_S| \quad \text{versus} \quad H_1 : \mu_1 > \mu_2 + |M_S| \\
H_0 : \mu_1 - \mu_2 &\leq |M_S| \quad \text{versus} \quad H_1 : \mu_1 - \mu_2 > |M_S| \\
H_0 : \delta &\leq |M_S| \quad \text{versus} \quad H_1 : \delta > |M_S| 
\end{align*}
\]

Case 2: High Values Bad
In this case, lower values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is less than the reference mean by at least the margin of superiority. The value of $\delta$ must be less than $-|e|$. The following are equivalent sets of hypotheses.

\[
\begin{align*}
H_0 : \mu_1 &\geq \mu_2 - |M_S| \quad \text{versus} \quad H_1 : \mu_1 < \mu_2 - |M_S| \\
H_0 : \mu_1 - \mu_2 &\geq -|M_S| \quad \text{versus} \quad H_1 : \mu_1 - \mu_2 < -|M_S| \\
H_0 : \delta &\geq -|M_S| \quad \text{versus} \quad H_1 : \delta < -|M_S| 
\end{align*}
\]

Example
A non-zero null test example will set the stage for the discussion of the terminology that follows. Suppose that a test is to be conducted to determine if a new cancer treatment substantially improves mean bone density. The adjusted mean bone density (AMBD) in the population of interest is 0.002300 gm/cm with a standard deviation of 0.000300 gm/cm. Clinicians decide that if the treatment increases AMBD by more than 5% (0.000115 gm/cm), it provides a significant health benefit.

The hypothesis of interest is whether the mean AMBD in the treated group is more than 0.000115 above that of the reference group. The statistical test will be set up so that if the null hypothesis is rejected, the conclusion will be that the new treatment is superior. The value 0.000115 gm/cm is called the margin of superiority.

Test Statistics
This section describes the test statistics that are available in this procedure.

Two-Sample T-Test
Under the null hypothesis, this test assumes that the two groups of data are simple random samples from a single population of normally-distributed values that all have the same mean and variance. This assumption implies that the data are continuous and their distribution is symmetric.
Tests for Two Means with Non-Zero Null using Differences

The calculation of the test statistic for the case when higher response values are good is as follows.

\[ t_{df} = \frac{(\bar{X}_1 - \bar{X}_2) - |\epsilon|}{s_{\bar{X}_1-\bar{X}_2}} \]

where

\[ \bar{X}_k = \frac{\sum_{i=1}^{N_k} X_{ki}}{N_k} \]

\[ s_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{\sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2 + \sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2}{N_1 + N_2 - 2} \left( \frac{1}{N_1} + \frac{1}{N_2} \right)} \]

\[ df = N_1 + N_2 - 2 \]

The null hypothesis is rejected if the computed p-value is less than a specified level (usually 0.05). Otherwise, no conclusion can be reached.

**Welch’s T-Test**

Welch (1938) proposed the following test when the two variances are not assumed to be equal.

\[ t_f^{*} = \frac{(\bar{X}_1 - \bar{X}_2) - |\epsilon|}{s_{\bar{X}_1-\bar{X}_2}^{*}} \]

where

\[ s_{\bar{X}_1-\bar{X}_2}^{*} = \sqrt{\left( \frac{\sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2}{N_1(N_1 - 1)} \right) + \left( \frac{\sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2}{N_2(N_2 - 1)} \right)} \]

\[ f = \frac{\left( \frac{s_1^2}{N_1} + \frac{s_2^2}{N_2} \right)^2}{\frac{s_1^4}{N_1^2(N_1 - 1)} + \frac{s_2^4}{N_2^2(N_2 - 1)}} \]

\[ s_1 = \sqrt{\frac{\sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2}{N_1 - 1}} \]

\[ s_2 = \sqrt{\frac{\sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2}{N_2 - 1}} \]
**Mann-Whitney U Test**

This test is the nonparametric substitute for the equal-variance t-test. Two key assumptions are that the distributions are at least ordinal and that they are identical under H0. This means that ties (repeated values) are not acceptable. When ties are present, you can use approximations, but the theoretic results no longer hold.

The Mann-Whitney test statistic is defined as follows in Gibbons (1985).

\[
W_1 = \sum_{k=1}^{N_1} \text{Rank}(X_{1k})
\]

The ranks are determined after combining the two samples. The standard deviation is calculated as

\[
s_w = \sqrt{\frac{N_1 N_2 (N_1 + N_2 + 1)}{12} - \frac{N_1 N_2 \sum_{i=1}^{t_i} (t_i - t_i)^3}{12(N_1 + N_2)(N_1 + N_2 - 1)}}
\]

where \( t_1 \) is the number of observations tied at value one, \( t_2 \) is the number of observations tied at some value two, and so forth.

The correction factor, \( C \), is 0.5 if the rest of the numerator is negative or -0.5 otherwise. The value of \( z \) is then compared to the normal distribution.

**Computing the Power**

**Standard Deviations Equal**

When \( \sigma_1 = \sigma_2 = \sigma \), the power of the \( t \) test is calculated as follows.

1. Find \( t_\alpha \) such that \( 1 - T_{df}(t_\alpha) = \alpha \), where \( T_{df}(t_\alpha) \) is the area under a central-\( t \) curve to the left of \( x \) and \( df = N_1 + N_2 - 2 \).

2. Calculate: \( \sigma_x = \sqrt{\frac{I}{N_1} + \frac{I}{N_2}} \)

3. Calculate the noncentrality parameter: \( \lambda = \frac{\delta}{\sigma_x} \)

4. Calculate: Power = \( 1 - T'_{df,\lambda}(t_\alpha) \), where \( T'_{df,\lambda}(x) \) is the area to the left of \( x \) under a noncentral-\( t \) curve with degrees of freedom \( df \) and noncentrality parameter \( \lambda \).
Standard Deviations Unequal

This case often recommends Welch’s test. When $\sigma_1 \neq \sigma_2$, the power is calculated as follows.

1. Calculate: $\sigma^*_x = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}$

2. Calculate: $f = \frac{\sigma_x^4}{\sigma_1^4} \left( N_1^2 \frac{1}{N_1+1} \right)^{2} \frac{\sigma_2^4}{N_2^2 (N_2+1)}$

   which is the adjusted degrees of freedom. Often, this is rounded to the next highest integer. Note that this is not the value of $f$ used in the computation of the actual test. Instead, this is the expected value of $f$.

3. Find $t_\alpha$ such that $1 - T_f(t_\alpha) = \alpha$, where $T_f(t_\alpha)$ is the area to the left of $x$ under a central-$t$ curve with $f$ degrees of freedom.

4. Calculate: $\lambda = \frac{\lambda}{\sigma^*_x}$, the noncentrality parameter.

5. Calculate: Power = $1 - T'_{f,\lambda}(t_\alpha)$, where $T'_{f,\lambda}(x)$ is the area to the left of $x$ under a noncentral-$t$ curve with degrees of freedom $f$ and noncentrality parameter $\lambda$.

Nonparametric Adjustment

When using the Mann-Whitney test rather than the $t$ test, results by Al-Sunduqchi and Guenther (1990) indicate that power calculations for the Mann-Whitney test may be made using the standard $t$ test formulations with a simple adjustment to the sample sizes. The size of the adjustment depends on the actual distribution of the data. They give sample size adjustment factors for four distributions. These are 1 for uniform, 2/3 for double exponential, 9 / $\pi^2$ for logistic, and $\pi / 3$ for normal distributions.
Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains most of the parameters and options that you will be concerned with.

Solve For

Find (Solve For)

This option specifies the parameter to be calculated from the values of the other parameters. Under most conditions, you would select either Power and Beta or $N_1$.

Select $N_1$ when you want to determine the sample size needed to achieve a given power and alpha.

Select Power and Beta when you want to calculate the power of an experiment that has already been run.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of inferiority when the null hypothesis should be rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when you reject the null hypothesis of inferiority when in fact the mean is not non-inferior.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.
Sample Size

N1 (Sample Size Group 1)
Enter a value (or range of values) for the sample size of group 1. Note that these values are ignored when you are solving for N1. You may enter a range of values such as 10 to 100 by 10.

N2 (Sample Size Group 2)
Enter a value (or range of values) for the sample size of group 2 or enter Use R to base N2 on the value of N1. You may enter a range of values such as 10 to 100 by 10.

- **Use R**
  When Use R is entered here, N2 is calculated using the formula
  \[ N2 = \lceil R(N1) \rceil \]
  where \( R \) is the Sample Allocation Ratio and the operator \( \lceil Y \rceil \) is the first integer greater than or equal to \( Y \). For example, if you want \( N1 = N2 \), select Use R and set \( R = 1 \).

R (Sample Allocation Ratio)
Enter a value (or range of values) for \( R \), the allocation ratio between samples. This value is only used when N2 is set to Use R.

When used, N2 is calculated from N1 using the formula: \( N2 = \lceil R(N1) \rceil \) where \( \lceil Y \rceil \) is the next integer greater than or equal to \( Y \). Note that setting \( R = 1.0 \) forces \( N2 = N1 \).

Effect Size – Mean Difference

SM (Superiority Margin)
This is the magnitude of the margin of superiority. It must be entered as a positive number.

When higher means are better, this value is the distance above the reference mean that is required to be considered superior. When higher means are worse, this value is the distance below the reference mean that is required to be considered superior.

D (True Difference, Trt Mean – Ref Mean)
This is the actual difference between the treatment mean and the reference mean at which the power is calculated.

When higher means are better, this value should be greater than SM. When higher means are worse, this value should be negative and greater in magnitude than SM.

Effect Size – Standard Deviations

S1 and S2 (Standard Deviations)
These options specify the values of the standard deviations for each group. When the S2 is set to S1, only S1 needs to be specified. The value of S1 will be copied into S2.

When these values are not known, you must supply estimates of them. Press the SD button to display the Standard Deviation Estimator window. This procedure will help you find appropriate values for the standard deviation.
Tests for Two Means with Non-Zero Null using Differences

**Test**

**Higher Means Are**
This option defines whether higher values of the response variable are to be considered better or worse. The choice here determines the direction of the test.

If Higher Means Are Better the null hypothesis is $\text{Diff} \leq \text{SM}$ and the alternative hypothesis is $\text{Diff} > \text{SM}$. If Higher Means Are Worse the null hypothesis is $\text{Diff} \geq -\text{SM}$ and the alternative hypothesis is $\text{Diff} < -\text{SM}$.

**Nonparametric Adjustment (Mann-Whitney Test)**
This option makes appropriate sample size adjustments for the Mann-Whitney test. Results by Al-Sunduqchi and Guenther (1990) indicate that power calculations for the Mann-Whitney test may be made using the standard $t$ test formulations with a simple adjustment to the sample size. The size of the adjustment depends upon the actual distribution of the data. They give sample size adjustment factors for four distributions. These are 1 for the uniform distribution, $2/3$ for the double exponential distribution, $9/\pi^2$ for the logistic distribution, and $\pi/3$ for the normal distribution.

The options are as follows:

- **Ignore**
  Do not make a Mann-Whitney adjustment. This indicates that you want to analyze a $t$ test, not the Wilcoxon test.

- **Uniform**
  Make the Mann-Whitney sample size adjustment assuming the uniform distribution. Since the factor is one, this option performs the same function as Ignore. It is included for completeness.

- **Double Exponential**
  Make the Mann-Whitney sample size adjustment assuming that the data actually follow the double exponential distribution.

- **Logistic**
  Make the Mann-Whitney sample size adjustment assuming that the data actually follow the logistic distribution.

- **Normal**
  Make the Mann-Whitney sample size adjustment assuming that the data actually follow the normal distribution.

**Example 1 – Power Analysis**

Suppose that a test is to be conducted to determine if a new cancer treatment improves bone density. The adjusted mean bone density (AMBD) in the population of interest is 0.002300 gm/cm with a standard deviation of 0.000300 gm/cm. Clinicians decide that if the treatment increases AMBD by more than 5% (0.000115 gm/cm), it generates a significant health benefit.
They also want to consider what would happen if the margin of superiority is set to 2.5%
(0.0000575 gm/cm).

The analysis will be a non-zero null test using the t-test at the 0.025 significance level. Power to
be calculated assuming that the new treatment has 7.5% improvement on AMBD. Several sample
sizes between 10 and 800 will be analyzed. The researchers want to achieve a power of at least
90%. All numbers have been multiplied by 10000 to make the reports and plots easier to read.

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from
the PASS Home window, load the Tests for Two Means with Non-Zero Null (Two-Sample T-
Test) [Differences] procedure window by expanding Means, then Two Means, then clicking on
Test (Non-Zero Null), and then clicking on Tests for Two Means with Non-Zero Null (Two-
Sample T-Test) [Differences]. You may then make the appropriate entries as listed below, or
open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.025</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>10 50 100 200 300 500 600 800</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>SM (Superiority Margin)</td>
<td>0.575 1.15</td>
</tr>
<tr>
<td>D (True Difference)</td>
<td>1.725</td>
</tr>
<tr>
<td>S1 (Standard Deviation Group 1)</td>
<td>3</td>
</tr>
<tr>
<td>S2 (Standard Deviation Group 2)</td>
<td>S1</td>
</tr>
<tr>
<td>Higher Means Are</td>
<td>Better</td>
</tr>
<tr>
<td>Nonparametric Adjustment</td>
<td>Ignore</td>
</tr>
</tbody>
</table>
**Tests for Two Means with Non-Zero Null using Differences  448-11**

**Annotated Output**

Click the Run button to perform the calculations and generate the following output.

**Numeric Results and Plots**

<table>
<thead>
<tr>
<th>Power</th>
<th>N1/N2</th>
<th>Superiority Margin (SM)</th>
<th>Actual Difference (D)</th>
<th>Significance Level (Alpha)</th>
<th>Beta</th>
<th>Standard Deviation1 (SD1)</th>
<th>Standard Deviation2 (SD2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.12553</td>
<td>10/10</td>
<td>0.575</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.87447</td>
<td>3.000</td>
<td>3.000</td>
</tr>
<tr>
<td>0.47524</td>
<td>50/50</td>
<td>0.575</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.52476</td>
<td>3.000</td>
<td>3.000</td>
</tr>
<tr>
<td>0.76957</td>
<td>100/100</td>
<td>0.575</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.23043</td>
<td>3.000</td>
<td>3.000</td>
</tr>
<tr>
<td>0.96885</td>
<td>200/200</td>
<td>0.575</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.03115</td>
<td>3.000</td>
<td>3.000</td>
</tr>
<tr>
<td>0.99681</td>
<td>300/300</td>
<td>0.575</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.00319</td>
<td>3.000</td>
<td>3.000</td>
</tr>
<tr>
<td>0.99998</td>
<td>500/500</td>
<td>0.575</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.00002</td>
<td>3.000</td>
<td>3.000</td>
</tr>
<tr>
<td>1.00000</td>
<td>600/600</td>
<td>0.575</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.00000</td>
<td>3.000</td>
<td>3.000</td>
</tr>
<tr>
<td>1.00000</td>
<td>800/800</td>
<td>0.575</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.00000</td>
<td>3.000</td>
<td>3.000</td>
</tr>
<tr>
<td>0.06013</td>
<td>10/10</td>
<td>1.150</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.93987</td>
<td>3.000</td>
<td>3.000</td>
</tr>
<tr>
<td>0.15601</td>
<td>50/50</td>
<td>1.150</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.84399</td>
<td>3.000</td>
<td>3.000</td>
</tr>
<tr>
<td>0.27052</td>
<td>100/100</td>
<td>1.150</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.72948</td>
<td>3.000</td>
<td>3.000</td>
</tr>
<tr>
<td>0.48089</td>
<td>200/200</td>
<td>1.150</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.51911</td>
<td>3.000</td>
<td>3.000</td>
</tr>
<tr>
<td>0.64940</td>
<td>300/300</td>
<td>1.150</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.35060</td>
<td>3.000</td>
<td>3.000</td>
</tr>
<tr>
<td>0.85769</td>
<td>500/500</td>
<td>1.150</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.14231</td>
<td>3.000</td>
<td>3.000</td>
</tr>
<tr>
<td>0.91295</td>
<td>600/600</td>
<td>1.150</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.08705</td>
<td>3.000</td>
<td>3.000</td>
</tr>
<tr>
<td>0.96943</td>
<td>800/800</td>
<td>1.150</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.03057</td>
<td>3.000</td>
<td>3.000</td>
</tr>
</tbody>
</table>

**Report Definitions**

Group 1 is the treatment group. Group 2 is the reference or standard group.

Power is the probability of rejecting a false null hypothesis.

N1 is the number of subjects in the first (treatment) group.

N2 is the number of subjects in the second (reference) group.

SM is the magnitude of the margin of superiority. Since higher means are better, this value is positive and is the distance above the reference mean that is required to be considered superior.

D is the mean difference at which the power is computed. D = Mean1 - Mean2 = treatment mean - reference mean.

Alpha is the probability of a false-positive result.

Beta is the probability of a false-negative result.

SD1 and SD2 are the standard deviations of groups 1 and 2, respectively.

**Summary Statements**

Group sample sizes of 10 and 10 achieve 13% power to detect superiority using a one-sided, two-sample t-test. The margin of superiority is 0.575. The true difference between the means is assumed to be 1.725. The significance level (alpha) of the test is 0.02500. The data are drawn from populations with standard deviations of 3.000 and 3.000.
The above report shows that for SM = 1.15, the sample size necessary to obtain 90% power is about 600 per group. However, if SM = 0.575, the required sample size is only about 180 per group.

Example 2 – Finding the Sample Size

Continuing with Example 1, the researchers want to know the exact sample size for each value of SM to achieve 90% power.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means with Non-Zero Null (Two-Sample T-Test) [Differences] procedure window by expanding Means, then Two Means, then clicking on Test (Non-Zero Null), and then clicking on Tests for Two Means with Non-Zero Null (Two-Sample T-Test) [Differences]. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.
Tests for Two Means with Non-Zero Null using Differences

Option | Value
--- | ---
Data Tab | 
Find (Solve For) | Power and Beta
Power | 0.90
Alpha | 0.025
N1 (Sample Size Group 1) | Ignored since this is the Find setting
N2 (Sample Size Group 2) | Use R
R (Sample Allocation Ratio) | 1.0
SM (Superiority Margin) | 0.575 1.15
D (True Difference) | 1.725
S1 (Standard Deviation Group 1) | 3
S2 (Standard Deviation Group 2) | S1
Higher Means Are | Better
Nonparametric Adjustment | Ignore

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N1/N2</th>
<th>Superiority Margin (SM)</th>
<th>Actual Difference (D)</th>
<th>Significance Level (Alpha)</th>
<th>Beta</th>
<th>Standard Deviation1 (SD1)</th>
<th>Standard Deviation2 (SD2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.90004</td>
<td>144/144</td>
<td>0.575</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.09996</td>
<td>3.000</td>
<td>3.000</td>
</tr>
<tr>
<td>0.90036</td>
<td>573/573</td>
<td>1.150</td>
<td>1.725</td>
<td>0.02500</td>
<td>0.09964</td>
<td>3.000</td>
<td>3.000</td>
</tr>
</tbody>
</table>

This report shows the exact sample size requirement for each value of SM.

Example 3 – Validation

This procedure uses the same mechanics as the Non-Inferiority Tests for Two Means using Differences procedure. We refer the user to Examples 3 and 4 of Chapter 450 for the validation.
Tests for Two Means with Non-Zero Null using Differences
Chapter 450

Non-Inferiority Tests for Two Means using Differences

Introduction

This procedure computes power and sample size for non-inferiority tests in two-sample designs in which the outcome is a continuous normal random variable. Measurements are made on individuals that have been randomly assigned to one of two groups. This is sometimes referred to as a parallel-groups design. This design is used in situations such as the comparison of the income level of two regions, the nitrogen content of two lakes, or the effectiveness of two drugs.

The two-sample t-test is commonly used with this situation. When the variances of the two groups are unequal, Welch’s t-test may be used. When the data are not normally distributed, the Mann-Whitney (Wilcoxon signed-ranks) U test may be used.

The details of sample size calculation for the two-sample design are presented in the Two-Sample T-Test chapter and they will not be duplicated here. This chapter only discusses those changes necessary for non-inferiority and superiority tests. Sample size formulas for non-inferiority and superiority tests of two means are presented in Chow et al. (2003) pages 57-59.

The Statistical Hypotheses

Both non-inferiority and superiority tests are examples of directional (one-sided) tests and their power and sample size could be calculated using the Two-Sample T-Test procedure. However, at the urging of our users, we have developed this module, which provides the input and output in formats that are convenient for these types of tests. This section will review the specifics of non-inferiority and superiority testing.
Remember that in the usual t-test setting, the null (H0) and alternative (H1) hypotheses for one-sided tests are defined as

\[ H_0: \mu_1 - \mu_2 \leq D \text{ versus } H_1: \mu_1 - \mu_2 > D \]

Rejecting this test implies that the mean difference is larger than the value \( D \). This test is called an \textit{upper-tailed test} because it is rejected in samples in which the difference between the sample means is larger than \( D \).

Following is an example of a \textit{lower-tailed test}.

\[ H_0: \mu_1 - \mu_2 \geq D \text{ versus } H_1: \mu_1 - \mu_2 < D \]

\textit{Non-inferiority} and \textit{superiority} tests are special cases of the above directional tests. It will be convenient to adopt the following specialized notation for the discussion of these tests.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PASS Input/Output</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu_1 )</td>
<td>Not used</td>
<td>\textit{Mean} of population 1. Population 1 is assumed to consist of those who have received the new treatment.</td>
</tr>
<tr>
<td>( \mu_2 )</td>
<td>Not used</td>
<td>\textit{Mean} of population 2. Population 2 is assumed to consist of those who have received the reference treatment.</td>
</tr>
<tr>
<td>( M_{NI} )</td>
<td>NIM</td>
<td>\textit{Margin of non-inferiority}. This is a tolerance value that defines the magnitude of the amount that is not of practical importance. This may be thought of as the largest change from the baseline that is considered to be trivial. The absolute value is shown to emphasize that this is a magnitude. The sign of the value will be determined by the specific design that is being used.</td>
</tr>
<tr>
<td>( \delta )</td>
<td>D</td>
<td>\textit{True difference}. This is the value of ( \mu_1 - \mu_2 ), the difference between the means. This is the value at which the power is calculated.</td>
</tr>
</tbody>
</table>

Note that the actual values of \( \mu_1 \) and \( \mu_2 \) are not needed. Only their difference is needed for power and sample size calculations.

\section*{Non-Inferiority Tests}

A \textit{non-inferiority test} tests that the treatment mean is not worse than the reference mean by more than the equivalence margin. The actual direction of the hypothesis depends on the response variable being studied.
Case 1: High Values Good, Non-Inferiority Test

In this case, higher values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is no less than a small amount below the reference mean. The value of $\delta$ is often set to zero. The following are equivalent sets of hypotheses.

$$
H_0: \mu_1 \leq \mu_2 - |M_{NI}| \quad \text{versus} \quad H_1: \mu_1 > \mu_2 - |M_{NI}|
$$

$$
H_0: \mu_1 - \mu_2 \leq -|M_{NI}| \quad \text{versus} \quad H_1: \mu_1 - \mu_2 > -|M_{NI}|
$$

$$
H_0: \delta \leq -|M_{NI}| \quad \text{versus} \quad H_1: \delta > -|M_{NI}|
$$

Case 2: High Values Bad, Non-Inferiority Test

In this case, lower values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is no more than a small amount above the reference mean. The value of $\delta$ is often set to zero. The following are equivalent sets of hypotheses.

$$
H_0: \mu_1 \geq \mu_2 + |M_{NI}| \quad \text{versus} \quad H_1: \mu_1 < \mu_2 + |M_{NI}|
$$

$$
H_0: \mu_1 - \mu_2 \geq |M_{NI}| \quad \text{versus} \quad H_1: \mu_1 - \mu_2 < |M_{NI}|
$$

$$
H_0: \delta \geq |M_{NI}| \quad \text{versus} \quad H_1: \delta < |M_{NI}|
$$

Example

A non-inferiority test example will set the stage for the discussion of the terminology that follows. Suppose that a test is to be conducted to determine if a new cancer treatment adversely affects mean bone density. The adjusted mean bone density (AMBD) in the population of interest is 0.002300 gm/cm with a standard deviation of 0.000300 gm/cm. Clinicians decide that if the treatment reduces AMBD by more than 5% (0.000115 gm/cm), it poses a significant health threat.

The hypothesis of interest is whether the mean AMBD in the treated group is more than 0.000115 below that of the reference group. The statistical test will be set up so that if the null hypothesis is rejected, the conclusion will be that the new treatment is non-inferior. The value 0.000115 gm/cm is called the margin of non-inferiority.

Test Statistics

This section describes the test statistics that are available in this procedure.

Two-Sample T-Test

Under the null hypothesis, this test assumes that the two groups of data are simple random samples from a single population of normally-distributed values that all have the same mean and variance. This assumption implies that the data are continuous and their distribution is symmetric. The calculation of the test statistic for the case when higher response values are good is as follows.
Non-Inferiority Tests for Two Means using Differences

\[ t_{df} = \frac{(\bar{X}_1 - \bar{X}_2) - |\epsilon|}{s_{\bar{X}_1-\bar{X}_2}} \]

where

\[ \bar{X}_k = \frac{\sum_{i=1}^{N_k} X_{ki}}{N_k} \]

\[ s_{\bar{X}_1-\bar{X}_2} = \sqrt{\frac{\sum_{i=1}^{N_1}(X_{1i} - \bar{X}_1)^2 + \sum_{i=1}^{N_2}(X_{2i} - \bar{X}_2)^2}{N_1 + N_2 - 2} \left( \frac{1}{N_1} + \frac{1}{N_2} \right) } \]

\[ df = N_1 + N_2 - 2 \]

The null hypothesis is rejected if the computed p-value is less than a specified level (usually 0.05). Otherwise, no conclusion can be reached.

**Welch’s T-Test**

Welch (1938) proposed the following test when the two variances are not assumed to be equal.

\[ t_f^* = \frac{(\bar{X}_1 - \bar{X}_2) - |\epsilon|}{s_{\bar{X}_1-\bar{X}_2}} \]

where

\[ s_{\bar{X}_1-\bar{X}_2}^* = \sqrt{\frac{\sum_{i=1}^{N_1}(X_{1i} - \bar{X}_1)^2}{N_1(N_1 - 1)} + \frac{\sum_{i=1}^{N_2}(X_{2i} - \bar{X}_2)^2}{N_2(N_2 - 1)}} \]

\[ f = \frac{\left( \frac{s_1^2}{N_1} + \frac{s_2^2}{N_2} \right)^2}{\frac{N_1^4}{(N_1^2 - 1)} + \frac{N_2^4}{(N_2^2 - 1)}} \]

\[ s_1 = \sqrt{\frac{\sum_{i=1}^{N_1}(X_{1i} - \bar{X}_1)^2}{N_1 - 1}}, \quad s_2 = \sqrt{\frac{\sum_{i=1}^{N_2}(X_{2i} - \bar{X}_2)^2}{N_2 - 1}} \]
**Mann-Whitney U Test**

This test is the nonparametric substitute for the equal-variance t-test. Two key assumptions are that the distributions are at least ordinal and that they are identical under H0. This means that ties (repeated values) are not acceptable. When ties are present, you can use approximations, but the theoretic results no longer hold.

The Mann-Whitney test statistic is defined as follows in Gibbons (1985).

\[ z = \frac{W_1 - \frac{N_1(N_1 + N_2 + 1)}{2} + C}{s_w} \]

where

\[ W_1 = \sum_{k=1}^{N_1} \text{Rank}(X_{1k}) \]

The ranks are determined after combining the two samples. The standard deviation is calculated as

\[ s_w = \sqrt{\frac{N_1N_2(N_1 + N_2 + 1)}{12} - \frac{N_1N_2 \sum_{i=1}^{N_1} (t_i^3 - t_i)}{12(N_1 + N_2)(N_1 + N_2 - 1)}} \]

where \( t_i \) is the number of observations tied at value one, \( t_2 \) is the number of observations tied at some value two, and so forth.

The correction factor, \( C \), is 0.5 if the rest of the numerator is negative or -0.5 otherwise. The value of \( z \) is then compared to the normal distribution.

### Computing the Power

#### Standard Deviations Equal

When \( \sigma_1 = \sigma_2 = \sigma \), the power of the \( t \) test is calculated as follows.

1. Find \( t_{\alpha} \) such that \( 1 - T_{df}(t_{\alpha}) = \alpha \), where \( T_{df}(t_{\alpha}) \) is the area under a central-\( t \) curve to the left of \( x \) and \( df = N_1 + N_2 - 2 \).

2. Calculate: \( \sigma_\eta = \sigma \sqrt{\frac{1}{N_1} + \frac{1}{N_2}} \)

3. Calculate the noncentrality parameter: \( \lambda = \frac{|\epsilon| - \delta}{\sigma_\eta} \)

4. Calculate: Power = \( 1 - T'_{df,\lambda}(t_{\alpha}) \), where \( T'_{df,\lambda}(x) \) is the area to the left of \( x \) under a noncentral-\( t \) curve with degrees of freedom \( df \) and noncentrality parameter \( \lambda \).
Standard Deviations Unequal

This case often recommends Welch’s test. When \( \sigma_1 \neq \sigma_2 \), the power is calculated as follows.

1. Calculate: \( \sigma_{\tau} = \sqrt[1]{\frac{\sigma_1^2 + \sigma_2^2}{N_1} \frac{1}{N_2}} \).

2. Calculate: \( f = \frac{\sigma_1^4}{\sigma_1^4} - \frac{\sigma_2^4}{\sigma_2^4} - 2 \frac{N_1^2(N_1+1)}{N_1^2(N_1+1)} \) \( + \frac{N_2^2(N_2+1)}{N_2^2(N_2+1)} \) which is the adjusted degrees of freedom. Often, this is rounded to the next highest integer. Note that this is not the value of \( f \) used in the computation of the actual test. Instead, this is the expected value of \( f \).

3. Find \( t_\alpha \) such that \( 1 - T_f(t_\alpha) = \alpha \), where \( T_f(t_\alpha) \) is the area to the left of \( x \) under a central-\( t \) curve with \( f \) degrees of freedom.

4. Calculate: \( \lambda = \frac{\mid \beta \mid}{\sigma_\tau} \), the noncentrality parameter.

5. Calculate: Power = \( 1 - T_{f,\lambda}(t_\alpha) \), where \( T_{f,\lambda}(x) \) is the area to the left of \( x \) under a noncentral-\( t \) curve with degrees of freedom \( f \) and noncentrality parameter \( \lambda \).

Nonparametric Adjustment

When using the Mann-Whitney test rather than the \( t \) test, results by Al-Sunduqchi and Guenther (1990) indicate that power calculations for the Mann-Whitney test may be made using the standard \( t \) test formulations with a simple adjustment to the sample sizes. The size of the adjustment depends on the actual distribution of the data. They give sample size adjustment factors for four distributions. These are 1 for uniform, 2/3 for double exponential, \( 9 / \pi^2 \) for logistic, and \( \pi / 3 \) for normal distributions.
Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains most of the parameters and options that you will be concerned with.

Solve For

Find (Solve For)
This option specifies the parameter to be calculated from the values of the other parameters. Under most conditions, you would select either Power and Beta or N1.

Select N1 when you want to determine the sample size needed to achieve a given power and alpha.

Select Power and Beta when you want to calculate the power of an experiment that has already been run.

Error Rates

Power or Beta
This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of inferiority when the null hypothesis should be rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

Alpha (Significance Level)
This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when you reject the null hypothesis of inferiority when in fact the mean is not non-inferior.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.
Sample Size

N1 (Sample Size Group 1)
Enter a value (or range of values) for the sample size of group 1. Note that these values are ignored when you are solving for N1. You may enter a range of values such as 10 to 100 by 10.

N2 (Sample Size Group 2)
Enter a value (or range of values) for the sample size of group 2 or enter Use R to base N2 on the value of N1. You may enter a range of values such as 10 to 100 by 10.

• Use R
  When Use R is entered here, N2 is calculated using the formula
  \[ N2 = \lceil R(N1) \rceil \]
  where R is the Sample Allocation Ratio and the operator \( \lceil Y \rceil \) is the first integer greater than or equal to Y. For example, if you want N1 = N2, select Use R and set R = 1.

R (Sample Allocation Ratio)
Enter a value (or range of values) for R, the allocation ratio between samples. This value is only used when N2 is set to Use R.

When used, N2 is calculated from N1 using the formula: \( N2 = \lceil R(N1) \rceil \) where \( \lceil Y \rceil \) is the next integer greater than or equal to Y. Note that setting R = 1.0 forces N2 = N1.

Effect Size – Mean Difference

NIM (Non-Inferiority Margin)
This is the magnitude of the margin of non-inferiority. It must be entered as a positive number.

When higher means are better, this value is the distance below the reference mean that is still considered non-inferior. When higher means are worse, this value is the distance above the reference mean that is still considered non-inferior.

D (True Difference)
This is the actual difference between the treatment mean and the reference mean at which power is calculated.

For non-inferiority tests, this value is often set to zero. When this value is non-zero, care should be taken that this value is consistent with whether higher means are better or worse.

Effect Size – Standard Deviations

S1 and S2 (Standard Deviations)
These options specify the values of the standard deviations for each group. When the S2 is set to S1, only S1 needs to be specified. The value of S1 will be copied into S2.

When these values are not known, you must supply estimates of them. Press the SD button to display the Standard Deviation Estimator window. This procedure will help you find appropriate values for the standard deviation.
**Non-Inferiority Tests for Two Means using Differences**

**Test**

**Higher Means Are**
This option defines whether higher values of the response variable are to be considered better or worse. The choice here determines the direction of the non-inferiority test.

If Higher Means Are Better the null hypothesis is $\text{Diff} <= -\text{NIM}$ and the alternative hypothesis is $\text{Diff} > -\text{NIM}$. If Higher Means Are Worse the null hypothesis is $\text{Diff} >= \text{NIM}$ and the alternative hypothesis is $\text{Diff} < \text{NIM}$.

**Nonparametric Adjustment (Mann-Whitney Test)**
This option makes appropriate sample size adjustments for the Mann-Whitney test. Results by Al-Sunduqchi and Guenther (1990) indicate that power calculations for the Mann-Whitney test may be made using the standard $t$ test formulations with a simple adjustment to the sample size. The size of the adjustment depends upon the actual distribution of the data. They give sample size adjustment factors for four distributions. These are $1$ for the uniform distribution, $2/3$ for the double exponential distribution, $9/\pi^2$ for the logistic distribution, and $\pi/3$ for the normal distribution.

The options are as follows:

- **Ignore**
  Do not make a Mann-Whitney adjustment. This indicates that you want to analyze a $t$ test, not the Wilcoxon test.

- **Uniform**
  Make the Mann-Whitney sample size adjustment assuming the uniform distribution. Since the factor is one, this option performs the same function as Ignore. It is included for completeness.

- **Double Exponential**
  Make the Mann-Whitney sample size adjustment assuming that the data actually follow the double exponential distribution.

- **Logistic**
  Make the Mann-Whitney sample size adjustment assuming that the data actually follow the logistic distribution.

- **Normal**
  Make the Mann-Whitney sample size adjustment assuming that the data actually follow the normal distribution.
Example 1 – Power Analysis

Suppose that a test is to be conducted to determine if a new cancer treatment adversely affects bone density. The adjusted mean bone density (AMBD) in the population of interest is 0.002300 gm/cm with a standard deviation of 0.000300 gm/cm. Clinicians decide that if the treatment reduces AMBD by more than 5% (0.000115 gm/cm), it poses a significant health threat. They also want to consider what would happen if the margin of equivalence is set to 2.5% (0.0000575 gm/cm).

Following accepted procedure, the analysis will be a non-inferiority test using the t-test at the 0.025 significance level. Power to be calculated assuming that the new treatment has no effect on AMBD. Several sample sizes between 10 and 800 will be analyzed. The researchers want to achieve a power of at least 90%. All numbers have been multiplied by 10000 to make the reports and plots easier to read.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Non-Inferiority for Two Means [Differences] procedure window by expanding Means, then Two Means, then clicking on Non-Inferiority, and then clicking on Non-Inferiority for Two Means [Differences]. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.025</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>10 50 100 200 300 500 600 800</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>NIM (Non-Inferiority Margin)</td>
<td>0.575 1.15</td>
</tr>
<tr>
<td>D (True Difference)</td>
<td>0</td>
</tr>
<tr>
<td>S1 (Standard Deviation Group 1)</td>
<td>3</td>
</tr>
<tr>
<td>S2 (Standard Deviation Group 2)</td>
<td>S1</td>
</tr>
<tr>
<td>Higher Means Are</td>
<td>Better</td>
</tr>
<tr>
<td>Nonparametric Adjustment</td>
<td>Ignore</td>
</tr>
</tbody>
</table>
Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results and Plots

<table>
<thead>
<tr>
<th>Power</th>
<th>N1/N2</th>
<th>Non-Inf Margin (-NIM)</th>
<th>Actual Difference (D)</th>
<th>Significance Level (Alpha)</th>
<th>Standard Deviation 1 (SD1)</th>
<th>Standard Deviation 2 (SD2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.06013</td>
<td>10/10</td>
<td>-0.575</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.93987</td>
<td>3.000</td>
</tr>
<tr>
<td>0.15601</td>
<td>50/50</td>
<td>-0.575</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.84399</td>
<td>3.000</td>
</tr>
<tr>
<td>0.27052</td>
<td>100/100</td>
<td>-0.575</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.72948</td>
<td>3.000</td>
</tr>
<tr>
<td>0.48089</td>
<td>200/200</td>
<td>-0.575</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.51911</td>
<td>3.000</td>
</tr>
<tr>
<td>0.64940</td>
<td>300/300</td>
<td>-0.575</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.35060</td>
<td>3.000</td>
</tr>
<tr>
<td>0.85769</td>
<td>500/500</td>
<td>-0.575</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.14231</td>
<td>3.000</td>
</tr>
<tr>
<td>0.91295</td>
<td>600/600</td>
<td>-0.575</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.08705</td>
<td>3.000</td>
</tr>
<tr>
<td>0.96943</td>
<td>800/800</td>
<td>-0.575</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.03057</td>
<td>3.000</td>
</tr>
<tr>
<td>0.12553</td>
<td>10/10</td>
<td>-1.150</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.87447</td>
<td>3.000</td>
</tr>
<tr>
<td>0.47524</td>
<td>50/50</td>
<td>-1.150</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.52476</td>
<td>3.000</td>
</tr>
<tr>
<td>0.76957</td>
<td>100/100</td>
<td>-1.150</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.23043</td>
<td>3.000</td>
</tr>
<tr>
<td>0.96885</td>
<td>200/200</td>
<td>-1.150</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.03115</td>
<td>3.000</td>
</tr>
<tr>
<td>0.99681</td>
<td>300/300</td>
<td>-1.150</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.00319</td>
<td>3.000</td>
</tr>
<tr>
<td>0.99998</td>
<td>500/500</td>
<td>-1.150</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.00002</td>
<td>3.000</td>
</tr>
<tr>
<td>1.00000</td>
<td>600/600</td>
<td>-1.150</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.00000</td>
<td>3.000</td>
</tr>
<tr>
<td>1.00000</td>
<td>800/800</td>
<td>-1.150</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.00000</td>
<td>3.000</td>
</tr>
</tbody>
</table>

Report Definitions

- **Group 1** is the treatment group. **Group 2** is the reference or standard group.
- **Power** is the probability of rejecting a false null hypothesis.
- **N1** is the number of subjects in the first (treatment) group.
- **N2** is the number of subjects in the second (reference) group.
- **-NIM** is the magnitude and direction of the margin of non-inferiority. Since higher means are better, this value is negative and is the distance below the reference mean that is still considered non-inferior.
- **D** is the mean difference at which the power is computed. \( D = \text{Mean1} - \text{Mean2} \), or Treatment Mean - Reference Mean.
- **Alpha** is the probability of a false-positive result.
- **Beta** is the probability of a false-negative result.
- **SD1** and **SD2** are the standard deviations of groups 1 and 2, respectively.

Summary Statements

- Group sample sizes of 10 and 10 achieve 6% power to detect non-inferiority using a one-sided, two-sample t-test. The margin of equivalence is 0.575. The true difference between the means is assumed to be 0.000. The significance level (alpha) of the test is 0.02500. The data are drawn from populations with standard deviations of 3.000 and 3.000.
The above report shows that for NIM = 1.15, the sample size necessary to obtain 90% power is about 150 per group. However, if NIM = 0.575, the required sample size is about 600 per group.

Example 2 – Finding the Sample Size

Continuing with Example 1, the researchers want to know the exact sample size for each value of NIM to achieve 90% power.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Non-Inferiority for Two Means [Differences] procedure window by expanding Means, then Two Means, then clicking on Non-Inferiority, and then clicking on Non-Inferiority for Two Means [Differences]. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.
Non-Inferiority Tests for Two Means using Differences

Option | Value
--- | ---
Data Tab | 
Find (Solve For) | N1
Power | 0.90
Alpha | 0.025
N1 (Sample Size Group 1) | Ignored since this is the Find setting
N2 (Sample Size Group 2) | Use R
R (Sample Allocation Ratio) | 1.0
NIM (Non-Inferiority Margin) | 0.575 1.15
D (True Difference) | 0
S1 (Standard Deviation Group 1) | 3
S2 (Standard Deviation Group 2) | S1
Higher Means Are | Better
Nonparametric Adjustment | Ignore

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Non-Inferiority Test (H0: Diff &lt;= -NIM; H1: Diff &gt; -NIM)</th>
<th>Higher Means are Better</th>
<th>Test Statistic: T-Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power</td>
<td>N1/N2</td>
<td>Non-Inf. Margin (NIM)</td>
</tr>
<tr>
<td>0.90036</td>
<td>573/573</td>
<td>-0.575</td>
</tr>
<tr>
<td>0.90004</td>
<td>144/144</td>
<td>-1.150</td>
</tr>
</tbody>
</table>

This report shows the exact sample size requirement for each value of NIM.

Example 3 – Validation using Chow

Chow, Shao, Wang (2003) page 62 has an example of a sample size calculation for a non-inferiority trial. Their example obtains a sample size of 51 in each group when D = 0, NIM = 0.05, S = 0.1, Alpha = 0.05, and Beta = 0.20.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Non-Inferiority for Two Means [Differences] procedure window by expanding Means, then Two Means, then clicking on Non-Inferiority, and then clicking on Non-Inferiority for Two Means [Differences]. You may then make the appropriate entries as listed below, or open Example 3 by going to the File menu and choosing Open Example Template.
Option | Value
---|---
Data Tab | 
Find (Solve For) | N1
Power | 0.80
Alpha | 0.05
N1 (Sample Size Group 1) | Ignored since this is the Find setting
N2 (Sample Size Group 2) | Use R
R (Sample Allocation Ratio) | 1.0
NIM (Non-Inferiority Margin) | 0.05
D (True Difference) | 0
S1 (Standard Deviation Group 1) | 0.1
S2 (Standard Deviation Group 2) | S1
Higher Means Are | Better
Nonparametric Adjustment | Ignore

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N1/N2</th>
<th>Non-Infer. Margin (-NIM)</th>
<th>Actual Difference (D)</th>
<th>Significance Level (Alpha)</th>
<th>Beta</th>
<th>Standard Deviation1 (SD1)</th>
<th>Standard Deviation2 (SD2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.80590</td>
<td>51/51</td>
<td>-0.050</td>
<td>0.000</td>
<td>0.05000</td>
<td>0.19410</td>
<td>0.100</td>
<td>0.100</td>
</tr>
</tbody>
</table>

PASS has also obtained a sample size of 51 per group.

Example 4 – Validation using Julious

Julious (2004) page 1950 gives an example of a sample size calculation for a parallel, non-inferiority design. His example obtains a sample size of 336 when D = 0, NIM = 10, S = 40, Alpha = 0.025, and Beta = 0.10.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Non-Inferiority for Two Means [Differences] procedure window by expanding Means, then Two Means, then clicking on Non-Inferiority, and then clicking on Non-Inferiority for Two Means [Differences]. You may then make the appropriate entries as listed below, or open Example 4 by going to the File menu and choosing Open Example Template.
### Option Value

#### Data Tab
- **Find (Solve For)**: \( N_1 \)
- **Power**: 0.90
- **Alpha**: 0.025
- **\( N_1 \) (Sample Size Group 1)**: Ignored since this is the Find setting
- **\( N_2 \) (Sample Size Group 2)**: Use \( R \)
- **\( R \) (Sample Allocation Ratio)**: 1.0
- **\( N_{IM} \) (Non-Inferiority Margin)**: 10
- **\( D \) (True Difference)**: 0
- **\( S_1 \) (Standard Deviation Group 1)**: 40
- **\( S_2 \) (Standard Deviation Group 2)**: \( S_1 \)
- **Higher Means Are**: Better
- **Nonparametric Adjustment**: Ignore

### Output

Click the Run button to perform the calculations and generate the following output.

### Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N1/N2</th>
<th>Non-Infer. Margin (-NIM)</th>
<th>Actual Difference (D)</th>
<th>Significance Level (Alpha)</th>
<th>Beta</th>
<th>Standard Deviation1 (SD1)</th>
<th>Standard Deviation2 (SD2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9045</td>
<td>337/337</td>
<td>-10.000</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.09955</td>
<td>40.000</td>
<td>40.000</td>
</tr>
</tbody>
</table>

\textit{PASS} obtained sample sizes of 337 in each group. The difference between 336 that Julious received and 337 that \textit{PASS} calculated is likely caused by rounding.
Chapter 453

Tests for Two Means with Non-Unity Null using Ratios

Introduction

This procedure calculates power and sample size for non-unity null t-tests from a parallel-groups design in which the logarithm of the outcome is a continuous normal random variable. This routine deals with the case in which the statistical hypotheses are expressed in terms of mean ratios instead of mean differences.

The details of testing the non-unity null hypothesis of two treatments using data from a two-group design are given in another chapter and they will not be repeated here. If the logarithm of the response can be assumed to follow a normal distribution, non-unity null hypotheses stated in terms of the ratio can be transformed into hypotheses about the difference. The details of this analysis are given in Julious (2004). They will only be summarized here.

Non-Inferiority Testing Using Ratios

It will be convenient to adopt the following specialized notation for the discussion of these tests.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PASS Input/Output</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_T$</td>
<td>Not used</td>
<td>Treatment mean. This is the treatment mean.</td>
</tr>
<tr>
<td>$\mu_R$</td>
<td>Not used</td>
<td>Reference mean. This is the mean of a reference population.</td>
</tr>
<tr>
<td>$M_s$</td>
<td>SM</td>
<td>Margin of superiority. This is a tolerance value that defines the magnitude of difference that is required for practical importance. This may be thought of as the smallest difference from the reference that is considered to be different.</td>
</tr>
</tbody>
</table>
Parameter | PASS Input/Output | Interpretation
---|---|---
φ | R1 | *True ratio.* This is the value of $\phi = \mu_T / \mu_R$ at which the power is calculated.

Note that the actual values of $\mu_T$ and $\mu_R$ are not needed. Only the ratio of these values is needed for power and sample size calculations.

The null hypothesis of non-superiority is

$$H_0: \phi \leq \phi_L \text{ where } \phi_L > 1.$$  

and the alternative hypothesis of superiority is

$$H_1: \phi > \phi_L.$$  

**Log-Transformation**

In many cases, hypotheses stated in terms of ratios are more convenient than hypotheses stated in terms of differences. This is because ratios can be interpreted as scale-less percentages, but differences must be interpreted as actual amounts in their original scale. Hence, it has become a common practice to take the following steps in hypothesis testing.

1. State the statistical hypotheses in terms of ratios.
2. Transform these into hypotheses about differences by taking logarithms.
3. Analyze the logged data—that is, do the analysis in terms of the difference.
4. Draw the conclusion in terms of the ratio.

The details of step 2 for the null hypothesis are as follows.

$$\phi_L \leq \phi$$

$$\Rightarrow \phi_L \leq \left\{ \frac{\mu_T}{\mu_R} \right\}$$

$$\Rightarrow \ln(\phi_L) \leq \left\{ \ln(\mu_T) - \ln(\mu_R) \right\}$$

Thus, a hypothesis about the ratio of the means on the original scale can be translated into a hypothesis about the difference of two means on the logged scale.

**Coefficient of Variation**

The coefficient of variation (COV) is the ratio of the standard deviation to the mean. This parameter can be used to represent the variation in the data because of a unique relationship that it has in the case of log-normal data.

Suppose the variable $X$ is the logarithm of the original variable $Y$. That is, $X = \ln(Y)$ and $Y = \exp(X)$. Label the mean and variance of $X$ as $\mu_X$ and $\sigma_X^2$, respectively. Similarly, label the mean and variance of $Y$ as $\mu_Y$ and $\sigma_Y^2$, respectively. If $X$ is normally distributed, then $Y$ is log-normally distributed. Julious (2004) presents the following well-known relationships between these two variables.
\[
\mu_Y = \left( e^{\mu_Y + \frac{\sigma_Y^2}{2}} \right)
\]
\[
\sigma_Y^2 = \mu_Y \left( e^{\sigma_Y^2} - 1 \right)
\]

From this relationship, the coefficient of variation of \( Y \) can be found to be
\[
COV_Y = \frac{\mu_Y \left( e^{\sigma_Y^2} - 1 \right)}{\mu_Y} = \sqrt{e^{\sigma_Y^2} - 1}
\]

Solving this relationship for \( \sigma_X^2 \), the standard deviation of \( X \) can be stated in terms of the coefficient of variation of \( Y \). This equation is
\[
\sigma_X = \sqrt{\ln(COV_Y^2 + 1)}
\]

Similarly, the mean of \( X \) is
\[
\mu_X = \frac{\mu_Y}{\ln(COV_Y^2 + 1)}
\]

One final note: for parallel-group designs, \( \sigma_X^2 \) equals \( \sigma_d^2 \), the average variance used in the t-test of the logged data.

Thus, the hypotheses can be stated in the original (Y) scale and then the power can be analyzed in the transformed (X) scale.

**Power Calculation**

As is shown above, the hypotheses can be stated in the original (Y) scale using ratios or the logged (X) scale using differences. In either case, the power and sample size calculations are made using the formulas for testing the difference in two means. These formulas are presented in another chapter and are not duplicated here.

**Procedure Options**

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

**Data Tab**

The Data tab contains the parameters associated with this test such as the means, sample sizes, alpha, and power.
**Solve For**

**Find (Solve For)**
This option specifies the parameter to be solved for from the other parameters. In most situations, you will select either *Power and Beta* for a power analysis or *N1* for sample size determination.

---

**Error Rates**

**Power or Beta**
This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of non-superiority when in fact the treatment mean is superior.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

**Alpha (Significance Level)**
This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when rejecting the null hypothesis of non-superiority when in fact the treatment group is not superior to the reference group.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.

---

**Sample Size**

**N1 (Sample Size Group 2)**
Enter a value (or range of values) for the sample size of group 1 (the reference group). Note that these values are ignored when you are solving for N1. You may enter a range of values such as 10 to 100 by 10.

**N2 (Sample Size Group 2)**
Enter a value (or range of values) for the sample size of group 2 (the treatment group) or enter *Use R* to base N2 on the value of N1. You may enter a range of values such as 10 to 100 by 10.

- **Use R**
  When *Use R* is entered here, N2 is calculated using the formula

  \[ N2 = \lceil R(N1) \rceil \]

  where R is the Sample Allocation Ratio and the operator \( \lceil Y \rceil \) is the first integer greater than or equal to Y. For example, if you want N1 = N2, select *Use R* and set R = 1.
R (Sample Allocation Ratio)
Enter a value (or range of values) for $R$, the allocation ratio between samples. This value is only used when $N_2$ is set to Use $R$.

When used, $N_2$ is calculated from $N_1$ using the formula: $N_2 = \lceil R(N_1) \rceil$ where $\lceil \cdot \rceil$ is the next integer greater than or equal to $Y$. Note that setting $R = 1.0$ forces $N_2 = N_1$.

Effect Size – Ratios

SM (Superiority Margin)
This is the magnitude of the margin of superiority. It must be entered as a positive number.

When higher means are better, this value is the distance above one that is required for the mean ratio (Treatment Mean / Reference Mean) to be considered superior. When higher means are worse, this value is the distance below one that is required for the mean ratio (Treatment Mean / Reference Mean) to be considered superior.

Note that the sign of this value is ignored. Only the magnitude is used.

Recommended values: 0.20 is a common value for this parameter.

R1 (True Ratio)
This is the value of the ratio of the two means (Treatment Mean / Reference Mean) at which the power is to be calculated.

When higher means are better, this value should be greater than $1+SM$. When higher means are worse, this value should be less than $1-SM$.

Effect Size – Coefficient of Variation

COV (Coefficient of Variation)
The coefficient of variation is the ratio of the standard deviation and the mean (SD/Mean). It is used to specify the variability (standard deviation). Note that this COV is defined on the original (not logarithmic) scale. This value must be determined from past experience or from a pilot study.

To be clear, consider the following definition. Suppose data on a response variable $Y$ are collected. This procedure assumes that the values of $X = \ln(Y)$ are analyzed using a two-sample t-test. Thus, there are two sets of means and standard deviations: those of $X$ labelled $M_X$ and $S_X$ and those of $Y$ labelled $M_Y$ and $S_Y$. The COV entered here is the COV of $Y = S_Y/M_Y$. For log-normal data, the following relationship exists: $\text{COV}(Y) = \text{SQR}((\exp(S_X*S_X)-1))$ where $S_X$ is the standard deviation of the log-transformed values.

Test

Higher Means Are
This option defines whether higher values of the response variable are to be considered better or worse. The choice here determines the direction of the test.

If Higher Means Are Better the null hypothesis is $R \leq 1+SM$ and the alternative hypothesis is $R > 1+SM$. If Higher Means Are Worse the null hypothesis is $R \geq 1-SM$ and the alternative hypothesis is $R < 1-SM$. 
Example 1 – Finding Power

A company has developed a drug for treating rheumatism and wants to show that it is superior to the standard drug by a certain amount. Responses following either treatment are known to follow a log normal distribution. A parallel-group design will be used and the logged data will be analyzed with a two-sample t-test.

Researchers have decided to set the margin of superiority at 0.20. Past experience leads the researchers to set the COV to 1.50. The significance level is 0.025. The power will be computed assuming that the true ratio is either 1.30 or 1.40. Sample sizes between 100 and 1000 will be included in the analysis.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means with Non-Unity Null (Two-Sample T-Test) [Ratios] procedure window by expanding Means, then Two Means, then clicking on Test (Non-Zero Null), and then clicking on Tests for Two Means with Non-Unity Null (Two-Sample T-Test) [Ratios]. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.025</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>100 to 1000 by 100</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>SM (Superiority Margin)</td>
<td>0.20</td>
</tr>
<tr>
<td>R1 (True Ratio)</td>
<td>1.30 1.40</td>
</tr>
<tr>
<td>COV (Coefficient of Variation)</td>
<td>1.50</td>
</tr>
<tr>
<td>Higher Means Are</td>
<td>Better</td>
</tr>
</tbody>
</table>
Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N1/N2</th>
<th>Superiority Margin (SM)</th>
<th>Superiority Bound (SB)</th>
<th>True Ratio (R1)</th>
<th>Significance Level (Alpha)</th>
<th>Coefficient of Variation (COV)</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.07477</td>
<td>100/100</td>
<td>0.20</td>
<td>1.20</td>
<td>1.30</td>
<td>0.02500</td>
<td>1.50</td>
<td>0.92523</td>
</tr>
<tr>
<td>0.11039</td>
<td>200/200</td>
<td>0.20</td>
<td>1.20</td>
<td>1.30</td>
<td>0.02500</td>
<td>1.50</td>
<td>0.88961</td>
</tr>
<tr>
<td>0.14493</td>
<td>300/300</td>
<td>0.20</td>
<td>1.20</td>
<td>1.30</td>
<td>0.02500</td>
<td>1.50</td>
<td>0.85507</td>
</tr>
<tr>
<td>0.17994</td>
<td>400/400</td>
<td>0.20</td>
<td>1.20</td>
<td>1.30</td>
<td>0.02500</td>
<td>1.50</td>
<td>0.82006</td>
</tr>
<tr>
<td>0.21389</td>
<td>500/500</td>
<td>0.20</td>
<td>1.20</td>
<td>1.30</td>
<td>0.02500</td>
<td>1.50</td>
<td>0.78611</td>
</tr>
<tr>
<td>0.24762</td>
<td>600/600</td>
<td>0.20</td>
<td>1.20</td>
<td>1.30</td>
<td>0.02500</td>
<td>1.50</td>
<td>0.75238</td>
</tr>
<tr>
<td>0.28099</td>
<td>700/700</td>
<td>0.20</td>
<td>1.20</td>
<td>1.30</td>
<td>0.02500</td>
<td>1.50</td>
<td>0.71901</td>
</tr>
<tr>
<td>0.31390</td>
<td>800/800</td>
<td>0.20</td>
<td>1.20</td>
<td>1.30</td>
<td>0.02500</td>
<td>1.50</td>
<td>0.68610</td>
</tr>
<tr>
<td>0.34624</td>
<td>900/900</td>
<td>0.20</td>
<td>1.20</td>
<td>1.30</td>
<td>0.02500</td>
<td>1.50</td>
<td>0.65376</td>
</tr>
<tr>
<td>0.37791</td>
<td>1000/1000</td>
<td>0.20</td>
<td>1.20</td>
<td>1.30</td>
<td>0.02500</td>
<td>1.50</td>
<td>0.62209</td>
</tr>
<tr>
<td>0.16832</td>
<td>100/100</td>
<td>0.20</td>
<td>1.20</td>
<td>1.40</td>
<td>0.02500</td>
<td>1.50</td>
<td>0.83168</td>
</tr>
<tr>
<td>0.29339</td>
<td>200/200</td>
<td>0.20</td>
<td>1.20</td>
<td>1.40</td>
<td>0.02500</td>
<td>1.50</td>
<td>0.70661</td>
</tr>
<tr>
<td>0.41147</td>
<td>300/300</td>
<td>0.20</td>
<td>1.20</td>
<td>1.40</td>
<td>0.02500</td>
<td>1.50</td>
<td>0.58853</td>
</tr>
<tr>
<td>0.51938</td>
<td>400/400</td>
<td>0.20</td>
<td>1.20</td>
<td>1.40</td>
<td>0.02500</td>
<td>1.50</td>
<td>0.48062</td>
</tr>
<tr>
<td>0.61229</td>
<td>500/500</td>
<td>0.20</td>
<td>1.20</td>
<td>1.40</td>
<td>0.02500</td>
<td>1.50</td>
<td>0.38771</td>
</tr>
<tr>
<td>0.69123</td>
<td>600/600</td>
<td>0.20</td>
<td>1.20</td>
<td>1.40</td>
<td>0.02500</td>
<td>1.50</td>
<td>0.30877</td>
</tr>
<tr>
<td>0.75687</td>
<td>700/700</td>
<td>0.20</td>
<td>1.20</td>
<td>1.40</td>
<td>0.02500</td>
<td>1.50</td>
<td>0.24313</td>
</tr>
<tr>
<td>0.81046</td>
<td>800/800</td>
<td>0.20</td>
<td>1.20</td>
<td>1.40</td>
<td>0.02500</td>
<td>1.50</td>
<td>0.18954</td>
</tr>
<tr>
<td>0.85355</td>
<td>900/900</td>
<td>0.20</td>
<td>1.20</td>
<td>1.40</td>
<td>0.02500</td>
<td>1.50</td>
<td>0.14645</td>
</tr>
<tr>
<td>0.88778</td>
<td>1000/1000</td>
<td>0.20</td>
<td>1.20</td>
<td>1.40</td>
<td>0.02500</td>
<td>1.50</td>
<td>0.11225</td>
</tr>
</tbody>
</table>

Report Definitions

H0 (null hypothesis) is that \( R \leq 1+SM \), where \( R = \) Treatment Mean / Reference Mean. H1 (alternative hypothesis) is that \( R > 1+SM \). Power is the probability of rejecting H0 when it is false. N1 is the number of subjects in the first (treatment) group. N2 is the number of subjects in the second (reference) group. SM is the magnitude of the margin of superiority. Since higher means are better, this value is positive and is the distance above one that is required to be considered superior. SB is the corresponding bound to the superiority margin, and equals 1 + SM. R1 is the mean ratio (treatment/reference) at which the power is evaluated. Alpha is the probability of falsely rejecting H0. COV is the coefficient of variation on the original scale. Beta is the probability of not rejecting H0 when it is false.

Summary Statements

Group sample sizes of 100 in the first group and 100 in the second group achieve 7% power to detect superiority using a one-sided, two-sample t-test. The margin of superiority is 0.20. The true ratio of the means at which the power is evaluated is 1.30. The significance level (alpha) of the test is 0.02500. The coefficients of variation of both groups are assumed to be 1.50.

This report shows the power for the indicated scenarios.
Plots Section

This plot shows the power versus the sample size.

Example 2 – Validation

We could not find a validation example for this procedure in the statistical literature. Therefore, we will show that this procedure gives the same results as the non-zero null test on differences—a procedure that has been validated. We will use the same settings as those given in Example 1. Since the output for this example is shown above, all that we need is the output from the procedure that uses differences.

To run the non-zero null test on differences, we need the values of SM and S1.

\[
S1 = \sqrt{\ln(COV^2 + 1)}
\]

\[
= \sqrt{\ln(1.5^2 + 1)}
\]

\[
= 1.085659
\]

\[
SM* = \ln(1 + SM)
\]

\[
= \ln(1.2)
\]

\[
= 0.182322
\]

\[
D = \ln(R1)
\]

\[
= \ln(1.3)
\]

\[
= 0.262364
\]
Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means with Non-Zero Null (Two-Sample T-Test) [Differences] procedure window by expanding Means, then Two Means, then clicking on Test (Non-Zero Null), and then clicking on Tests for Two Means with Non-Zero Null (Two-Sample T-Test) [Differences]. You may then make the appropriate entries as listed below, or open Example 1b by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.025</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>100 to 1000 by 100</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>SM (Superiority Margin)</td>
<td>0.182322</td>
</tr>
<tr>
<td>D (True Difference)</td>
<td>0.262364</td>
</tr>
<tr>
<td>S1 (Standard Deviation Group 1)</td>
<td>1.085659</td>
</tr>
<tr>
<td>S2 (Standard Deviation Group 2)</td>
<td>S1</td>
</tr>
<tr>
<td>Higher Means Are</td>
<td>Better</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N1/N2</th>
<th>Superiority Margin (SM)</th>
<th>Actual Difference (D)</th>
<th>Significance Level (Alpha)</th>
<th>Beta</th>
<th>Standard Deviation1 (SD1)</th>
<th>Standard Deviation2 (SD2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.07477</td>
<td>100/100</td>
<td>0.182</td>
<td>0.262</td>
<td>0.02500</td>
<td>0.92523</td>
<td>1.086</td>
<td>1.086</td>
</tr>
<tr>
<td>0.11039</td>
<td>200/200</td>
<td>0.182</td>
<td>0.262</td>
<td>0.02500</td>
<td>0.88961</td>
<td>1.086</td>
<td>1.086</td>
</tr>
<tr>
<td>0.14493</td>
<td>300/300</td>
<td>0.182</td>
<td>0.262</td>
<td>0.02500</td>
<td>0.85507</td>
<td>1.086</td>
<td>1.086</td>
</tr>
<tr>
<td>0.17994</td>
<td>400/400</td>
<td>0.182</td>
<td>0.262</td>
<td>0.02500</td>
<td>0.82006</td>
<td>1.086</td>
<td>1.086</td>
</tr>
<tr>
<td>0.21389</td>
<td>500/500</td>
<td>0.182</td>
<td>0.262</td>
<td>0.02500</td>
<td>0.78611</td>
<td>1.086</td>
<td>1.086</td>
</tr>
<tr>
<td>0.24761</td>
<td>600/600</td>
<td>0.182</td>
<td>0.262</td>
<td>0.02500</td>
<td>0.75239</td>
<td>1.086</td>
<td>1.086</td>
</tr>
<tr>
<td>0.28099</td>
<td>700/700</td>
<td>0.182</td>
<td>0.262</td>
<td>0.02500</td>
<td>0.71901</td>
<td>1.086</td>
<td>1.086</td>
</tr>
<tr>
<td>0.31390</td>
<td>800/800</td>
<td>0.182</td>
<td>0.262</td>
<td>0.02500</td>
<td>0.68610</td>
<td>1.086</td>
<td>1.086</td>
</tr>
<tr>
<td>0.34624</td>
<td>900/900</td>
<td>0.182</td>
<td>0.262</td>
<td>0.02500</td>
<td>0.65376</td>
<td>1.086</td>
<td>1.086</td>
</tr>
<tr>
<td>0.37790</td>
<td>1000/1000</td>
<td>0.182</td>
<td>0.262</td>
<td>0.02500</td>
<td>0.62210</td>
<td>1.086</td>
<td>1.086</td>
</tr>
</tbody>
</table>

You can compare these power values with those shown above in Example 1 to validate the procedure. You will find that the power values are identical.
453-10 Tests for Two Means with Non-Unity Null using Ratios
Chapter 455

Non-Inferiority Tests for Two Means using Ratios

Introduction

This procedure calculates power and sample size for non-inferiority t-tests from a parallel-groups design in which the logarithm of the outcome is a continuous normal random variable. This routine deals with the case in which the statistical hypotheses are expressed in terms of mean ratios instead of mean differences.

The details of testing the non-inferiority of two treatments using data from a two-group design are given in another chapter and they will not be repeated here. If the logarithm of the response can be assumed to follow a normal distribution, hypotheses about non-inferiority stated in terms of the ratio can be transformed into hypotheses about the difference. The details of this analysis are given in Julious (2004). They will only be summarized here.

Non-Inferiority Testing Using Ratios

It will be convenient to adopt the following specialized notation for the discussion of these tests.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PASS Input/Output</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_T$</td>
<td>Not used</td>
<td>Treatment mean. This is the treatment mean.</td>
</tr>
<tr>
<td>$\mu_R$</td>
<td>Not used</td>
<td>Reference mean. This is the mean of a reference population.</td>
</tr>
<tr>
<td>$M_{NI}$</td>
<td>NIM</td>
<td>Margin of non-inferiority. This is a tolerance value that defines the maximum amount that is not of practical importance. This is the largest change in the mean ratio from the baseline value (usually one) that is still considered to be trivial.</td>
</tr>
</tbody>
</table>
Non-Inferiority Tests for Two Means using Ratios

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PASS Input/Output</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\phi$</td>
<td>R1</td>
<td>True ratio. This is the value of $\phi = \mu_T / \mu_R$ at which the power is calculated.</td>
</tr>
</tbody>
</table>

Note that the actual values of $\mu_T$ and $\mu_R$ are not needed. Only the ratio of these values is needed for power and sample size calculations.

The null hypothesis of inferiority is

$$H_0: \phi \leq \phi_L \quad \text{where} \quad \phi_L < 1.$$ 

and the alternative hypothesis of non-inferiority is

$$H_1: \phi > \phi_L.$$ 

**Log-Transformation**

In many cases, hypotheses stated in terms of ratios are more convenient than hypotheses stated in terms of differences. This is because ratios can be interpreted as scale-less percentages, but differences must be interpreted as actual amounts in their original scale. Hence, it has become a common practice to take the following steps in hypothesis testing.

1. State the statistical hypotheses in terms of ratios.
2. Transform these into hypotheses about differences by taking logarithms.
3. Analyze the logged data—that is, do the analysis in terms of the difference.
4. Draw the conclusion in terms of the ratio.

The details of step 2 for the null hypothesis are as follows.

$$\phi_L \leq \phi \quad \Rightarrow \quad \phi_L \leq \left\{ \frac{\mu_T}{\mu_R} \right\}$$

$$\Rightarrow \quad \ln(\phi_L) \leq \left\{ \ln(\mu_T) - \ln(\mu_R) \right\}$$

Thus, a hypothesis about the ratio of the means on the original scale can be translated into a hypothesis about the difference of two means on the logged scale.

**Coefficient of Variation**

The coefficient of variation (COV) is the ratio of the standard deviation to the mean. This parameter can be used to represent the variation in the data because of a unique relationship that it has in the case of log-normal data.

Suppose the variable $X$ is the logarithm of the original variable $Y$. That is, $X = \ln(Y)$ and $Y = \exp(X)$. Label the mean and variance of $X$ as $\mu_X$ and $\sigma_X^2$, respectively. Similarly, label the mean and variance of $Y$ as $\mu_Y$ and $\sigma_Y^2$, respectively. If $X$ is normally distributed, then $Y$ is log-normally distributed. Julious (2004) presents the following well-known relationships between these two variables.
\[ \mu_Y = \left( e^{\mu_Y^2 + \frac{\sigma_Y^2}{2}} \right) \]

\[ \sigma_Y^2 = \mu_Y^2 \left( e^{\sigma_Y^2} - 1 \right) \]

From this relationship, the coefficient of variation of \( Y \) can be found to be

\[ COV_Y = \sqrt{\frac{\mu_Y^2 \left( e^{\sigma_Y^2} - 1 \right)}{\mu_Y}} = \sqrt{e^{\sigma_Y^2} - 1} \]

Solving this relationship for \( \sigma_X^2 \), the standard deviation of \( X \) can be stated in terms of the coefficient of variation of \( Y \). This equation is

\[ \sigma_X = \sqrt{\ln \left( COV_Y^2 + 1 \right)} \]

Similarly, the mean of \( X \) is

\[ \mu_X = \frac{\mu_Y}{\ln \left( COV_Y^2 + 1 \right)} \]

One final note: for parallel-group designs, \( \sigma_X^2 \) equals \( \sigma_d^2 \), the average variance used in the t-test of the logged data.

Thus, the hypotheses can be stated in the original (Y) scale and then the power can be analyzed in the transformed (X) scale.

**Power Calculation**

As is shown above, the hypotheses can be stated in the original (Y) scale using ratios or the logged (X) scale using differences. In either case, the power and sample size calculations are made using the formulas for testing the difference in two means. These formulas are presented in another chapter and are not duplicated here.

**Procedure Options**

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

**Data Tab**

The Data tab contains the parameters associated with this test such as the means, sample sizes, alpha, and power.
**Solve For**

**Find (Solve For)**
This option specifies the parameter to be solved for from the other parameters. In most situations, you will select either Power and Beta for a power analysis or $N1$ for sample size determination.

**Error Rates**

**Power or Beta**
This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of inferiority when in fact the treatment mean is non-inferior.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as $0.8 \text{ to } 0.95 \text{ by } 0.05$ may be entered.

**Alpha (Significance Level)**
This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when rejecting the null hypothesis of inferiority when in fact the treatment group is not inferior to the reference group.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as $0.01 \text{ to } 0.10 \text{ by } 0.01$.

**Sample Size**

**N1 (Sample Size Group 2)**
Enter a value (or range of values) for the sample size of group 1 (the reference group). Note that these values are ignored when you are solving for $N1$. You may enter a range of values such as $10 \text{ to } 100 \text{ by } 10$.

**N2 (Sample Size Group 2)**
Enter a value (or range of values) for the sample size of group 2 (the treatment group) or enter Use R to base N2 on the value of N1. You may enter a range of values such as $10 \text{ to } 100 \text{ by } 10$.

- **Use R**
  When Use R is entered here, N2 is calculated using the formula

  \[ N2 = \lceil R(N1) \rceil \]

  where R is the Sample Allocation Ratio and the operator \([Y]\) is the first integer greater than or equal to Y. For example, if you want N1 = N2, select Use R and set R = 1.
**R (Sample Allocation Ratio)**
Enter a value (or range of values) for $R$, the allocation ratio between samples. This value is only used when $N_2$ is set to Use $R$.

When used, $N_2$ is calculated from $N_1$ using the formula: $N_2 = \lceil R(N_1) \rceil$ where $\lceil Y \rceil$ is the next integer greater than or equal to $Y$. Note that setting $R = 1.0$ forces $N_2 = N_1$.

---

**Effect Size – Ratios**

**NIM (Non-Inferiority Margin)**
This is the magnitude of the margin of non-inferiority. It must be entered as a positive number.

When higher means are better, this value is the distance below one for which the mean ratio (Treatment Mean / Reference Mean) still indicates non-inferiority of the treatment mean. E.g., a value of 0.2 here specifies that mean ratios greater than 0.8 indicate non-inferiority of the treatment mean.

When higher means are worse, this value is the distance above one for which the mean ratio (Treatment Mean / Reference Mean) still indicates non-inferiority of the treatment mean. E.g., a value of 0.2 here specifies that mean ratios less than 1.2 indicate non-inferiority of the treatment mean.

Note that the sign of this value is ignored. Only the magnitude is used.

Recommended values:
0.20 is a common value for this parameter.

**R1 (True Ratio)**
This is the value of the ratio of the two means (Treatment Mean / Reference Mean) at which the power is to be calculated.

Often, the ratio will be set to one. However, some authors recommend using a ratio slightly different than one, such as 0.95 (when higher means are "better") or 1.05 (when higher means are "worse"), since this will require a larger sample size.

---

**Effect Size – Coefficient of Variation**

**COV (Coefficient of Variation)**
The coefficient of variation is used to specify the variability (standard deviation). It is important to realize that this is the COV defined on the original (not logged) scale. This value must be determined from past experience or from a pilot study. See the discussion above for more details on the definition of the coefficient of variation.

---

**Test**

**Higher Means Are**
This option defines whether higher values of the response variable are to be considered better or worse. The choice here determines the direction of the non-inferiority test.

If Higher Means Are Better the null hypothesis is $R \leq 1 - \text{NIM}$ and the alternative hypothesis is $R > 1 - \text{NIM}$. If Higher Means Are Worse the null hypothesis is $R \geq 1 + \text{NIM}$ and the alternative hypothesis is $R < 1 + \text{NIM}$.
Example 1 – Finding Power

A company has developed a generic drug for treating rheumatism and wants to show that it is not inferior to the standard drug. Responses following either treatment are known to follow a log normal distribution. A parallel-group design will be used and the logged data will be analyzed with a two-sample t-test.

Researchers have decided to set the margin of equivalence at 0.20. Past experience leads the researchers to set the COV to 1.50. The significance level is 0.025. The power will be computed assuming that the true ratio is either 0.95 or 1.00. Sample sizes between 100 and 1000 will be included in the analysis.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Non-Inferiority for Two Means [Ratios] procedure window by expanding Means, then Two Means, then clicking on Non-Inferiority, and then clicking on Non-Inferiority for Two Means [Ratios]. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.025</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>100 to 1000 by 100</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>E (Equivalence Margin)</td>
<td>0.20</td>
</tr>
<tr>
<td>R1 (True Ratio)</td>
<td>0.95 1.0</td>
</tr>
<tr>
<td>COV (Coefficient of Variation)</td>
<td>1.50</td>
</tr>
<tr>
<td>Test Type</td>
<td>Non-Inferiority</td>
</tr>
<tr>
<td>Higher Is</td>
<td>Good</td>
</tr>
</tbody>
</table>
Annotated Output
Click the Run button to perform the calculations and generate the following output.

### Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N1/N2</th>
<th>Non-Inferiority Margin (NIM)</th>
<th>Non-Inferiority Bound (NIB)</th>
<th>True Ratio (R1)</th>
<th>Significance Level (Alpha)</th>
<th>Coefficient of Variation (COV)</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1987</td>
<td>100/100</td>
<td>0.20</td>
<td>0.80</td>
<td>0.95</td>
<td>0.0250</td>
<td>1.50</td>
<td>0.8013</td>
</tr>
<tr>
<td>0.3539</td>
<td>200/200</td>
<td>0.20</td>
<td>0.80</td>
<td>0.95</td>
<td>0.0250</td>
<td>1.50</td>
<td>0.6461</td>
</tr>
<tr>
<td>0.4918</td>
<td>300/300</td>
<td>0.20</td>
<td>0.80</td>
<td>0.95</td>
<td>0.0250</td>
<td>1.50</td>
<td>0.5082</td>
</tr>
<tr>
<td>0.6098</td>
<td>400/400</td>
<td>0.20</td>
<td>0.80</td>
<td>0.95</td>
<td>0.0250</td>
<td>1.50</td>
<td>0.3902</td>
</tr>
<tr>
<td>0.7064</td>
<td>500/500</td>
<td>0.20</td>
<td>0.80</td>
<td>0.95</td>
<td>0.0250</td>
<td>1.50</td>
<td>0.2936</td>
</tr>
<tr>
<td>0.7827</td>
<td>600/600</td>
<td>0.20</td>
<td>0.80</td>
<td>0.95</td>
<td>0.0250</td>
<td>1.50</td>
<td>0.2173</td>
</tr>
<tr>
<td>0.8416</td>
<td>700/700</td>
<td>0.20</td>
<td>0.80</td>
<td>0.95</td>
<td>0.0250</td>
<td>1.50</td>
<td>0.1584</td>
</tr>
<tr>
<td>0.8860</td>
<td>800/800</td>
<td>0.20</td>
<td>0.80</td>
<td>0.95</td>
<td>0.0250</td>
<td>1.50</td>
<td>0.1140</td>
</tr>
<tr>
<td>0.9189</td>
<td>900/900</td>
<td>0.20</td>
<td>0.80</td>
<td>0.95</td>
<td>0.0250</td>
<td>1.50</td>
<td>0.0811</td>
</tr>
<tr>
<td>0.9428</td>
<td>1000/1000</td>
<td>0.20</td>
<td>0.80</td>
<td>0.95</td>
<td>0.0250</td>
<td>1.50</td>
<td>0.0572</td>
</tr>
<tr>
<td>0.3038</td>
<td>100/100</td>
<td>0.20</td>
<td>0.80</td>
<td>1.00</td>
<td>0.0250</td>
<td>1.50</td>
<td>0.6962</td>
</tr>
<tr>
<td>0.5384</td>
<td>200/200</td>
<td>0.20</td>
<td>0.80</td>
<td>1.00</td>
<td>0.0250</td>
<td>1.50</td>
<td>0.4616</td>
</tr>
<tr>
<td>0.7113</td>
<td>300/300</td>
<td>0.20</td>
<td>0.80</td>
<td>1.00</td>
<td>0.0250</td>
<td>1.50</td>
<td>0.2887</td>
</tr>
<tr>
<td>0.8280</td>
<td>400/400</td>
<td>0.20</td>
<td>0.80</td>
<td>1.00</td>
<td>0.0250</td>
<td>1.50</td>
<td>0.1720</td>
</tr>
<tr>
<td>0.9013</td>
<td>500/500</td>
<td>0.20</td>
<td>0.80</td>
<td>1.00</td>
<td>0.0250</td>
<td>1.50</td>
<td>0.0987</td>
</tr>
<tr>
<td>0.9451</td>
<td>600/600</td>
<td>0.20</td>
<td>0.80</td>
<td>1.00</td>
<td>0.0250</td>
<td>1.50</td>
<td>0.0549</td>
</tr>
<tr>
<td>0.9702</td>
<td>700/700</td>
<td>0.20</td>
<td>0.80</td>
<td>1.00</td>
<td>0.0250</td>
<td>1.50</td>
<td>0.0298</td>
</tr>
<tr>
<td>0.9842</td>
<td>800/800</td>
<td>0.20</td>
<td>0.80</td>
<td>1.00</td>
<td>0.0250</td>
<td>1.50</td>
<td>0.0158</td>
</tr>
<tr>
<td>0.9918</td>
<td>900/900</td>
<td>0.20</td>
<td>0.80</td>
<td>1.00</td>
<td>0.0250</td>
<td>1.50</td>
<td>0.0082</td>
</tr>
<tr>
<td>0.9958</td>
<td>1000/1000</td>
<td>0.20</td>
<td>0.80</td>
<td>1.00</td>
<td>0.0250</td>
<td>1.50</td>
<td>0.0042</td>
</tr>
</tbody>
</table>

**Report Definitions**
- **H0** (null hypothesis) is that R <= 1-NIM, where R = Treatment Mean / Reference Mean.
- **H1** (alternative hypothesis) is that R > 1-NIM.
- Power is the probability of rejecting H0 when it is false.
- N1 is the number of subjects in the first (treatment) group.
- N2 is the number of subjects in the second (reference) group.
- NIM is the magnitude and direction of the margin of non-inferiority. Since higher means are better, this value is negative and is the distance below one that is still considered non-inferior.
- NIB is the corresponding bound to the non-inferiority margin, and equals 1 - NIM.
- R1 is the mean ratio (treatment/reference) at which the power is computed.
- Alpha is the probability of falsely rejecting H0.
- COV is the coefficient of variation on the original scale.
- Beta is the probability of not rejecting H0 when it is false.

**Summary Statements**
- Group sample sizes of 100 in the first group and 100 in the second group achieve 20% power to detect non-inferiority using a one-sided, two-sample t-test. The margin of non-inferiority is 0.20.
- The true ratio of the means at which the power is evaluated is 0.95. The significance level (alpha) of the test is 0.0250. The coefficients of variation of both groups are assumed to be...
Plots Section

This plot shows the power versus the sample size.

Example 2 – Validation

We could not find a validation example for this procedure in the statistical literature. Therefore, we will show that this procedure gives the same results as the non-inferiority test on differences—a procedure that has been validated. We will use the same settings as those given in Example 1. Since the output for this example is shown above, all that we need is the output from the procedure that uses differences.

To run the inferiority test on differences, we need the values of NIM and S1.

\[
S1 = \sqrt{\ln(COV^2 + 1)}
\]

\[
= \sqrt{\ln(1.5^2 + 1)}
\]

\[
= 1.085659
\]

\[
NIM^* = \ln(1 - NIM)
\]

\[
= \ln(0.8)
\]

\[
= 0.223144
\]

\[
D = \ln(R1)
\]

\[
= \ln(0.95)
\]

\[
= -0.051293
\]
Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Non-Inferiority for Two Means [Differences] procedure window by expanding Means, then Two Means, then clicking on Non-Inferiority, and then clicking on Non-Inferiority for Two Means [Differences]. You may then make the appropriate entries as listed below, or open Example 1b by going to the File menu and choosing Open Example Template.

Option | Value
--- | ---
Data Tab | Power and Beta
Find (Solve For) | Power and Beta
Power | Ignored since this is the Find setting
Alpha | 0.025
N1 (Sample Size Group 1) | 100 to 1000 by 100
N2 (Sample Size Group 2) | Use R
R (Sample Allocation Ratio) | 1.0
|E| (Equivalence Margin) | 0.223144
D (True Difference) | -0.051293 0.0
S1 (Standard Deviation Group 1) | 1.085659
S2 (Standard Deviation Group 2) | S1
Higher Means Are | Better

Output

Click the Run button to perform the calculations and generate the following output.

**Numeric Results**

<table>
<thead>
<tr>
<th>Power</th>
<th>N1/N2</th>
<th>Non-Infer. Margin Difference (D)</th>
<th>Actual Difference (D)</th>
<th>Significance Level (Alpha)</th>
<th>Beta</th>
<th>Standard Deviation1 (SD1)</th>
<th>Standard Deviation2 (SD2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1987</td>
<td>100/100</td>
<td>-0.223 -0.051</td>
<td>-0.223 -0.051</td>
<td>0.025</td>
<td>0.8013</td>
<td>1.086</td>
<td>1.086</td>
</tr>
<tr>
<td>0.3539</td>
<td>200/200</td>
<td>-0.223 -0.051</td>
<td>-0.223 -0.051</td>
<td>0.025</td>
<td>0.6461</td>
<td>1.086</td>
<td>1.086</td>
</tr>
<tr>
<td>0.4918</td>
<td>300/300</td>
<td>-0.223 -0.051</td>
<td>-0.223 -0.051</td>
<td>0.025</td>
<td>0.5082</td>
<td>1.086</td>
<td>1.086</td>
</tr>
<tr>
<td>0.6098</td>
<td>400/400</td>
<td>-0.223 -0.051</td>
<td>-0.223 -0.051</td>
<td>0.025</td>
<td>0.3902</td>
<td>1.086</td>
<td>1.086</td>
</tr>
<tr>
<td>0.7064</td>
<td>500/500</td>
<td>-0.223 -0.051</td>
<td>-0.223 -0.051</td>
<td>0.025</td>
<td>0.2936</td>
<td>1.086</td>
<td>1.086</td>
</tr>
<tr>
<td>0.7828</td>
<td>600/600</td>
<td>-0.223 -0.051</td>
<td>-0.223 -0.051</td>
<td>0.025</td>
<td>0.2172</td>
<td>1.086</td>
<td>1.086</td>
</tr>
<tr>
<td>0.8416</td>
<td>700/700</td>
<td>-0.223 -0.051</td>
<td>-0.223 -0.051</td>
<td>0.025</td>
<td>0.1584</td>
<td>1.086</td>
<td>1.086</td>
</tr>
<tr>
<td>0.8860</td>
<td>800/800</td>
<td>-0.223 -0.051</td>
<td>-0.223 -0.051</td>
<td>0.025</td>
<td>0.1140</td>
<td>1.086</td>
<td>1.086</td>
</tr>
<tr>
<td>0.9189</td>
<td>900/900</td>
<td>-0.223 -0.051</td>
<td>-0.223 -0.051</td>
<td>0.025</td>
<td>0.0811</td>
<td>1.086</td>
<td>1.086</td>
</tr>
<tr>
<td>0.9428</td>
<td>1000/1000</td>
<td>-0.223 -0.051</td>
<td>-0.223 -0.051</td>
<td>0.025</td>
<td>0.0572</td>
<td>1.086</td>
<td>1.086</td>
</tr>
<tr>
<td>0.3038</td>
<td>100/100</td>
<td>-0.223 0.000</td>
<td>-0.223 0.000</td>
<td>0.025</td>
<td>0.6962</td>
<td>1.086</td>
<td>1.086</td>
</tr>
<tr>
<td>0.5384</td>
<td>200/200</td>
<td>-0.223 0.000</td>
<td>-0.223 0.000</td>
<td>0.025</td>
<td>0.4616</td>
<td>1.086</td>
<td>1.086</td>
</tr>
<tr>
<td>0.7113</td>
<td>300/300</td>
<td>-0.223 0.000</td>
<td>-0.223 0.000</td>
<td>0.025</td>
<td>0.2887</td>
<td>1.086</td>
<td>1.086</td>
</tr>
<tr>
<td>0.8280</td>
<td>400/400</td>
<td>-0.223 0.000</td>
<td>-0.223 0.000</td>
<td>0.025</td>
<td>0.1720</td>
<td>1.086</td>
<td>1.086</td>
</tr>
<tr>
<td>0.9013</td>
<td>500/500</td>
<td>-0.223 0.000</td>
<td>-0.223 0.000</td>
<td>0.025</td>
<td>0.0987</td>
<td>1.086</td>
<td>1.086</td>
</tr>
<tr>
<td>0.9451</td>
<td>600/600</td>
<td>-0.223 0.000</td>
<td>-0.223 0.000</td>
<td>0.025</td>
<td>0.0549</td>
<td>1.086</td>
<td>1.086</td>
</tr>
<tr>
<td>0.9922</td>
<td>700/700</td>
<td>-0.223 0.000</td>
<td>-0.223 0.000</td>
<td>0.025</td>
<td>0.0238</td>
<td>1.086</td>
<td>1.086</td>
</tr>
<tr>
<td>0.9842</td>
<td>800/800</td>
<td>-0.223 0.000</td>
<td>-0.223 0.000</td>
<td>0.025</td>
<td>0.0158</td>
<td>1.086</td>
<td>1.086</td>
</tr>
<tr>
<td>0.9918</td>
<td>900/900</td>
<td>-0.223 0.000</td>
<td>-0.223 0.000</td>
<td>0.025</td>
<td>0.0082</td>
<td>1.086</td>
<td>1.086</td>
</tr>
<tr>
<td>0.9958</td>
<td>1000/1000</td>
<td>-0.223 0.000</td>
<td>-0.223 0.000</td>
<td>0.025</td>
<td>0.0042</td>
<td>1.086</td>
<td>1.086</td>
</tr>
</tbody>
</table>
You can compare these power values with those shown above in Example 1 to validate the procedure. You will find that the power values are identical.
Chapter 460

Equivalence Tests for Two Means using Differences

Introduction

This procedure allows you to study the power and sample size of equivalence tests of the means of two independent groups using the two-sample t-test. Schuirmann’s (1987) two one-sided tests (TOST) approach is used to test equivalence. Only a brief introduction to the subject will be given here. For a comprehensive discussion, refer to Chow and Liu (1999).

Measurements are made on individuals that have been randomly assigned to one of two groups. This parallel-groups design may be analyzed by a TOST equivalence test to show that the means of the two groups do not differ by more than a small amount, called the margin of equivalence.

The definition of equivalence has been refined in recent years using the concepts of prescribability and switchability. Prescribability refers to ability of a physician to prescribe either of two drugs at the beginning of the treatment. However, once prescribed, no other drug can be substituted for it. Switchability refers to the ability of a patient to switch from one drug to another during treatment without adverse effects. Prescribability is associated with equivalence of location and variability. Switchability is associated with the concept of individual equivalence. This procedure analyzes average equivalence. Thus, it partially analyzes prescribability. It does not address equivalence of variability or switchability.

Parallel-Group Design

In a parallel-group design, subjects are assigned at random to either of two groups. Group 1 is the treatment group and group 2 is the reference group.

Outline of an Equivalence Test

PASS follows the two one-sided tests approach described by Schuirmann (1987) and Phillips (1990). Let $\mu_2 = \mu_T$ be the test group mean, $\mu_1 = \mu_R$ the reference group mean, and $\varepsilon_L$ and $\varepsilon_U$ the lower and upper bounds on $D = \mu_2 - \mu_1 = \mu_T - \mu_R$ that define the region of equivalence. The null hypothesis of non-equivalence is
460-2 Equivalence Tests for Two Means using Differences

\[ H_0: D \leq \varepsilon_L \quad \text{or} \quad H_0: D \geq \varepsilon_U \]

and the alternative hypothesis of equivalence is

\[ H_1: \varepsilon_L < D < \varepsilon_U. \]

**Two-Sample T-Test**

This test assumes that the two groups of normally-distributed values have the same variance. The calculation of the two one-sided test statistics uses the following equations.

\[ T_L = \frac{(\bar{X}_2 - \bar{X}_1) - \varepsilon_L}{s_{\bar{X}_1 - \bar{X}_2}} \quad \text{and} \quad T_U = \frac{(\bar{X}_2 - \bar{X}_1) - \varepsilon_U}{s_{\bar{X}_1 - \bar{X}_2}} \]

where

\[ \bar{X}_k = \frac{\sum_{i=1}^{N_k} X_{ki}}{N_k} \]

\[ s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2 + \sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2}{N_1 + N_2 - 2} \left( \frac{1}{N_1} + \frac{1}{N_2} \right)} \]

\[ df = N_1 + N_2 - 2 \]

The null hypothesis is rejected if \( T_L \) and \( -T_U \) are greater than or equal to \( t_{1-\alpha, N_1+N_2-2} \).

The power of this test is given by

\[ \Pr(T_L \geq t_{1-\alpha, N_1+N_2-2} \quad \text{and} \quad T_U \leq -t_{1-\alpha, N_1+N_2-2} | \mu_L, \mu_R, \sigma^2) 1 \]

where \( T_L \) and \( T_U \) are distributed as the bivariate, noncentral \( t \) distribution with noncentrality parameters \( \Delta_L \) and \( \Delta_U \) given by

\[ \Delta_L = \frac{D - \varepsilon_L}{\sigma \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \]

\[ \Delta_U = \frac{D - \varepsilon_U}{\sigma \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}} \]
Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains the parameters associated with this test such as the means, sample sizes, alpha, and power.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for from the other parameters. Under most situations, you will select either Power and Beta for a power analysis or N1 for sample size determination.

Select N1 when you want to calculate the sample size needed to achieve a given power and alpha level.

Select Power and Beta when you want to calculate the power of an experiment that has already been run.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of nonequivalent means when in fact the means are equivalent.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when you reject the null hypothesis of non-equivalent means when in fact the means are nonequivalent.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.
Sample Size

N1 (Sample Size Reference Group)
Specify the number of subjects in the reference group. The total number of subjects in the experiment is equal to $N1 + N2$.

You may enter a range of values such as 10 to 100 by 10.

N2 (Sample Size Treatment Group)
Specify one or more values for the number of subjects in the treatment group. Alternatively, enter Use R to base N2 on the value of N1. You may also enter a range of values such as 10 to 100 by 10.

• Use R
  When Use R is entered here, N2 is calculated using the formula
  \[ N2 = \lceil R(N1) \rceil \]
  where $R$ is the Sample Allocation Ratio and $\lceil Y \rceil$ means take the first integer greater than or equal to $Y$. For example, if you want $N1 = N2$, select Use R and set $R = 1$.

R (Sample Allocation Ratio)
Enter a value (or range of values) for $R$, the allocation ratio between samples. This value is only used when N2 is set to Use R.

When used, N2 is calculated from N1 using the formula: $N2 = \lceil R(N1) \rceil$ where $\lceil Y \rceil$ means take the next integer greater than or equal to $Y$. Note that setting $R = 1.0$ forces $N1 = N2$.

Effect Size – Equivalence Limits

|EU| Upper Equivalence Limit
This value gives upper limit on equivalence. Differences outside EL and EU are not considered equivalent. Differences between them are considered equivalent.

Note that EL<0 and EU>0. Also, you must have EL<D<EU.

-|EL| Lower Equivalence Limit
This value gives lower limit on equivalence. Differences outside EL and EU are not considered equivalent. Differences between them are.

If you want symmetric limits, enter -UPPER LIMIT for EL to force EL = -|EU|.

Note that EL<0 and EU>0. Also, you must have EL<D<EU. Finally, the scale of these numbers must match the scale of S.

Effect Size – True Mean Difference

D (True Difference)
This is the true difference between the two means at which the power is to be computed. Often this value is set to zero, but it can be non-zero as long as it is between the equivalence limits EL and EU.
Effect Size – Standard Deviation

S (Standard Deviation)
Specify the within-group standard deviation, \( \sigma \). The standard deviation is assumed to be the same for both groups.

Example 1 – Parallel-Group Design
A parallel-group is to be used to compare influence of two drugs on diastolic blood pressure. The diastolic blood pressure is known to be close to 96 mmHg with the reference drug and is thought to be 92 mmHg with the experimental drug. Based on similar studies, the within-group standard deviation is set to 18 mmHg. Following FDA guidelines, the researchers want to show that the diastolic blood pressure with the experimental drug is within 20% of the diastolic blood pressure with the reference drug. Note that 20% of 96 is 19.2. They decide to calculate the power for a range of sample sizes between 3 and 60. The significance level is 0.05.

Setup
This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for Two Means [Differences]** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for Two Means [Differences]**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Tab</strong></td>
<td><strong>Power and Beta</strong></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>3 5 8 10 15 20 30 40 50 60</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>(</td>
<td>EU</td>
</tr>
<tr>
<td>(-</td>
<td>EL</td>
</tr>
<tr>
<td>D (True Difference)</td>
<td>-4</td>
</tr>
<tr>
<td>S (Standard Deviation)</td>
<td>18</td>
</tr>
</tbody>
</table>
## Numeric Results

### Numeric Results for Testing Equivalence Using a Parallel-Group Design

<table>
<thead>
<tr>
<th>Power</th>
<th>Reference Group Sample Size (N1)</th>
<th>Treatment Group Sample Size (N2)</th>
<th>Lower Equiv. Limit</th>
<th>Upper Equiv. Limit</th>
<th>True Difference</th>
<th>Standard Deviation</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0386</td>
<td>3</td>
<td>3</td>
<td>-19.20</td>
<td>19.20</td>
<td>-4.00</td>
<td>18.00</td>
<td>0.0500</td>
<td>0.9614</td>
</tr>
<tr>
<td>0.0928</td>
<td>5</td>
<td>5</td>
<td>-19.20</td>
<td>19.20</td>
<td>-4.00</td>
<td>18.00</td>
<td>0.0500</td>
<td>0.9072</td>
</tr>
<tr>
<td>0.2887</td>
<td>8</td>
<td>8</td>
<td>-19.20</td>
<td>19.20</td>
<td>-4.00</td>
<td>18.00</td>
<td>0.0500</td>
<td>0.7113</td>
</tr>
<tr>
<td>0.4391</td>
<td>10</td>
<td>10</td>
<td>-19.20</td>
<td>19.20</td>
<td>-4.00</td>
<td>18.00</td>
<td>0.0500</td>
<td>0.5609</td>
</tr>
<tr>
<td>0.6934</td>
<td>15</td>
<td>15</td>
<td>-19.20</td>
<td>19.20</td>
<td>-4.00</td>
<td>18.00</td>
<td>0.0500</td>
<td>0.3066</td>
</tr>
<tr>
<td>0.8266</td>
<td>20</td>
<td>20</td>
<td>-19.20</td>
<td>19.20</td>
<td>-4.00</td>
<td>18.00</td>
<td>0.0500</td>
<td>0.1734</td>
</tr>
<tr>
<td>0.9433</td>
<td>30</td>
<td>30</td>
<td>-19.20</td>
<td>19.20</td>
<td>-4.00</td>
<td>18.00</td>
<td>0.0500</td>
<td>0.0567</td>
</tr>
<tr>
<td>0.9820</td>
<td>40</td>
<td>40</td>
<td>-19.20</td>
<td>19.20</td>
<td>-4.00</td>
<td>18.00</td>
<td>0.0500</td>
<td>0.0180</td>
</tr>
<tr>
<td>0.9946</td>
<td>50</td>
<td>50</td>
<td>-19.20</td>
<td>19.20</td>
<td>-4.00</td>
<td>18.00</td>
<td>0.0500</td>
<td>0.0054</td>
</tr>
<tr>
<td>0.9984</td>
<td>60</td>
<td>60</td>
<td>-19.20</td>
<td>19.20</td>
<td>-4.00</td>
<td>18.00</td>
<td>0.0500</td>
<td>0.0016</td>
</tr>
</tbody>
</table>

### Report Definitions

- **Power** is the probability of rejecting non-equivalence when they are equivalent.
- **N1** is the number of subjects in the reference group.
- **N2** is the number of subjects in the treatment group.
- The **Upper & Lower Limits** are the maximum allowable differences that result in equivalence.
- **True Difference** is the anticipated actual difference between the means.
- **Standard Deviation** is the average S.D. within the two groups.
- **Alpha** is the probability of rejecting non-equivalence when they are non-equivalent.
- **Beta** is the probability of accepting non-equivalence when they are equivalent.

### Summary Statements

An equivalence test of means using two one-sided tests on data from a parallel-group design with sample sizes of 3 in the reference group and 3 in the treatment group achieves 4% power at a 5% significance level when the true difference between the means is -4.00, the standard deviation is 18.00, and the equivalence limits are -19.20 and 19.20.

This report shows the power for the indicated parameter configurations. Note that when the parameters are specified as percentages, they are displayed in the output with percent signs. Note that the desired 80% power occurs for a per group sample size between 15 and 20.

### Plot Section

This plot shows the power versus the sample size.
Example 2 – Parallel-Group Validation using Machin

Machin et al. (1997) page 107 present an example of determining the sample size for a parallel-group design in which the reference mean is 96, the treatment mean is 94, the standard deviation is 8, the limits are plus or minus 5, the power is 80%, and the significance level is 0.05. They calculate the sample size to be 88. It is important to note that Machin et al. use an approximation, so their results should not be expected to exactly match the results obtained using PASS.

We will now set up this example in PASS.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Equivalence Tests for Two Means [Differences] procedure window by expanding Means, then Two Independent Means, then clicking on Equivalence, and then clicking on Equivalence Tests for Two Means [Differences]. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N1</td>
</tr>
<tr>
<td>Power</td>
<td>0.80</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>Ignored</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>EU</td>
</tr>
<tr>
<td></td>
<td>-</td>
</tr>
<tr>
<td>D (True Difference)</td>
<td>-2</td>
</tr>
<tr>
<td>S (Standard Deviation)</td>
<td>8</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

### Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>Reference Group Sample Size (N1)</th>
<th>Treatment Group Sample Size (N2)</th>
<th>Lower Equiv. Limit</th>
<th>Upper Equiv. Limit</th>
<th>True Difference</th>
<th>Standard Deviation</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8015</td>
<td>89</td>
<td>89</td>
<td>-5.00</td>
<td>5.00</td>
<td>-2.00</td>
<td>8.00</td>
<td>0.0500</td>
<td>0.1985</td>
</tr>
</tbody>
</table>

Note that PASS has obtained a sample size of 89 which is very close to the approximate value of 88 that Machin calculated.
460-8 Equivalence Tests for Two Means using Differences
Equivalence Tests for Two Means (Simulation)

Introduction

This procedure allows you to study the power and sample size of an equivalence test comparing two means from independent groups. Schuirmann’s (1987) two one-sided tests (TOST) approach is used to test equivalence. The t-test is commonly used in this situation, but other tests have been developed for use when the t-test assumptions are not met. These additional tests include the Mann-Whitney U test, Welch’s unequal variance test, and trimmed versions of the t-test and the Welch test.

Measurements are made on individuals that have been randomly assigned to, or randomly chosen from, one of two groups. This parallel-groups design may be analyzed by a TOST equivalence test to show that the means of the two groups do not differ by more than a small amount, called the margin of equivalence.

The two-sample t-test is commonly used in this situation. When the variances of the two groups are unequal, Welch’s t-test is often used. When the data are not normally distributed, the Mann-Whitney (Wilcoxon signed-ranks) U test and, less frequently, the trimmed t-test may be used.

The details of the power analysis of equivalence test using analytic techniques are presented in another PASS chapter and they won’t be duplicated here. This chapter will only consider power analysis using computer simulation.

Technical Details

Computer simulation allows us to estimate the power and significance level that is actually achieved by a test procedure in situations that are not mathematically tractable. Computer simulation was once limited to mainframe computers. But, in recent years, as computer speeds have increased, simulation studies can be completed on desktop and laptop computers in a reasonable period of time.

The steps to a simulation study are as follows.

1. Specify how the test is carried out. This includes indicating how the test statistic is calculated and how the significance level is specified.
2. Generate random samples from the distributions specified by the alternative hypothesis. Calculate the test statistics from the simulated data and determine if the null hypothesis is accepted or rejected. Tabulate the number of rejections and use this to calculate the test’s power.

3. Generate random samples from the distributions specified by the null hypothesis. Calculate each test statistic from the simulated data and determine if the null hypothesis is accepted or rejected. Tabulate the number of rejections and use this to calculate the test’s significance level.

4. Repeat steps 2 and 3 several thousand times, tabulating the number of times the simulated data leads to a rejection of the null hypothesis. The power is the proportion of simulated samples in step 2 that lead to rejection. The significance level is the proportion of simulated samples in step 3 that lead to rejection.

Generating Random Distributions

Two methods are available in **PASS** to simulate random samples. The first method generates the random variates directly, one value at a time. The second method generates a large pool (over 10,000) of random values and then draw the random numbers from this pool. This second method can cut the running time of the simulation by 70%.

As mentioned above, the second method begins by generating a large pool of random numbers from the specified distributions. Each of these pools is evaluated to determine if its mean is within a small relative tolerance (0.0001) of the target mean. If the actual mean is not within the tolerance of the target mean, individual members of the population are replaced with new random numbers if the new random number moves the mean towards its target. Only a few hundred such swaps are required to bring the actual mean to within tolerance of the target mean. This population is then sampled with replacement using the uniform distribution. We have found that this method works well as long as the size of the pool is the maximum of twice the number of simulated samples desired and 10,000.

Simulating Data for an Equivalence Test

Simulating equivalence data is more complex than simulating data for a regular two-sided test. An equivalence test essentially reverses the roles of the null and alternative hypothesis. The null hypothesis becomes

\[ H_0: (\mu_1 - \mu_2) \leq -D \text{ or } (\mu_1 - \mu_2) \geq D \]

where \( D \) is the margin of equivalence. Thus the null hypothesis is made up of two simple hypotheses:

\[ H_{01}: (\mu_1 - \mu_2) \leq -D \]
\[ H_{02}: (\mu_1 - \mu_2) \geq D \]

The additional complexity comes in deciding which of the two null hypotheses are used to simulate data for the null hypothesis situation. The choice becomes more problematic when asymmetric equivalence limits are chosen. In this case, you may want to try simulating using each simple null hypothesis in turn.
To generate data for the null hypotheses, generate data for each group. The difference in the means of these two groups will become one of the equivalence limits. The other equivalence limit will be determined by symmetry and will always have a sign that is the opposite of the first equivalence limit.

### Test Statistics

This section describes the test statistics that are available in this procedure. Note that these test statistics are computed on the differences. Thus, when the equation refers to an X value, this X value is assumed to be a difference between two individual variates.

#### Two-Sample T-Test

The t-test assumes that the data are simple random samples from populations of normally-distributed values that have the same mean and variance. This assumption implies that the data are continuous and their distribution is symmetric. The calculation of the t statistic is as follows.

\[
t_{df} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{x}_1 - \bar{x}_2}}
\]

where

\[
\bar{X}_k = \frac{\sum_{i=1}^{N_k} X_{ki}}{N_k}
\]

\[
s_{\bar{x}_1 - \bar{x}_2} = \sqrt{\frac{\sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2 + \sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2}{N_1 + N_2 - 2} \left(\frac{1}{N_1} + \frac{1}{N_2}\right)}
\]

\[df = N_1 + N_2 - 2\]

The significance of the test statistic is determined by computing a p-value which is based on the t distribution with appropriate degrees of freedom. If this p-value is less than a specified level (often 0.05), the null hypothesis is rejected. Otherwise, no conclusion can be reached.

#### Welch’s T-Test

Welch (1938) proposed the following test for use when the two variances are not assumed to be equal.

\[
t^*_{df} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{x}_1 - \bar{x}_2}}
\]
where
\[
S_{X_1 - X_2}^* = \sqrt{\frac{N_1}{N_1(N_1 - 1)} \left( \sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2 \right) + \frac{N_2}{N_2(N_2 - 1)} \left( \sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2 \right)}
\]

\[
f = \frac{\left( \frac{s_1^2}{N_1} + \frac{s_2^2}{N_2} \right)^2}{\frac{s_1^4}{N_1(N_1 - 1)} + \frac{s_2^4}{N_2(N_2 - 1)}}
\]

\[
s_1 = \sqrt{\frac{\sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2}{N_1 - 1}}
\]

\[
s_2 = \sqrt{\frac{\sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2}{N_2 - 1}}
\]

**Trimmed T-Test assuming Equal Variances**

The notion of trimming off a small proportion of possibly outlying observations and using the remaining data to form a t-test was first proposed for one sample by Tukey and McLaughlin (1963). Tukey and Dixon (1968) consider a slight modification of this test, called *Winsorization*, which replaces the trimmed data with the nearest remaining value. The two-sample trimmed t-test was proposed by Yuen and Dixon (1973).

Assume that the data values have been sorted from lowest to highest. The *trimmed mean* is defined as

\[
\bar{X}_{tg} = \frac{\sum_{k=g+1}^{N-g} X_k}{h}
\]

where \( h = N - 2g \) and \( g = \lceil N(G/100) \rceil \). Here we use \( \lceil Z \rceil \) to mean the largest integer smaller than \( Z \) with the modification that if \( G \) is non-zero, the value of \( \lceil N(G/100) \rceil \) is at least one. \( G \) is the percent trimming and should usually be less than 25%, often between 5% and 10%. Thus, the \( g \) smallest and \( g \) largest observation are omitted in the calculation.

To calculate the modified t-test, calculate the *Winsorized mean* and the *Winsorized* sum of squared deviations as follows.

\[
\bar{X}_{wg} = \frac{g(X_{g+1} + X_{N-g}) + \sum_{k=g+1}^{N-g} X_k}{N}
\]

\[
SSD_{wg} = \frac{g(X_{g+1} - \bar{X}_{wg})^2 + g(X_{N-g} - \bar{X}_{wg})^2 + \sum_{k=g+1}^{N-g} (X_k - \bar{X}_{wg})^2}{N}
\]
Using the above definitions, the two-sample trimmed t-test is given by

$$T_{tg} = \frac{(\bar{X}_{1tg} - \bar{X}_{2tg}) - (\mu_1 - \mu_2)}{\sqrt{\frac{SSD_{1wg} + SSD_{2wg}}{h_1 + h_2 - 2} \left( \frac{1}{h_1} + \frac{1}{h_2} \right)}}$$

The distribution of this $t$ statistic is approximately that of a $t$ distribution with degrees of freedom equal to $h_1 + h_2 - 2$. This approximation is often reasonably accurate if both sample sizes are greater than 6.

**Trimmed T-Test assuming Unequal Variances**

Yuen (1974) combines trimming (see above) with Welch’s (1938) test. The resulting trimmed Welch test is resistant to outliers and seems to alleviate some of the problems that occur because of skewness in the underlying distributions. Extending the results from above, the trimmed version of Welch’s $t$-test is given by

$$T_{tg}^* = \frac{(\bar{X}_{1tg} - \bar{X}_{2tg}) - (\mu_1 - \mu_2)}{\sqrt{\frac{SSD_{1wg}}{h_1(h_1 - 1)} + \frac{SSD_{2wg}}{h_2(h_2 - 1)}}}$$

with degrees of freedom $f$ given by

$$\frac{1}{f} = \frac{c^2}{h_1 - 1} + \frac{1 - c^2}{h_2 - 1}$$

where

$$c = \frac{SSD_{1wg}}{SSD_{1wg}/h_1(h_1 - 1) + SSD_{2wg}/h_2(h_2 - 1)}$$

**Mann-Whitney U Test**

This test is the nonparametric substitute for the equal-variance $t$-test. Two key assumptions for this test are that the distributions are at least ordinal and that they are identical under H0. This means that ties (repeated values) are not acceptable. When ties are present, an approximation can be used, but the theoretic results no longer hold.

The Mann-Whitney test statistic is defined as follows in Gibbons (1985).

$$z = \frac{W_1 - \frac{N_1(N_1 + N_2 + 1)}{2} + C}{s_w}$$
where

\[ W_1 = \sum_{k=1}^{N_1} Rank(X_{1k}) \]

The ranks are determined after combining the two samples. The standard deviation is calculated as

\[ s_{WR} = \sqrt{\frac{N_1N_2(N_1 + N_2 + 1)}{12} \left( \frac{N_1N_2}{12(N_1 + N_2)(N_1 + N_2 - 1)} \sum_{i=1}^t (t_i^2 - t_i) \right)} \]

where \( t_i \) is the number of observations tied at value one, \( t_2 \) is the number of observations tied at some value two, and so forth.

The correction factor, \( C \), is 0.5 if the rest of the numerator of \( z \) is negative or -0.5 otherwise. The value of \( z \) is then compared to the standard normal distribution.

**Standard Deviations**

Care must be used when either the null or alternative distribution is not normal. In these cases, the standard deviation is usually not specified directly. For example, you might use a gamma distribution with a shape parameter of 1.5 and a mean of 4 as the null distribution and a gamma distribution with the same shape parameter and a mean of 5 as the alternative distribution. This allows you to compare the two means. However, although the shape parameters are constant, the standard deviations, which are based on both the shape parameter and the mean, are not. Thus the distributions not only have different means, but different standard deviations!

**Procedure Options**

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

**Data Tab**

The Data tab contains most of the parameters and options that you will be concerned with.

**Solve For**

**Find (Solve For)**

This option specifies the parameter to be calculated using the values of the other parameters. Under most conditions, you would select either Power or N1.

Select Power when you want to estimate the power for a specific scenario.

Select N1 when you want to determine the sample size needed to achieve a given power and alpha level. This option is computationally intensive and may take a long time to complete.
**Error Rates**

**Power or Beta**
This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of nonequivalent means when in fact the means are equivalent.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

**Alpha (Significance Level)**
This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when you reject the null hypothesis of non-equivalent means when in fact the means are nonequivalent.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.

**Sample Size**

**N1 (Sample Size Group 1)**
Enter a value (or range of values) for the sample size of group 1. Note that these values are ignored when you are solving for N1. You may enter a range of values such as 10 to 100 by 10.

**N2 (Sample Size Group 2)**
Enter a value (or range of values) for the sample size of group 2 or enter Use R to base N2 on the value of N1. You may enter a range of values such as 10 to 100 by 10.

**Use R**
When Use R is entered here, N2 is calculated using the formula

\[ N2 = \lceil R(N1) \rceil \]

where R is the Sample Allocation Ratio and the operator \( \lceil \cdot \rceil \) is the first integer greater than or equal to Y. For example, if you want \( N1 = N2 \), select Use R and set \( R = 1 \).

**R (Sample Allocation Ratio)**
Enter a value (or range of values) for R, the allocation ratio between samples. This value is only used when N2 is set to Use R.

When used, N2 is calculated from N1 using the formula: \( N2 = \lceil R(N1) \rceil \) where \( \lceil \cdot \rceil \) is the next integer greater than or equal to Y. Note that setting \( R = 1.0 \) forces \( N2 = N1 \).
Effect Size

Group 1 (and 2) Distribution | H0
These options specify the distributions of the two groups under the null hypothesis, H0. The difference between the means of these two distributions is the value of one of the equivalence limits.

Group 1 is often called the reference (or standard) distribution. Group 2 is often called the treatment distribution. These options specify these two distributions under the null hypothesis, H0. The difference between the means of these two distributions is, by definition, one of the equivalence limits. Thus, you set the equivalence limit by specifying the two means.

The parameters of each distribution are specified using numbers or letters. If letters are used, their values are specified in the boxes below. The characters $M_0$ and $M_1$ are to be used for the means of the distributions of groups 1 and 2 under H0, respectively. An equivalence limit is then $M_0 - M_1$, which must be non-zero.

For example, suppose you entered $N(M_0 S)$ for group 1 and $N(M_1 S)$ for group 2. Also, you set $M_0$ equal to 5 and $M_1$ equal to 4. The upper (positive) equivalence limit would be $5 - 4 = 1$.

Following is a list of the distributions that are available and the syntax used to specify them. Note that, except for the multinomial, the distributions are parameterized so that the mean is entered first.

- Beta=$A(M_0,A,B,Minimum)$
- Binomial=$B(M_0,N)$
- Cauchy=$C(M_0,Scale)$
- Constant=$K(Value)$
- Exponential=$E(M_0)$
- F=$F(M_0,DF_1)$
- Gamma=$G(M_0,A)$
- Multinomial=$M(P_1,P_2,....,P_k)$
- Normal=$N(M_0,SD)$
- Poisson=$P(M_0)$
- Student's T=$T(M_0,D)$
- Tukey's Lambda=$L(M_0,S,Skewness,Elongation)$
- Uniform=$U(M_0,Minimum)$
- Weibull=$W(M_0,B)$

Details of writing mixture distributions, combined distributions, and compound distributions are found in the chapter on Data Simulation and will not be repeated here.

Finding the Value of the Mean of a Specified Distribution
The distributions have been parameterized in terms of their means, since this is the parameter being tested. The mean of a distribution created as a linear combination of other distributions is found by applying the linear combination to the individual means. However, the mean of a distribution created by multiplying or dividing other distributions is not necessarily equal to applying the same function to the individual means. For example, the mean of $4N(4, 5) + 2N(5, 6)$ is $4*4 + 2*5 = 26$, but the mean of $4N(4, 5) * 2N(5, 6)$ is not exactly $4*4*2*5 = 160$ (although it is close).
**Group 1 (and 2) Distribution|H1**

These options specify the distributions of the two groups under the alternative hypothesis, H1. The difference between the means of these two distributions is the difference that is assumed to be the true value of the difference.

Usually, the mean difference is specified by entering $M0$ for the mean parameter in the distribution expression for group 1 and $M1$ for the mean parameter in the distribution expression for group 2. The mean difference under H1 then becomes the value of $M0 - M0 = 0$. If you want a non-zero value, you specify it by specifying unequal values for the two distribution means. For example, you could enter $A$ for the mean of group 2. The mean difference will then be $M0 - A$.

The parameters of each distribution are specified using numbers or letters. If letters are used, their values are specified in the boxes below. The value $M1$ is reserved for the value of the mean of group 2 under the alternative hypothesis.

Following is a list of the distributions that are available and the syntax used to specify them. Note that, except for the multinomial, the distributions are parameterized so that the mean, $M1$, is entered first.

- Beta = A(M1,A,B,Minimum)
- Binomial = B(M1,N)
- Cauchy = C(M1,Scale)
- Constant = K(Value)
- Exponential = E(M1)
- F = F(M1,DF1)
- Gamma = G(M1,A)
- Multinomial = M(P1,P2,…,Pk)
- Normal = N(M1,SD)
- Poisson = P(M1)
- Student's T = T(M1,D)
- Tukey's Lambda = L(M1,S,Skewness,Elongation)
- Uniform = U(M1,Minimum)
- Weibull = W(M1,B)

Details of writing mixture distributions, combined distributions, and compound distributions are found in the chapter on Data Simulation and will not be repeated here.

**Effect Size – Distribution Parameters**

**M0 (Mean|H0)**

These values are substituted for $M0$ in the distribution specifications given above. $M0$ is intended to be the value of the mean hypothesized by the null hypothesis, H0.

You can enter a list of values using the syntax 0 1 2 3 or 0 to 3 by 1.

**M1 (Mean|H1)**

These values are substituted for $M1$ in the distribution specifications given above. Although it can be used wherever you want, $M1$ is intended to be the value of the mean hypothesized by the alternative hypothesis, H1.

You can enter a list of values using the syntax 0 1 2 3 or 0 to 3 by 1.
Parameter Values (S, A, B)
Enter the numeric value(s) of parameter listed above. These values are substituted for the corresponding letter in all four distribution specifications.

You can enter a list of values using the syntax ‘0 2 3’ or ‘0 to 3 by 1.’

You can also change the letter than is used as the name of this parameter.

Test and Simulations

Equivalence Limit

Equivalence limits are defined as the positive and negative limits around zero that define a zone of equivalence. This zone of equivalence is a set of difference values that define a region in which the two means are ‘close enough’ so that they are considered to be the same for practical purposes.

Rather than define these limits explicitly, they are set implicitly. This is done as follows. One limit is found by subtracting the Group 2 Dist’n|H0 mean from the Group 1 Dist’n|H0 mean. If the limits are symmetric, the other limit is this difference times -1. To obtain symmetric limits, enter ‘Symmetric’ here.

If asymmetric limits are desired, a numerical value is specified here. It is given the sign (+ or -) that is opposite the difference of the means discussed above.

For example, if the mean of group 1 under H0 is 5, the mean of group 2 under H0 is 4, and Symmetric is entered here, the equivalence limits will be 5 - 4 = 1 and -1. However, if the value 1.25 is entered here, the equivalence limits are 1 and -1.25.

If you do not have a specific value in mind for the equivalence limit, a common value for an equivalence limit is 20% or 25% of the group 1 (reference) mean.

Test Type

Specify which test statistic is to be used in the simulation. Although the t-test is the most commonly used test statistic, it is based on assumptions that may not be viable in many situations. For your data, you may find that one of the other tests are more accurate (actual alpha = target alpha) and more precise (better power).

Simulations

This option specifies the number of iterations, \( M \), used in the simulation. The larger the number of iterations, the longer the running time, and, the more accurate the results.

The precision of the simulated power estimates are calculated from the binomial distribution. Thus, confidence intervals may be constructed for various power values. The following table gives an estimate of the precision that is achieved for various simulation sizes when the power is either 0.50 or 0.95. The table values are interpreted as follows: a 95% confidence interval of the true power is given by the power reported by the simulation plus and minus the ‘Precision’ amount given in the table.
<table>
<thead>
<tr>
<th>Simulation Size M</th>
<th>Precision when Power = 0.50</th>
<th>Precision when Power = 0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.100</td>
<td>0.044</td>
</tr>
<tr>
<td>500</td>
<td>0.045</td>
<td>0.019</td>
</tr>
<tr>
<td>1000</td>
<td>0.032</td>
<td>0.014</td>
</tr>
<tr>
<td>2000</td>
<td>0.022</td>
<td>0.010</td>
</tr>
<tr>
<td>5000</td>
<td>0.014</td>
<td>0.006</td>
</tr>
<tr>
<td>10000</td>
<td>0.010</td>
<td>0.004</td>
</tr>
<tr>
<td>50000</td>
<td>0.004</td>
<td>0.002</td>
</tr>
<tr>
<td>100000</td>
<td>0.003</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Notice that a simulation size of 1000 gives a precision of plus or minus 0.01 when the true power is 0.95. Also note that as the simulation size is increased beyond 5000, there is only a small amount of additional accuracy achieved.

**Options Tab**

The Options tab contains limits on the number of iterations and various options about individual tests.

**Maximum Iterations**

**Maximum Iterations Before Search Termination**
Specify the maximum number of iterations before the search for the sample size, N1, is aborted. When the maximum number of iterations is reached without convergence, the sample size is left blank. We recommend a value of at least 500.

**Random Numbers**

**Random Number Pool Size**
This is the size of the pool of values from which the random samples will be drawn. Pools should be at least the maximum of 10,000 and twice the number of simulations. You can enter Automatic and an appropriate value will be calculated.

If you do not want to draw numbers from a pool, enter 0 here.

**Trimmed T-Test**

**Percent Trimmed at Each End**
Specify the percent of each end of the sorted data that is to be trimmed (constant G above) when using the trimmed means procedures. This percentage is applied to the sample size to determine how many of the lowest and highest data values are to be trimmed by the procedure. For example, if the sample size (N1) is 27 and you specify 10 here, then \([27 \times 10/100] = 2\) observations will be trimmed at the bottom and the top. For any percentage, at least one observation is trimmed from each end of the sorted dataset.

The range of possible values is 0 to 25.
Example 1 – Power at Various Sample Sizes

Researchers are planning an experiment to determine if the response to a new drug is equivalent to the response to the standard drug. The average response level to the standard drug is known to be 63 with a standard deviation of 5. The researchers decide that if the average response level to the new drug is between 60 and 66, they will consider it to be equivalent to the standard drug.

The researchers decide to use a parallel-group design. The response level for the standard drug will be measured for each subject. They will analyze the data using an equivalence test based on the t-test with an alpha level of 0.05. They want to compare the power at sample sizes of 10, 30, 50, and 70. They assume that the data are normally distributed and that the true difference between the mean response of the two drugs is zero. Since this is an exploratory analysis, the number of simulation iterations is set to 2000.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Equivalence Tests for Two Means (Simulation) procedure window by expanding Means, then Two Independent Means, then clicking on Equivalence, and then clicking on Equivalence Tests for Two Means (Simulation). You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>10 30 50 70</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Group 1 Distribution</td>
<td>H0</td>
</tr>
<tr>
<td>Group 2 Distribution</td>
<td>H0</td>
</tr>
<tr>
<td>Group 1 Distribution</td>
<td>H1</td>
</tr>
<tr>
<td>Group 2 Distribution</td>
<td>H1</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H1)</td>
</tr>
<tr>
<td>S</td>
<td>5</td>
</tr>
<tr>
<td>Equivalence Limit</td>
<td>Symmetric</td>
</tr>
<tr>
<td>Simulations</td>
<td>2000</td>
</tr>
<tr>
<td>Test Type</td>
<td>T-Test</td>
</tr>
</tbody>
</table>
**Annotated Output**

Click the Run button to perform the calculations and generate the following output.

**Numeric Results and Plots**

<table>
<thead>
<tr>
<th>Power</th>
<th>N1/N2</th>
<th>Diff1</th>
<th>Lower Equiv. Limit</th>
<th>Upper Equiv. Limit</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>Beta</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.009</td>
<td>10/10</td>
<td>0.0</td>
<td>-3.0</td>
<td>3.0</td>
<td>0.050</td>
<td>0.005</td>
<td>0.991</td>
<td>63.0</td>
<td>66.0</td>
<td>5.0</td>
</tr>
<tr>
<td>(0.004)</td>
<td>[0.005, 0.013]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.003)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.477</td>
<td>30/30</td>
<td>0.0</td>
<td>-3.0</td>
<td>3.0</td>
<td>0.050</td>
<td>0.053</td>
<td>0.524</td>
<td>63.0</td>
<td>66.0</td>
<td>5.0</td>
</tr>
<tr>
<td>(0.022)</td>
<td>[0.455, 0.498]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.010)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.816</td>
<td>50/50</td>
<td>0.0</td>
<td>-3.0</td>
<td>3.0</td>
<td>0.050</td>
<td>0.061</td>
<td>0.184</td>
<td>63.0</td>
<td>66.0</td>
<td>5.0</td>
</tr>
<tr>
<td>(0.017)</td>
<td>[0.799, 0.833]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.010)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.944</td>
<td>70/70</td>
<td>0.0</td>
<td>-3.0</td>
<td>3.0</td>
<td>0.050</td>
<td>0.056</td>
<td>0.056</td>
<td>63.0</td>
<td>66.0</td>
<td>5.0</td>
</tr>
<tr>
<td>(0.010)</td>
<td>[0.934, 0.954]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.040)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:

**Summary Statements**

Group sample sizes of 10 and 10 achieve 1% power to detect equivalence when the margin of equivalence is from -3.0 to 3.0 and the actual mean difference is 0.0. The significance level (alpha) is 0.050 using two one-sided T-Tests. These results are based on 2000 Monte Carlo samples from the null distributions: Normal(M0 S) and Normal(M1 S), and the alternative distributions: Normal(M0 S) and Normal(M0 S).

**Chart Section**

This report shows the estimated power for each scenario. The first row shows the parameter settings and the estimated power and significance level (Actual Alpha). The second row shows two 95% confidence intervals in brackets: the first for the power and the second for the significance level. Half the width of each confidence interval is given in parentheses as a fundamental measure of the accuracy of the simulation. As the number of simulations is increased, the width of the confidence intervals will decrease.
Example 2 – Finding the Sample Size

Continuing with Example 1, the researchers want to determine how large a sample is needed to obtain a power of 0.90.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Equivalence Tests for Two Means (Simulation) procedure window by expanding Means, then Two Independent Means, then clicking on Equivalence, and then clicking on Equivalence Tests for Two Means (Simulation). You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td>Find (Solve For) ...................... N1</td>
</tr>
<tr>
<td></td>
<td>Power ........................................ 0.90</td>
</tr>
<tr>
<td></td>
<td>Alpha ........................................... 0.05</td>
</tr>
<tr>
<td></td>
<td>N1 (Sample Size Group 1) ............... Ignored since this is the Find setting</td>
</tr>
<tr>
<td></td>
<td>N2 (Sample Size Group 2) .............. Use R</td>
</tr>
<tr>
<td></td>
<td>R (Sample Allocation Ratio) ........... 1.0</td>
</tr>
<tr>
<td></td>
<td>Group 1 Distribution</td>
</tr>
<tr>
<td></td>
<td>Group 2 Distribution</td>
</tr>
<tr>
<td></td>
<td>Group 1 Distribution</td>
</tr>
<tr>
<td></td>
<td>Group 2 Distribution</td>
</tr>
<tr>
<td></td>
<td>M0 (Mean</td>
</tr>
<tr>
<td></td>
<td>M1 (Mean</td>
</tr>
<tr>
<td></td>
<td>S ........................................ 5</td>
</tr>
<tr>
<td></td>
<td>Equivalence Limit ..................... Symmetric</td>
</tr>
<tr>
<td></td>
<td>Simulations ......................... 2000</td>
</tr>
<tr>
<td></td>
<td>Test Type .................................... T-Test</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N1/N2</th>
<th>H1 Lower Limit</th>
<th>H1 Upper Limit</th>
<th>Target Equiv Limit</th>
<th>Actual Alpha</th>
<th>Actual Beta</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.911</td>
<td>61/61</td>
<td>0.0</td>
<td>-3.0</td>
<td>3.0</td>
<td>0.045</td>
<td>0.089</td>
<td>63.0</td>
<td>66.0</td>
<td>5.0</td>
</tr>
</tbody>
</table>

The required sample size is 61 per group.
Example 3 – Comparative Results when the Data Contain Outliers

Continuing Example 1, this example will investigate the impact of outliers on the characteristics of the various test statistics. The two-sample t-test is known to be robust to the violation of some assumptions, but it is susceptible to inaccuracy when the data contains outliers. This example will investigate the impact of outliers on the power and precision of the five test statistics available in PASS.

A mixture of two normal distributions will be used to randomly generate outliers. The mixture will draw 95% of the data from a standard distribution. The other 5% of the data will come from a normal distribution with the same mean but with a standard deviation that is one, five, and ten times larger than that of the standard.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Equivalence Tests for Two Means (Simulation) procedure window by expanding Means, then Two Independent Means, then clicking on Equivalence, and then clicking on Equivalence Tests for Two Means (Simulation). You may then make the appropriate entries as listed below, or open Example 3 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Option</td>
<td>Power</td>
</tr>
<tr>
<td>Data Tab</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Data Tab</td>
<td>Power</td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power</td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power</td>
</tr>
<tr>
<td>Power</td>
<td>Power</td>
</tr>
<tr>
<td>Power</td>
<td>Power</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>40</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Group 1 Distribution</td>
<td>N(M0 S)[95];N(M0 A)[5]</td>
</tr>
<tr>
<td>Group 1 Distribution</td>
<td>N(M0 S)[95];N(M0 A)[5]</td>
</tr>
<tr>
<td>Group 1 Distribution</td>
<td>N(M0 S)[95];N(M0 A)[5]</td>
</tr>
<tr>
<td>Group 2 Distribution</td>
<td>N(M1 S)[95];N(M1 A)[5]</td>
</tr>
<tr>
<td>Group 2 Distribution</td>
<td>N(M1 S)[95];N(M1 A)[5]</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H1)</td>
</tr>
<tr>
<td>S</td>
<td>5</td>
</tr>
<tr>
<td>A</td>
<td>5 25 50</td>
</tr>
<tr>
<td>Equivalence Limit</td>
<td>Symmetric</td>
</tr>
<tr>
<td>Simulations</td>
<td>2000</td>
</tr>
<tr>
<td>Test Type</td>
<td>T-Test</td>
</tr>
</tbody>
</table>

Reports Tab

Show Comparative Reports | Checked |
Show Comparative Plots | Checked |
Output

Click the Run button to perform the calculations and generate the following output.

Power Comparison for Testing Equivalence. Hypotheses: H0: Diff>=|Diff0|; H1: Diff<|Diff0|
H0 Dist's: Normal(M0 S)[95]; Normal(M0 A)[5] & Normal(M1 S)[95]; Normal(M1 A)[5]
H1 Dist's: Normal(M0 S)[95]; Normal(M0 A)[5] & Normal(M0 S)[95]; Normal(M0 A)[5]

<table>
<thead>
<tr>
<th>H1 Diff</th>
<th>Lower Equiv. Limit</th>
<th>Upper Equiv. Limit</th>
<th>Target Alpha</th>
<th>T-Test Power</th>
<th>Welch T-Test Power</th>
<th>Trim. T-Test Power</th>
<th>Mann Whit'y Power</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>40/40</td>
<td>-3.0</td>
<td>3.0</td>
<td>0.050</td>
<td>0.708</td>
<td>0.657</td>
<td>0.656</td>
<td>0.543</td>
<td>0.539</td>
<td>63.0</td>
<td>66.0</td>
<td>5.0</td>
</tr>
<tr>
<td>40/40</td>
<td>-3.0</td>
<td>3.0</td>
<td>0.050</td>
<td>0.247</td>
<td>0.543</td>
<td>0.509</td>
<td>0.508</td>
<td>0.510</td>
<td>63.0</td>
<td>66.0</td>
<td>5.0</td>
</tr>
</tbody>
</table>

Alpha Comparison for Testing Equivalence. Hypotheses: H0: Diff>=|Diff0|; H1: Diff<|Diff0|

<table>
<thead>
<tr>
<th>H1 Diff</th>
<th>Lower Equiv. Limit</th>
<th>Upper Equiv. Limit</th>
<th>Target Alpha</th>
<th>T-Test Power</th>
<th>Welch T-Test Power</th>
<th>Trim. T-Test Power</th>
<th>Mann Whit'y Power</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>40/40</td>
<td>-3.0</td>
<td>3.0</td>
<td>0.050</td>
<td>0.050</td>
<td>0.050</td>
<td>0.058</td>
<td>0.056</td>
<td>63.0</td>
<td>66.0</td>
<td>5.0</td>
<td>5.0</td>
</tr>
<tr>
<td>40/40</td>
<td>-3.0</td>
<td>3.0</td>
<td>0.050</td>
<td>0.030</td>
<td>0.041</td>
<td>0.041</td>
<td>0.044</td>
<td>63.0</td>
<td>66.0</td>
<td>5.0</td>
<td>25.0</td>
</tr>
</tbody>
</table>


When A = 5, there are no outliers and the power of the nonparametric test and the trimmed tests are a little less than that of the t-test. When A = 25, the distortion of the t-test caused by the outliers becomes apparent. In this case, the powers of the standard t-test and Welch’s t-test are 0.247, but the powers of the nonparametric Mann-Whitney test and the trimmed tests are about 0.54. When A = 50, the standard t-test only achieves a power of 0.073, but the trimmed and nonparametric tests achieve powers of about 0.51!

Looking at the second table, we see that the true significance level of the t-test is distorted by the outliers, while the significance levels of the other tests remain close to the target value.
Example 4 – Selecting a Test Statistic when the Data Are Skewed

Continuing Example 3, this example will investigate the impact of skewness in the underlying distribution on the characteristics of the various test statistics.

Tukey’s lambda distribution will be used because it allows the amount of skewness to be gradually increased.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Equivalence Tests for Two Means (Simulation) procedure window by expanding Means, then Two Independent Means, then clicking on Equivalence, and then clicking on Equivalence Tests for Two Means (Simulation). You may then make the appropriate entries as listed below, or open Example 4 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>40</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Group 1 Distribution</td>
<td>H0 L(M0 S G 0)</td>
</tr>
<tr>
<td>Group 2 Distribution</td>
<td>H0 L(M1 S G 0)</td>
</tr>
<tr>
<td>Group 1 Distribution</td>
<td>H1 L(M0 S G 0)</td>
</tr>
<tr>
<td>Group 2 Distribution</td>
<td>H1 L(M0 S G 0)</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H1)</td>
</tr>
<tr>
<td>S</td>
<td>5</td>
</tr>
<tr>
<td>G</td>
<td>0.0 0.5 0.9</td>
</tr>
<tr>
<td>Equivalence Limit</td>
<td>Symmetric</td>
</tr>
<tr>
<td>Simulations</td>
<td>2000</td>
</tr>
<tr>
<td>Test Type</td>
<td>T-Test</td>
</tr>
</tbody>
</table>

Reports Tab

Show Comparative Reports .......... Checked
Show Comparative Plots ............ Checked

---

Tukey's lambda distribution will be used because it allows the amount of skewness to be gradually increased.
Output

Click the Run button to perform the calculations and generate the following output.

<table>
<thead>
<tr>
<th>H1 Lower Upper</th>
<th>Trim. Mann Power</th>
<th>Power</th>
<th>Power</th>
<th>Power</th>
<th>Power</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>N1/N2 (Diff1)</td>
<td>Equiv. Limit</td>
<td>Equiv. Limit</td>
<td>Target Alpha</td>
<td>T-Test Alpha</td>
<td>Trim. T-Test Power</td>
<td>Trim. Welch Power</td>
<td>Mann Whit'y Power</td>
<td></td>
<td></td>
</tr>
<tr>
<td>40/40</td>
<td>0.0</td>
<td>3.0</td>
<td>0.050</td>
<td>0.708</td>
<td>0.747</td>
<td>0.746</td>
<td>0.940</td>
<td>0.939</td>
<td>0.996</td>
</tr>
<tr>
<td>40/40</td>
<td>-3.0</td>
<td>3.0</td>
<td>0.050</td>
<td>0.708</td>
<td>0.773</td>
<td>0.772</td>
<td>0.893</td>
<td>0.835</td>
<td>0.996</td>
</tr>
<tr>
<td>40/40</td>
<td>0.0</td>
<td>3.0</td>
<td>0.050</td>
<td>0.747</td>
<td>0.940</td>
<td>0.939</td>
<td>0.996</td>
<td>63.0</td>
<td>66.0</td>
</tr>
</tbody>
</table>

We see that as the degree of skewness is increased, the power of the t-test increases slightly, but the powers of the trimmed and nonparametric tests improve dramatically. The significance levels do not appear to be adversely impacted.
Example 5 – Validation using Machin

Machin et al. (1997) page 107 present an example of determining the sample size for a parallel-group design in which the reference mean is 96, the treatment mean is 94, the standard deviation is 8, the limits are plus or minus 5, the power is 80%, and the significance level is 0.05. They calculate the sample size to be 88. It is important to note that Machin et al. use an approximation, so their results cannot be expected to exactly match those of PASS.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Equivalence Tests for Two Means (Simulation) procedure window by expanding Means, then Two Independent Means, then clicking on Equivalence, and then clicking on Equivalence Tests for Two Means (Simulation). You may then make the appropriate entries as listed below, or open Example 5 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N1</td>
</tr>
<tr>
<td>Power</td>
<td>0.80</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Group 1 Distribution</td>
<td>H0</td>
</tr>
<tr>
<td>Group 2 Distribution</td>
<td>H0</td>
</tr>
<tr>
<td>Group 1 Distribution</td>
<td>H1</td>
</tr>
<tr>
<td>Group 2 Distribution</td>
<td>H1</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H1)</td>
</tr>
<tr>
<td>S</td>
<td>8</td>
</tr>
<tr>
<td>Equivalence Limit</td>
<td>Symmetric</td>
</tr>
<tr>
<td>Simulations</td>
<td>2000</td>
</tr>
<tr>
<td>Test Type</td>
<td>T-Test</td>
</tr>
</tbody>
</table>
Output

Click the Run button to perform the calculations and generate the following output.

<table>
<thead>
<tr>
<th>Power</th>
<th>N1/N2</th>
<th>H1 Diff1</th>
<th>Lower Equiv. Limit</th>
<th>Upper Equiv. Limit</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>Beta</th>
<th>M0</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.807</td>
<td>87/87</td>
<td>2.0</td>
<td>-5.0</td>
<td>5.0</td>
<td>0.050</td>
<td>0.049</td>
<td>0.193</td>
<td>96.0</td>
<td>8.0</td>
</tr>
</tbody>
</table>

(0.017) [0.790 0.824] (0.009) [0.039 0.058]

Notes:

The sample size of 87 per group is reasonably close to the analytic answer of 88.
Chapter 470

Equivalence Tests for Two Means using Ratios

Introduction

This procedure calculates power and sample size of statistical tests for equivalence tests from parallel-group design with two groups. This routine deals with the case in which the statistical hypotheses are expressed in terms of mean ratios rather than mean differences.

The details of testing the equivalence of two treatments using a parallel-group design are given in the chapter entitled “Equivalence Tests for Two Means using Differences” and will not be repeated here. If the logarithms of the responses can be assumed to follow a normal distribution, hypotheses about equivalence in terms of the ratio can be transformed into hypotheses about the difference. The details of this analysis are given in Julious (2004).

Equivalence Testing Using Ratios

It will be convenient to adopt the following specialize notation for the discussion of these tests.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PASS Input/Output</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu_T )</td>
<td>Not used</td>
<td>Treatment mean. This is the treatment mean.</td>
</tr>
<tr>
<td>( \mu_R )</td>
<td>Not used</td>
<td>Reference mean. This is the mean of a reference population.</td>
</tr>
<tr>
<td>( \phi_L, \phi_U )</td>
<td>RL, RU</td>
<td>Margin of equivalence. These limits define an interval of the ratio of the means in which their difference is so small that it may be ignored.</td>
</tr>
<tr>
<td>( \phi )</td>
<td>R1</td>
<td>True ratio. This is the value of ( \phi = \mu_T / \mu_R ) at which the power is calculated.</td>
</tr>
</tbody>
</table>

Note that the actual values of \( \mu_T \) and \( \mu_R \) are not needed. Only their ratio is needed for power and sample size calculations.
The null hypothesis of non-equivalence is

\[ H_0: \phi \leq \phi_L \text{ or } \phi \geq \phi_U \text{ where } \phi_L < 1, \phi_U > 1. \]

and the alternative hypothesis of equivalence is

\[ H_1: \phi_L < \phi < \phi_U \]

**Log-Transformation**

In many cases, hypotheses stated in terms of ratios are more convenient than hypotheses stated in terms of differences. This is because ratios can be interpreted as scale-less percentages, but differences must be interpreted as actual amounts in their original scale. Hence, it has become a common practice to take the following steps in hypothesis testing.

1. State the statistical hypotheses in terms of ratios.
2. Transform these into hypotheses about differences by taking logarithms.
3. Analyze the logged data—that is, do the analysis in terms of the difference.
4. Draw the conclusion in terms of the ratio.

The details of step 2 for the null hypothesis are as follows.

\[ \phi_L \leq \phi \leq \phi_U \]

\[ \Rightarrow \phi_L \leq \left\{ \frac{\mu_L}{\mu_R} \right\} \leq \phi_U \]

\[ \Rightarrow \ln(\phi_L) \leq \left\{ \ln(\mu_L) - \ln(\mu_R) \right\} \leq \ln(\phi_U) \]

Thus, a hypothesis about the ratio of the means on the original scale can be translated into a hypothesis about the difference of two means on the logged scale.

When performing an equivalence test on the difference between means, the usual procedure is to set the equivalence limits symmetrically above and below zero. Thus the equivalence limits will be plus or minus an appropriate amount. The common practice is to do the same when the data are being analyzed on the log scale. However, when symmetric limits are set on the log scale, they do not translate to symmetric limits on the original scale. Instead, they translate to limits that are the inverses of each other.

Perhaps these concepts can best be understood by considering an example. Suppose the researchers have determined that the lower equivalence limit should be 80% on the original scale. Since they are planning to use a log scale for their analysis, they transform this limit to the log scale by taking the logarithm of 0.80. The result is -0.223144. Wanting symmetric limits, they set the upper equivalence limit to 0.223144. Exponentiating this value, they find that \(\exp(0.223144) = 1.25\). Note that \(1/(0.80) = 1.25\). Thus, the limits on the original scale are 80% and 125%, not 80% and 120%.

Using this procedure, appropriate equivalence limits for the ratio of two means can be easily determined. Here are a few sets of equivalence limits.
Equivalence Tests for Two Means using Ratios  470-3

<table>
<thead>
<tr>
<th>Specified Percent Change</th>
<th>Lower Limit Original Scale</th>
<th>Upper Limit Original Scale</th>
<th>Lower Limit Log Scale</th>
<th>Upper Limit Log Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>-25%</td>
<td>75.0%</td>
<td>133.3%</td>
<td>-0.287682</td>
<td>0.287682</td>
</tr>
<tr>
<td>+25%</td>
<td>80.0%</td>
<td>125.0%</td>
<td>-0.223144</td>
<td>0.223144</td>
</tr>
<tr>
<td>-20%</td>
<td>80.0%</td>
<td>125.0%</td>
<td>-0.223144</td>
<td>0.223144</td>
</tr>
<tr>
<td>+20%</td>
<td>83.3%</td>
<td>120.0%</td>
<td>-0.182322</td>
<td>0.182322</td>
</tr>
<tr>
<td>-10%</td>
<td>90.0%</td>
<td>111.1%</td>
<td>-0.105361</td>
<td>0.105361</td>
</tr>
<tr>
<td>+10%</td>
<td>90.9%</td>
<td>110.0%</td>
<td>-0.095310</td>
<td>0.095310</td>
</tr>
</tbody>
</table>

Note that negative percent-change values specify the lower limit first, while positive percent-change values specify the upper limit first. After the first limit is found, the other limit is calculated as its inverse.

**Coefficient of Variation**

The coefficient of variation (COV) is the ratio of the standard deviation to the mean. This parameter can be used to represent the variation in the data because of a unique relationship that it has in the case of log-normal data.

Suppose the variable $X$ is the logarithm of the original variable $Y$. That is, $X = \ln(Y)$ and $Y = \exp(X)$. Label the mean and variance of $X$ as $\mu_X$ and $\sigma_X^2$, respectively. Similarly, label the mean and variance of $Y$ as $\mu_Y$ and $\sigma_Y^2$, respectively. If $X$ is normally distributed, then $Y$ is log-normally distributed. Julious (2004) presents the following well-known relationships between these two variables

$$
\mu_Y = \left( e^{\mu_X + \frac{\sigma_X^2}{2}} \right)
$$

$$
\sigma_Y^2 = \mu_Y \left( e^{\sigma_X^2} - 1 \right)
$$

From this relationship, the coefficient of variation of $Y$ can be expressed as

$$
COV_Y = \frac{\sqrt{\mu_Y \left( e^{\sigma_X^2} - 1 \right)}}{\mu_Y} = \sqrt{e^{\sigma_X^2} - 1}
$$

Solving this relationship for $\sigma_X^2$, the standard deviation of $X$ can be stated in terms of the coefficient of variation of $Y$. This equation is

$$
\sigma_X = \sqrt{\ln(COV_Y^2 + 1)}
$$

Similarly, the mean of $X$ is

$$
\mu_X = \frac{\mu_Y}{\ln(COV_Y^2 + 1)}
$$
One final note: for parallel-group designs, $\sigma_X^2 = \sigma_d^2$, the average variance used in the t-test of the logged data.

Thus, the hypotheses can be stated in the original (Y) scale and then the power can be analyzed in the transformed (X) scale.

---

### Power Calculation

As is shown above, the hypotheses can be stated in the original (Y) scale using ratios or the logged (X) scale using differences. Either way, the power and sample size calculations are made using the formulas for testing the equivalence of the difference in two means. These formulas are presented another chapter and are not duplicated here.

---

### Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

#### Data Tab

The Data tab contains the parameters associated with this test such as the means, sample sizes, alpha, and power.

#### Solve For

**Find (Solve For)**

This option specifies the parameter to be solved for from the other parameters. Under most situations, you will select either Beta for a power analysis or $N1$ for sample size determination.

#### Error Rates

**Power or Beta**

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of nonequivalent means when in fact the means are equivalent.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as $0.8$ to $0.95$ by $0.05$ may be entered.
**Alpha (Significance Level)**
This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when you reject the null hypothesis of non-equivalent means when in fact the means are nonequivalent.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as \(0.01, 0.05, 0.10\) or \(0.01\) to \(0.10\) by \(0.01\).

You can enter a range of values such as \(0.05, 0.10, 0.15\) or \(0.05\) to \(0.15\) by \(0.01\).

---

**Sample Size**

**N1 (Sample Size Reference Group)**
Enter a value (or range of values) for the sample size of group 1 (the reference group). Note that these values are ignored when you are solving for N1. You may enter a range of values such as 10 to 100 by 10.

**N2 (Sample Size Treatment Group)**
Enter a value (or range of values) for the sample size of group 2 (the treatment group) or enter Use R to base N2 on the value of N1. You may enter a range of values such as 10 to 100 by 10.

- **Use R**
  
  When Use R is entered here, N2 is calculated using the formula:
  
  \[ N2 = \lceil R(N1) \rceil \]

  where \( R \) is the Sample Allocation Ratio and \( \lceil Y \rceil \) means take the first integer greater than or equal to \( Y \). For example, if you want \( N1 = N2 \), select Use R and set \( R = 1 \).

**R (Sample Allocation Ratio)**
Enter a value (or range of values) for \( R \), the allocation ratio between samples. This value is only used when N2 is set to Use R.

When used, N2 is calculated from N1 using the formula: \( N2 = \lceil R(N1) \rceil \) where \( \lceil Y \rceil \) is the next integer greater than or equal to \( Y \). Note that setting \( R = 1.0 \) forces \( N2 = N1 \).

---

**Effect Size – Equivalence Limits**

**RU (Upper Equivalence Limit)**
Enter the upper equivalence limit for the ratio of the two means. When the ratio of the means is between this value and RL, the two means are said to be equivalent. The value must be greater than one. A popular choice is 1.25. Note that this value is not a percentage.

If you enter \(1/RL\), then \(1/RL\) will be calculated and used here. This choice is commonly used because RL and \(1/RL\) give limits that are of equal magnitude on the log scale.
RL (Lower Equivalence Limit)
Enter the lower equivalence limit for the ratio of the two means. When the ratio of the means is between this value and RU, the two means are said to be equivalent. The value must be less than one. A popular choice is 0.80. Note that this value is not a percentage.

If you enter $1/RU$, then $1/RU$ will be calculated and used here. This choice is commonly used because RU and $1/RU$ give limits that are of equal magnitude on the log scale.

Effect Size – True Ratio

R1 (True Ratio)
This is the value of the ratio of the two means at which the power is to be calculated. Usually, the ratio will be assumed to be one. However, some authors recommend calculating the power using a ratio of 1.05 since this will require a larger sample size.

Effect Size – Coefficient of Variation

COV (Coefficient of Variation)
The coefficient of variation is used to specify the variability (standard deviation). It is important to realize that this is the COV defined on the original (not logged) scale. This value must be determined from past experience or from a pilot study. It is most easily calculated from the within mean-square error of the analysis of variance of the logged data using the relationship

$$COV_Y = \sqrt{e^{\sigma^2_d}} - 1.$$  

If prior studies used a t-test to analyze the logged data, you will not have a direct estimate of $\sigma^2_w$. However, the two variances, $\sigma^2_d$ and $\sigma^2_w$, are functionally related. The relationship between these quantities is $\sigma^2_d = 2\sigma^2_w$.  

Example 1 – Finding Power

A company has developed a generic drug for treating rheumatism and wants to show that it is equivalent to the standard drug. A parallel-group design will be used to test the equivalence of the two drugs.

Researchers have decided to set the lower limit of equivalence at 0.80. Past experience leads the researchers to set the COV to 1.50. The significance level is 0.05. The power will be computed assuming that the true ratio is either 1.00 or 1.05. Sample sizes between 50 and 550 will be included in the analysis.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Equivalence Tests for Two Means [Ratios] procedure window by expanding Means, then Two Independent Means, then clicking on Equivalence, and then clicking on Equivalence Tests for Two Means [Ratios]. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Find (Solve For)..............</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power.........................</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha.........................</td>
<td>0.05</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1).....</td>
<td>50 to 550 by 100</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2).....</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1</td>
</tr>
<tr>
<td>RU (Upper Equivalence Limit)</td>
<td>1/RL</td>
</tr>
<tr>
<td>RL (Lower Equivalence Limit)</td>
<td>0.80</td>
</tr>
<tr>
<td>R1 (True Ratio)................</td>
<td>1.0 1.05</td>
</tr>
<tr>
<td>COV (Coefficient of Variation)</td>
<td>1.50</td>
</tr>
</tbody>
</table>
Annotated Output

Click the Run button to perform the calculations and generate the following output.

### Numeric Results for Testing Equivalence Using a Parallel-Group Design

<table>
<thead>
<tr>
<th>Reference Group Sample Size (N1)</th>
<th>Treatment Group Sample Size (N2)</th>
<th>Lower Equiv. Limit (RL)</th>
<th>Upper Equiv. Limit (RU)</th>
<th>True Ratio (R1)</th>
<th>Coefficient of Variation (COV)</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0000</td>
<td>50</td>
<td>0.80</td>
<td>0.80</td>
<td>1.00</td>
<td>1.50</td>
<td>0.0500</td>
<td>1.0000</td>
</tr>
<tr>
<td>0.1049</td>
<td>150</td>
<td>0.80</td>
<td>0.80</td>
<td>1.00</td>
<td>1.50</td>
<td>0.0500</td>
<td>0.8951</td>
</tr>
<tr>
<td>0.4843</td>
<td>250</td>
<td>0.80</td>
<td>0.80</td>
<td>1.00</td>
<td>1.50</td>
<td>0.0500</td>
<td>0.5157</td>
</tr>
<tr>
<td>0.7170</td>
<td>350</td>
<td>0.80</td>
<td>0.80</td>
<td>1.00</td>
<td>1.50</td>
<td>0.0500</td>
<td>0.2830</td>
</tr>
<tr>
<td>0.8494</td>
<td>450</td>
<td>0.80</td>
<td>0.80</td>
<td>1.00</td>
<td>1.50</td>
<td>0.0500</td>
<td>0.1506</td>
</tr>
<tr>
<td>0.9221</td>
<td>550</td>
<td>0.80</td>
<td>0.80</td>
<td>1.00</td>
<td>1.50</td>
<td>0.0500</td>
<td>0.0779</td>
</tr>
<tr>
<td>0.0000</td>
<td>50</td>
<td>0.80</td>
<td>0.80</td>
<td>1.05</td>
<td>1.50</td>
<td>0.0500</td>
<td>1.0000</td>
</tr>
<tr>
<td>0.1010</td>
<td>150</td>
<td>0.80</td>
<td>0.80</td>
<td>1.05</td>
<td>1.50</td>
<td>0.0500</td>
<td>0.9890</td>
</tr>
<tr>
<td>0.4360</td>
<td>250</td>
<td>0.80</td>
<td>0.80</td>
<td>1.05</td>
<td>1.50</td>
<td>0.0500</td>
<td>0.5640</td>
</tr>
<tr>
<td>0.6366</td>
<td>350</td>
<td>0.80</td>
<td>0.80</td>
<td>1.05</td>
<td>1.50</td>
<td>0.0500</td>
<td>0.3634</td>
</tr>
<tr>
<td>0.7602</td>
<td>450</td>
<td>0.80</td>
<td>0.80</td>
<td>1.05</td>
<td>1.50</td>
<td>0.0500</td>
<td>0.2398</td>
</tr>
<tr>
<td>0.8396</td>
<td>550</td>
<td>0.80</td>
<td>0.80</td>
<td>1.05</td>
<td>1.50</td>
<td>0.0500</td>
<td>0.1604</td>
</tr>
</tbody>
</table>

**Report Definitions**

Power is the probability of rejecting non-equivalence when they are equivalent.
N1 is the number of subjects in the first group.
N2 is the number of subjects in the second group.
RU & RL are the maximum allowable ratios that result in equivalence.
R1 is the ratio of the means at which the power is computed.
COV is the coefficient of variation on the original scale.
Alpha is the probability of accepting non-equivalence when the means are equivalent.
Beta is the probability of rejecting non-equivalence when the means are non-equivalent.

**Summary Statements**

An equivalence test of means using two one-sided tests on data from a parallel-group design with sample sizes of 50 in the reference group and 50 in the treatment group achieves 0% power at a 5% significance level when the true ratio of the means is 1.00, the coefficient of variation on the original, unlogged scale is 1.50, and the equivalence limits of the mean ratio are 0.80 and 1.25.

This report shows the power for the indicated scenarios.

**Plot Section**

This plot shows the power versus the sample size.
Example 2 – Validation using Julious

Julious (2004) page 1971 presents an example of determining the sample size for a parallel-group design in which the actual ratio is 1.0, the coefficient of variation is 0.80, the equivalence limits are 0.80 and 1.25, the power is 90%, and the significance level is 0.05. He calculates the per group sample size to be 216.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Equivalence Tests for Two Means [Ratios] procedure window by expanding Means, then Two Independent Means, then clicking on Equivalence, and then clicking on Equivalence Tests for Two Means [Ratios]. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N1</td>
</tr>
<tr>
<td>Power</td>
<td>0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1</td>
</tr>
<tr>
<td>RU (Upper Equivalence Limit)</td>
<td>1/RL</td>
</tr>
<tr>
<td>RL (Lower Equivalence Limit)</td>
<td>0.80</td>
</tr>
<tr>
<td>R1 (True Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>COV (Coefficient of Variation)</td>
<td>0.80</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Reference Group Sample Size (N1)</th>
<th>Treatment Group Sample Size (N2)</th>
<th>Lower Equiv. Limit (RL)</th>
<th>Upper Equiv. Limit (RU)</th>
<th>True Ratio (R1)</th>
<th>Coefficient of Variation (COV)</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power 0.9004</td>
<td>216</td>
<td>0.80</td>
<td>1.25</td>
<td>1.00</td>
<td>0.80</td>
<td>0.0500</td>
<td>0.0996</td>
</tr>
</tbody>
</table>

PASS has also calculated the per group sample size to be 216, which matches Julious’s result.
470-10  Equivalence Tests for Two Means using Ratios
Chapter 471

Confidence Intervals for the Difference Between Two Means

Introduction

This procedure calculates the sample size necessary to achieve a specified distance from the difference in sample means to the confidence limit(s) at a stated confidence level for a confidence interval about the difference in means when the underlying data distribution is normal.

Caution: This procedure assumes that the standard deviations of the future samples will be the same as the standard deviations that are specified. If the standard deviation to be used in the procedure is estimated from a previous sample or represents the population standard deviation, the Confidence Intervals for the Difference between Two Means with Tolerance Probability procedure should be considered. That procedure controls the probability that the distance from the difference in means to the confidence limits will be less than or equal to the value specified.

Technical Details

There are two formulas for calculating a confidence interval for the difference between two population means. The different formulas are based on whether the standard deviations are assumed to be equal or unequal.

For each of the cases below, let the means of the two populations be represented by $\mu_1$ and $\mu_2$, and let the standard deviations of the two populations be represented as $\sigma_1$ and $\sigma_2$. 
Case 1 – Standard Deviations Assumed Equal

When \( \sigma_1 = \sigma_2 = \sigma \) are unknown, the appropriate two-sided confidence interval for \( \mu_1 - \mu_2 \) is

\[
X_1 - X_2 \pm t_{1-\alpha/2, n_1+n_2-2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}
\]

where

\[
s_p = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}}
\]

Upper and lower one-sided confidence intervals can be obtained by replacing \( \alpha/2 \) with \( \alpha \).

The required sample size for a given precision, D, can be found by solving the following equation iteratively

\[
D = t_{1-\alpha/2, n_1+n_2-2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}
\]

This equation can be used to solve for D or \( n_1 \) or \( n_2 \) based on the values of the remaining parameters.

Case 2 – Standard Deviations Assumed Unequal

When \( \sigma_1 \neq \sigma_2 \) are unknown, the appropriate two-sided confidence interval for \( \mu_1 - \mu_2 \) is

\[
X_1 - X_2 \pm t_{1-\alpha/2, \nu} \sqrt{\frac{s_1^2 + s_2^2}{n_1}}
\]

where

\[
\nu = \frac{\left(\frac{s_1^2 + s_2^2}{n_1} \right)^2}{\frac{s_1^4}{n_1^2(n_1 - 1)} + \frac{s_2^4}{n_2^2(n_2 - 1)}}
\]

In this case t is an approximate t and the method is known as the Welch-Satterthwaite method.

Upper and lower one-sided confidence intervals can be obtained by replacing \( \alpha/2 \) with \( \alpha \).

The required sample size for a given precision, D, can be found by solving the following equation iteratively

\[
D = t_{1-\alpha/2, \nu} \sqrt{\frac{s_1^2 + s_2^2}{n_1}}
\]

This equation can be used to solve for D or \( n_1 \) or \( n_2 \) based on the values of the remaining parameters.
Confidence Intervals for the Difference Between Two Means

**Confidence Level**

The confidence level, \( 1 - \alpha \), has the following interpretation. If thousands of samples of \( n_1 \) and \( n_2 \) items are drawn from populations using simple random sampling and a confidence interval is calculated for each sample, the proportion of those intervals that will include the true population mean difference is \( 1 - \alpha \).

Notice that is a long term statement about many, many samples.

**Procedure Options**

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

**Data Tab**

The Data tab contains most of the parameters and options that you will be concerned with.

**Solve For**

**Find (Solve For)**

This option specifies the parameter to be solved for from the other parameters.

**Confidence**

**Confidence Level (1 – Alpha)**

The confidence level, \( 1 - \alpha \), has the following interpretation. If thousands of samples of \( n_1 \) and \( n_2 \) items are drawn from populations using simple random sampling and a confidence interval is calculated for each sample, the proportion of those intervals that will include the true population mean difference is \( 1 - \alpha \).

Often, the values 0.95 or 0.99 are used. You can enter single values or a range of values such as 0.90, 0.95 or 0.90 to 0.99 by 0.01.

**Sample Size**

**N1 (Sample Size Group 1)**

Enter a value (or range of values) for the sample size of this group. Note that these values are ignored when you are solving for \( N1 \). You may enter a range of values such as 10 to 100 by 10.

**N2 (Sample Size Group 2)**

Enter a value (or range of values) for the sample size of group 2 or enter Use R to base \( N2 \) on the value of \( N1 \). You may enter a range of values such as 10 to 100 by 10.

- **Use R**

  When \( \text{Use R} \) is entered here, \( N2 \) is calculated using the formula

  \[ N2 = [R(N1)] \]
where $R$ is the Sample Allocation Ratio and the operator $[Y]$ is the first integer greater than or equal to $Y$. For example, if you want $N_1 = N_2$, select Use R and set $R = 1$.

**R (Sample Allocation Ratio)**

Enter a value (or range of values) for $R$, the allocation ratio between samples. This value is only used when $N_2$ is set to Use $R$.

When used, $N_2$ is calculated from $N_1$ using the formula: $N_2 = [R(N_1)]$ where $[Y]$ is the next integer greater than or equal to $Y$. Note that setting $R = 1.0$ forces $N_2 = N_1$.

---

**One-Sided or Two-Sided Interval**

**Interval Type**

Specify whether the interval to be used will be a one-sided or a two-sided confidence interval.

---

**Precision**

**Distance from Mean Difference to Limit(s)**

This is the distance from the confidence limit(s) to the difference in means. For two-sided intervals, it is also known as the precision, half-width, or margin of error.

You can enter a single value or a list of values. The value(s) must be greater than zero.

---

**Standard Deviations**

**S1 and S2 (Standard Deviations)**

Enter an estimate of the standard deviation of group 1 or 2. The standard deviation must be a positive number.

Caution: The sample size estimates for this procedure assume that the standard deviation that is achieved when the confidence interval is produced is the same as the standard deviation entered here.

Press the 'Standard Deviation Estimator' button to obtain help on estimating the standard deviation.

You can enter a range of values such as 1, 2, 3 or 1 to 10 by 1.

---

**Standard Deviation Equality Assumption**

Specify whether the standard deviations are assumed to be the same or different. The choice will determine which of the two common confidence interval formulas for estimating the difference in population means will be used.

- **Assume S1 and S2 are Unequal**
  
  When the standard deviations are assumed to be unequal, the variances are not pooled and an approximate method is used for the confidence interval formula. This approximate method is sometimes called the Welch-Satterthwaite method.

- **Assume S1 and S2 are Equal**
  
  When the standard deviations are assumed to be equal, the pooled variance formula is used in the calculation of the confidence interval. The degrees of freedom are $N_1 + N_2 – 2$. 
Recommendation: Because the standard deviations of two populations are rarely equal, it is recommended that the standard deviations are assumed to be unequal. The Welch-Satterthwaite confidence interval calculation is generally accepted and commonly used.

**Iterations Tab**

This tab sets an option used in the iterative procedures.

**Maximum Iterations**

**Maximum Iterations Before Search Termination**

Specify the maximum number of iterations allowed before the search for the criterion of interest is aborted. When the maximum number of iterations is reached without convergence, the criterion is left blank. A value of 500 is recommended.

**Example 1 – Calculating Sample Size**

Suppose a study is planned in which the researcher wishes to construct a two-sided 95% confidence interval for the difference between two population means such that the width of the interval is no wider than 20 units. The confidence level is set at 0.95, but 0.99 is included for comparative purposes. The standard deviation estimates, based on the range of data values, are 32 for Population 1 and 38 for Population 2. Instead of examining only the interval half-width of 10, a series of half-widths from 5 to 15 will also be considered.

The goal is to determine the necessary sample size for each group.

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Confidence Intervals for the Difference Between Two Means procedure window by expanding Means, then Two Independent Means, then clicking on Confidence Interval, and then clicking on Confidence Intervals for the Difference Between Two Means. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N1</td>
</tr>
<tr>
<td>Confidence Level</td>
<td>0.95 0.99</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Distance from Mean Diff to Limit(s)</td>
<td>5 to 15 by 1</td>
</tr>
<tr>
<td>S1 (Standard Deviation Group 1)</td>
<td>32</td>
</tr>
<tr>
<td>S2 (Standard Deviation Group 2)</td>
<td>38</td>
</tr>
<tr>
<td>SD Equality Assumption</td>
<td>Assume S1 and S2 are Unequal</td>
</tr>
<tr>
<td>Interval Type</td>
<td>Two-Sided</td>
</tr>
</tbody>
</table>
471-6 Confidence Intervals for the Difference Between Two Means

Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>N1</th>
<th>N2</th>
<th>Allocation Ratio</th>
<th>Target Dist from Mean Diff to Limits</th>
<th>Actual Dist from Mean Diff to Limits</th>
<th>S1</th>
<th>S2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.95</td>
<td>380</td>
<td>380</td>
<td>1.00</td>
<td>5.000</td>
<td>4.995</td>
<td>32.00</td>
<td>38.00</td>
</tr>
<tr>
<td>0.95</td>
<td>265</td>
<td>265</td>
<td>1.00</td>
<td>6.000</td>
<td>5.995</td>
<td>32.00</td>
<td>38.00</td>
</tr>
<tr>
<td>0.95</td>
<td>195</td>
<td>195</td>
<td>1.00</td>
<td>7.000</td>
<td>6.995</td>
<td>32.00</td>
<td>38.00</td>
</tr>
<tr>
<td>0.95</td>
<td>150</td>
<td>150</td>
<td>1.00</td>
<td>8.000</td>
<td>7.984</td>
<td>32.00</td>
<td>38.00</td>
</tr>
<tr>
<td>0.95</td>
<td>119</td>
<td>119</td>
<td>1.00</td>
<td>9.000</td>
<td>8.973</td>
<td>32.00</td>
<td>38.00</td>
</tr>
<tr>
<td>0.95</td>
<td>97</td>
<td>97</td>
<td>1.00</td>
<td>10.000</td>
<td>9.951</td>
<td>32.00</td>
<td>38.00</td>
</tr>
<tr>
<td>0.95</td>
<td>80</td>
<td>80</td>
<td>1.00</td>
<td>11.000</td>
<td>10.973</td>
<td>32.00</td>
<td>38.00</td>
</tr>
<tr>
<td>0.95</td>
<td>68</td>
<td>68</td>
<td>1.00</td>
<td>12.000</td>
<td>11.918</td>
<td>32.00</td>
<td>38.00</td>
</tr>
<tr>
<td>0.95</td>
<td>58</td>
<td>58</td>
<td>1.00</td>
<td>13.000</td>
<td>12.926</td>
<td>32.00</td>
<td>38.00</td>
</tr>
<tr>
<td>0.95</td>
<td>50</td>
<td>50</td>
<td>1.00</td>
<td>14.000</td>
<td>13.947</td>
<td>32.00</td>
<td>38.00</td>
</tr>
<tr>
<td>0.95</td>
<td>44</td>
<td>44</td>
<td>1.00</td>
<td>15.000</td>
<td>14.895</td>
<td>32.00</td>
<td>38.00</td>
</tr>
<tr>
<td>0.99</td>
<td>655</td>
<td>655</td>
<td>1.00</td>
<td>5.000</td>
<td>5.000</td>
<td>32.00</td>
<td>38.00</td>
</tr>
<tr>
<td>0.99</td>
<td>455</td>
<td>455</td>
<td>1.00</td>
<td>6.000</td>
<td>5.999</td>
<td>32.00</td>
<td>38.00</td>
</tr>
<tr>
<td>0.99</td>
<td>335</td>
<td>335</td>
<td>1.00</td>
<td>7.000</td>
<td>6.991</td>
<td>32.00</td>
<td>38.00</td>
</tr>
<tr>
<td>0.99</td>
<td>258</td>
<td>258</td>
<td>1.00</td>
<td>8.000</td>
<td>7.997</td>
<td>32.00</td>
<td>38.00</td>
</tr>
<tr>
<td>0.99</td>
<td>205</td>
<td>205</td>
<td>1.00</td>
<td>9.000</td>
<td>8.981</td>
<td>32.00</td>
<td>38.00</td>
</tr>
<tr>
<td>0.99</td>
<td>166</td>
<td>166</td>
<td>1.00</td>
<td>10.000</td>
<td>9.991</td>
<td>32.00</td>
<td>38.00</td>
</tr>
<tr>
<td>0.99</td>
<td>138</td>
<td>138</td>
<td>1.00</td>
<td>11.000</td>
<td>10.972</td>
<td>32.00</td>
<td>38.00</td>
</tr>
<tr>
<td>0.99</td>
<td>116</td>
<td>116</td>
<td>1.00</td>
<td>12.000</td>
<td>11.983</td>
<td>32.00</td>
<td>38.00</td>
</tr>
<tr>
<td>0.99</td>
<td>99</td>
<td>99</td>
<td>1.00</td>
<td>13.000</td>
<td>12.991</td>
<td>32.00</td>
<td>38.00</td>
</tr>
<tr>
<td>0.99</td>
<td>86</td>
<td>86</td>
<td>1.00</td>
<td>14.000</td>
<td>13.960</td>
<td>32.00</td>
<td>38.00</td>
</tr>
<tr>
<td>0.99</td>
<td>75</td>
<td>75</td>
<td>1.00</td>
<td>15.000</td>
<td>14.975</td>
<td>32.00</td>
<td>38.00</td>
</tr>
</tbody>
</table>

References


Report Definitions

- Confidence level is the proportion of confidence intervals (constructed with this same confidence level, sample size, etc.) that would contain the true difference in population means.
- N1 and N2 are the sample sizes drawn from the two populations.
- Allocation Ratio is the ratio of the sample sizes, N2/N1.
- Dist from Mean Diff to Limit is the distance from the confidence limit(s) to the difference in sample means.
  - For two-sided intervals, it is also know as the precision, half-width, or margin of error.
- Target Dist from Mean Diff to Limit is the value of the distance that is entered into the procedure.
- Actual Dist from Mean Diff to Limit is the value of the distance that is obtained from the procedure.
- S1 and S2 are the standard deviations upon which the distance from mean difference to limit calculations are based.

Summary Statements

Group sample sizes of 380 and 380 produce a two-sided 95% confidence interval with a distance from the difference in means to the limits that is equal to 4.995 when the estimated standard deviations are 32.00 and 38.00.

This report shows the calculated sample size for each of the scenarios.
Plots Section

This plot shows the sample size of each group versus the precision for the two confidence levels.

Example 2 – Validation using Ostle and Malone

Ostle and Malone (1988) page 150 give an example of a precision calculation for a confidence interval for the difference between two means when the confidence level is 95%, the two standard deviations are 6.2185 and 16.06767, and the sample sizes are 7 and 6. The precision is 13.433 (when df = 6.257, not 6).

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Confidence Intervals for the Difference Between Two Means procedure window by expanding Means, then Two Independent Means, then clicking on Confidence Interval, and then clicking on Confidence Intervals for the Difference Between Two Means. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Find (Solve For)</td>
<td>N1</td>
</tr>
<tr>
<td>Confidence Level</td>
<td>0.90</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>6</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>Ignored</td>
</tr>
<tr>
<td>Distance from Mean Diff to Limit(s)</td>
<td>13.433</td>
</tr>
<tr>
<td>S1 (Standard Deviation Group 1)</td>
<td>6.2185</td>
</tr>
<tr>
<td>S2 (Standard Deviation Group 2)</td>
<td>16.06767</td>
</tr>
<tr>
<td>SD Equality Assumption</td>
<td>Assume S1 and S2 are Unequal</td>
</tr>
<tr>
<td>Interval Type</td>
<td>Two-Sided</td>
</tr>
</tbody>
</table>
Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>N1</th>
<th>N2</th>
<th>Allocation Ratio</th>
<th>Target Dist from Mean Diff to Limits</th>
<th>Actual Dist from Mean Diff to Limits</th>
<th>S1</th>
<th>S2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.90</td>
<td>7</td>
<td>6</td>
<td>0.857</td>
<td>13.433</td>
<td>13.433</td>
<td>6.22</td>
<td>16.07</td>
</tr>
</tbody>
</table>

PASS also calculated the sample size in Group 1 to be 7.

Example 3 – Validation using Zar

Zar (1984) page 132 gives an example of a precision calculation for a confidence interval for the difference between two means when the confidence level is 95%, the pooled standard deviation estimate is 0.7206, and the sample sizes are 6 and 7. The precision is 0.88.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Confidence Intervals for the Difference Between Two Means procedure window by expanding Means, then Two Independent Means, then clicking on Confidence Interval, and then clicking on Confidence Intervals for the Difference Between Two Means. You may then make the appropriate entries as listed below, or open Example 3 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Distance from Mean Difference to Limit</td>
</tr>
<tr>
<td>Confidence Level</td>
<td>0.95</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>6</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>7</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>Ignored</td>
</tr>
<tr>
<td>Distance from Mean Diff to Limit(s)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>S1 (Standard Deviation Group 1)</td>
<td>0.7206</td>
</tr>
<tr>
<td>S2 (Standard Deviation Group 2)</td>
<td>S1</td>
</tr>
<tr>
<td>SD Equality Assumption</td>
<td>Assume S1 and S2 are Equal</td>
</tr>
<tr>
<td>Interval Type</td>
<td>Two-Sided</td>
</tr>
</tbody>
</table>
Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>N1</th>
<th>N2</th>
<th>Allocation Ratio</th>
<th>Target Dist from Mean Diff to Limits</th>
<th>Actual Dist from Mean Diff to Limits</th>
<th>S1</th>
<th>S2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.95</td>
<td>6</td>
<td>7</td>
<td>1.167</td>
<td></td>
<td>0.882</td>
<td>0.72</td>
<td>0.72</td>
</tr>
</tbody>
</table>

*PASS* also calculated the precision to be 0.88.
Confidence Intervals for the Difference Between Two Means
Chapter 472

Confidence Intervals for the Difference Between Two Means with Tolerance Probability

Introduction

This procedure calculates the sample size necessary to achieve a specified distance from the difference in sample means to the confidence limit(s) with a given tolerance probability at a stated confidence level for a confidence interval about the difference in means when the underlying data distribution is normal.

Sample sizes are calculated only for the case where the standard deviations are assumed to be equal, wherein the pooled standard deviation formula is used.

Technical Details

Let the means of the two populations be represented by $\mu_1$ and $\mu_2$, and let the standard deviations of the two populations be represented as $\sigma_1$ and $\sigma_2$.

When $\sigma_1 = \sigma_2 = \sigma$ are unknown, the appropriate two-sided confidence interval for $\mu_1 - \mu_2$ is

$$
\bar{X}_1 - \bar{X}_2 \pm t_{1-\alpha/2, n_1 + n_2 - 2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}
$$
where

\[ s_p = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}} \]

Upper and lower one-sided confidence intervals can be obtained by replacing \( \alpha / 2 \) with \( \alpha \).

The required sample size for a given precision, \( D \), can be found by solving the following equation iteratively

\[ D = t_{1-\alpha/2,n_1+n_2-2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \]

This equation can be used to solve for \( D \) or \( n_1 \) or \( n_2 \) based on the values of the remaining parameters.

There is an additional subtlety that arises when the standard deviation is to be chosen for estimating sample size. The sample sizes determined from the formula above produce confidence intervals with the specified widths only when the future samples have a pooled standard deviation that is no greater than the value specified.

As an example, suppose that 15 individuals are sampled from each population in a pilot study, and a pooled standard deviation estimate of 5.4 is obtained from the sample. The purpose of a later study is to estimate the difference in means within 10 units. Suppose further that the sample size needed is calculated to be 62 per group using the formula above with 5.4 as the estimate for the pooled standard deviation. The samples of size 62 are then obtained from each population, but the pooled standard deviation turns out to be 6.3 rather than 5.4. The confidence interval is computed and the distance from the difference in means to the confidence limits is greater than 10 units.

This example illustrates the need for an adjustment to adjust the sample size such that the distance from the difference in means to the confidence limits will be below the specified value with known probability.

Such an adjustment for situations where a previous sample is used to estimate the standard deviation is derived by Harris, Horvitz, and Mood (1948) and discussed in Zar (1984). The adjustment is

\[ D = t_{1-\gamma/2,m_1,n_2-2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \sqrt{F_{1-\gamma,m_1,n_2-2}} \]

where \( 1 - \gamma \) is the probability that the distance from the difference in means to the confidence limit(s) will be below the specified value, and \( m_1 \) and \( m_2 \) are the sample sizes in the previous samples that were used to estimate the pooled standard deviation.

The corresponding adjustment when no previous sample is available is discussed in Kupper and Hafner (1989). The adjustment in this case is

\[ D = t_{1-\alpha/2,n_1+n_2-2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} \sqrt{\frac{\chi^2_{1-\gamma,n_1+n_2-2}}{n_1 + n_2 - 2}} \]
where, again, $1 - \gamma$ is the probability that the distance from the difference in means to the confidence limit(s) will be below the specified value.

Each of these adjustments accounts for the variability in a future estimate of the pooled standard deviation. In the first adjustment formula (Harris, Horvitz, and Mood, 1948), the distribution of the pooled standard deviation is based on the estimate from previous samples. In the second adjustment formula, the distribution of the pooled standard deviation is based on a specified value that is assumed to be the population pooled standard deviation.

### Confidence Level

The confidence level, $1 - \alpha$, has the following interpretation. If thousands of samples of $n_1$ and $n_2$ items are drawn from populations using simple random sampling and a confidence interval is calculated for each sample, the proportion of those intervals that will include the true population mean difference is $1 - \alpha$.

Notice that is a long term statement about many, many samples.

### Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

### Data Tab

The Data tab contains most of the parameters and options that you will be concerned with.

**Solve For**

**Find (Solve For)**

This option specifies the parameter to be solved for from the other parameters.

**Confidence and Tolerance**

**Confidence Level (1 – Alpha)**

The confidence level, $1 - \alpha$, has the following interpretation. If thousands of samples of $n_1$ and $n_2$ items are drawn from populations using simple random sampling and a confidence interval is calculated for each sample, the proportion of those intervals that will include the true population mean difference is $1 - \alpha$.

Often, the values 0.95 or 0.99 are used. You can enter single values or a range of values such as 0.90 to 0.99 by 0.01.

**Tolerance Probability**

This is the probability that a future interval with sample sizes N1 and N2 and the specified confidence level will have a distance from the difference in means to the limit(s) that is less than or equal to the distance specified.
If a tolerance probability is not used, as in the 'Confidence Intervals for the Difference between Two Means' procedure, the sample size is calculated for the expected distance from the difference in means to the limit(s), which assumes that the future standard deviation will also be the one specified.

Using a tolerance probability implies that the standard deviation of the future sample will not be known in advance, and therefore, an adjustment is made to the sample size formula to account for the variability in the standard deviation. Use of a tolerance probability is similar to using an upper bound for the standard deviation in the 'Confidence Intervals for the Difference between Two Means' procedure.

The range of values that can be entered here is values between 0 and 1.

You can enter a range of values such as .70 .80 .90 or .70 to .95 by .05.

---

**Sample Size**

**N1 (Sample Size Group 1)**

Enter a value (or range of values) for the sample size of this group. Note that these values are ignored when you are solving for \( N1 \). You may enter a range of values such as 10 to 100 by 10.

**N2 (Sample Size Group 2)**

Enter a value (or range of values) for the sample size of group 2 or enter Use R to base \( N2 \) on the value of \( N1 \). You may enter a range of values such as 10 to 100 by 10.

- **Use R**
  
  When Use R is entered here, \( N2 \) is calculated using the formula
  
  \[ N2 = \lceil R(N1) \rceil \]
  
  where \( R \) is the Sample Allocation Ratio and the operator \( \lceil Y \rceil \) is the first integer greater than or equal to \( Y \). For example, if you want \( N1 = N2 \), select Use R and set \( R = 1 \).

**R (Sample Allocation Ratio)**

Enter a value (or range of values) for \( R \), the allocation ratio between samples. This value is only used when \( N2 \) is set to Use R.

When used, \( N2 \) is calculated from \( N1 \) using the formula: \( N2 = \lceil R(N1) \rceil \) where \( \lceil Y \rceil \) is the next integer greater than or equal to \( Y \). Note that setting \( R = 1.0 \) forces \( N2 = N1 \).

---

**One-Sided or Two-Sided Interval**

**Interval Type**

Specify whether the interval to be used will be a one-sided or a two-sided confidence interval.

---

**Precision**

**Distance from Mean Difference to Limit(s)**

This is the distance from the confidence limit(s) to the difference in means. For two-sided intervals, it is also known as the precision, half-width, or margin of error.

You can enter a single value or a list of values. The value(s) must be greater than zero.
Pooled Standard Deviation

Standard Deviation Source
This procedure permits two sources for estimates of the pooled standard deviation:

- **S is a Population Standard Deviation**
  This option should be selected if there are no previous samples that can be used to obtain an estimate of the pooled standard deviation. In this case, the algorithm assumes that the future sample obtained will be from a population with standard deviation S.

- **S from a Previous Sample**
  This option should be selected if the estimate of the pooled standard deviation is obtained from previous random samples from the same distributions as those to be sampled. The total sample size of the previous samples must also be entered under 'Total Sample Size of Previous Sample'.

Pooled Standard Deviation – S is a Population Standard Deviation

S (Standard Deviation)
Enter an estimate of the pooled standard deviation (must be positive). In this case, the algorithm assumes that future samples obtained will be from a population with pooled standard deviation S.

One common method for estimating the standard deviation is the range divided by 4, 5, or 6.

You can enter a range of values such as 1 2 3 or 1 to 10 by 1.

Press the Standard Deviation Estimator button to load the Standard Deviation Estimator window.

Pooled Standard Deviation – S from a Previous Sample

S (Standard Deviation)
Enter an estimate of the pooled standard deviation from a previous (or pilot) study. This value must be positive.

A range of values may be entered.

Press the Standard Deviation Estimator button to load the Standard Deviation Estimator window.

Total Sample Size of Previous Sample
Enter the total sample size that was used to estimate the pooled standard deviation entered in S (SD Estimated from a Previous Sample). The total sample size should be the total of the two sample sizes \((m_1 + m_2)\) that were used to estimate the pooled standard deviation.

If the previous sample used for the estimate of the pooled standard deviation is a single sample rather than two samples, enter the sample size of the previous sample plus one.

This value is entered only when 'Standard Deviation Source:' is set to 'S from a Previous Sample'.

Iterations Tab
This tab sets an option used in the iterative procedures.

Maximum Iterations

Maximum Iterations Before Search Termination
Specify the maximum number of iterations allowed before the search for the criterion of interest is aborted. When the maximum number of iterations is reached without convergence, the criterion is left blank. A value of 500 is recommended.

Example 1 – Calculating Sample Size
Suppose a study is planned in which the researcher wishes to construct a two-sided 95% confidence interval for the difference between two population means. It is very important that the mean weight is estimated within 10 units. The pooled standard deviation estimate, based on the range of data values, is 25.6. Instead of examining only the interval half-width of 10, a series of half-widths from 5 to 15 will also be considered.

The goal is to determine the sample size necessary to obtain a two-sided confidence interval such that the difference in means is estimated within 10 units. Tolerance probabilities of 0.70 to 0.95 will be examined.

Setup
This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Confidence Intervals for the Difference between Two Means with Tolerance Probability procedure window by expanding Means, then Two Independent Means, then clicking on Confidence Interval, and then clicking on Confidence Intervals for the Difference between Two Means with Tolerance Probability. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N1</td>
</tr>
<tr>
<td>Confidence Level</td>
<td>0.95</td>
</tr>
<tr>
<td>Tolerance Probability</td>
<td>0.70 to 0.95 by 0.05</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Distance from Mean Diff to Limit(s)</td>
<td>10</td>
</tr>
<tr>
<td>Standard Deviation Source</td>
<td>S is a Population Standard Deviation</td>
</tr>
<tr>
<td>S</td>
<td>25.6</td>
</tr>
<tr>
<td>Interval Type</td>
<td>Two-Sided</td>
</tr>
</tbody>
</table>
Confidence Intervals for the Diff. Between Two Means with Tolerance Probability

Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>N1</th>
<th>N2</th>
<th>Allocation Ratio</th>
<th>Target Dist from Mean Diff to Limits</th>
<th>Actual Dist from Mean Diff to Limits</th>
<th>Pooled Standard Deviation</th>
<th>Tolerance Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.95</td>
<td>55</td>
<td>55</td>
<td>1.00</td>
<td>10.000</td>
<td>9.994</td>
<td>25.60</td>
<td>0.70</td>
</tr>
<tr>
<td>0.95</td>
<td>56</td>
<td>56</td>
<td>1.00</td>
<td>10.000</td>
<td>9.998</td>
<td>25.60</td>
<td>0.75</td>
</tr>
<tr>
<td>0.95</td>
<td>58</td>
<td>58</td>
<td>1.00</td>
<td>10.000</td>
<td>9.919</td>
<td>25.60</td>
<td>0.80</td>
</tr>
<tr>
<td>0.95</td>
<td>59</td>
<td>59</td>
<td>1.00</td>
<td>10.000</td>
<td>9.951</td>
<td>25.60</td>
<td>0.85</td>
</tr>
<tr>
<td>0.95</td>
<td>61</td>
<td>61</td>
<td>1.00</td>
<td>10.000</td>
<td>9.921</td>
<td>25.60</td>
<td>0.90</td>
</tr>
<tr>
<td>0.95</td>
<td>63</td>
<td>63</td>
<td>1.00</td>
<td>10.000</td>
<td>9.962</td>
<td>25.60</td>
<td>0.95</td>
</tr>
</tbody>
</table>

References


Report Definitions

- Confidence level is the proportion of confidence intervals (constructed with this same confidence level, sample size, etc.) that would contain the true difference in population means.
- N1 and N2 are the sample sizes drawn from the two populations.
- Allocation Ratio is the ratio of the sample sizes, N2/N1.
- Dist from Mean Diff to Limit(s) is the distance from the confidence limit(s) to the difference in sample means. For two-sided intervals, it is also known as the precision, half-width, or margin of error.
- Target Dist from Mean Diff to Limit(s) is the value of the distance that is entered into the procedure.
- Actual Dist from Mean Diff to Limit(s) is the value of the distance that is obtained from the procedure.
- Pooled Standard Deviation is the standard deviation upon which the distance from mean difference to limit calculations are based.
- Tolerance Probability is the probability that a future interval with sample size N and corresponding confidence level will have a distance from the mean to the limit(s) that is less than or equal to the specified distance.

Summary Statements

The probability is 0.70 that group sample sizes of 55 and 55 will produce a two-sided 95% confidence interval with a distance from the difference in means to the limits that is less than or equal to 9.994 if the pooled standard deviation is 25.60.

This report shows the calculated sample size for each of the scenarios.

Plots Section

This plot shows the sample size of each group versus the precision for the two confidence levels.
Example 2 – Validation using Zar

Zar (1984) pages 133-134 gives an example of a precision calculation for a confidence interval for the difference between two means when the confidence level is 95%, the pooled standard deviation is 0.720625 from a total sample size of 13, the precision is 0.5, and the tolerance probability is 0.90. The sample size for each group is determined to be 34.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Confidence Intervals for the Difference between Two Means with Tolerance Probability procedure window by expanding Means, then Two Independent Means, then clicking on Confidence Interval, and then clicking on Confidence Intervals for the Difference between Two Means with Tolerance Probability. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N1</td>
</tr>
<tr>
<td>Confidence Level</td>
<td>0.95</td>
</tr>
<tr>
<td>Tolerance Probability</td>
<td>0.90</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Distance from Mean Diff to Limit(s)</td>
<td>0.5</td>
</tr>
<tr>
<td>Standard Deviation Source</td>
<td>S from a Previous Sample</td>
</tr>
<tr>
<td>S</td>
<td>0.720625</td>
</tr>
<tr>
<td>Total Sample Size of Previous Sample</td>
<td>13</td>
</tr>
<tr>
<td>Interval Type</td>
<td>Two-Sided</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Numeric Results for Two-Sided Confidence Intervals for the Difference in Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidence Level</td>
</tr>
<tr>
<td>------------------</td>
</tr>
<tr>
<td>0.95</td>
</tr>
</tbody>
</table>

Total sample size for estimate of pooled standard deviation from previous samples = 13.

PASS also calculated the sample size in each group to be 34.
Chapter 475

Group-Sequential Tests for Two Means

Introduction

Clinical trials are longitudinal. They accumulate data sequentially through time. The participants cannot be enrolled and randomized on the same day. Instead, they are enrolled as they enter the study. It may take several years to enroll enough patients to meet sample size requirements. Because clinical trials are long term studies, it is in the interest of both the participants and the researchers to monitor the accumulating information for early convincing evidence of either harm or benefit. This permits early termination of the trial.

Group sequential methods allow statistical tests to be performed on accumulating data while a phase III clinical trial is ongoing. Statistical theory and practical experience with these designs have shown that making four or five interim analyses is almost as effective in detecting large differences between treatment groups as performing a new analysis after each new data value. Besides saving time and resources, such a strategy can reduce the experimental subject’s exposure to an inferior treatment and make superior treatments available sooner.

When repeated significance testing occurs on the same data, adjustments have to be made to the hypothesis testing procedure to maintain overall significance and power levels. The landmark paper of Lan & DeMets (1983) provided the theory behind the alpha spending function approach to group sequential testing. This paper built upon the earlier work of Armitage, McPherson, & Rowe (1969), Pocock (1977), and O’Brien & Fleming (1979). PASS implements the methods given in Reboussin, DeMets, Kim, & Lan (1992) to calculate the power and sample sizes of various group sequential designs.

This module calculates sample size and power for group sequential designs used to compare two treatment means. Other modules perform similar analyses for the comparison of proportions and survival functions. The program allows you to vary the number and times of interim tests, the type of alpha spending function, and the test boundaries. It also gives you complete flexibility in solving for power, significance level, sample size, or effect size. The results are displayed in both numeric reports and informative graphics.
Technical Details

Suppose the means of two samples of $N_1$ and $N_2$ individuals will be compared at various stages of a trial using the $z_k$ statistic:

$$z_k = \frac{\bar{X}_{1k} - \bar{X}_{2k}}{\sqrt{\frac{S_{1k}^2}{N_{1k}} + \frac{S_{2k}^2}{N_{2k}}}}$$

The subscript $k$ indicates that the computations use all data that are available at the time of the $k$th interim analysis or $k$th look ($k$ goes from 1 to $K$). This formula computes the standard $z$ test that is appropriate when the variances of the two groups are different. The statistic, $z_k$, is assumed to be normally distributed.

Spending Functions

Lan and DeMets (1983) introduced alpha spending functions, $\alpha(\tau)$, that determine a set of boundaries $b_1, b_2, \ldots, b_K$ for the sequence of test statistics $z_1, z_2, \ldots, z_K$. These boundaries are the critical values of the sequential hypothesis tests. That is, after each interim test, the trial is continued as long as $z_k < b_k$. When $z_k \geq b_k$, the hypothesis of equal means is rejected and the trial is stopped early.

The time argument $\tau$ either represents the proportion of elapsed time to the maximum duration of the trial or the proportion of the sample that has been collected. When elapsed time is being used it is referred to as calendar time. When time is measured in terms of the sample, it is referred to as information time. Since it is a proportion, $\tau$ can only vary between zero and one.

Alpha spending functions have the characteristics:

$$\alpha(0) = 0$$
$$\alpha(1) = \alpha$$

The last characteristic guarantees a fixed $\alpha$ level when the trial is complete. That is,

$$\Pr(\{z_1 \geq b_1 \text{ or } z_2 \geq b_2 \text{ or } \cdots \text{ or } z_K \geq b_K\}) = \alpha(\tau)$$

This methodology is very flexible since neither the times nor the number of analyses must be specified in advance. Only the functional form of $\alpha(\tau)$ must be specified.

PASS provides five popular spending functions plus the ability to enter and analyze your own boundaries. These are calculated as follows:
1. **O’Brien-Fleming** \[ 2 - 2\Phi\left(\frac{Z_{\alpha/2}}{\sqrt{\tau}}\right) \]

![O'Brien-Fleming Boundaries with Alpha = 0.05](image)

2. **Pocock** \[ \alpha \ln(1 + (e - 1)\tau) \]

![Pocock Boundaries with Alpha = 0.05](image)

3. **Alpha * time** \[ \alpha \tau \]

![Alpha(Time) Boundaries with Alpha = 0.05](image)
4. **Alpha * time^1.5  \( \alpha \tau^{3/2} \)**

![Graph showing (Alpha)(Time^1.5) Boundaries with Alpha = 0.05]

5. **Alpha * time^2  \( \alpha \tau^{2} \)**

![Graph showing (Alpha)(Time^2.0) Boundaries with Alpha = 0.05]

6. **User Supplied**

   A custom set of boundaries may be entered.

The O’Brien-Fleming boundaries are commonly used because they do not significantly increase the overall sample size and because they are conservative early in the trial. Conservative in the sense that the means must be extremely different before statistical significance is indicated. The Pocock boundaries are nearly equal for all times. The Alpha*t boundaries use equal amounts of alpha when the looks are equally spaced. You can enter your own set of boundaries using the User Supplied option.

**Theory**

A detailed account of the methodology is contained in Lan & DeMets (1983), DeMets & Lan (1984), Lan & Zucker (1993), and DeMets & Lan (1994). The theoretical basis of the method will be presented here.

Group sequential procedures for interim analysis are based on their equivalence to discrete boundary crossing of a Brownian motion process with drift parameter \( \theta \). The test statistics \( z_k \) follow the multivariate normal distribution with means \( \theta \sqrt{\tau_k} \) and, for \( j \leq k \), covariances \( \sqrt{\tau_k / \tau_j} \). The drift parameter is related to the parameters of the z-test through the equation
\[
\theta = \frac{\mu_1 - \mu_2}{\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}}}
\]

Hence, the algorithm is as follows:

1. Compute boundary values based on a specified spending function and alpha value.
2. Calculate the drift parameter based on those boundary values and a specified power value.
3. Use the drift parameter and estimates of the other parameters in the above equation to calculate the appropriate sample size.

**Procedure Tabs**

This section describes the options that are specific to this procedure. These are located on the Data and Options tabs. For more information about the options of other tabs, go to the Procedure Window chapter.

**Data Tab**

The Data tab contains the parameters associated with the \( z \) test such as the means, variances, sample sizes, alpha, and power.

**Solve For**

**Find (Solve For)**

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are \( \text{Mean1}, \text{Mean2}, \text{Alpha}, \text{Power and Beta}, \text{N1} \) or \( \text{N2} \). Under most situations, you will select either \( \text{Power and Beta} \) or \( \text{N1} \).

Select \( \text{N1} \) when you want to calculate the sample size needed to achieve a given power and alpha level.

Select \( \text{Power and Beta} \) when you want to calculate the power of an experiment that has already been run since power is equal to one minus beta.

**Error Rates**

**Power or Beta**

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of equal means when in fact they are different.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.
A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

**Alpha**
This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. For this procedure, a type-I error occurs when you reject the null hypothesis of equal means when in fact they are equal.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.

---

**Sample Size**

**N1 (Sample Size Group 1)**
Enter a value (or range of values) for the sample size of this group. Note that these values are ignored when you are solving for N1. You may enter a range of values such as 10 to 100 by 10.

**N2 (Sample Size Group 2)**
Enter a value (or range of values) for the sample size of group 2 or enter Use R to base N2 on the value of N1. You may enter a range of values such as 10 to 100 by 10.

- **Use R**
  
  When Use R is entered here, N2 is calculated using the formula

  \[ N2 = \lceil R \times N1 \rceil \]

  where R is the Sample Allocation Ratio and \( \lceil Y \rceil \) is the first integer greater than or equal to Y. For example, if you want \( N1 = N2 \), select Use R and set \( R = 1 \).

**R (Sample Allocation Ratio)**
Enter a value (or range of values) for R, the allocation ratio between samples. This value is only used when N2 is set to Use R.

When used, N2 is calculated from N1 using the formula: \( N2 = \lceil R \times N1 \rceil \) where \( \lceil Y \rceil \) is the next integer greater than or equal to Y. Note that setting \( R = 1.0 \) forces \( N2 = N1 \).

---

**Effect Size**

**Mean1**
Enter value(s) for the mean of the first group under both hypotheses and the mean of the second group under the null hypothesis of equal means. Note that only the difference between the two means is used in the calculations. You may enter a range of values such as 10,20,30 or 0 to 100 by 25.

If you want to use a single difference rather than the two means, enter the value of the difference as Mean2 and zero for Mean1 (or vice versa).
Mean2
Enter value(s) for the mean of the second group under the alternative hypothesis. Note that only the difference between the two means is used in the calculations. You may enter a range of values such as 10,20,30 or 0 to 100 by 25.

If you want to use a single difference rather than the two means, enter the value of the difference as Mean2 and zero for Mean1 (or vice versa).

S1 (SD, Group 1)
Enter an estimate of the standard deviation of group 1. The standard deviation must be a positive number. Refer to the chapter on Estimating the Standard Deviation for more information on estimating the standard deviation. Press the SD button to obtain a special window designed to help you obtain a realistic value for the standard deviation.

Above all else, remember that the experience of consulting statisticians is that researchers tend to underestimate the standard deviation!

S2 (SD, Group 2)
Enter an estimate of the standard deviation of group 2. The standard deviation must be a positive number. Refer to the chapter on Estimating the Standard Deviation for more information on estimating the standard deviation. Press the SD button to obtain a special window designed to help you obtain a realistic value for the standard deviation.

You can enter S1 here if you want to assume that the standard deviations are equal and use the value entered for S1.

Look Details
This box contains the parameters associated with Group Sequential Design such as the type of spending function, the times, and so on.

Number of Looks
This is the number of interim analyses (including the final analysis). For example, a five here means that four interim analyses will be run in addition to the final analysis.

Boundary Truncation
You can truncate the boundary values at a specified value. For example, you might decide that no boundaries should be larger than 4.0. If you want to implement a boundary limit, enter the value here.

If you do not want a boundary limit, enter None here.

Spending Function
Specify which alpha spending function to use. The most popular is the O'Brien-Fleming boundary that makes early tests very conservative. Select User Specified if you want to enter your own set of boundaries.

Max Time
This is the total running time of the trial. It is used to convert the values in the Times box to fractions. The units (months or years) do not matter, as long as they are consistent with those entered in the Times box.

For example, suppose Max Time = 3 and Times = 1, 2, 3. Interim analyses would be assumed to have occurred at 0.33, 0.67, and 1.00.
Times
Enter a list of time values here at which the interim analyses will occur. These values are scaled according to the value of the Max Time option.

For example, suppose a 48-month trial calls for interim analyses at 12, 24, 36, and 48 months. You could set Max Time to 48 and enter 12,24,36,48 here or you could set Max Time to 1.0 and enter 0.25,0.50,0.75,1.00 here.

The number of times entered here must match the value of the Number of Looks.

• Equally Spaced
If you are planning to conduct the interim analyses at equally spaced points in time, you can enter Equally Spaced and the program will generate the appropriate time values for you.

Informations
You can weight the interim analyses on the amount of information obtained at each time point rather than on actual calendar time. If you would like to do this, enter the information amounts here. Usually, these values are the sample sizes obtained up to the time of the analysis.

For example, you might enter 50, 76, 103, 150 to indicate that 50 individuals were included in the first interim analysis, 76 in the second, and so on.

Upper and Lower Boundaries (Spending = User)
If the Spending Function is set to User Supplied you can enter a set of lower test boundaries, one for each interim analysis. The lower boundaries should be negative and the upper boundaries should be positive. Typical entries are 4,3,3,3,2 and 4,3,2,2,2.

• Symmetric
If you only want to enter the upper boundaries and have them copied with a change in sign to the lower boundaries, enter Symmetric for the lower boundaries.

Test
Alternative Hypothesis
Specify whether the test is one-sided or two-sided. When a two-sided hypothesis is selected, the value of alpha is halved. Everything else remains the same.

Note that the accepted procedure is to use Two Sided option unless you can justify using a one-sided test.

Options Tab
The Options tab controls the convergence of the various iterative algorithms used in the calculations.

Maximum Iterations
Maximum Iterations Before Search Termination
Specify the maximum number of iterations to be run before the search for the criterion of interest (Alpha, Beta, etc.) is aborted. When the maximum number of iterations is reached without convergence, the criterion is left blank.
Recommended: 500 (or more).

**Maximum Iterations (Lan-Demets algorithm)**
This is the maximum number of iterations used in the Lan-DeMets algorithm during its search routine. We recommend a value of at least 200.

---

**Tolerance**

**Probability Tolerance**
During the calculation of the probabilities associated with a set of boundary values, probabilities less than this are assumed to be zero.

We suggest a value of 0.00000000001.

**Power Tolerance**
This is the convergence level for the search for the spending function values that achieve a certain power. Once the iteration changes are less than this amount, convergence is assumed. We suggest a value of 0.0000001.

If the search is too time consuming, you might try increasing this value.

**Alpha Tolerance**
This is the convergence level for the search for a given alpha value. Once the changes in the computed alpha value are less than this amount, convergence is assumed and iterations stop. We suggest a value of 0.0001.

This option is only used when you are searching for alpha.

If the search is too time consuming, you can try increasing this value.
**Example 1 – Finding the Sample Size**

A clinical trial is to be conducted over a two-year period to compare the mean response of a new treatment with the current treatment. The current mean is 127 with a standard deviation of 55.88. The health community will be interested in the new treatment if the mean response rate is increased by 20%. So that the sample size requirements for different effect sizes can be compared, it is also of interest to compute the sample size at 10%, 30%, 40%, 50%, 60%, and 70% increases in the response rates.

Testing will be done at the 0.05 significance level and the power should be set to 0.10. A total of four tests are going to be performed on the data as they are obtained. The O’Brien-Fleming boundaries will be used.

Find the necessary sample sizes and test boundaries assuming equal sample sizes per arm and two-sided hypothesis tests.

We could enter these amounts directly into the Group Sequential Means window. Since the base mean is 127, a 20% increase would translate to a new mean response of 127(120/100) = 152.4. The other mean response rates could be computed similarly. However, to make the results more meaningful, we will scale the input by dividing by the current mean. The scaled standard deviation will be 100(55.88)/127 = 44.00. We set Mean1 to zero since we are only interested in the changes in Mean2. The values of Mean2 will then be 10, 20, 30, 40, 50, 60, and 70.

---

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Group-Sequential Tests for Two Means** procedure window by expanding Means, then Two Independent Means, then clicking on Group-Sequential, and then clicking on **Group-Sequential Tests for Two Means**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N1</td>
</tr>
<tr>
<td>Power</td>
<td>0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>Ignored</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Mean1 (Mean of Group 1)</td>
<td>0</td>
</tr>
<tr>
<td>Mean2 (Mean of Group 2)</td>
<td>10 to 70 by 10</td>
</tr>
<tr>
<td>S1 (Standard Deviation Group 1)</td>
<td>44</td>
</tr>
<tr>
<td>S2 (Standard Deviation Group 2)</td>
<td>S1</td>
</tr>
<tr>
<td>Number of Looks</td>
<td>4</td>
</tr>
<tr>
<td>Spending Function</td>
<td>O’Brien-Fleming</td>
</tr>
<tr>
<td>Max Time</td>
<td>2</td>
</tr>
<tr>
<td>Times</td>
<td>Equally Spaced</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Two-Sided</td>
</tr>
</tbody>
</table>
Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N1</th>
<th>N2</th>
<th>Alpha</th>
<th>Beta</th>
<th>Mean1</th>
<th>Mean2</th>
<th>S1</th>
<th>S2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9005</td>
<td>415</td>
<td>415</td>
<td>0.0500</td>
<td>0.0995</td>
<td>0.00</td>
<td>10.00</td>
<td>44.00</td>
<td>44.00</td>
</tr>
<tr>
<td>0.9012</td>
<td>104</td>
<td>104</td>
<td>0.0500</td>
<td>0.0988</td>
<td>0.00</td>
<td>20.00</td>
<td>44.00</td>
<td>44.00</td>
</tr>
<tr>
<td>0.9058</td>
<td>47</td>
<td>47</td>
<td>0.0500</td>
<td>0.0942</td>
<td>0.00</td>
<td>30.00</td>
<td>44.00</td>
<td>44.00</td>
</tr>
<tr>
<td>0.9012</td>
<td>26</td>
<td>26</td>
<td>0.0500</td>
<td>0.0988</td>
<td>0.00</td>
<td>40.00</td>
<td>44.00</td>
<td>44.00</td>
</tr>
<tr>
<td>0.9071</td>
<td>17</td>
<td>17</td>
<td>0.0500</td>
<td>0.0929</td>
<td>0.00</td>
<td>50.00</td>
<td>44.00</td>
<td>44.00</td>
</tr>
<tr>
<td>0.9116</td>
<td>12</td>
<td>12</td>
<td>0.0500</td>
<td>0.0884</td>
<td>0.00</td>
<td>60.00</td>
<td>44.00</td>
<td>44.00</td>
</tr>
<tr>
<td>0.9170</td>
<td>9</td>
<td>9</td>
<td>0.0500</td>
<td>0.0830</td>
<td>0.00</td>
<td>70.00</td>
<td>44.00</td>
<td>44.00</td>
</tr>
</tbody>
</table>

Report Definitions

Power is the probability of rejecting a false null hypothesis. Power should be close to one.
N1 and N2 are the number of items sampled from groups 1 and 2.
Alpha is the probability of rejecting a true null hypothesis in at least one of the sequential tests.
Beta is the probability of accepting a false null hypothesis at the conclusion of all tests.
Mean1 is the mean of populations 1 and 2 under the null hypothesis of equality.
Mean2 is the mean of population 2 under the alternative hypothesis. The mean of population 1 is unchanged.
S1 and S2 are the population standard deviations of groups 1 and 2.

Summary Statements

Sample sizes of 415 and 415 achieve 90% power to detect a difference of 10.00 between the group means with standard deviations of 44.00 and 44.00 at a significance level (alpha) of 0.0500 using a two-sided z-test. These results assume that 4 sequential tests are made using the O'Brien-Fleming spending function to determine the test boundaries.

This report shows the values of each of the parameters, one scenario per row. Note that 104 participants in each arm of the study are required to meet the 90% power requirement when the mean increase is 20%.

The values from this table are in the chart below. Note that this plot actually occurs further down in the report.

Plots Section

This plot shows that a large increase in sample size is necessary to test mean differences below 20%.
Details Section

<table>
<thead>
<tr>
<th>Look</th>
<th>Time</th>
<th>Lower Bndry</th>
<th>Upper Bndry</th>
<th>Nominal Alpha</th>
<th>Inc Alpha</th>
<th>Total Alpha</th>
<th>Inc Power</th>
<th>Total Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.50</td>
<td>-4.33263</td>
<td>4.33263</td>
<td>0.000015</td>
<td>0.000015</td>
<td>0.000015</td>
<td>0.003512</td>
<td>0.003512</td>
</tr>
<tr>
<td>2</td>
<td>1.00</td>
<td>-2.96311</td>
<td>2.96311</td>
<td>0.003045</td>
<td>0.003036</td>
<td>0.003051</td>
<td>0.254998</td>
<td>0.258510</td>
</tr>
<tr>
<td>3</td>
<td>1.50</td>
<td>-2.35902</td>
<td>2.35902</td>
<td>0.018323</td>
<td>0.016248</td>
<td>0.019299</td>
<td>0.427601</td>
<td>0.698111</td>
</tr>
<tr>
<td>4</td>
<td>2.00</td>
<td>-2.01406</td>
<td>2.01406</td>
<td>0.044003</td>
<td>0.030701</td>
<td>0.050000</td>
<td>0.214371</td>
<td>0.900483</td>
</tr>
</tbody>
</table>

Drift 3.27383

This report shows information about the individual interim tests. One report is generated for each scenario.

**Look**
These are the sequence numbers of the interim tests.

**Time**
These are the time points at which the interim tests are conducted. Since the Max Time was set to 2 (for two years), these time values are in years. Hence, the first interim test is at half a year, the second at one year, and so on.

We could have set Max Time to 24 so that the time scale was in months.

**Lower and Upper Boundary**
These are the test boundaries. If the computed value of the test statistic $z$ is between these values, the trial should continue. Otherwise, the trial can be stopped.

**Nominal Alpha**
This is the value of alpha for these boundaries if they were used for a single, standalone, test. Hence, this is the significance level that must be found for this look in a standard statistical package that does not adjust for multiple looks.

**Inc Alpha**
This is the amount of alpha that is spent by this interim test. It is close to, but not equal to, the value of alpha that would be achieved if only a single test was conducted. For example, if we lookup the third value, 2.35902, in normal probability tables, we find that this corresponds to a (two-sided) alpha of 0.0183. However, the entry is 0.0162. The difference is due to the correction that must be made for multiple tests.

**Total Alpha**
This is the total amount of alpha that is used up to and including the current test.

**Inc Power**
These are the amounts that are added to the total power at each interim test. They are often called the exit probabilities because they give the probability that significance is found and the trial is stopped, given the alternative hypothesis.

**Total Power**
These are the cumulative power values. They are also the cumulative exit probabilities. That is, they are the probability that the trial is stopped at or before the corresponding time.

**Drift**
This is the value of the Brownian motion drift parameter.
Boundary Plots

This plot shows the interim boundaries for each look. This plot shows very dramatically that the results must be extremely significant at early looks, but that they are near the single test boundary (1.96 and -1.96) at the last look.

Example 2 – Finding the Power

A clinical trial is to be conducted over a two-year period to compare the mean response of a new treatment with the current treatment. The current mean is 127 with a standard deviation of 55.88. The health community will be interested in the new treatment if the mean response rate is increased by 20%. The researcher wishes to calculate the power of the design at sample sizes 20, 60, 100, 140, 180, and 220. Testing will be done at the 0.01, 0.05, 0.10 significance levels and the overall power will be set to 0.10. A total of four tests are going to be performed on the data as they are obtained. The O’Brien-Fleming boundaries will be used. Find the power of these sample sizes and test boundaries assuming equal sample sizes per arm and two-sided hypothesis tests.

Proceeding as in Example 1, we decide to translate the mean and standard deviation into a percent of mean scale.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Group-Sequential Tests for Two Means procedure window by expanding Means, then Two Independent Means, then clicking on Group-Sequential, and then clicking on Group-Sequential Tests for Two Means. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.01 0.05 0.10</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>20 to 220 by 40</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
</tbody>
</table>
Data Tab (continued)
R (Sample Allocation Ratio) .................... 1.0
Mean1 (Mean of Group 1) ....................... 0
Mean2 (Mean of Group 2) ....................... 20
S1 (Standard Deviation Group 1) ............ 44
S2 (Standard Deviation Group 2) ............ S1
Number of Looks ................................. 4
Spending Function .............................. O'Brien-Fleming
Max Time .......................................... 2
Times ................................................. Equally Spaced
Alternative Hypothesis ....................... Two-Sided

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N1</th>
<th>N2</th>
<th>Alpha</th>
<th>Beta</th>
<th>Mean1</th>
<th>Mean2</th>
<th>S1</th>
<th>S2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1256</td>
<td>20</td>
<td>20</td>
<td>0.0100</td>
<td>0.8744</td>
<td>0.00</td>
<td>20.00</td>
<td>44.00</td>
<td>44.00</td>
</tr>
<tr>
<td>0.4605</td>
<td>60</td>
<td>60</td>
<td>0.0100</td>
<td>0.5395</td>
<td>0.00</td>
<td>20.00</td>
<td>44.00</td>
<td>44.00</td>
</tr>
<tr>
<td>0.7335</td>
<td>100</td>
<td>100</td>
<td>0.0100</td>
<td>0.2665</td>
<td>0.00</td>
<td>20.00</td>
<td>44.00</td>
<td>44.00</td>
</tr>
<tr>
<td>0.8871</td>
<td>140</td>
<td>140</td>
<td>0.0100</td>
<td>0.1129</td>
<td>0.00</td>
<td>20.00</td>
<td>44.00</td>
<td>44.00</td>
</tr>
<tr>
<td>0.9572</td>
<td>180</td>
<td>180</td>
<td>0.0100</td>
<td>0.0428</td>
<td>0.00</td>
<td>20.00</td>
<td>44.00</td>
<td>44.00</td>
</tr>
<tr>
<td>0.9851</td>
<td>220</td>
<td>220</td>
<td>0.0100</td>
<td>0.0149</td>
<td>0.00</td>
<td>20.00</td>
<td>44.00</td>
<td>44.00</td>
</tr>
<tr>
<td>0.2948</td>
<td>20</td>
<td>20</td>
<td>0.0500</td>
<td>0.7052</td>
<td>0.00</td>
<td>20.00</td>
<td>44.00</td>
<td>44.00</td>
</tr>
<tr>
<td>0.6929</td>
<td>60</td>
<td>60</td>
<td>0.0500</td>
<td>0.3071</td>
<td>0.00</td>
<td>20.00</td>
<td>44.00</td>
<td>44.00</td>
</tr>
<tr>
<td>0.8897</td>
<td>100</td>
<td>100</td>
<td>0.0500</td>
<td>0.1103</td>
<td>0.00</td>
<td>20.00</td>
<td>44.00</td>
<td>44.00</td>
</tr>
<tr>
<td>0.9650</td>
<td>140</td>
<td>140</td>
<td>0.0500</td>
<td>0.0350</td>
<td>0.00</td>
<td>20.00</td>
<td>44.00</td>
<td>44.00</td>
</tr>
<tr>
<td>0.9898</td>
<td>180</td>
<td>180</td>
<td>0.0500</td>
<td>0.0102</td>
<td>0.00</td>
<td>20.00</td>
<td>44.00</td>
<td>44.00</td>
</tr>
<tr>
<td>0.9972</td>
<td>220</td>
<td>220</td>
<td>0.0500</td>
<td>0.0028</td>
<td>0.00</td>
<td>20.00</td>
<td>44.00</td>
<td>44.00</td>
</tr>
<tr>
<td>0.4094</td>
<td>20</td>
<td>20</td>
<td>0.1000</td>
<td>0.5906</td>
<td>0.00</td>
<td>20.00</td>
<td>44.00</td>
<td>44.00</td>
</tr>
<tr>
<td>0.7909</td>
<td>60</td>
<td>60</td>
<td>0.1000</td>
<td>0.2091</td>
<td>0.00</td>
<td>20.00</td>
<td>44.00</td>
<td>44.00</td>
</tr>
<tr>
<td>0.9368</td>
<td>100</td>
<td>100</td>
<td>0.1000</td>
<td>0.0632</td>
<td>0.00</td>
<td>20.00</td>
<td>44.00</td>
<td>44.00</td>
</tr>
<tr>
<td>0.9827</td>
<td>140</td>
<td>140</td>
<td>0.1000</td>
<td>0.0173</td>
<td>0.00</td>
<td>20.00</td>
<td>44.00</td>
<td>44.00</td>
</tr>
<tr>
<td>0.9956</td>
<td>180</td>
<td>180</td>
<td>0.1000</td>
<td>0.0044</td>
<td>0.00</td>
<td>20.00</td>
<td>44.00</td>
<td>44.00</td>
</tr>
<tr>
<td>0.9989</td>
<td>220</td>
<td>220</td>
<td>0.1000</td>
<td>0.0011</td>
<td>0.00</td>
<td>20.00</td>
<td>44.00</td>
<td>44.00</td>
</tr>
</tbody>
</table>

These data show the power for various sample sizes and alphas. It is interesting to note that once the sample size is greater than 150, the value of alpha makes little difference on the value of power.
Example 3 – Effect of Number of Looks

Continuing with examples one and two, it is interesting to determine the impact of the number of looks on power. **PASS** allows only one value for the Number of Looks parameter per run, so it will be necessary to run several analyses. To conduct this study, set alpha to 0.05, \( N_1 \) to 100, and leave the other parameters as before. Run the analysis with Number of Looks equal to 1, 2, 3, 4, 6, 8, 10, and 20. Record the power for each run.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Group-Sequential Tests for Two Means** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Group-Sequential**, and then clicking on **Group-Sequential Tests for Two Means**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>( N_1 ) (Sample Size Group 1)</td>
<td>100</td>
</tr>
<tr>
<td>( N_2 ) (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>( R ) (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Mean1 (Mean of Group 1)</td>
<td>0</td>
</tr>
<tr>
<td>Mean2 (Mean of Group 2)</td>
<td>20</td>
</tr>
<tr>
<td>( S_1 ) (Standard Deviation Group 1)</td>
<td>44</td>
</tr>
<tr>
<td>( S_2 ) (Standard Deviation Group 2)</td>
<td>( S_1 )</td>
</tr>
<tr>
<td>Number of Looks</td>
<td>1 (Also run with 2, 3, 4, 6, 8, 10, and 20)</td>
</tr>
<tr>
<td>Spending Function</td>
<td>O’Brien-Fleming</td>
</tr>
<tr>
<td>Max Time</td>
<td>2</td>
</tr>
<tr>
<td>Times</td>
<td>Equally Spaced</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Two-Sided</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.
Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N1</th>
<th>N2</th>
<th>Alpha</th>
<th>Beta</th>
<th>Mean1</th>
<th>Mean2</th>
<th>S1</th>
<th>S2</th>
<th>Looks</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8951</td>
<td>100</td>
<td>100</td>
<td>0.0500</td>
<td>0.1049</td>
<td>0.00</td>
<td>20.00</td>
<td>44.00</td>
<td>44.00</td>
<td>1</td>
</tr>
<tr>
<td>0.8941</td>
<td>100</td>
<td>100</td>
<td>0.0500</td>
<td>0.1059</td>
<td>0.00</td>
<td>20.00</td>
<td>44.00</td>
<td>44.00</td>
<td>2</td>
</tr>
<tr>
<td>0.8916</td>
<td>100</td>
<td>100</td>
<td>0.0500</td>
<td>0.1084</td>
<td>0.00</td>
<td>20.00</td>
<td>44.00</td>
<td>44.00</td>
<td>3</td>
</tr>
<tr>
<td>0.8897</td>
<td>100</td>
<td>100</td>
<td>0.0500</td>
<td>0.1103</td>
<td>0.00</td>
<td>20.00</td>
<td>44.00</td>
<td>44.00</td>
<td>4</td>
</tr>
<tr>
<td>0.8871</td>
<td>100</td>
<td>100</td>
<td>0.0500</td>
<td>0.1129</td>
<td>0.00</td>
<td>20.00</td>
<td>44.00</td>
<td>44.00</td>
<td>6</td>
</tr>
<tr>
<td>0.8856</td>
<td>100</td>
<td>100</td>
<td>0.0500</td>
<td>0.1144</td>
<td>0.00</td>
<td>20.00</td>
<td>44.00</td>
<td>44.00</td>
<td>8</td>
</tr>
<tr>
<td>0.8845</td>
<td>100</td>
<td>100</td>
<td>0.0500</td>
<td>0.1155</td>
<td>0.00</td>
<td>20.00</td>
<td>44.00</td>
<td>44.00</td>
<td>10</td>
</tr>
<tr>
<td>0.8820</td>
<td>100</td>
<td>100</td>
<td>0.0500</td>
<td>0.1180</td>
<td>0.00</td>
<td>20.00</td>
<td>44.00</td>
<td>44.00</td>
<td>20</td>
</tr>
</tbody>
</table>

This analysis shows how little the number of looks impacts the power of the design. The power of a study with no interim looks is 0.8951. When twenty interim looks are made, the power falls just 0.0131, to 0.8820—a very small change.

Example 4 – Studying a Boundary Set

Continuing with the previous examples, suppose that you are presented with a set of boundaries and want to find the quality of the design (as measured by alpha and power). This is easy to do with PASS. Suppose that the analysis is to be run with five interim looks at equally spaced time points. The upper boundaries to be studied are 3.5, 3.5, 3.0, 2.5, 2.0. The lower boundaries are symmetric. The analysis would be run as follows.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Group-Sequential Tests for Two Means procedure window by expanding Means, then Two Independent Means, then clicking on Group-Sequential, and then clicking on Group-Sequential Tests for Two Means. You may then make the appropriate entries as listed below, or open Example 4 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td></td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05 (will be calculated from boundaries)</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>100</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Mean1 (Mean of Group 1)</td>
<td>0</td>
</tr>
<tr>
<td>Mean2 (Mean of Group 2)</td>
<td>20</td>
</tr>
<tr>
<td>S1 (Standard Deviation Group 1)</td>
<td>44</td>
</tr>
<tr>
<td>S2 (Standard Deviation Group 2)</td>
<td>S1</td>
</tr>
<tr>
<td>Number of Looks</td>
<td>5</td>
</tr>
<tr>
<td>Spending Function</td>
<td>User Supplied</td>
</tr>
<tr>
<td>Max Time</td>
<td>2</td>
</tr>
</tbody>
</table>
Data Tab (continued)
Times…………………………………………………Equally Spaced
Upper Boundaries……………………………………3.5 3.5 3.0 2.5 2.0
Lower Boundaries……………………………………Symmetric
Alternative Hypothesis……………………………Two-Sided

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N1</th>
<th>N2</th>
<th>Alpha</th>
<th>Beta</th>
<th>Mean1</th>
<th>Mean2</th>
<th>S1</th>
<th>S2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8898</td>
<td>100</td>
<td>100</td>
<td>0.0482</td>
<td>0.1102</td>
<td>0.00</td>
<td>20.00</td>
<td>44.00</td>
<td>44.00</td>
</tr>
</tbody>
</table>

Details when Spending = User Supplied, N1 = 100, N2 = 100, S1 = 44.00, S2 = 44.00, Diff = -20.00

<table>
<thead>
<tr>
<th>Look</th>
<th>Time</th>
<th>Lower Bndry</th>
<th>Upper Bndry</th>
<th>Nominal Alpha</th>
<th>Inc Alpha</th>
<th>Total Alpha</th>
<th>Inc Power</th>
<th>Total Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.40</td>
<td>-3.50000</td>
<td>3.50000</td>
<td>0.000465</td>
<td>0.000465</td>
<td>0.000930</td>
<td>0.019576</td>
<td>0.019576</td>
</tr>
<tr>
<td>2</td>
<td>0.80</td>
<td>-3.50000</td>
<td>3.50000</td>
<td>0.000465</td>
<td>0.000408</td>
<td>0.000874</td>
<td>0.058835</td>
<td>0.078411</td>
</tr>
<tr>
<td>3</td>
<td>1.20</td>
<td>-3.00000</td>
<td>3.00000</td>
<td>0.002700</td>
<td>0.002410</td>
<td>0.005110</td>
<td>0.232486</td>
<td>0.310897</td>
</tr>
<tr>
<td>4</td>
<td>1.60</td>
<td>-2.50000</td>
<td>2.50000</td>
<td>0.012419</td>
<td>0.010331</td>
<td>0.022750</td>
<td>0.339966</td>
<td>0.650863</td>
</tr>
<tr>
<td>5</td>
<td>2.00</td>
<td>-2.00000</td>
<td>2.00000</td>
<td>0.045500</td>
<td>0.034542</td>
<td>0.080042</td>
<td>0.238928</td>
<td>0.889791</td>
</tr>
</tbody>
</table>

Drift 3.21412

The power for this design is about 0.89. This value depends on both the boundaries and the sample size. The alpha level is 0.048157. This value only depends on the boundaries.

Example 5 – Validation using O’Brien-Fleming Boundaries

Reboussin (1992) presents an example for normally distributed data for a design with two-sided O’Brien-Fleming boundaries, looks = 5, alpha = 0.05, beta = 0.10, Mean1 = 220, Mean2 = 200, standard deviation = 30. They compute a drift of 3.28 and a sample size of 48.41 per group. The upper boundaries are: 4.8769, 3.3569, 2.6803, 2.2898, 2.0310.

To test that PASS provides the same result, enter the following.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Group-Sequential Tests for Two Means procedure window by expanding Means, then Two Independent Means, then clicking on Group-Sequential, and then clicking on Group-Sequential Tests for Two Means. You may then make the appropriate entries as listed below, or open Example 5 by going to the File menu and choosing Open Example Template.
Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N1</th>
<th>N2</th>
<th>Alpha</th>
<th>Beta</th>
<th>Mean1</th>
<th>Mean2</th>
<th>S1</th>
<th>S2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.905</td>
<td>49</td>
<td>49</td>
<td>0.05</td>
<td>0.10</td>
<td>220</td>
<td>200</td>
<td>30</td>
<td>30</td>
</tr>
</tbody>
</table>

Details when Spending = O'Brien-Fleming, N1 = 49, N2 = 49, S1 = 30.00, S2 = 30.00, Diff = 20.00

<table>
<thead>
<tr>
<th>Look</th>
<th>Time</th>
<th>Lower Bndry</th>
<th>Upper Bndry</th>
<th>Nominal Alpha</th>
<th>Inc Power</th>
<th>Total Power</th>
<th>Lower Bndry</th>
<th>Upper Bndry</th>
<th>Nominal Alpha</th>
<th>Inc Power</th>
<th>Total Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.20</td>
<td>-4.87688</td>
<td>4.87688</td>
<td>0.000001</td>
<td>0.000001</td>
<td>0.000336</td>
<td>0.000336</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.40</td>
<td>-3.35695</td>
<td>3.35695</td>
<td>0.000788</td>
<td>0.000787</td>
<td>0.101727</td>
<td>0.102062</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.60</td>
<td>-2.68026</td>
<td>2.68026</td>
<td>0.007357</td>
<td>0.007616</td>
<td>0.350673</td>
<td>0.452735</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.80</td>
<td>-2.28979</td>
<td>2.28979</td>
<td>0.022034</td>
<td>0.024424</td>
<td>0.751921</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.00</td>
<td>-2.03100</td>
<td>2.03100</td>
<td>0.042255</td>
<td>0.050000</td>
<td>0.903623</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Drift 3.29983

The slight difference in the power and the drift parameter is attributable to the rounding of the sample size from 48.41 to 49.

Example 6 – Validation with Pocock Boundaries

Reboussin (1992) presents an example for normally distributed data for a design with two-sided Pocock boundaries, looks = 5, alpha = 0.05, beta = 0.10, Mean1 = 220, Mean2 = 200, standard deviation = 30. They compute a drift of 3.55 and a sample size of 56.71 per group. The upper boundaries are: 2.4380, 2.4268, 2.4101, 2.3966, and 2.3859.

To test that PASS provides the same result, enter the following.
Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Group-Sequential Tests for Two Means procedure window by expanding Means, then Two Independent Means, then clicking on Group-Sequential, and then clicking on Group-Sequential Tests for Two Means. You may then make the appropriate entries as listed below, or open Example 6 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N1</td>
</tr>
<tr>
<td>Power</td>
<td>0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Sample Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Mean1 (Mean of Group 1)</td>
<td>220</td>
</tr>
<tr>
<td>Mean2 (Mean of Group 2)</td>
<td>200</td>
</tr>
<tr>
<td>S1 (Standard Deviation Group 1)</td>
<td>30</td>
</tr>
<tr>
<td>S2 (Standard Deviation Group 2)</td>
<td>S1</td>
</tr>
<tr>
<td>Number of Looks</td>
<td>5</td>
</tr>
<tr>
<td>Spending Function</td>
<td>Pocock</td>
</tr>
<tr>
<td>Max Time</td>
<td>1</td>
</tr>
<tr>
<td>Times</td>
<td>Equally Spaced</td>
</tr>
<tr>
<td>Upper Boundaries</td>
<td>Ignored</td>
</tr>
<tr>
<td>Lower Boundaries</td>
<td>Ignored</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Two-Sided</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N1</th>
<th>N2</th>
<th>Alpha</th>
<th>Beta</th>
<th>Mean1</th>
<th>Mean2</th>
<th>S1</th>
<th>S2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.903263</td>
<td>57</td>
<td>57</td>
<td>0.050000</td>
<td>0.096737</td>
<td>220.00</td>
<td>200.00</td>
<td>30.00</td>
<td>30.00</td>
</tr>
</tbody>
</table>

Details when Spending = O'Brien-Fleming, N1 = 49, N2 = 49, S1 = 30.00, S2 = 30.00, Diff = 20.00

<table>
<thead>
<tr>
<th>Look</th>
<th>Time</th>
<th>Lower Bndry</th>
<th>Upper Bndry</th>
<th>Nominal Alpha</th>
<th>Inc Alpha</th>
<th>Total Alpha</th>
<th>Inc Power</th>
<th>Total Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.20</td>
<td>-2.43798</td>
<td>2.43798</td>
<td>0.014770</td>
<td>0.014770</td>
<td>0.198712</td>
<td>0.198712</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>0.40</td>
<td>-2.42677</td>
<td>2.42677</td>
<td>0.015234</td>
<td>0.011387</td>
<td>0.260597</td>
<td>0.459308</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>0.60</td>
<td>-2.41014</td>
<td>2.41014</td>
<td>0.015946</td>
<td>0.009269</td>
<td>0.214118</td>
<td>0.673426</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>0.80</td>
<td>-2.39658</td>
<td>2.39658</td>
<td>0.016549</td>
<td>0.007816</td>
<td>0.143792</td>
<td>0.817218</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>1.00</td>
<td>-2.38591</td>
<td>2.38591</td>
<td>0.017037</td>
<td>0.006758</td>
<td>0.086045</td>
<td>0.903263</td>
<td></td>
</tr>
<tr>
<td>Drift</td>
<td></td>
<td>3.55903</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The slight difference in the power and the drift parameter is attributable to the rounding of the sample size from 56.71 to 57.
475-20 Group-Sequential Tests for Two Means
Chapter 476

Group-Sequential Tests for Two Means (Simulation)

Introduction

This procedure can be used to determine power, sample size and/or boundaries for group sequential tests comparing the means of two groups. The common two-sample T-Test and the Mann-Whitney U test can be simulated in this procedure. For two-sided tests, significance (efficacy) boundaries can be generated. For one-sided tests, significance and futility boundaries can be produced. The spacing of the looks can be equal or custom specified. Boundaries can be computed based on popular alpha- and beta-spending functions (O’Brien-Fleming, Pocock, Hwang-Shih-DeCani Gamma family, linear) or custom spending functions. Boundaries can also be input directly to verify alpha- and/or beta-spending properties. Futility boundaries can be binding or non-binding. Maximum and average (expected) sample sizes are reported as well as the alpha and/or beta spent and incremental power at each look. Corresponding P-Value boundaries are also given for each boundary statistic. Plots of boundaries are also produced.

The distributions of each of the groups under the null and alternative hypotheses can be specified directly using over ten distributions including normal, exponential, Gamma, Uniform, Beta, and Cauchy.

Technical Details

This section outlines many of the technical details of the techniques used in this procedure including the simulation summary, the test statistic details, and the use of spending functions.

An excellent text for the background and details of many group-sequential methods is Jennison and Turnbull (2000).
Simulation Procedure

In this procedure, a large number of simulations are used to calculate boundaries and power using the following steps:

1. Based on the specified distributions, random samples of size $N_1$ and $N_2$ are generated under the null distribution and under the alternative distribution. These are simulated samples as though the final look is reached.

2. For each sample, test statistics for each look are produced. For example, if $N_1$ and $N_2$ are 100 and there are 5 equally spaced looks, test statistics are generated from the random samples at $N_1 = N_2 = 20$, $N_1 = N_2 = 40$, $N_1 = N_2 = 60$, $N_1 = N_2 = 80$, and $N_1 = N_2 = 100$ for both null and alternative samples.

3. To generate the first significance boundary, the null distribution statistics of the first look (e.g., at $N_1 = N_2 = 20$) are ordered and the percent of alpha to be spent at the first look is determined (using either the alpha-spending function or the input value). The statistic for which the percent of statistics above (or below, as the case may be) that value is equal to the percent of alpha to be spent at the first look is the boundary statistic. It is seen here how important a large number of simulations is to the precision of the boundary estimates.

4. All null distribution samples that are outside the first significance boundary at the first look are removed from consideration for the second look. If binding futility boundaries are also being computed, all null distribution samples with statistics that are outside the first futility boundary are also removed from consideration for the second look. If non-binding futility boundaries are being computed, null distribution samples with statistics outside the first futility boundary are not removed.

5. To generate the second significance boundary, the remaining null distribution statistics of the second look (e.g., at $N_1 = N_2 = 40$) are ordered and the percent of alpha to be spent at the second look is determined (again, using either the alpha-spending function or the input value). The percent of alpha to be spent at the second look is multiplied by the total number of simulations to determine the number of the statistic that is to be the second boundary statistic. The statistic for which that number of statistics is above it (or below, as the case may be) is the second boundary statistic. For example, suppose there are initially 1000 simulated samples, with 10 removed at the first look (from, say, alpha spent at Look 1 equal to 0.01), leaving 990 samples considered for the second look. Suppose further that the alpha to be spent at the second look is 0.02. This is multiplied by 1000 to give 20. The 990 still-considered statistics are ordered and the 970th (20 in from 990) statistic is the second boundary.

6. All null distribution samples that are outside the second significance boundary and the second futility boundary, if binding, at the second look are removed from consideration for the third look (e.g., leaving 970 statistics computed at $N_1 = N_2 = 60$ to be considered at the third look). Steps 4 and 5 are repeated until the final look is reached.

Futility boundaries are computed in a similar manner using the desired beta-spending function or custom beta-spending values and the alternative hypothesis simulated statistics at each look. For both binding and non-binding futility boundaries, samples for which alternative hypothesis statistics are outside either the significance or futility boundaries of the previous look are excluded from current and future looks.

Because the final futility and significance boundaries are required to be the same, futility boundaries are computed beginning at a small value of beta (e.g., 0.0001) and incrementing beta by that amount until the futility and significance boundaries meet.
When boundaries are entered directly, this procedure uses the null hypothesis and alternative hypothesis simulations to determine the number of test statistics that are outside the boundaries at each look. The cumulative proportion of alternative hypothesis statistics that are outside the significance boundaries is the overall power of the study.

**Generating Random Distributions**

Two methods are available in PASS to simulate random samples. The first method generates the random variates directly, one value at a time. The second method generates a large pool (over 10,000) of random values and then draws the random numbers from this pool. This second method can cut the running time of the simulation by 70%.

As mentioned above, the second method begins by generating a large pool of random numbers from the specified distributions. Each of these pools is evaluated to determine if its mean is within a small relative tolerance (0.0001) of the target mean. If the actual mean is not within the tolerance of the target mean, individual members of the population are replaced with new random numbers if the new random number moves the mean towards its target. Only a few hundred such swaps are required to bring the actual mean to within tolerance of the target mean. This population is then sampled with replacement using the uniform distribution. We have found that this method works well as long as the size of the pool is the maximum of twice the number of simulated samples desired and 10,000.

**Test Statistics**

This section describes the test statistics that are available in this procedure.

**Two-Sample T-Test**

The two-sample t-test assumes that the data are a simple random sample from a population of normally-distributed values that all have the same mean and variance. This assumption implies that the data are continuous and their distribution is symmetric. The calculation of the t statistic is as follows

\[ t_{df} = \frac{\bar{X}_1 - \bar{X}_2 - (\mu_1 - \mu_2)}{s_{\bar{X}_1 - \bar{X}_2}} \]

where

\[ \bar{X}_k = \frac{\sum_{i=1}^{N_k} X_{ki}}{N_k} \]

\[ s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2 + \sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2}{N_1 + N_2 - 2} \left( \frac{1}{N_1} + \frac{1}{N_2} \right)} \]

\[ df = N_1 + N_2 - 2 \]
The significance of the test statistic is determined by computing the p-value based on the \( t \) distribution with degrees of freedom \( df \). If this p-value is less than a specified level (often 0.05), the null hypothesis is rejected. Otherwise, no conclusion can be reached.

**Mann-Whitney U Test**

This test is the nonparametric substitute for the equal-variance t-test. Two key assumptions for this test are that the distributions are at least ordinal and that they are identical under \( H_0 \). This implies that ties (repeated values) are not acceptable. When ties are present, the approximation provided can be used, but know that the theoretic results no longer hold.

The Mann-Whitney test statistic is defined as follows in Gibbons (1985).

\[
z = \frac{W_1 - \frac{N_1(N_1 + N_2 + 1)}{2}}{s_w} + C
\]

where

\[
W_1 = \sum_{k=1}^{N_1} \text{Rank}(X_{1k})
\]

The ranks are determined after combining the two samples. The standard deviation is calculated as

\[
s_w = \sqrt{\frac{N_1N_2(N_1 + N_2 + 1)}{12} - \frac{N_1N_2\sum_{i=1}^{N_1}(t_i^3 - t_i)}{12(N_1 + N_2)(N_1 + N_2 - 1)}}
\]

where \( t_i \) is the number of observations tied at value one, \( t_2 \) is the number of observations tied at some value two, and so forth.

The correction factor, \( C \), is 0.5 if the rest of the numerator of \( z \) is negative or -0.5 otherwise. The value of \( z \) is then compared to the standard normal distribution.

**Standard Deviations**

Care must be used when either the null or alternative distribution is not normal. In these cases, the standard deviation is usually not specified directly. For example, you might use a gamma distribution with a shape parameter of 1.5 and a mean of 4 as the null distribution and a gamma distribution with the same shape parameter and a mean of 5 as the alternative distribution. This allows you to compare the two means. However, although the shape parameters are constant, the standard deviations, which are based on both the shape parameter and the mean, are not. Thus the distributions not only have different means, but also different standard deviations.
Spending Functions

Spending functions can be used in this procedure to specify the proportion of alpha or beta that is spent at each look without having to specify the proportion directly.

Spending functions have the characteristics that they are increasing and that

\[ \alpha(0) = 0 \]
\[ \alpha(1) = \alpha \]

The last characteristic guarantees a fixed \( \alpha \) level when the trial is complete. This methodology is very flexible since neither the times nor the number of analyses must be specified in advance. Only the functional form of \( \alpha(\tau) \) must be specified.

**PASS** provides several popular spending functions plus the ability to enter and analyze your own percents of alpha or beta spent. These are calculated as follows (beta may be substituted for alpha for beta-spending functions):

1. Hwang-Shih-DeCani (gamma family)
   \[ \alpha \left[ \frac{1 - e^{-\gamma \tau}}{1 - e^{-\gamma}} \right], \gamma \neq 0; \ \alpha, \gamma = 0 \]

2. O’Brien-Fleming Analog
   \[ 2 - 2\Phi \left( \frac{Z_{\alpha/2}}{\sqrt{t}} \right) \]
3. **Pocock Analog** \( \alpha \cdot \ln(1 + (e - 1)t) \)

![Pocock Boundaries with Alpha = 0.05](image1)

4. **Alpha * time** \( \alpha \cdot t \)

![Alpha*(Time) Boundaries with Alpha = 0.05](image2)

5. **Alpha * time^1.5** \( \alpha \cdot t^{3/2} \)

![Alpha*(Time^1.5) Boundaries with Alpha = 0.05](image3)
6.  \( \text{Alpha} \times \text{time}^2 \quad \alpha \cdot t^2 \)

\[
\begin{array}{c|c|c|c}
\text{Look} & \text{Upper} & \text{Lower} \\
1 & -1.7 & 1.7 \\
2 & -3.3 & 3.3 \\
3 & -5.0 & 5.0 \\
4 & 0.0 & \\
5 & 1.7 & \\
6 & 3.3 & \\
7 & 5.0 & \\
\end{array}
\]

7.  \( \text{Alpha} \times \text{time}^C \quad \alpha \cdot t^C \)

\[
\begin{array}{c|c|c|c}
\text{Look} & \text{Upper} & \text{Lower} \\
1 & -1.7 & 1.7 \\
2 & -3.3 & 3.3 \\
3 & -5.0 & 5.0 \\
4 & 0.0 & \\
5 & 1.7 & \\
6 & 3.3 & \\
7 & 5.0 & \\
\end{array}
\]

8.  \text{User Supplied Percents}

A custom set of percents of alpha to be spent at each look may be input directly.

The O'Brien-Fleming Analog spends very little alpha or beta at the beginning and much more at
the final looks. The Pocock Analog and (Alpha or Beta)(Time) spending functions spend alpha or
beta more evenly across the looks. The Hwang-Shih-DeCani (C) (gamma family) spending
functions and (Alpha or Beta)(Time\(^C\)) spending functions are flexible spending functions that
can be used to spend more alpha or beta early or late or evenly, depending on the choice of C.

**Procedure Options**

This section describes the options that are specific to this procedure. These are located on the
Data, Looks & Boundaries, Enter Boundaries, and Options tabs. For more information about the
options of other tabs, go to the Procedure Window chapter.

**Data Tab**

The Data tab contains most of the parameters and options for the general setup of the procedure.
Solve For

Find (Solve For)
Solve for either power, sample size, or enter the boundaries directly and solve for power and alpha. When solving for power or sample size, the look and boundary details are specified on the "Looks & Boundaries" tab and the "Enter Boundaries" tab is ignored. When entering the boundaries directly and solving for power and alpha, the boundaries are input on the "Enter Boundaries" tab and the "Looks & Boundaries" tab is ignored.

When solving for power or N1, the early-stopping boundaries are also calculated. High accuracy for early-stopping boundaries requires a very large number of simulations (Recommended 100,000 to 10,000,000).

The parameter selected here is the parameter displayed on the vertical axis of the plot.
Because this is a simulation based procedure, the search for the sample size may take several minutes or hours to complete. You may find it quicker and more informative to solve for Power for a range of sample sizes.

Error Rates

Power or Beta
Power is the probability of rejecting the null hypothesis when it is false. Power is equal to 1-Beta, so specifying power implicitly specifies beta.

Beta is the probability obtaining a false negative on the statistical test. That is, it is the probability of accepting a false null hypothesis.

In the context of simulated group sequential trials, the power is the proportion of the alternative hypothesis simulations that cross any one of the significance (efficacy) boundaries.

The valid range is between 0 to 1.

Different disciplines and protocols have different standards for setting power. A common choice is 0.90, but 0.80 is also popular.

You can enter a range of values such as 0.70 0.80 0.90 or 0.70 to 0.90 by 0.1.

Alpha (Significance Level)
Alpha is the probability of obtaining a false positive on the statistical test. That is, it is the probability of rejecting a true null hypothesis.

The null hypothesis is usually that the parameters (the means, proportions, etc.) are all equal.

In the context of simulated group sequential trials, alpha is the proportion of the null hypothesis simulations that cross any one of the significance (efficacy) boundaries.

Since Alpha is a probability, it is bounded by 0 and 1. Commonly, it is between 0.001 and 0.250. Alpha is often set to 0.05 for two-sided tests and to 0.025 for one-sided tests.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.
Sample Size

N1 (Sample Size Group 1)
Enter a value for the sample size, N1. This is the number of subjects in the first group of the study at the final look.

You may enter a range such as 10 to 100 by 10 or a list of values separated by commas or blanks. You might try entering the same number two or three times to get an idea of the variability in your results. For example, you could enter "10 10 10".

N2 (Sample Size Group 2)
Enter a value (or range of values) for the sample size of group 2. This is the number of subjects in the second group of the study at the final look.

- Use R
  When Use R is entered here, N2 is calculated using the formula
  \[ N2 = \lceil R(N1) \rceil \]
  where R is the Sample Allocation Ratio and the operator \( \lceil Y \rceil \) is the first integer greater than or equal to Y. For example, if you want \( N1 = N2 \), select Use R and set \( R = 1 \).

R (Sample Allocation Ratio)
Enter a value (or range of values) for R, the allocation ratio between samples. This value is only used when N2 is set to Use R.

\[ R = \frac{N2}{N1} \]

When used, N2 is calculated from N1 using the formula: \( N2 = \lceil R(N1) \rceil \) where \( \lceil Y \rceil \) is the next integer greater than or equal to Y. Note that setting \( R = 1.0 \) forces \( N2 = N1 \).

Effect Size

Group 1 (and 2) Distribution|H0
These options specify the distributions of the two groups under the null hypothesis, H0. The difference between the means of these two distributions is the difference that is tested, Diff0.

Usually, these two distributions will be identical and Diff0 = 0. However, if you are planning a non-inferiority test, the means will be different.

All of the distributions are parameterized so that the mean is entered first. For example, if you wanted to specify that the mean of a normally-distributed variable is to be five, you could enter N(5, S) or N(M0, S) here and M0 = 5 later.

The parameters of each distribution are specified using numbers or letters. If letters are used, their values are specified in the boxes below. The value M0 is reserved for the value of the mean under the null hypothesis.

Following is a list of the distributions that are available and the syntax used to specify them. Note that, except for the multinomial, the distributions are parameterized so that the mean is entered first.

- Beta=A(M0,A,B,Minimum)
- Binomial=B(M0,N)
- Cauchy=C(M0,Scale)
Group-Sequential Tests for Two Means (Simulation)

- **Constant** = K(Value)
- **Exponential** = E(M0)
- **F** = F(M0, DF1)
- **Gamma** = G(M0, A)
- **Multinomial** = M(P1, P2, ..., Pk)
- **Normal** = N(M0, SD)
- **Poisson** = P(M0)
- **Student's T** = T(M0, D)
- **Tukey's Lambda** = L(M0, S, Skewness, Elongation)
- **Uniform** = U(M0, Minimum)
- **Weibull** = W(M0, B)

Details of writing mixture distributions, combined distributions, and compound distributions are found in the chapter on Data Simulation and are not repeated here.

**Finding the Value of the Mean of a Specified Distribution**

Except for the multinomial distribution, the distributions have been parameterized in terms of their means, since this is the parameter being tested. The mean of a distribution created as a linear combination of other distributions is found by applying the linear combination to the individual means. However, the mean of a distribution created by multiplying or dividing other distributions is not necessarily equal to applying the same function to the individual means. For example, the mean of 4N(4, 5) + 2N(5, 6) is 4*4 + 2*5 = 26, but the mean of 4N(4, 5) * 2N(5, 6) is not exactly 4*4*2*5 = 160 (although it is close).

**Group 1 (and 2) Distribution[H1]**

These options specify the distributions of the two groups under the alternative hypothesis, H1. The difference between the means of these two distributions is the difference that is assumed to be the true value of the difference. That is, this is the difference at which the power is computed.

Usually, the mean difference is specified by entering $M0$ for the mean parameter in the distribution expression for group 1 and $M1$ for the mean parameter in the distribution expression for group 2. The mean difference under H1 then becomes the value of $M0 - M1$.

The parameters of each distribution are specified using numbers or letters. If letters are used, their values are specified in the boxes below. The value $M1$ is reserved for the value of the mean of group 2 under the alternative hypothesis.

Following is a list of the distributions that are available and the syntax used to specify them. Note that, except for the multinomial, the distributions are parameterized so that the mean, $M1$, is entered first.

- **Beta** = A(M1, A, B, Minimum)
- **Binomial** = B(M1, N)
- **Cauchy** = C(M1, Scale)
- **Constant** = K(Value)
- **Exponential** = E(M1)
- **F** = F(M1, DF1)
- **Gamma** = G(M1, A)
- **Multinomial** = M(P1, P2, ..., Pk)
- **Normal** = N(M1, SD)
- **Poisson** = P(M1)
- **Student's T** = T(M1, D)
Tukey's Lambda = L(M1, S, Skewness, Elongation)
Uniform = U(M1, Minimum)
Weibull = W(M1, B)

Details of writing mixture distributions, combined distributions, and compound distributions are found in the chapter on Data Simulation and will not be repeated here.

Effect Size – Distribution Parameters

M0 (Mean|H0)
These values are substituted for $M0$ in the distribution specifications given above. $M0$ is intended to be the value of the mean hypothesized by the null hypothesis, $H0$.

You can enter a list of values using the syntax 0 1 2 3 or 0 to 3 by 1.

M1 (Mean|H1)
These values are substituted for $M1$ in the distribution specifications given above. Although it can be used wherever you want, $M1$ is intended to be the value of the mean hypothesized by the alternative hypothesis, $H1$.

You can enter a list of values using the syntax 0 1 2 3 or 0 to 3 by 1.

Parameter Values (S, A, B)
Enter the numeric value(s) of the parameters listed above. These values are substituted for the corresponding letter in all four distribution specifications.

You can enter a list of values for each letter using the syntax 0 1 2 3 or 0 to 3 by 1.

You can also change the letter that is used as the name of this parameter using the pull-down menu to the side.

Test and Simulations

Alternative Hypothesis
Specify the alternative hypothesis of the test. Since the null hypothesis is the opposite, specifying the alternative is the only hypothesis needed.

When a two-sided alternative is selected, futility boundaries are not permitted. If you wish to do a two-sided test with futility boundaries you may consider two one-sided tests.

In each case, Mean1 - Mean2 is the true difference between population means. Diff0 is the difference in population means under the null hypothesis, and is usually 0 for a superiority/significance/efficacy test.

Care need be taken in the choice of direction of the test relative to the choice of $M0$ and $M1$. For example, for a one-sided test to show that the true mean for group 1 (control) is less than the true mean for group 2 (treatment), $M1$ should be larger than $M0$ and the alternative hypothesis chosen should be Mean1 - Mean2 < Diff0.

Test Type
Specify which test statistic is to be reported on.

The T-Test is the standard T-test based on the pooled variance and SS1 + SS2 - 2 degrees of freedom, where SS1 and SS2 are the sample sizes of groups 1 and 2 at the corresponding look.
The Mann-Whitney Test is the common non-parametric alternative and is based on the ranks of the observations. It is sometimes called the Mann-Whitney U, Mann-Whitney-Wilcoxon, Wilcoxon Rank-Sum, or Wilcoxon-Mann-Whitney test.

Simulations
This option specifies the number of iterations, $M$, used in the simulation. As the number of iterations is increased, the accuracy and running time of the simulation will be increased also.

The precision of the simulated power estimates are calculated from the binomial distribution. Thus, confidence intervals may be constructed for various power values. The following table gives an estimate of the precision that is achieved for various simulation sizes when the power is either 0.50 or 0.95. The table values are interpreted as follows: a 95% confidence interval of the true power is given by the power reported by the simulation plus and minus the ‘Precision’ amount given in the table.

<table>
<thead>
<tr>
<th>Simulation Size M</th>
<th>Precision when Power = 0.50</th>
<th>Precision when Power = 0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.100</td>
<td>0.044</td>
</tr>
<tr>
<td>500</td>
<td>0.045</td>
<td>0.019</td>
</tr>
<tr>
<td>1000</td>
<td>0.032</td>
<td>0.014</td>
</tr>
<tr>
<td>2000</td>
<td>0.022</td>
<td>0.010</td>
</tr>
<tr>
<td>5000</td>
<td>0.014</td>
<td>0.006</td>
</tr>
<tr>
<td>10000</td>
<td>0.010</td>
<td>0.004</td>
</tr>
<tr>
<td>50000</td>
<td>0.004</td>
<td>0.002</td>
</tr>
<tr>
<td>100000</td>
<td>0.003</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Notice that a simulation size of 1000 gives a precision of plus or minus 0.01 when the true power is 0.95. Also note that as the simulation size is increased beyond 5000, there is only a small amount of additional accuracy achieved.

However, when solving for Power or N1, the simulations are also used to calculate the look boundaries. To obtain precise boundary estimates, the number of simulations needs to be high. This consideration competes with the length of time to complete the simulation. When solving for power, a large number of simulations (100,000 or 1,000,000) will finish in several minutes. When solving for N1, perhaps 10,000 simulations can be run for each iteration. Then, a final run with the resulting N1 solving for power can be run with more simulations.

Looks & Boundaries Tab
The Data tab contains most of the parameters and options for the general setup of the procedure.

Looks and Boundaries
Specification of Looks and Boundaries
Choose whether spending functions will be used to divide alpha and beta for each look (Simple Specification), or whether the percents of alpha and beta to be spent at each look will be specified directly (Custom Specification).

Under Simple Specification, the looks are automatically considered to be equally spaced. Under Custom Specification, the looks may be equally spaced or custom defined based on the percent of accumulated information.
Group-Sequential Tests for Two Means (Simulation) 476-13

**Looks and Boundaries – Simple Specification**

**Number of Equally Spaced Looks**
Select the total number of looks that will be used if the study is not stopped early for the crossing of a boundary.

**Alpha Spending Function**
Specify the type of alpha spending function to use.

The O'Brien-Fleming Analog spends very little alpha at the beginning and much more at the final looks. The Pocock Analog and (Alpha)(Time) spending functions spend alpha more evenly across the looks. The Hwang-Shih-DeCani (C) (sometimes called the gamma family) spending functions and (Alpha)(Time^C) spending functions are flexible spending functions that can be used to spend more alpha early or late or evenly, depending on the choice of C.

**C (Alpha Spending)**
C is used to define the Hwang-Shih-DeCani (C) or (Alpha)(Time^C) spending functions.

For the Hwang-Shih-DeCani (C) spending function, negative values of C spend more of alpha at later looks, values near 0 spend alpha evenly, and positive values of C spend more of alpha at earlier looks.

For the (Alpha)(Time^C) spending function, only positive values for C are permitted. Values of C near zero spend more of alpha at earlier looks, values near 1 spend alpha evenly, and larger values of C spend more of alpha at later looks.

**Type of Futility Boundary**
This option determines whether or not futility boundaries will be created, and if so, whether they are binding or non-binding.

Futility boundaries are boundaries such that, if crossed at a given look, stop the study in favor of H0.

Binding futility boundaries are computed in concert with significance boundaries. They are called binding because they require the stopping of a trial if they are crossed. If the trial is not stopped, the probability of a false positive will exceed alpha.

When Non-binding futility boundaries are computed, the significance boundaries are first computed, ignoring the futility boundaries. The futility boundaries are then computed. These futility boundaries are non-binding because continuing the trial after they are crossed will not affect the overall probability of a false positive declaration.

**Number of Skipped Futility Looks**
In some trials it may be desirable to wait a number of looks before examining the trial for futility. This option allows the beta to begin being spent after a specified number of looks.

The Number of Skipped Futility Looks should be less than the number of looks.

**Beta Spending Function**
Specify the type of beta spending function to use.

The O'Brien-Fleming Analog spends very little beta at the beginning and much more at the final looks. The Pocock Analog and (Beta)(Time) spending functions spend beta more evenly across the looks. The Hwang-Shih-DeCani (C) (sometimes called the gamma family) spending functions
and (Beta)(Time^C) spending functions are flexible spending functions that can be used to spend more beta early or late or evenly, depending on the choice of C.

**C (Beta Spending)**

C is used to define the Hwang-Shih-DeCani (C) or (Beta)(Time^C) spending functions.

For the Hwang-Shih-DeCani (C) spending function, negative values of C spend more of beta at later looks, values near 0 spend beta evenly, and positive values of C spend more of beta at earlier looks.

For the (Beta)(Time^C) spending function, only positive values for C are permitted. Values of C near zero spend more of beta at earlier looks, values near 1 spend beta evenly, and larger values of C spend more of beta at later looks.

---

**Looks and Boundaries – Custom Specification**

**Number of Looks**

This is the total number of looks of either type (significance or futility or both).

**Equally Spaced**

If this box is checked, the Accumulated Information boxes are ignored and the accumulated information is evenly spaced.

**Type of Futility Boundary**

This option determines whether or not futility boundaries will be created, and if so, whether they are binding or non-binding.

Futility boundaries are boundaries such that, if crossed at a given look, stop the study in favor of H0.

Binding futility boundaries are computed in concert with significance boundaries. They are called binding because they require the stopping of a trial if they are crossed. If the trial is not stopped, the probability of a false positive will exceed alpha.

When Non-binding futility boundaries are computed, the significance boundaries are first computed, ignoring the futility boundaries. The futility boundaries are then computed. These futility boundaries are non-binding because continuing the trial after they are crossed will not affect the overall probability of a false positive declaration.

**Accumulated Information**

The accumulated information at each look defines the proportion or percent of the sample size that is used at that look.

These values are accumulated information values so they must be increasing.

Proportions, percents, or sample sizes may be entered. All proportions, percents, or sample sizes will be divided by the value at the final look to create an accumulated information proportion for each look.

**Percent of Alpha Spent**

This is the percent of the total alpha that is spent at the corresponding look. It is not the cumulative value.
Percents, proportions, or alphas may be entered here. Each of the values is divided by the sum of the values to obtain the proportion of alpha that is used at the corresponding look.

**Percent of Beta Spent**
This is the percent of the total beta (1-power) that is spent at the corresponding look. It is not the cumulative value.

Percents, proportions, or betas may be entered here. Each of the values is divided by the sum of the values to obtain the proportion of beta that is used at the corresponding look.

---

**Enter Boundaries Tab**
The Data tab contains most of the parameters and options for the general setup of the procedure.

---

**Looks and Boundaries**

**Number of Looks**
This is the total number of looks of either type (significance or futility or both).

**Equally Spaced**
If this box is checked, the Accumulated Information boxes are ignored and the accumulated information is evenly spaced.

**Types of Boundaries**
This option determines whether or not futility boundaries will be entered.

Futility boundaries are boundaries such that, if crossed at a given look, stop the study in favor of H0.

**Accumulated Information**
The accumulated information at each look defines the proportion or percent of the sample size that is used at that look.

These values are accumulated information values so they must be increasing.

Proportions, percents, or sample sizes may be entered. All proportions, percents, or sample sizes will be divided by the value at the final look to create an accumulated information proportion for each look.

**Significance Boundary**
Enter the value of the significance boundary corresponding to the chosen test statistic. These are sometimes called efficacy boundaries.

**Futility Boundary**
Enter the value of the futility boundary corresponding to the chosen test statistic.
Options Tab

The Options tab contains limits on the number of iterations and various options about individual tests.

Maximum Iterations

Maximum Iterations Before Search Termination
Specify the maximum N1 before the search for N1 is aborted.

Since simulations for large sample sizes are very computationally intensive and hence time-consuming, this value can be used to stop searches when N1 is larger than reasonable sample sizes for the study.

This applies only when "Find (Solve For)" is set to N1.

The procedure uses a binary search when searching for N1. If a value for N1 is tried that exceeds this value, and the power is not reached, a warning message will be shown on the output indicating the desired power was not reached.

We recommend a value of at least 20000.

Random Numbers

Random Number Pool Size
This is the size of the pool of random values from which the random samples will be drawn. Pools should be at least the maximum of 10,000 and twice the number of simulations. You can enter Automatic and an appropriate value will be calculated.

If you do not want to draw numbers from a pool, enter 0 here.

Matching Boundaries at Final Look

Beta Search Increment
For each simulation, when futility bounds are computed, the appropriate beta is found by searching from 0 to 1 by this increment. Smaller increments are more refined, but the search takes longer.

We recommend 0.001 or 0.0001.
Example 1 – Power and Output

A clinical trial is to be conducted over a two-year period to compare the mean response of a new treatment to that of the current treatment. The current response mean is 108. Although the researchers do not know the true mean of the new treatment, they would like to examine the power that is achieved if the mean of the new treatment is 113. The standard deviation for both groups is assumed to be 25. The sample size at the final look is to be 500 per group. Testing will be done at the 0.05 significance level. A total of five tests are going to be performed on the data as they are obtained. The O’Brien-Fleming (Analog) boundaries will be used.

Find the power and test boundaries assuming equal sample sizes per arm and two-sided hypothesis tests.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Group-Sequential Tests for Two Means (Simulation) procedure window by expanding Means, then Two Independent Means, then clicking on Group-Sequential, and then clicking on Group-Sequential Tests for Two Means (Simulation). You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td>Find (Solve For) Power</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>500</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Group 1 Dist’n</td>
<td>H0</td>
</tr>
<tr>
<td>Group 2 Dist’n</td>
<td>H0</td>
</tr>
<tr>
<td>Group 1 Dist’n</td>
<td>H1</td>
</tr>
<tr>
<td>Group 2 Dist’n</td>
<td>H1</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H1)</td>
</tr>
<tr>
<td>S</td>
<td>25</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Mean1 - Mean2 ≠ Diff0</td>
</tr>
<tr>
<td>Simulations</td>
<td>20000</td>
</tr>
<tr>
<td>Test Type</td>
<td>T-Test</td>
</tr>
</tbody>
</table>

Looks and Boundaries Tab
 Specification of Looks and Boundaries Simple
 Number of Equally Spaced Looks 5
 Alpha Spending Function O’Brien-Fleming Analog
Output

Click the Run button to perform the calculations and generate the following output.

**Numeric Results and Plots**

**Scenario 1 Numeric Results for Group Sequential Testing Mean Difference = Diff0.**

Hypotheses: H0: Diff1=Diff0; H1: Diff1<>Diff0
H0 Dist's: Normal(M0 S) & Normal(M0 S)
H1 Dist's: Normal(M0 S) & Normal(M1 S)
Test Statistic: T-Test
Alpha-Spending Function: O'Brien-Fleming Analog
Beta-Spending Function: None
Futility Boundary Type: None
Number of Looks: 5
Simulations: 20000
Pool Size: 40000

**Numeric Summary for Scenario 1**

<table>
<thead>
<tr>
<th>Value</th>
<th>95% LCL</th>
<th>95% UCL</th>
<th>Target</th>
<th>Actual</th>
<th>95% LCL</th>
<th>95% UCL</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8770</td>
<td>0.8724</td>
<td>0.8816</td>
<td>0.0500</td>
<td>0.0482</td>
<td>0.0452</td>
<td>0.0512</td>
<td>0.1230</td>
</tr>
</tbody>
</table>


--- Average Sample Size ---

<table>
<thead>
<tr>
<th>N1</th>
<th>N2</th>
<th>Grp1</th>
<th>Grp2</th>
<th>Grp1</th>
<th>Grp2</th>
<th>Diff0</th>
<th>Diff1</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>500</td>
<td>497</td>
<td>497</td>
<td>380</td>
<td>380</td>
<td>0.0</td>
<td>-5.0</td>
<td>108.0</td>
<td>113.0</td>
<td>25.0</td>
</tr>
</tbody>
</table>

**Report Definitions**

- **Power** is the probability of rejecting a false null hypothesis at one of the looks. It is the total proportion of alternative hypothesis simulations that are outside the significance boundaries.
- **Power 95% LCL and UCL** are the lower and upper confidence limits for the power estimate. The width of the interval is based on the number of simulations.
- **Target Alpha** is the user-specified probability of rejecting a true null hypothesis. It is the total alpha spent.
- **Alpha or Actual Alpha** is the alpha level that was actually achieved by the experiment. It is the total proportion of null hypothesis simulations that are outside the significance boundaries.
- **Alpha 95% LCL and UCL** are the lower and upper confidence limits for the actual alpha estimate. The width of the interval is based on the number of simulations.
- **Beta** is the probability of accepting a false null hypothesis. It is the total proportion of alternative hypothesis simulations that do not cross the significance boundaries.
- **N1 and N2** are the sample sizes of each group if the study reaches the final look.
- **Average Sample Size** Given H0 Grp1 and Grp2 are the average or expected sample sizes of each group if H0 is true. These are based on the proportion of null hypothesis simulations that cross the significance or futility boundaries at each look.
- **Average Sample Size** Given H1 Grp1 and Grp2 are the average or expected sample sizes of each group if H1 is true. These are based on the proportion of alternative hypothesis simulations that cross the significance or futility boundaries at each look.
- **H0 Diff0** is the mean difference between groups (Grp1 - Grp2) assuming the null hypothesis, H0.
- **H1 Diff1** is the mean difference between groups (Grp1 - Grp2) assuming the alternative hypothesis, H1.
- The parameters to the right of H1 Diff1 are the parameters that were set by the user to define the null and alternative simulation distributions.

**Summary Statements**

Group sequential trials with sample sizes of 500 and 500 at the final look achieve 88% power to detect a difference of 5.0 between the null hypothesis mean difference of 0.0 and the actual mean difference of -5.0 at the 0.0496 significance level (alpha) using a two-sided T-Test. These results are based on 10000 Monte Carlo samples from the null distributions: Normal(M0 S) and Normal(M0 S), and the alternative distributions: Normal(M0 S) and Normal(M1 S).
Accumulated Information Details for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>Percent</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20.0</td>
<td>100</td>
<td>100</td>
<td>200</td>
</tr>
<tr>
<td>2</td>
<td>40.0</td>
<td>200</td>
<td>200</td>
<td>400</td>
</tr>
<tr>
<td>3</td>
<td>60.0</td>
<td>300</td>
<td>300</td>
<td>600</td>
</tr>
<tr>
<td>4</td>
<td>80.0</td>
<td>400</td>
<td>400</td>
<td>800</td>
</tr>
<tr>
<td>5</td>
<td>100.0</td>
<td>500</td>
<td>500</td>
<td>1000</td>
</tr>
</tbody>
</table>

Accumulated Information Details Definitions
Look is the number of the look.
Accumulated Information Percent is the percent of the sample size accumulated up to the corresponding look.
Accumulated Sample Size Group 1 is total number of individuals in group 1 at the corresponding look.
Accumulated Sample Size Group 2 is total number of individuals in group 2 at the corresponding look.
Accumulated Sample Size Total is total number of individuals in the study (group 1 + group 2) at the corresponding look.

Boundaries for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>T-Value Scale</th>
<th>P-Value Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+/- 4.0302</td>
<td>0.0001</td>
</tr>
<tr>
<td>2</td>
<td>+/- 3.3336</td>
<td>0.0009</td>
</tr>
<tr>
<td>3</td>
<td>+/- 2.7016</td>
<td>0.0071</td>
</tr>
<tr>
<td>4</td>
<td>+/- 2.2941</td>
<td>0.0220</td>
</tr>
<tr>
<td>5</td>
<td>+/- 2.0350</td>
<td>0.0421</td>
</tr>
</tbody>
</table>

Boundaries Definitions
Look is the number of the look.
Significance Boundary T-Value Scale is the value such that statistics outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. They are sometimes called efficacy boundaries.
Significance Boundary P-Value Scale is the value such that P-Values outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. This P-Value corresponds to the T-Value Boundary and is sometimes called the nominal alpha.

Boundary Plot

![Boundary Plot](image-url)
Boundary Plot - P-Value

Boundary Plot

Significance Boundaries with 95% Simulation Confidence Intervals for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>T-Value Boundary Value</th>
<th>95% LCL</th>
<th>95% UCL</th>
<th>P-Value Boundary Value</th>
<th>95% LCL</th>
<th>95% UCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+/- 4.0302</td>
<td>0.0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>+/- 3.3336</td>
<td>-3.5614</td>
<td>-3.1578</td>
<td>0.0009</td>
<td>0.0004</td>
<td>0.0017</td>
</tr>
<tr>
<td>3</td>
<td>+/- 2.7016</td>
<td>-2.8012</td>
<td>-2.6176</td>
<td>0.0071</td>
<td>0.0053</td>
<td>0.0091</td>
</tr>
<tr>
<td>4</td>
<td>+/- 2.2941</td>
<td>-2.3677</td>
<td>-2.2288</td>
<td>0.0220</td>
<td>0.0181</td>
<td>0.0261</td>
</tr>
<tr>
<td>5</td>
<td>+/- 2.0350</td>
<td>-2.0937</td>
<td>-1.9996</td>
<td>0.0421</td>
<td>0.0365</td>
<td>0.0458</td>
</tr>
</tbody>
</table>

Significance Boundary Confidence Limit Definitions

Look is the number of the look.

T-Value Boundary Value is the value such that statistics outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. They are sometimes called efficacy boundaries.

P-Value Boundary Value is the value such that P-Values outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. This P-Value corresponds to the T-Value Boundary and is sometimes called the nominal alpha.

95% LCL and UCL are the lower and upper confidence limits for the boundary at the given look. The width of the interval is based on the number of simulations.

Alpha-Spending and Null Hypothesis Simulation Details for Scenario 1

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+/- 4.0302</td>
<td>0.0001</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0043</td>
<td>0.0043</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>+/- 3.3336</td>
<td>0.0009</td>
<td>0.0008</td>
<td>0.0008</td>
<td>0.0009</td>
<td>0.0880</td>
<td>0.0922</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>+/- 2.7016</td>
<td>0.0071</td>
<td>0.0068</td>
<td>0.0076</td>
<td>0.0061</td>
<td>0.3075</td>
<td>0.3997</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>+/- 2.2941</td>
<td>0.0220</td>
<td>0.0168</td>
<td>0.0244</td>
<td>0.0167</td>
<td>0.3080</td>
<td>0.7077</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>+/- 2.0350</td>
<td>0.0421</td>
<td>0.0256</td>
<td>0.0500</td>
<td>0.0246</td>
<td>0.1694</td>
<td>0.8770</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Alpha-Spending Details Definitions
Look is the number of the look.
Significance Boundary T-Value Scale is the value such that statistics outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. They are sometimes called efficacy boundaries.
Significance Boundary P-Value Scale is the value such that P-Values outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. This P-Value corresponds to the Significance T-Value Boundary and is sometimes called the nominal alpha.
Spending Function Alpha is the intended portion of alpha allocated to the particular look based on the alpha-spending function.
Cumulative Spending Function Alpha is the intended accumulated alpha allocated to the particular look. It is the sum of the Spending Function Alpha up to the corresponding look.
Alpha Spent is the proportion of the null hypothesis simulations resulting in statistics outside the Significance Boundary at this look.
Cumulative Alpha Spent is the proportion of the null hypothesis simulations resulting in Significance Boundary termination up to and including this look. It is the sum of the Alpha Spent up to the corresponding look.
Proportion H1 Sims Outside Significance Boundary is the proportion of the alternative hypothesis simulations resulting in statistics outside the Significance Boundary at this look. It may be thought of as the incremental power.
Cumulative H1 Sims Outside Significance Boundary is the proportion of the alternative hypothesis simulations resulting in Significance Boundary termination up to and including this look. It is the sum of the Proportion H1 Sims Outside Significance Boundary up to the corresponding look.

Run Time: 69.77 seconds.

The values obtained from any given run of this example will vary slightly due to the variation in simulations.

Example 2 – Power for One-Sided Test with Futility Boundaries
Suppose researchers would like to compare two treatments with a one-sided test at each look. Further, suppose they would like to terminate the study early when it can be deemed highly unlikely that the new treatment is better than the standard. Suppose the control group mean is 108. The researchers wish to know the power of the test if the treatment group mean is 113. The sample size at the final look is to be 500 per group. Testing will be done at the 0.05 significance level. A total of five tests are going to be performed on the data as they are obtained. The O’Brien-Fleming (Analog) boundaries will be used for both significance and futility boundaries.

Find the power and test boundaries assuming equal sample sizes per arm and one-sided hypothesis tests.

Setup
This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Group-Sequential Tests for Two Means (Simulation) procedure window by expanding Means, then Two Independent Means, then clicking on Group-Sequential, and then clicking on Group-Sequential Tests for Two Means (Simulation). You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.
**Option** | **Value**
--- | ---
**Data Tab**
Find (Solve For) | Power
Power | Ignored since this is the Find setting
Alpha | 0.05
N1 (Sample Size Group 1) | 500
N2 (Sample Size Group 2) | Use R
R (Allocation Ratio) | 1.0
Group 1 Dist'n | H0 | N(M0 S)
Group 2 Dist'n | H0 | N(M0 S)
Group 1 Dist'n | H1 | N(M0 S)
Group 2 Dist'n | H1 | N(M1 S)
M0 (Mean|H0) | 108
M1 (Mean|H1) | 113
S | 25
Alternative Hypothesis | Mean1 - Mean2 < Diff0
Simulations | 20000
Test Type | T-Test
**Looks and Boundaries Tab**
Specification of Looks and Boundaries | Simple
Number of Equally Spaced Looks | 5
Alpha Spending Function | O'Brien-Fleming Analog
Type of Futility Boundary | Non-Binding
Number of Skipped Futility Looks | 0
Beta Spending Function | O'Brien-Fleming Analog

---

**Output**

Click the Run button to perform the calculations and generate the following output.

**Numeric Results and Plots**

<table>
<thead>
<tr>
<th>Scenario 1 Numeric Results for Group Sequential Testing Mean Difference = Diff0.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypotheses: H0: Diff1=Diff0; H1: Diff1&lt;Diff0</td>
</tr>
<tr>
<td>H0 Dist's: Normal(M0 S) &amp; Normal(M0 S)</td>
</tr>
<tr>
<td>H1 Dist's: Normal(M0 S) &amp; Normal(M1 S)</td>
</tr>
<tr>
<td>Test Statistic: T-Test</td>
</tr>
<tr>
<td>Alpha-Spending Function: O'Brien-Fleming Analog</td>
</tr>
<tr>
<td>Beta-Spending Function: O'Brien-Fleming Analog</td>
</tr>
<tr>
<td>Futility Boundary Type: Non-binding</td>
</tr>
<tr>
<td>Number of Looks: 5</td>
</tr>
<tr>
<td>Simulations: 20000</td>
</tr>
<tr>
<td>Pool Size: 40000</td>
</tr>
</tbody>
</table>

**Numeric Summary for Scenario 1**

<table>
<thead>
<tr>
<th>Value</th>
<th>Power</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9143</td>
<td>0.9104</td>
<td>0.0500</td>
</tr>
</tbody>
</table>
----- Average Sample Size ----
-- Given H0 --  -- Given H1 --

<table>
<thead>
<tr>
<th>N1</th>
<th>N2</th>
<th>Grp1</th>
<th>Grp2</th>
<th>Grp1</th>
<th>Grp2</th>
<th>Diff0</th>
<th>Diff1</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>500</td>
<td>302</td>
<td>302</td>
<td>335</td>
<td>335</td>
<td>0.0</td>
<td>-5.0</td>
<td>108.0</td>
<td>113.0</td>
<td>25.0</td>
</tr>
</tbody>
</table>

Accumulated Information Details for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>Percent</th>
<th>Accumulated Information</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20.0</td>
<td></td>
<td>100</td>
<td>100</td>
<td>200</td>
</tr>
<tr>
<td>2</td>
<td>40.0</td>
<td></td>
<td>200</td>
<td>200</td>
<td>400</td>
</tr>
<tr>
<td>3</td>
<td>60.0</td>
<td></td>
<td>300</td>
<td>300</td>
<td>600</td>
</tr>
<tr>
<td>4</td>
<td>80.0</td>
<td></td>
<td>400</td>
<td>400</td>
<td>800</td>
</tr>
<tr>
<td>5</td>
<td>100.0</td>
<td></td>
<td>500</td>
<td>500</td>
<td>1000</td>
</tr>
</tbody>
</table>

Boundaries for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>T-Value</th>
<th>P-Value</th>
<th>T-Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-4.5851</td>
<td>0.0000</td>
<td>2.1482</td>
<td>0.9835</td>
</tr>
<tr>
<td>2</td>
<td>-2.8811</td>
<td>0.0021</td>
<td>0.4558</td>
<td>0.6756</td>
</tr>
<tr>
<td>3</td>
<td>-2.3230</td>
<td>0.00103</td>
<td>-0.5014</td>
<td>0.3081</td>
</tr>
<tr>
<td>4</td>
<td>-1.9357</td>
<td>0.0266</td>
<td>-1.1580</td>
<td>0.1236</td>
</tr>
<tr>
<td>5</td>
<td>-1.7361</td>
<td>0.0414</td>
<td>-1.7361</td>
<td>0.0414</td>
</tr>
</tbody>
</table>

Boundary Plot
Boundary Plot - P-Value

Significance Boundaries with 95% Simulation Confidence Intervals for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>T-Value Boundary</th>
<th>P-Value Boundary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Value</td>
<td>95% LCL</td>
</tr>
<tr>
<td>1</td>
<td>-4.5851</td>
<td>-2.8960</td>
</tr>
<tr>
<td>2</td>
<td>-2.8811</td>
<td>-2.3879</td>
</tr>
<tr>
<td>3</td>
<td>-2.3230</td>
<td>-2.2690</td>
</tr>
<tr>
<td>4</td>
<td>-1.9357</td>
<td>-1.9122</td>
</tr>
<tr>
<td>5</td>
<td>-1.7361</td>
<td>-1.7013</td>
</tr>
</tbody>
</table>

Futility Boundaries with 95% Simulation Confidence Intervals for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>T-Value Boundary</th>
<th>P-Value Boundary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Value</td>
<td>95% LCL</td>
</tr>
<tr>
<td>1</td>
<td>2.1482</td>
<td>-0.5014</td>
</tr>
<tr>
<td>2</td>
<td>0.4558</td>
<td>-0.5014</td>
</tr>
<tr>
<td>3</td>
<td>-0.5014</td>
<td>-0.5346</td>
</tr>
<tr>
<td>4</td>
<td>-1.1580</td>
<td>-1.1286</td>
</tr>
<tr>
<td>5</td>
<td>-1.7361</td>
<td>-1.7055</td>
</tr>
</tbody>
</table>

Alpha-Spending and Null Hypothesis Simulation Details for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>T-Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Signif. Boundary</td>
<td>Target</td>
</tr>
<tr>
<td></td>
<td>Spending Function</td>
<td>Actual</td>
</tr>
<tr>
<td></td>
<td>Cum. Alpha Spent</td>
<td>Proportion H0 Sims</td>
</tr>
<tr>
<td></td>
<td>Cum. H0 Sims</td>
<td>Cum. H0 Sims</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Look</th>
<th>T-Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Signif. Boundary</td>
<td>Target</td>
</tr>
<tr>
<td></td>
<td>Spending Function</td>
<td>Actual</td>
</tr>
<tr>
<td></td>
<td>Cum. Alpha Spent</td>
<td>Proportion H0 Sims</td>
</tr>
<tr>
<td></td>
<td>Cum. H0 Sims</td>
<td>Cum. H0 Sims</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Look</th>
<th>T-Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Signif. Boundary</td>
<td>Target</td>
</tr>
<tr>
<td></td>
<td>Spending Function</td>
<td>Actual</td>
</tr>
<tr>
<td></td>
<td>Cum. Alpha Spent</td>
<td>Proportion H0 Sims</td>
</tr>
<tr>
<td></td>
<td>Cum. H0 Sims</td>
<td>Cum. H0 Sims</td>
</tr>
</tbody>
</table>
### Example 3 – Enter Boundaries

With a set-up similar to Example 2, suppose we wish to investigate the properties of a set of significance (-3, -3, -3, -2, -1) and futility (2, 1, 0, 0, -1) boundaries.

#### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Group-Sequential Tests for Two Means (Simulation)** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Group-Sequential**, and then clicking on **Group-Sequential Tests for Two Means (Simulation)**. You may then make the appropriate entries as listed below, or open Example 3 by going to the **File** menu and choosing **Open Example Template**.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Tab</strong></td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Alpha and Power (Enter Boundaries)</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>500</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Group 1 Dist’n</td>
<td>N(M0 S)</td>
</tr>
<tr>
<td>Group 2 Dist’n</td>
<td>N(M0 S)</td>
</tr>
<tr>
<td>Group 1 Dist’n</td>
<td>N(M0 S)</td>
</tr>
<tr>
<td>Group 2 Dist’n</td>
<td>N(M1 S)</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H1)</td>
</tr>
<tr>
<td>S</td>
<td>25</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Mean1 - Mean2 &lt; Diff0</td>
</tr>
<tr>
<td>Simulations</td>
<td>20000</td>
</tr>
<tr>
<td>Test Type</td>
<td>T-Test</td>
</tr>
</tbody>
</table>

The values obtained from any given run of this example will vary slightly due to the variation in simulations.

#### Beta-Spending and Alternative Hypothesis Simulation Details for Scenario 1

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.1482</td>
<td>0.9835</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0001</td>
<td>0.0009</td>
<td>0.0009</td>
<td>0.0009</td>
</tr>
<tr>
<td>2</td>
<td>0.4558</td>
<td>0.6756</td>
<td>0.0065</td>
<td>0.0066</td>
<td>0.0066</td>
<td>0.0066</td>
<td>0.0066</td>
<td>0.1872</td>
<td>0.1872</td>
<td>0.1872</td>
</tr>
<tr>
<td>3</td>
<td>-0.5014</td>
<td>0.3081</td>
<td>0.0200</td>
<td>0.0267</td>
<td>0.0267</td>
<td>0.0267</td>
<td>0.0266</td>
<td>0.3640</td>
<td>0.3640</td>
<td>0.5521</td>
</tr>
<tr>
<td>4</td>
<td>-1.1580</td>
<td>0.1236</td>
<td>0.0283</td>
<td>0.0549</td>
<td>0.0549</td>
<td>0.0549</td>
<td>0.2671</td>
<td>0.8192</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-1.7361</td>
<td>0.0414</td>
<td>0.0311</td>
<td>0.0860</td>
<td>0.0860</td>
<td>0.0857</td>
<td>0.0952</td>
<td>0.9143</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Run Time: 1.99 minutes.
Enter Boundaries Tab
Number of Looks.................................5
Types of Boundaries..............................Significance and Futility Boundaries
Significance Boundary............................-3 -3 -3 -2 -1 (for looks 1 through 5)
Futility Boundary.................................2 1 0 0 -1 (for looks 1 through 5)

Output
Click the Run button to perform the calculations and generate the following output.

Numeric Results and Plots

Scenario 1 Numeric Results for Group Sequential Testing Mean Difference = Diff0.
Hypotheses: H0: Diff1=Diff0; H1: Diff1<Diff0
H0 Dist's: Normal(M0 S) & Normal(M0 S)
H1 Dist's: Normal(M0 S) & Normal(M1 S)
Test Statistic: T-Test
Type of Boundaries: Significance and Futility Boundaries
Number of Looks: 5
Simulations: 20000
Pool Size: 40000

Numeric Summary for Scenario 1

<table>
<thead>
<tr>
<th>Value</th>
<th>95% LCL</th>
<th>95% UCL</th>
<th>Value</th>
<th>95% LCL</th>
<th>95% UCL</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9801</td>
<td>0.9781</td>
<td>0.9820</td>
<td>0.1529</td>
<td>0.1479</td>
<td>0.1578</td>
<td>0.0199</td>
</tr>
</tbody>
</table>

----- Average Sample Size ----
-- Given H0 --    -- Given H1 --
N1   N2   Grp1 Grp2 Grp1 Grp2 Diff0 Diff1 M0  M1  S
500  500  370 370 364 364   0.0  -5.0  108.0 113.0 25.0

Accumulated Information Details for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>Accumulated Information</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20.0</td>
<td>100</td>
<td>100</td>
<td>200</td>
</tr>
<tr>
<td>2</td>
<td>40.0</td>
<td>200</td>
<td>200</td>
<td>400</td>
</tr>
<tr>
<td>3</td>
<td>60.0</td>
<td>300</td>
<td>300</td>
<td>600</td>
</tr>
<tr>
<td>4</td>
<td>80.0</td>
<td>400</td>
<td>400</td>
<td>800</td>
</tr>
<tr>
<td>5</td>
<td>100.0</td>
<td>500</td>
<td>500</td>
<td>1000</td>
</tr>
</tbody>
</table>

Boundaries for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>T-Value Scale</th>
<th>P-Value Scale</th>
<th>T-Value Scale</th>
<th>P-Value Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-3.0000</td>
<td>0.0015</td>
<td>2.0000</td>
<td>0.9766</td>
</tr>
<tr>
<td>2</td>
<td>-3.0000</td>
<td>0.0014</td>
<td>1.0000</td>
<td>0.8410</td>
</tr>
<tr>
<td>3</td>
<td>-3.0000</td>
<td>0.0014</td>
<td>0.0000</td>
<td>0.5000</td>
</tr>
<tr>
<td>4</td>
<td>-2.0000</td>
<td>0.0229</td>
<td>0.0000</td>
<td>0.5000</td>
</tr>
<tr>
<td>5</td>
<td>-1.0000</td>
<td>0.1588</td>
<td>-1.0000</td>
<td>0.1588</td>
</tr>
</tbody>
</table>
Group-Sequential Tests for Two Means (Simulation)  476-27

Boundary Plot

Boundary Plot - P-Value
Example 4 – Validation Using Simulation

With a set-up similar to Example 1, we examine the power and alpha generated by the set of two-sided significance boundaries (+/- 4.0302, +/- 3.3336, +/- 2.7016, +/- 2.2941, +/- 2.0350).

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Group-Sequential Tests for Two Means (Simulation) procedure window by expanding Means, then Two Independent Means, then clicking on Group-Sequential, and then clicking on Group-Sequential Tests for Two Means (Simulation). You may then make the appropriate entries as listed below, or open Example 4 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Alpha and Power (Enter Boundaries)</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>500</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Group 1 Dist’n</td>
<td>N(M0 S)</td>
</tr>
<tr>
<td>Group 2 Dist’n</td>
<td>N(M0 S)</td>
</tr>
</tbody>
</table>
### Data Tab (continued)

Group 1 Dist'n | H1 ......................... N(M0 S)
Group 2 Dist'n | H1 ......................... N(M1 S)
M0 (Mean|H0) ......................................... 108
M1 (Mean|H1) ......................................... 113
S ............................................................. 25
Alternative Hypothesis .................................. Mean1 - Mean2 ≠ Diff0
Simulations ............................................. 20000
Test Type ................................................ T-Test

### Enter Boundaries Tab

Number of Looks ..................................... 5
Types of Boundaries .................................. Significance Boundaries
Significance Boundary ............................... 4.0302, 3.3336, 2.7016, 2.2941, 2.0350

---

### Output

Click the Run button to perform the calculations and generate the following output.

**Numeric Results and Plots**

---

**Scenario 1 Numeric Results for Group Sequential Testing Mean Difference = Diff0.**

Hypotheses: H0: Diff1=Diff0; H1: Diff1<>Diff0

H0 Dist's: Normal(M0 S) & Normal(M0 S)
H1 Dist's: Normal(M0 S) & Normal(M1 S)

Test Statistic: T-Test

Type of Boundaries: Significance Boundaries Only
Number of Looks: 5
Simulations: 20000
Pool Size: 40000

**Numeric Summary for Scenario 1**

<table>
<thead>
<tr>
<th>Value</th>
<th>95% LCL</th>
<th>95% UCL</th>
<th>Value</th>
<th>95% LCL</th>
<th>95% UCL</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8827</td>
<td>0.8782</td>
<td>0.8872</td>
<td>0.0530</td>
<td>0.0498</td>
<td>0.0561</td>
<td>0.1173</td>
</tr>
</tbody>
</table>

---

Average Sample Size

<table>
<thead>
<tr>
<th>N1</th>
<th>N2</th>
<th>Grp1</th>
<th>Grp2</th>
<th>Grp1</th>
<th>Grp2</th>
<th>Diff0</th>
<th>Diff1</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>500</td>
<td>496</td>
<td>496</td>
<td>378</td>
<td>378</td>
<td>0.0</td>
<td>-5.0</td>
<td>108</td>
<td>113</td>
<td>25.0</td>
</tr>
</tbody>
</table>

Run Time: 1.90 minutes.

The values obtained from any given run of this example will vary slightly due to the variation in simulations. The power and alpha generated with these boundaries are very close to the values of Example 1.
Chapter 477

Group-Sequential Tests for Two Means Assuming Normality (Simulation)

Introduction

This procedure can be used to determine power, sample size and/or boundaries for group sequential tests comparing the means of two groups. The common two-sample T-Test and the Mann-Whitney U test can be simulated in this procedure. For two-sided tests, significance (efficacy) boundaries can be generated. For one-sided tests, significance and futility boundaries can be produced. The spacing of the looks can be equal or custom specified. Boundaries can be computed based on popular alpha- and beta-spending functions (O’Brien-Fleming, Pocock, Hwang-Shih-DeCani Gamma family, linear) or custom spending functions. Boundaries can also be input directly to verify alpha- and/or beta-spending properties. Futility boundaries can be binding or non-binding. Maximum and average (expected) sample sizes are reported as well as the alpha and/or beta spent and incremental power at each look. Corresponding P-Value boundaries are also given for each boundary statistic. Plots of boundaries are also produced.

Technical Details

This section outlines many of the technical details of the techniques used in this procedure including the simulation summary, the test statistic details, and the use of spending functions.

An excellent text for the background and details of many group-sequential methods is Jennison and Turnbull (2000).
Simulation Procedure

In this procedure, a large number of simulations are used to calculate boundaries and power using the following steps

1. Based on the specified distributions, random samples of size $N_1$ and $N_2$ are generated under the null distribution and under the alternative distribution. These are simulated samples as though the final look is reached.

2. For each sample, test statistics for each look are produced. For example, if $N_1$ and $N_2$ are 100 and there are 5 equally spaced looks, test statistics are generated from the random samples at $N_1 = N_2 = 20$, $N_1 = N_2 = 40$, $N_1 = N_2 = 60$, $N_1 = N_2 = 80$, and $N_1 = N_2 = 100$ for both null and alternative samples.

3. To generate the first significance boundary, the null distribution statistics of the first look (e.g., at $N_1 = N_2 = 20$) are ordered and the percent of alpha to be spent at the first look is determined (using either the alpha-spending function or the input value). The statistic for which the percent of statistics above (or below, as the case may be) that value is equal to the percent of alpha to be spent at the first look is the boundary statistic. It is seen here how important a large number of simulations is to the precision of the boundary estimates.

4. All null distribution samples that are outside the first significance boundary at the first look are removed from consideration for the second look. If binding futility boundaries are also being computed, all null distribution samples with statistics that are outside the first futility boundary are also removed from consideration for the second look. If non-binding futility boundaries are being computed, null distribution samples with statistics outside the first futility boundary are not removed.

5. To generate the second significance boundary, the remaining null distribution statistics of the second look (e.g., at $N_1 = N_2 = 40$) are ordered and the percent of alpha to be spent at the second look is determined (again, using either the alpha-spending function or the input value). The percent of alpha to be spent at the second look is multiplied by the total number of simulations to determine the number of the statistic that is to be the second boundary statistic. The statistic for which that number of statistics is above it (or below, as the case may be) is the second boundary statistic. For example, suppose there are initially 1000 simulated samples, with 10 removed at the first look (from, say, alpha spent at Look 1 equal to 0.01), leaving 990 samples considered for the second look. Suppose further that the alpha to be spent at the second look is 0.02. This is multiplied by 1000 to give 20. The 990 still-considered statistics are ordered and the 970th (20 in from 990) statistic is the second boundary.

6. All null distribution samples that are outside the second significance boundary and the second futility boundary, if binding, at the second look are removed from consideration for the third look (e.g., leaving 970 statistics computed at $N_1 = N_2 = 60$ to be considered at the third look). Steps 4 and 5 are repeated until the final look is reached.

Futility boundaries are computed in a similar manner using the desired beta-spending function or custom beta-spending values and the alternative hypothesis simulated statistics at each look. For both binding and non-binding futility boundaries, samples for which alternative hypothesis statistics are outside either the significance or futility boundaries of the previous look are excluded from current and future looks.

Because the final futility and significance boundaries are required to be the same, futility boundaries are computed beginning at a small value of beta (e.g., 0.0001) and incrementing beta by that amount until the futility and significance boundaries meet.
When boundaries are entered directly, this procedure uses the null hypothesis and alternative hypothesis simulations to determine the number of test statistics that are outside the boundaries at each look. The cumulative proportion of alternative hypothesis statistics that are outside the significance boundaries is the overall power of the study.

**Generating Random Distributions**

Two methods are available in PASS to simulate random samples. The first method generates the random variates directly, one value at a time. The second method generates a large pool (over 10,000) of random values and then draws the random numbers from this pool. This second method can cut the running time of the simulation by 70%.

As mentioned above, the second method begins by generating a large pool of random numbers from the specified distributions. Each of these pools is evaluated to determine if its mean is within a small relative tolerance (0.0001) of the target mean. If the actual mean is not within the tolerance of the target mean, individual members of the population are replaced with new random numbers if the new random number moves the mean towards its target. Only a few hundred such swaps are required to bring the actual mean to within tolerance of the target mean. This population is then sampled with replacement using the uniform distribution. We have found that this method works well as long as the size of the pool is the maximum of twice the number of simulated samples desired and 10,000.

**Test Statistics**

This section describes the test statistics that are available in this procedure.

**Two-Sample T-Test**

The two-sample t-test assumes that the data are a simple random sample from a population of normally-distributed values that all have the same mean and variance. This assumption implies that the data are continuous and their distribution is symmetric. The calculation of the t statistic is as follows

\[
t_{df} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{X}_1 - \bar{X}_2}}
\]

where

\[
\bar{X}_k = \frac{\sum_{i=1}^{N_k} X_{ki}}{N_k}
\]

\[
s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2 + \sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2}{N_1 + N_2 - 2} \left( \frac{1}{N_1} + \frac{1}{N_2} \right)}
\]

\[
df = N_1 + N_2 - 2
\]
The significance of the test statistic is determined by computing the p-value based on the t distribution with degrees of freedom $df$. If this p-value is less than a specified level (often 0.05), the null hypothesis is rejected. Otherwise, no conclusion can be reached.

**Mann-Whitney U Test**

This test is the nonparametric substitute for the equal-variance t-test. Two key assumptions for this test are that the distributions are at least ordinal and that they are identical under $H_0$. This implies that ties (repeated values) are not acceptable. When ties are present, the approximation provided can be used, but know that the theoretic results no longer hold.

The Mann-Whitney test statistic is defined as follows in Gibbons (1985).

$$ z = \frac{W_1 - \frac{N_1(N_1 + N_2 + 1)}{2} + C}{s_W} $$

where

$$ W_1 = \sum_{k=1}^{N_1} \text{Rank}(X_{1k}) $$

The ranks are determined after combining the two samples. The standard deviation is calculated as

$$ s_W = \sqrt{\frac{N_1 N_2 (N_1 + N_2 + 1)}{12} - \frac{N_1 N_2 \sum_{i=1}^{t_i} (t_i^3 - t_i)}{12(N_1 + N_2)(N_1 + N_2 - 1)}} $$

where $t_i$ is the number of observations tied at value one, $t_i$ is the number of observations tied at some value two, and so forth.

The correction factor, $C$, is 0.5 if the rest of the numerator of $z$ is negative or -0.5 otherwise. The value of $z$ is then compared to the standard normal distribution.

**Spending Functions**

Spending functions can be used in this procedure to specify the proportion of alpha or beta that is spent at each look without having to specify the proportion directly.

Spending functions have the characteristics that they are increasing and that

$$ \alpha(0) = 0 $$
$$ \alpha(1) = \alpha $$

The last characteristic guarantees a fixed $\alpha$ level when the trial is complete. This methodology is very flexible since neither the times nor the number of analyses must be specified in advance. Only the functional form of $\alpha(\tau)$ must be specified.

PASS provides several popular spending functions plus the ability to enter and analyze your own percents of alpha or beta spent. These are calculated as follows (beta may be substituted for alpha for beta-spending functions):
1. Hwang-Shih-DeCani (gamma family) 
\[ \alpha \left( \frac{1 - e^{-\gamma}}{1 - e^{-\gamma_0}} \right), \gamma \neq 0; \ \alpha, \gamma = 0 \]

2. O’Brien-Fleming Analog 
\[ 2 - 2 \Phi \left( \sqrt{\frac{Z_{\alpha/2}}{t}} \right) \]

3. Pocock Analog 
\[ \alpha \cdot \ln(1 + (e - 1)\gamma) \]
477-6  Group-Sequential Tests for Two Means Assuming Normality (Simulation)

4. Alpha * time $\alpha \cdot t$

5. Alpha * time$^{1.5}$ $\alpha \cdot t^{1/2}$

6. Alpha * time$^2$ $\alpha \cdot t^2$
7. \( \text{Alpha} \times \text{time}^C = \alpha \cdot t^C \)

8. **User Supplied Percents**

A custom set of percents of alpha to be spent at each look may be input directly.

The O'Brien-Fleming Analog spends very little alpha or beta at the beginning and much more at the final looks. The Pocock Analog and (Alpha or Beta)(Time) spending functions spend alpha or beta more evenly across the looks. The Hwang-Shih-DeCani (C) (gamma family) spending functions and (Alpha or Beta)(Time^C) spending functions are flexible spending functions that can be used to spend more alpha or beta early or late or evenly, depending on the choice of C.

**Procedure Options**

This section describes the options that are specific to this procedure. These are located on the Data, Looks & Boundaries, Enter Boundaries, and Options tabs. For more information about the options of other tabs, go to the Procedure Window chapter.

**Data Tab**

The Data tab contains most of the parameters and options for the general setup of the procedure.

**Solve For**

**Find (Solve For)**

Solve for either power, sample size, or enter the boundaries directly and solve for power and alpha. When solving for power or sample size, the look and boundary details are specified on the "Looks & Boundaries" tab and the "Enter Boundaries" tab is ignored. When entering the boundaries directly and solving for power and alpha, the boundaries are input on the "Enter Boundaries" tab and the "Looks & Boundaries" tab is ignored.

When solving for power or N1, the early-stopping boundaries are also calculated. High accuracy for early-stopping boundaries requires a very large number of simulations (Recommended 100,000 to 10,000,000).

The parameter selected here is the parameter displayed on the vertical axis of the plot.

Because this is a simulation based procedure, the search for the sample size may take several minutes or hours to complete. You may find it quicker and more informative to solve for Power for a range of sample sizes.
Error Rates

Power or Beta
Power is the probability of rejecting the null hypothesis when it is false. Power is equal to 1-Beta, so specifying power implicitly specifies beta.

Beta is the probability obtaining a false negative on the statistical test. That is, it is the probability of accepting a false null hypothesis.

In the context of simulated group sequential trials, the power is the proportion of the alternative hypothesis simulations that cross any one of the significance (efficacy) boundaries.

The valid range is between 0 to 1.

Different disciplines and protocols have different standards for setting power. A common choice is 0.90, but 0.80 is also popular.

You can enter a range of values such as 0.70 0.80 0.90 or 0.70 to 0.90 by 0.1.

Alpha (Significance Level)
Alpha is the probability of obtaining a false positive on the statistical test. That is, it is the probability of rejecting a true null hypothesis.

The null hypothesis is usually that the parameters (the means, proportions, etc.) are all equal.

In the context of simulated group sequential trials, alpha is the proportion of the null hypothesis simulations that cross any one of the significance (efficacy) boundaries.

Since Alpha is a probability, it is bounded by 0 and 1. Commonly, it is between 0.001 and 0.250.

Alpha is often set to 0.05 for two-sided tests and to 0.025 for one-sided tests.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.

Sample Size

N1 (Sample Size Group 1)
Enter a value for the sample size, N1. This is the number of subjects in the first group of the study at the final look.

You may enter a range such as 10 to 100 by 10 or a list of values separated by commas or blanks.

You might try entering the same number two or three times to get an idea of the variability in your results. For example, you could enter "10 10 10".

N2 (Sample Size Group 2)
Enter a value (or range of values) for the sample size of group 2. This is the number of subjects in the second group of the study at the final look.

- Use R
  When Use R is entered here, N2 is calculated using the formula

  \[ N2 = \lceil R(N1) \rceil \]

  where \( R \) is the Sample Allocation Ratio and the operator \( \lceil Y \rceil \) is the first integer greater than or equal to \( Y \). For example, if you want \( N1 = N2 \), select Use R and set \( R = 1 \).
**R (Sample Allocation Ratio)**
Enter a value (or range of values) for R, the allocation ratio between samples. This value is only used when \( N_2 \) is set to *Use R.*

\[ R = \frac{N_2}{N_1} \]

When used, \( N_2 \) is calculated from \( N_1 \) using the formula: \( N_2 = \lceil R(N_1) \rceil \) where \( \lceil Y \rceil \) is the next integer greater than or equal to \( Y \). Note that setting \( R = 1.0 \) forces \( N_2 = N_1 \).

---

**Effect Size**

**Mean1 (Mean of Group 1, Control)**
Enter a value for the hypothesized mean of this group. The difference between the two means and the direction of the difference relative to the alternative hypothesis is of primary importance here.

For the simulations under the null and alternative hypothesis, the test statistics are calculated as mean of group 1 minus mean of group 2. For the purposes of simulation, if a two-sided alternative is selected, Mean1 should be larger than Mean2.

If you want to analyze the difference, enter the value of the difference for one of the means and zero for the other.

You may enter a range of values such as 10 20 30 or 0 to 100 by 25.

**Mean2 (Mean of Group 2, Treatment | H1)**
Enter a value for the hypothesized mean of this group. The difference between the two means and the direction of the difference relative to the alternative hypothesis is of primary importance here.

For the simulations under the null and alternative hypothesis, the test statistics are calculated as mean of group 1 minus mean of group 2. For the purposes of simulation, if a two-sided alternative is selected, Mean1 should be larger than Mean2.

If you want to analyze the difference, enter the value of the difference for one of the means and zero for the other.

You may enter a range of values such as 10 20 30 or 0 to 100 by 25.

**Standard Deviation**
Enter the desired value for the standard deviation of the population from which simulations will be generated. The standard deviation must be a positive number.

This standard deviation is the standard deviation of each group.

Press the "SD" button to obtain help on estimating the standard deviation.

You can enter a range of values such as 1 2 3 or 1 to 10 by 1.
Test and Simulations

Alternative Hypothesis

Specify the alternative hypothesis of the test. Since the null hypothesis is the opposite, specifying the alternative is the only hypothesis needed. When a two-sided alternative is selected, futility boundaries are not permitted. If you wish to do a two-sided test with futility boundaries you may consider two one-sided tests.

Care need be taken in the choice of direction of the test relative to the choice of Mean1 and Mean2. For example, for a one-sided test to show that the true mean for group 1 (control) is less than the true mean for group 2 (treatment), Mean2 should be larger than Mean1 and the alternative hypothesis chosen should be Mean1 < Mean2.

Test Type

Specify which test statistic is to be reported on.

The T-Test is the standard T-test based on the pooled variance and SS1 + SS2 - 2 degrees of freedom, where SS1 and SS2 are the sample sizes of groups 1 and 2 at the corresponding look.

The Mann-Whitney Test is the common non-parametric alternative and is based on the ranks of the observations. It is sometimes called the Mann-Whitney U, Mann-Whitney-Wilcoxon, Wilcoxon Rank-Sum, or Wilcoxon-Mann-Whitney test.

Simulations

This option specifies the number of iterations, M, used in the simulation. As the number of iterations is increased, the accuracy and running time of the simulation will be increased also.

The precision of the simulated power estimates are calculated from the binomial distribution. Thus, confidence intervals may be constructed for various power values. The following table gives an estimate of the precision that is achieved for various simulation sizes when the power is either 0.50 or 0.95. The table values are interpreted as follows: a 95% confidence interval of the true power is given by the power reported by the simulation plus and minus the ‘Precision’ amount given in the table.

<table>
<thead>
<tr>
<th>Simulation Size M</th>
<th>Precision when Power = 0.50</th>
<th>Precision when Power = 0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.100</td>
<td>0.044</td>
</tr>
<tr>
<td>500</td>
<td>0.045</td>
<td>0.019</td>
</tr>
<tr>
<td>1000</td>
<td>0.032</td>
<td>0.014</td>
</tr>
<tr>
<td>2000</td>
<td>0.022</td>
<td>0.010</td>
</tr>
<tr>
<td>5000</td>
<td>0.014</td>
<td>0.006</td>
</tr>
<tr>
<td>10000</td>
<td>0.010</td>
<td>0.004</td>
</tr>
<tr>
<td>50000</td>
<td>0.004</td>
<td>0.002</td>
</tr>
<tr>
<td>100000</td>
<td>0.003</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Notice that a simulation size of 1000 gives a precision of plus or minus 0.01 when the true power is 0.95. Also note that as the simulation size is increased beyond 5000, there is only a small amount of additional accuracy achieved.

However, when solving for Power or N1, the simulations are also used to calculate the look boundaries. To obtain precise boundary estimates, the number of simulations needs to be high. This consideration competes with the length of time to complete the simulation. When solving for power, a large number of simulations (100,000 or 1,000,000) will finish in several minutes. When
solving for N1, perhaps 10,000 simulations can be run for each iteration. Then, a final run with the resulting N1 solving for power can be run with more simulations.

**Looks & Boundaries Tab**

The Data tab contains most of the parameters and options for the general setup of the procedure.

**Looks and Boundaries**

**Specification of Looks and Boundaries**
Choose whether spending functions will be used to divide alpha and beta for each look (Simple Specification), or whether the percents of alpha and beta to be spent at each look will be specified directly (Custom Specification).

Under Simple Specification, the looks are automatically considered to be equally spaced. Under Custom Specification, the looks may be equally spaced or custom defined based on the percent of accumulated information.

**Looks and Boundaries – Simple Specification**

**Number of Equally Spaced Looks**
Select the total number of looks that will be used if the study is not stopped early for the crossing of a boundary.

**Alpha Spending Function**
Specify the type of alpha spending function to use.

The O'Brien-Fleming Analog spends very little alpha at the beginning and much more at the final looks. The Pocock Analog and (Alpha)(Time) spending functions spend alpha more evenly across the looks. The Hwang-Shih-DeCani (C) (sometimes called the gamma family) spending functions and (Alpha)(Time^C) spending functions are flexible spending functions that can be used to spend more alpha early or late or evenly, depending on the choice of C.

**C (Alpha Spending)**
C is used to define the Hwang-Shih-DeCani (C) or (Alpha)(Time^C) spending functions.

For the Hwang-Shih-DeCani (C) spending function, negative values of C spend more of alpha at later looks, values near 0 spend alpha evenly, and positive values of C spend more of alpha at earlier looks.

For the (Alpha)(Time^C) spending function, only positive values for C are permitted. Values of C near zero spend more of alpha at earlier looks, values near 1 spend alpha evenly, and larger values of C spend more of alpha at later looks.

**Type of Futility Boundary**
This option determines whether or not futility boundaries will be created, and if so, whether they are binding or non-binding.

Futility boundaries are boundaries such that, if crossed at a given look, stop the study in favor of H0.
Binding futility boundaries are computed in concert with significance boundaries. They are called binding because they require the stopping of a trial if they are crossed. If the trial is not stopped, the probability of a false positive will exceed alpha.

When Non-binding futility boundaries are computed, the significance boundaries are first computed, ignoring the futility boundaries. The futility boundaries are then computed. These futility boundaries are non-binding because continuing the trial after they are crossed will not affect the overall probability of a false positive declaration.

**Number of Skipped Futility Looks**

In some trials it may be desirable to wait a number of looks before examining the trial for futility. This option allows the beta to begin being spent after a specified number of looks.

The Number of Skipped Futility Looks should be less than the number of looks.

**Beta Spending Function**

Specify the type of beta spending function to use.

The O'Brien-Fleming Analog spends very little beta at the beginning and much more at the final looks. The Pocock Analog and (Beta)(Time) spending functions spend beta more evenly across the looks. The Hwang-Shih-DeCani (C) (sometimes called the gamma family) spending functions and (Beta)(Time^C) spending functions are flexible spending functions that can be used to spend more beta early or late or evenly, depending on the choice of C.

**C (Beta Spending)**

C is used to define the Hwang-Shih-DeCani (C) or (Beta)(Time^C) spending functions.

For the Hwang-Shih-DeCani (C) spending function, negative values of C spend more of beta at later looks, values near 0 spend beta evenly, and positive values of C spend more of beta at earlier looks.

For the (Beta)(Time^C) spending function, only positive values for C are permitted. Values of C near zero spend more of beta at earlier looks, values near 1 spend beta evenly, and larger values of C spend more of beta at later looks.

**Looks and Boundaries – Custom Specification**

**Number of Looks**

This is the total number of looks of either type (significance or futility or both).

**Equally Spaced**

If this box is checked, the Accumulated Information boxes are ignored and the accumulated information is evenly spaced.

**Type of Futility Boundary**

This option determines whether or not futility boundaries will be created, and if so, whether they are binding or non-binding.

Futility boundaries are boundaries such that, if crossed at a given look, stop the study in favor of H0.

Binding futility boundaries are computed in concert with significance boundaries. They are called binding because they require the stopping of a trial if they are crossed. If the trial is not stopped,
the probability of a false positive will exceed alpha.

When Non-binding futility boundaries are computed, the significance boundaries are first computed, ignoring the futility boundaries. The futility boundaries are then computed. These futility boundaries are non-binding because continuing the trial after they are crossed will not affect the overall probability of a false positive declaration.

**Accumulated Information**

The accumulated information at each look defines the proportion or percent of the sample size that is used at that look.

These values are accumulated information values so they must be increasing.

Proportions, percents, or sample sizes may be entered. All proportions, percents, or sample sizes will be divided by the value at the final look to create an accumulated information proportion for each look.

**Percent of Alpha Spent**

This is the percent of the total alpha that is spent at the corresponding look. It is not the cumulative value.

Percentages, proportions, or alphas may be entered here. Each of the values is divided by the sum of the values to obtain the proportion of alpha that is used at the corresponding look.

**Percent of Beta Spent**

This is the percent of the total beta (1-power) that is spent at the corresponding look. It is not the cumulative value.

Percentages, proportions, or betas may be entered here. Each of the values is divided by the sum of the values to obtain the proportion of beta that is used at the corresponding look.

**Enter Boundaries Tab**

The Data tab contains most of the parameters and options for the general setup of the procedure.

---

**Looks and Boundaries**

**Number of Looks**

This is the total number of looks of either type (significance or futility or both).

**Equally Spaced**

If this box is checked, the Accumulated Information boxes are ignored and the accumulated information is evenly spaced.

**Types of Boundaries**

This option determines whether or not futility boundaries will be entered.

Futility boundaries are boundaries such that, if crossed at a given look, stop the study in favor of H0.
Accumulated Information
The accumulated information at each look defines the proportion or percent of the sample size that is used at that look.
These values are accumulated information values so they must be increasing.
Proportions, percents, or sample sizes may be entered. All proportions, percents, or sample sizes will be divided by the value at the final look to create an accumulated information proportion for each look.

Significance Boundary
Enter the value of the significance boundary corresponding to the chosen test statistic. These are sometimes called efficacy boundaries.

Futility Boundary
Enter the value of the futility boundary corresponding to the chosen test statistic.

Options Tab
The Options tab contains limits on the number of iterations and various options about individual tests.

Maximum Iterations

Maximum Iterations Before Search Termination
Specify the maximum N1 before the search for N1 is aborted.
Since simulations for large sample sizes are very computationally intensive and hence time-consuming, this value can be used to stop searches when N1 is larger than reasonable sample sizes for the study.
This applies only when "Find (Solve For)" is set to N1.
The procedure uses a binary search when searching for N1. If a value for N1 is tried that exceeds this value, and the power is not reached, a warning message will be shown on the output indicating the desired power was not reached.
We recommend a value of at least 20000.

Random Numbers

Random Number Pool Size
This is the size of the pool of random values from which the random samples will be drawn. Pools should be at least the maximum of 10,000 and twice the number of simulations. You can enter Automatic and an appropriate value will be calculated.
If you do not want to draw numbers from a pool, enter 0 here.
Matching Boundaries at Final Look

Beta Search Increment
For each simulation, when futility bounds are computed, the appropriate beta is found by searching from 0 to 1 by this increment. Smaller increments are more refined, but the search takes longer.

We recommend 0.001 or 0.0001.

Example 1 – Power and Output

A clinical trial is to be conducted over a two-year period to compare the mean response of a new treatment to that of the current treatment. The current response mean is 108. Although the researchers do not know the true mean of the new treatment, they would like to examine the power that is achieved if the mean of the new treatment is 113. The standard deviation for both groups is assumed to be 25. The sample size at the final look is to be 500 per group. Testing will be done at the 0.05 significance level. A total of five tests are going to be performed on the data as they are obtained. The O’Brien-Fleming (Analog) boundaries will be used.

Find the power and test boundaries assuming equal sample sizes per arm and two-sided hypothesis tests.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Group-Sequential Tests for Two Means Assuming Normality (Simulation) procedure window by expanding Means, then Two Independent Means, then clicking on Group-Sequential, and then clicking on Group-Sequential Tests for Two Means Assuming Normality (Simulation). You may then make the entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>500</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Mean1 (Control)</td>
<td>108</td>
</tr>
<tr>
<td>Mean2 (Treatment</td>
<td>H1)</td>
</tr>
<tr>
<td>S</td>
<td>25</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Mean1 ≠ Mean2</td>
</tr>
<tr>
<td>Test Type</td>
<td>T-Test</td>
</tr>
<tr>
<td>Simulations</td>
<td>20000</td>
</tr>
</tbody>
</table>

| Looks and Boundaries Tab        |                            |
| Specification of Looks and Boundaries | Simple                  |
| Number of Equally Spaced Looks  | 5                          |
| Alpha Spending Function         | O’Brien-Fleming Analog     |
Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results and Plots

Scenario 1 Numeric Results for Group Sequential Testing Mean Difference = 0.
Hypotheses: H0: Mean1=Mean2; H1: Mean1≠Mean2
Test Statistic: T-Test
Alpha-Spending Function: O'Brien-Fleming Analog
Beta-Spending Function: None
Futility Boundary Type: None
Number of Looks: 5
Simulations: 20000
Pool Size: 40000

Numeric Summary for Scenario 1

<table>
<thead>
<tr>
<th>Value</th>
<th>Power 95% LCL</th>
<th>Power 95% UCL</th>
<th>Target Power</th>
<th>Actual Power</th>
<th>Alpha 95% LCL</th>
<th>Alpha 95% UCL</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.887</td>
<td>0.883</td>
<td>0.892</td>
<td>0.050</td>
<td>0.052</td>
<td>0.049</td>
<td>0.055</td>
<td>0.113</td>
</tr>
</tbody>
</table>

----- Average Sample Size ----

<table>
<thead>
<tr>
<th>N1</th>
<th>N2</th>
<th>Grp1</th>
<th>Grp2</th>
<th>Grp1</th>
<th>Grp2</th>
<th>Diff0</th>
<th>Diff1</th>
<th>P1</th>
<th>H1</th>
<th>P2</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>500</td>
<td>496</td>
<td>496</td>
<td>379</td>
<td>379</td>
<td>-5.0</td>
<td>108.0</td>
<td>113.0</td>
<td>25.0</td>
<td></td>
</tr>
</tbody>
</table>

Report Definitions

Power is the probability of rejecting a false null hypothesis at one of the looks. It is the total proportion of alternative hypothesis simulations that are outside the significance boundaries.

Power 95% LCL and UCL are the lower and upper confidence limits for the power estimate. The width of the interval is based on the number of simulations.

Target Alpha is the user-specified probability of rejecting a true null hypothesis. It is the total alpha spent.

Actual Alpha is the alpha level that was actually achieved by the experiment. It is the total proportion of the null hypothesis simulations that are outside the significance boundaries.

Alpha 95% LCL and UCL are the lower and upper confidence limits for the actual alpha estimate. The width of the interval is based on the number of simulations.

Beta is the probability of accepting a false null hypothesis. It is the total proportion of alternative hypothesis simulations that do not cross the significance boundaries.

N1 and N2 are the sample sizes of each group if the study reaches the final look.

Average Sample Size Given H0 Grp1 and Grp2 are the average or expected sample sizes of each group if H0 is true. These are based on the proportion of null hypothesis simulations that cross the significance or futility boundaries at each look.

Average Sample Size Given H1 Grp1 and Grp2 are the average or expected sample sizes of each group if H1 is true. These are based on the proportion of alternative hypothesis simulations that cross the significance or futility boundaries at each look.

Diff is the mean difference between groups (Grp1 - Grp2) assuming the alternative hypothesis, H1.

Mean1, Mean2, and Std Dev are the parameters that were set by the user to define the null and alternative simulation distributions.

Summary Statements

Group sequential trials with group sample sizes of 500 and 500 at the final look achieve 89% power to detect a difference of -5.0 at the 0.052 significance level (alpha) using a two-sided T-Test.
Accumulated Information Details for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>Percent</th>
<th>Accumulated Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Group 1</td>
</tr>
<tr>
<td>1</td>
<td>20.0</td>
<td>100</td>
</tr>
<tr>
<td>2</td>
<td>40.0</td>
<td>200</td>
</tr>
<tr>
<td>3</td>
<td>60.0</td>
<td>300</td>
</tr>
<tr>
<td>4</td>
<td>80.0</td>
<td>400</td>
</tr>
<tr>
<td>5</td>
<td>100.0</td>
<td>500</td>
</tr>
</tbody>
</table>

Accumulated Information Details Definitions

Look is the number of the look.
Accumulated Information Percent is the percent of the sample size accumulated up to the corresponding look.
Accumulated Sample Size Group 1 is total number of individuals in group 1 at the corresponding look.
Accumulated Sample Size Group 2 is total number of individuals in group 2 at the corresponding look.
Accumulated Sample Size Total is total number of individuals in the study (group 1 + group 2) at the corresponding look.

Boundaries for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>Significance Boundary T-Value Scale</th>
<th>P-Value Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+/- 4.026</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>+/- 3.396</td>
<td>0.001</td>
</tr>
<tr>
<td>3</td>
<td>+/- 2.682</td>
<td>0.008</td>
</tr>
<tr>
<td>4</td>
<td>+/- 2.302</td>
<td>0.022</td>
</tr>
<tr>
<td>5</td>
<td>+/- 1.990</td>
<td>0.047</td>
</tr>
</tbody>
</table>

Boundaries Definitions

Look is the number of the look.
Significance Boundary T-Value Scale is the value such that statistics outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. They are sometimes called efficacy boundaries.
Significance Boundary P-Value Scale is the value such that P-Values outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. This P-Value corresponds to the T-Value Boundary and is sometimes called the nominal alpha.

Boundary Plot
Significance Boundaries with 95% Simulation Confidence Intervals for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>T-Value Boundary Value</th>
<th>95% LCL</th>
<th>95% UCL</th>
<th>P-Value Boundary Value</th>
<th>95% LCL</th>
<th>95% UCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+/- 4.026</td>
<td>0.000</td>
<td></td>
<td>0.000</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>+/- 3.396</td>
<td>-3.560</td>
<td>-3.264</td>
<td>0.001</td>
<td>0.000</td>
<td>0.001</td>
</tr>
<tr>
<td>3</td>
<td>+/- 2.682</td>
<td>-2.738</td>
<td>-2.645</td>
<td>0.008</td>
<td>0.006</td>
<td>0.008</td>
</tr>
<tr>
<td>4</td>
<td>+/- 2.302</td>
<td>-2.332</td>
<td>-2.240</td>
<td>0.022</td>
<td>0.020</td>
<td>0.025</td>
</tr>
<tr>
<td>5</td>
<td>+/- 1.990</td>
<td>-2.036</td>
<td>-1.960</td>
<td>0.047</td>
<td>0.042</td>
<td>0.050</td>
</tr>
</tbody>
</table>

Significance Boundary Confidence Limit Definitions

- **Look** is the number of the look.
- **Look** is the number of the look.
- **T-Value Boundary Value** is the value such that statistics outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. They are sometimes called efficacy boundaries.
- **P-Value Boundary Value** is the value such that P-Values outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. This P-Value corresponds to the T-Value Boundary and is sometimes called the nominal alpha.
- **95% LCL and UCL** are the lower and upper confidence limits for the boundary at the given look. The width of the interval is based on the number of simulations.

Alpha-Spending and Null Hypothesis Simulation Details for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>T-Value Scale</th>
<th>P-Value Scale</th>
<th>Alpha Function</th>
<th>Alpha Spent</th>
<th>Cum. H1 Sims</th>
<th>Proportion H1 Sims</th>
<th>Cum. H1 Sims</th>
<th>H1 Sims Outside Signif. Boundary</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>+/- 4.026</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>+/- 3.396</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
<td>0.001</td>
</tr>
<tr>
<td>3</td>
<td>+/- 2.682</td>
<td>0.008</td>
<td>0.007</td>
<td>0.008</td>
<td>0.009</td>
<td>0.009</td>
<td>0.007</td>
<td>0.003</td>
</tr>
<tr>
<td>4</td>
<td>+/- 2.302</td>
<td>0.022</td>
<td>0.017</td>
<td>0.024</td>
<td>0.026</td>
<td>0.026</td>
<td>0.026</td>
<td>0.017</td>
</tr>
<tr>
<td>5</td>
<td>+/- 1.990</td>
<td>0.047</td>
<td>0.026</td>
<td>0.050</td>
<td>0.052</td>
<td>0.052</td>
<td>0.052</td>
<td>0.017</td>
</tr>
</tbody>
</table>
Example 2 – Power for One-Sided Test with Futility Boundaries

Suppose researchers would like to compare two treatments with a one-sided test at each look. Further, suppose they would like to terminate the study early when it can be deemed highly unlikely that the new treatment is better than the standard. Suppose the control group mean is 108. The researchers wish to know the power of the test if the treatment group mean is 113. The sample size at the final look is to be 500 per group. Testing will be done at the 0.05 significance level. A total of five tests are going to be performed on the data as they are obtained. The O’Brien-Fleming (Analog) boundaries will be used for both significance and futility boundaries.

Find the power and test boundaries assuming equal sample sizes per arm and one-sided hypothesis tests.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Group-Sequential Tests for Two Means Assuming Normality (Simulation) procedure window by expanding Means, then Two Independent Means, then clicking on Group-Sequential, and then clicking on Group-Sequential Tests for Two Means Assuming Normality (Simulation). You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.
Option | Value
--- | ---
**Data Tab**
Find (Solve For) | Power
Power | Ignored since this is the Find setting
Alpha | 0.05
N1 (Sample Size Group 1) | 500
N2 (Sample Size Group 2) | Use R
R (Allocation Ratio) | 1.0
Mean1 (Control) | 108
Mean2 (Treatment | H1) | 113
S | 25
Alternative Hypothesis | Mean1 < Mean2
Test Type | T-Test
Simulations | 20000

**Looks and Boundaries Tab**
Specification of Looks and Boundaries | Simple
Number of Equally Spaced Looks | 5
Alpha Spending Function | O'Brien-Fleming Analog
Type of Futility Boundary | Non-Binding
Number of Skipped Futility Looks | 0
Beta Spending Function | O'Brien-Fleming Analog

**Output**

Click the Run button to perform the calculations and generate the following output.

**Numeric Results and Plots**

| Scenario 1 Numeric Results for Group Sequential Testing Mean Difference = 0. |
|---|---|
| Hypotheses: H0: Mean1=Mean2; H1: Mean1<Mean2 |
| Test Statistic: T-Test |
| Alpha-Spending Function: O'Brien-Fleming Analog |
| Beta-Spending Function: O'Brien-Fleming Analog |
| Futility Boundary Type: Non-binding |
| Number of Looks: 5 |
| Simulations: 20000 |
| Pool Size: 40000 |

**Numeric Summary for Scenario 1**

<table>
<thead>
<tr>
<th>Value</th>
<th>Power</th>
<th>Target</th>
<th>Actual</th>
<th>Alpha</th>
<th>95% LCL</th>
<th>95% UCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.910</td>
<td>0.906</td>
<td>0.914</td>
<td>0.050</td>
<td>0.045</td>
<td>0.042</td>
<td>0.048</td>
</tr>
</tbody>
</table>
### Accumulated Information Details for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>Percent</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20.0</td>
<td>100</td>
<td>100</td>
<td>200</td>
</tr>
<tr>
<td>2</td>
<td>40.0</td>
<td>200</td>
<td>200</td>
<td>400</td>
</tr>
<tr>
<td>3</td>
<td>60.0</td>
<td>300</td>
<td>300</td>
<td>600</td>
</tr>
<tr>
<td>4</td>
<td>80.0</td>
<td>400</td>
<td>400</td>
<td>800</td>
</tr>
<tr>
<td>5</td>
<td>100.0</td>
<td>500</td>
<td>500</td>
<td>1000</td>
</tr>
</tbody>
</table>

### Boundaries for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>T-Value</th>
<th>P-Value</th>
<th>T-Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-3.850</td>
<td>0.000</td>
<td>2.368</td>
<td>0.991</td>
</tr>
<tr>
<td>2</td>
<td>-2.906</td>
<td>0.002</td>
<td>0.445</td>
<td>0.672</td>
</tr>
<tr>
<td>3</td>
<td>-2.335</td>
<td>0.010</td>
<td>-0.517</td>
<td>0.303</td>
</tr>
<tr>
<td>4</td>
<td>-1.939</td>
<td>0.026</td>
<td>-1.175</td>
<td>0.120</td>
</tr>
<tr>
<td>5</td>
<td>-1.754</td>
<td>0.040</td>
<td>-1.754</td>
<td>0.040</td>
</tr>
</tbody>
</table>

### Boundary Plot

[Boundary Plot Image]
477-22 Group-Sequential Tests for Two Means Assuming Normality (Simulation)

Boundary Plot - P-Value

Significance Boundaries with 95% Simulation Confidence Intervals for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>T-Value Boundary</th>
<th>P-Value Boundary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Value</td>
<td>95% LCL</td>
</tr>
<tr>
<td>1</td>
<td>-3.850</td>
<td>0.000</td>
</tr>
<tr>
<td>2</td>
<td>-2.906</td>
<td>-2.990</td>
</tr>
<tr>
<td>3</td>
<td>-2.335</td>
<td>-2.368</td>
</tr>
<tr>
<td>4</td>
<td>-1.939</td>
<td>-1.974</td>
</tr>
<tr>
<td>5</td>
<td>-1.754</td>
<td>-1.782</td>
</tr>
</tbody>
</table>

Futility Boundaries with 95% Simulation Confidence Intervals for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>T-Value Boundary</th>
<th>P-Value Boundary</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Value</td>
<td>95% LCL</td>
</tr>
<tr>
<td>1</td>
<td>2.368</td>
<td>1.911</td>
</tr>
<tr>
<td>2</td>
<td>0.445</td>
<td>0.383</td>
</tr>
<tr>
<td>3</td>
<td>-0.517</td>
<td>-0.548</td>
</tr>
<tr>
<td>4</td>
<td>-1.175</td>
<td>-1.203</td>
</tr>
<tr>
<td>5</td>
<td>-1.754</td>
<td>-1.780</td>
</tr>
</tbody>
</table>

Alpha-Spending and Null Hypothesis Simulation Details for Scenario 1

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T-Value Scale</td>
<td>Scale</td>
<td>Alpha Scale</td>
<td>Alpha</td>
<td>Alpha Spent</td>
<td>H0 Sims</td>
<td>Outside</td>
<td>H0 Sims</td>
<td>Outside</td>
<td>Futility Boundary</td>
</tr>
<tr>
<td>1</td>
<td>-3.850</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.000</td>
<td>0.009</td>
<td>0.009</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-2.906</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.002</td>
<td>0.315</td>
<td>0.324</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-2.335</td>
<td>0.010</td>
<td>0.011</td>
<td>0.011</td>
<td>0.011</td>
<td>0.011</td>
<td>0.011</td>
<td>0.377</td>
<td>0.701</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>-1.939</td>
<td>0.026</td>
<td>0.017</td>
<td>0.028</td>
<td>0.017</td>
<td>0.028</td>
<td>0.017</td>
<td>0.189</td>
<td>0.890</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>-1.754</td>
<td>0.040</td>
<td>0.022</td>
<td>0.050</td>
<td>0.017</td>
<td>0.045</td>
<td>0.017</td>
<td>0.065</td>
<td>0.955</td>
<td></td>
</tr>
</tbody>
</table>
Group-Sequential Tests for Two Means Assuming Normality (Simulation) 477-23

Beta-Spending and Alternative Hypothesis Simulation Details for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>Futility Boundary</th>
<th>Target</th>
<th>Actual</th>
<th>Proportion H1 Sims</th>
<th>Cum. H1 Sims Outside</th>
<th>Signif. Boundary</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.008</td>
<td>0.008</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.176</td>
<td>0.184</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.366</td>
<td>0.550</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.267</td>
<td>0.817</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>0.093</td>
<td>0.910</td>
</tr>
</tbody>
</table>

Run Time: 2.22 minutes.

The values obtained from any given run of this example will vary slightly due to the variation in simulations.

Example 3 – Enter Boundaries

With a set-up similar to Example 2, suppose we wish to investigate the properties of a set of significance (-3, -3, -3, -2, -1) and futility (2, 1, 0, 0, -1) boundaries.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Group-Sequential Tests for Two Means Assuming Normality (Simulation) procedure window by expanding Means, then Two Independent Means, then clicking on Group-Sequential, and then clicking on Group-Sequential Tests for Two Means Assuming Normality (Simulation). You may then make the entries as listed below, or open Example 3 by going to the File menu and choosing Open Example Template.

Option | Value
--- | ---
Data Tab | Alpha and Power (Enter Boundaries)
Find (Solve For) | Alpha and Power (Enter Boundaries)
Power | Ignored since this is the Find setting
Alpha | Ignored since this is the Find setting
N1 (Sample Size Group 1) | 500
N2 (Sample Size Group 2) | Use R
R (Allocation Ratio) | 1.0
Mean1 (Control) | 108
Mean2 (Treatment | H1) | 113
S | 25
Alternative Hypothesis | Mean1 < Mean2
Test Type | T-Test
Simulations | 20000

Enter Boundaries Tab

Number of Looks | 5
Types of Boundaries | Significance and Futility Boundaries
Significance Boundary | -3 -3 -3 -2 -1 (for looks 1 through 5)
Futility Boundary | 2 1 0 0 -1 (for looks 1 through 5)
Output

Click the Run button to perform the calculations and generate the following output.

### Numeric Results and Plots

**Scenario 1 Numeric Results for Group Sequential Testing Mean Difference = 0.**

Hypotheses: H0: Mean1=Mean2; H1: Mean1<Mean2  
Test Statistic: T-Test  
Type of Boundaries: Significance and Futility Boundaries  
Number of Looks: 5  
Simulations: 20000  
Pool Size: 40000

#### Numeric Summary for Scenario 1

<table>
<thead>
<tr>
<th>Value</th>
<th>95% LCL</th>
<th>95% UCL</th>
<th>Value</th>
<th>95% LCL</th>
<th>95% UCL</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.980</td>
<td>0.978</td>
<td>0.982</td>
<td>0.150</td>
<td>0.145</td>
<td>0.155</td>
<td>0.020</td>
</tr>
</tbody>
</table>

--- Average Sample Size ---

<table>
<thead>
<tr>
<th>N1</th>
<th>N2</th>
<th>Grp1</th>
<th>Grp2</th>
<th>Grp1</th>
<th>Grp2</th>
<th>Diff0</th>
<th>Diff1</th>
<th>P1</th>
<th>H1</th>
<th>P2</th>
</tr>
</thead>
<tbody>
<tr>
<td>500</td>
<td>500</td>
<td>368</td>
<td>368</td>
<td>364</td>
<td>364</td>
<td>-5.0</td>
<td>108.0</td>
<td>113.0</td>
<td>25.0</td>
<td></td>
</tr>
</tbody>
</table>

#### Accumulated Information Details for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>Information Percent</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20.0</td>
<td>100</td>
<td>100</td>
<td>200</td>
</tr>
<tr>
<td>2</td>
<td>40.0</td>
<td>200</td>
<td>200</td>
<td>400</td>
</tr>
<tr>
<td>3</td>
<td>60.0</td>
<td>300</td>
<td>300</td>
<td>600</td>
</tr>
<tr>
<td>4</td>
<td>80.0</td>
<td>400</td>
<td>400</td>
<td>800</td>
</tr>
<tr>
<td>5</td>
<td>100.0</td>
<td>500</td>
<td>500</td>
<td>1000</td>
</tr>
</tbody>
</table>

#### Boundaries for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>T-Value</th>
<th>P-Value</th>
<th>T-Value</th>
<th>P-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-3.000</td>
<td>0.002</td>
<td>2.000</td>
<td>0.977</td>
</tr>
<tr>
<td>2</td>
<td>-3.000</td>
<td>0.001</td>
<td>1.000</td>
<td>0.841</td>
</tr>
<tr>
<td>3</td>
<td>-3.000</td>
<td>0.001</td>
<td>0.000</td>
<td>0.500</td>
</tr>
<tr>
<td>4</td>
<td>-2.000</td>
<td>0.023</td>
<td>0.000</td>
<td>0.500</td>
</tr>
<tr>
<td>5</td>
<td>-1.000</td>
<td>0.159</td>
<td>-1.000</td>
<td>0.159</td>
</tr>
</tbody>
</table>
Example 4 – Validation Using Simulation

With a set-up similar to Example 1, we examine the power and alpha generated by the set of two-sided significance boundaries (+/- 4.026, +/- 3.396, +/- 2.682, +/- 2.302, +/- 1.990).

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Group-Sequential Tests for Two Means Assuming Normality (Simulation) procedure window by expanding Means, then Two Independent Means, then clicking on Group-Sequential, and then clicking on Group-Sequential Tests for Two Means Assuming Normality (Simulation). You may then make the entries as listed below, or open Example 4 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Alpha and Power (Enter Boundaries)</td>
</tr>
<tr>
<td></td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td></td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>500</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Mean1 (Control)</td>
<td>108</td>
</tr>
<tr>
<td>Mean2 (Treatment</td>
<td>H1)</td>
</tr>
</tbody>
</table>
S ..............................................................25
Alternative Hypothesis .......................... Mean1 ≠ Mean2
Test Type ................................................. T-Test
Simulations ............................................... 20000

Enter Boundaries Tab
Number of Looks ................................. 5
Types of Boundaries .............................. Significance Boundaries
Significance Boundary ......................... 4.026, 3.396, 2.682, 2.302, 1.990

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results and Plots

Scenario 1 Numeric Results for Group Sequential Testing Mean Difference = 0.
Hypotheses: H0: Mean1=Mean2; H1: Mean1≠Mean2
Test Statistic: T-Test
Type of Boundaries: Significance Boundaries Only
Number of Looks: 5
Simulations: 20000
Pool Size: 40000

Numeric Summary for Scenario 1

<table>
<thead>
<tr>
<th></th>
<th>Power</th>
<th>Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value</td>
<td>95% LCL</td>
<td>95% UCL</td>
</tr>
<tr>
<td>0.889</td>
<td>0.884</td>
<td>0.893</td>
</tr>
</tbody>
</table>

----- Average Sample Size ----

```
N1    N2    Grp1   Grp2    Grp1   Grp2   Diff0    Diff1   P1|H1   P2
500   500   497    497    379    379   -5.0     108.0   113.0   25.0
```

Run Time: 2.16 minutes.

The values obtained from any given run of this example will vary slightly due to the variation in simulations. The power and alpha generated with these boundaries are very close to the values of Example 1.
477-28  Group-Sequential Tests for Two Means Assuming Normality (Simulation)
Chapter 478

Group-Sequential Non-Inferiority Tests for Two Means (Simulation)

Introduction

This procedure can be used to determine power, sample size and/or boundaries for group sequential non-inferiority tests comparing the means of two groups. The common two-sample T-Test and the Mann-Whitney U test can be simulated in this procedure. Significance and futility boundaries can be produced. The spacing of the looks can be equal or custom specified. Boundaries can be computed based on popular alpha- and beta-spending functions (O’Brien-Fleming, Pocock, Hwang-Shih-DeCani Gamma family, linear) or custom spending functions. Boundaries can also be input directly to verify alpha- and/or beta-spending properties. Futility boundaries can be binding or non-binding. Maximum and average (expected) sample sizes are reported as well as the alpha and/or beta spent and incremental power at each look. Corresponding P-Value boundaries are also given for each boundary statistic. Plots of boundaries are also produced.

Technical Details

This section outlines many of the technical details of the techniques used in this procedure including the simulation summary, the test statistic details, and the use of spending functions.

An excellent text for the background and details of many group-sequential methods is Jennison and Turnbull (2000).
Simulation Procedure

In this procedure, a large number of simulations are used to calculate boundaries and power using the following steps

1. Based on the specified distributions, random samples of size N1 and N2 are generated under the null distribution and under the alternative distribution. These are simulated samples as though the final look is reached.

2. For each sample, test statistics for each look are produced. For example, if N1 and N2 are 100 and there are 5 equally spaced looks, test statistics are generated from the random samples at N1 = N2 = 20, N1 = N2 = 40, N1 = N2 = 60, N1 = N2 = 80, and N1 = N2 = 100 for both null and alternative samples.

3. To generate the first significance boundary, the null distribution statistics of the first look (e.g., at N1 = N2 = 20) are ordered and the percent of alpha to be spent at the first look is determined (using either the alpha-spending function or the input value). The statistic for which the percent of statistics above (or below, as the case may be) that value is equal to the percent of alpha to be spent at the first look is the boundary statistic. It is seen here how important a large number of simulations is to the precision of the boundary estimates.

4. All null distribution samples that are outside the first significance boundary at the first look are removed from consideration for the second look. If binding futility boundaries are also being computed, all null distribution samples with statistics that are outside the first futility boundary are also removed from consideration for the second look. If non-binding futility boundaries are being computed, null distribution samples with statistics outside the first futility boundary are not removed.

5. To generate the second significance boundary, the remaining null distribution statistics of the second look (e.g., at N1 = N2 = 40) are ordered and the percent of alpha to be spent at the second look is determined (again, using either the alpha-spending function or the input value). The percent of alpha to be spent at the second look is multiplied by the total number of simulations to determine the number of the statistic that is to be the second boundary statistic. The statistic for which that number of statistics is above it (or below, as the case may be) is the second boundary statistic. For example, suppose there are initially 1000 simulated samples, with 10 removed at the first look (from, say, alpha spent at Look 1 equal to 0.01), leaving 990 samples considered for the second look. Suppose further that the alpha to be spent at the second look is 0.02. This is multiplied by 1000 to give 20. The 990 still-considered statistics are ordered and the 970th (20 in from 990) statistic is the second boundary.

6. All null distribution samples that are outside the second significance boundary and the second futility boundary, if binding, at the second look are removed from consideration for the third look (e.g., leaving 970 statistics computed at N1 = N2 = 60 to be considered at the third look). Steps 4 and 5 are repeated until the final look is reached.

Futility boundaries are computed in a similar manner using the desired beta-spending function or custom beta-spending values and the alternative hypothesis simulated statistics at each look. For both binding and non-binding futility boundaries, samples for which alternative hypothesis statistics are outside either the significance or futility boundaries of the previous look are excluded from current and future looks.

Because the final futility and significance boundaries are required to be the same, futility boundaries are computed beginning at a small value of beta (e.g., 0.0001) and incrementing beta by that amount until the futility and significance boundaries meet.
When boundaries are entered directly, this procedure uses the null hypothesis and alternative hypothesis simulations to determine the number of test statistics that are outside the boundaries at each look. The cumulative proportion of alternative hypothesis statistics that are outside the significance boundaries is the overall power of the study.

**Generating Random Distributions**

Two methods are available in PASS to simulate random samples. The first method generates the random variates directly, one value at a time. The second method generates a large pool (over 10,000) of random values and then draws the random numbers from this pool. This second method can cut the running time of the simulation by 70%.

As mentioned above, the second method begins by generating a large pool of random numbers from the specified distributions. Each of these pools is evaluated to determine if its mean is within a small relative tolerance (0.0001) of the target mean. If the actual mean is not within the tolerance of the target mean, individual members of the population are replaced with new random numbers if the new random number moves the mean towards its target. Only a few hundred such swaps are required to bring the actual mean to within tolerance of the target mean. This population is then sampled with replacement using the uniform distribution. We have found that this method works well as long as the size of the pool is the maximum of twice the number of simulated samples desired and 10,000.

**The Statistical Hypotheses**

This section will review the specifics of non-inferiority testing.

Remember that in the usual t-test setting, the null (H0) and alternative (H1) hypotheses for one-sided tests are defined as

\[ H_0: \mu_1 - \mu_2 \leq D \quad \text{versus} \quad H_1: \mu_1 - \mu_2 > D \]

Rejecting this test implies that the mean difference is larger than the value \( D \). This test is called an upper-tailed test because it is rejected in samples in which the difference between the sample means is larger than \( D \).

Following is an example of a lower-tailed test.

\[ H_0: \mu_1 - \mu_2 \geq D \quad \text{versus} \quad H_1: \mu_1 - \mu_2 < D \]

Non-inferiority tests are special cases of the above directional tests. It will be convenient to adopt the following specialized notation for the discussion of these tests.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PASS Input/Output</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu_1 )</td>
<td>Mean1</td>
<td>Mean of population 1. Population 1 is assumed to consist of those who have received the new treatment.</td>
</tr>
<tr>
<td>( \mu_2 )</td>
<td>Mean2</td>
<td>Mean of population 2. Population 2 is assumed to consist of those who have received the reference treatment.</td>
</tr>
</tbody>
</table>
Margin of non-inferiority. This is a tolerance value that defines the magnitude of the amount that is not of practical importance. This may be thought of as the largest change from the baseline that is considered to be trivial. The absolute value is shown to emphasize that this is a magnitude. The sign of the value will be determined by the specific design that is being used.

True difference. This is the value of $\mu_1 - \mu_2$, the difference between the means. This is the value at which the power is calculated.

Note that the actual values of $\mu_1$ and $\mu_2$ are not needed. Only their difference is needed for power and sample size calculations.

Non-Inferiority Tests

A non-inferiority test tests that the treatment mean is not worse than the reference mean by more than the equivalence margin. The actual direction of the hypothesis depends on the response variable being studied.

Case 1: High Values Good, Non-Inferiority Test

In this case, higher values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is no less than a small amount below the reference mean. The value of $\delta$ is often set to zero. The following are equivalent sets of hypotheses.

$$H_0 : \mu_1 \geq \mu_2 + |M_{NI}| \quad \text{versus} \quad H_1 : \mu_1 < \mu_2 + |M_{NI}|$$

$$H_0 : \mu_1 - \mu_2 \geq |M_{NI}| \quad \text{versus} \quad H_1 : \mu_1 - \mu_2 < |M_{NI}|$$

$$H_0 : \delta \geq |M_{NI}| \quad \text{versus} \quad H_1 : \delta < |M_{NI}|$$

Case 2: High Values Bad, Non-Inferiority Test

In this case, lower values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is no more than a small amount above the reference mean. The value of $\delta$ is often set to zero. The following are equivalent sets of hypotheses.

$$H_0 : \mu_1 \leq \mu_2 - |M_{NI}| \quad \text{versus} \quad H_1 : \mu_1 > \mu_2 - |M_{NI}|$$

$$H_0 : \mu_1 - \mu_2 \leq -|M_{NI}| \quad \text{versus} \quad H_1 : \mu_1 - \mu_2 > -|M_{NI}|$$

$$H_0 : \delta \leq -|M_{NI}| \quad \text{versus} \quad H_1 : \delta > -|M_{NI}|$$
Test Statistics

This section describes the test statistics that are available in this procedure.

Two-Sample T-Test

The two-sample t-test assumes that the data are a simple random sample from a population of normally-distributed values that all have the same mean and variance. This assumption implies that the data are continuous and their distribution is symmetric. The calculation of the t statistic is as follows

\[ t_{df} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{s_{\bar{X}_1 - \bar{X}_2}} \]

where

\[ \bar{X}_k = \frac{\sum_{i=1}^{N_k} X_{ki}}{N_k} \]

\[ s_{\bar{X}_1 - \bar{X}_2} = \sqrt{\frac{\sum_{i=1}^{N_1} (X_{1i} - \bar{X}_1)^2 + \sum_{i=1}^{N_2} (X_{2i} - \bar{X}_2)^2}{N_1 + N_2 - 2} \left( \frac{1}{N_1} + \frac{1}{N_2} \right)} \]

\[ df = N_1 + N_2 - 2 \]

The significance of the test statistic is determined by computing the p-value based on the t distribution with degrees of freedom df. If this p-value is less than a specified level (often 0.05), the null hypothesis is rejected. Otherwise, no conclusion can be reached.

Mann-Whitney U Test

This test is the nonparametric substitute for the equal-variance t-test. Two key assumptions for this test are that the distributions are at least ordinal and that they are identical under H0. This implies that ties (repeated values) are not acceptable. When ties are present, the approximation provided can be used, but know that the theoretic results no longer hold.

The Mann-Whitney test statistic is defined as follows in Gibbons (1985).

\[ z = \frac{W_1 - N_1(N_1 + N_2 + 1)/2 + C}{s_w} \]

where

\[ W_1 = \sum_{k=1}^{N_1} \text{Rank}(X_{1k}) \]

The ranks are determined after combining the two samples. The standard deviation is calculated as
Group-Sequential Non-Inferiority Tests for Two Means (Simulation)

\[ s_{wi} = \sqrt{\frac{N_1N_2(N_1 + N_2 + 1)}{12}} - \frac{N_1N_2\sum_{i=1}^{t_2} (t_i^3 - t_i)}{12(N_1 + N_2)(N_1 + N_2 - 1)} \]

where \( t_i \) is the number of observations tied at value one, \( t_2 \) is the number of observations tied at some value two, and so forth.

The correction factor, \( C \), is 0.5 if the rest of the numerator of \( z \) is negative or -0.5 otherwise. The value of \( z \) is then compared to the standard normal distribution.

**Spending Functions**

Spending functions can be used in this procedure to specify the proportion of alpha or beta that is spent at each look without having to specify the proportion directly.

Spending functions have the characteristics that they are increasing and that

\[ \alpha(0) = 0 \]
\[ \alpha(1) = \alpha \]

The last characteristic guarantees a fixed \( \alpha \) level when the trial is complete. This methodology is very flexible since neither the times nor the number of analyses must be specified in advance. Only the functional form of \( \alpha(\tau) \) must be specified.

**PAS** provides several popular spending functions plus the ability to enter and analyze your own percents of alpha or beta spent. These are calculated as follows (beta may be substituted for alpha for beta-spending functions):

1. **Hwang-Shih-DeCani (gamma family)**

\[ \alpha \left[ \frac{1 - e^{-\gamma}}{1 - e^{-\gamma}} \right], \quad \gamma \neq 0; \quad \alpha, \gamma = 0 \]
2. **O’Brien-Fleming Analog** 

\[ 2 - 2 \Phi \left( \frac{Z_{\alpha/2}}{\sqrt{t}} \right) \]

<table>
<thead>
<tr>
<th>Look</th>
<th>Z Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1.7</td>
</tr>
<tr>
<td>2</td>
<td>-3.3</td>
</tr>
<tr>
<td>3</td>
<td>-5.0</td>
</tr>
<tr>
<td>4</td>
<td>0.0</td>
</tr>
<tr>
<td>5</td>
<td>1.7</td>
</tr>
<tr>
<td>6</td>
<td>3.3</td>
</tr>
<tr>
<td>7</td>
<td>5.0</td>
</tr>
</tbody>
</table>

3. **Pocock Analog** 

\[ \alpha \cdot \ln(1 + (e - 1)t) \]

<table>
<thead>
<tr>
<th>Look</th>
<th>Z Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-1.7</td>
</tr>
<tr>
<td>2</td>
<td>-3.3</td>
</tr>
<tr>
<td>3</td>
<td>-5.0</td>
</tr>
<tr>
<td>4</td>
<td>0.0</td>
</tr>
<tr>
<td>5</td>
<td>1.7</td>
</tr>
<tr>
<td>6</td>
<td>3.3</td>
</tr>
<tr>
<td>7</td>
<td>5.0</td>
</tr>
</tbody>
</table>

4. **Alpha * time** 

\[ \alpha \cdot t \]
5. **Alpha * time^1.5  \( \alpha \cdot t^{1/2} \)**

![Graph](image1.png)

6. **Alpha * time^2  \( \alpha \cdot t^2 \)**

![Graph](image2.png)

7. **Alpha * time^C  \( \alpha \cdot t^C \)**

![Graph](image3.png)

8. **User Supplied Percents**

A custom set of percents of alpha to be spent at each look may be input directly.

The O'Brien-Fleming Analog spends very little alpha or beta at the beginning and much more at the final looks. The Pocock Analog and (Alpha or Beta)(Time) spending functions spend alpha or beta more evenly across the looks. The Hwang-Shih-DeCani (C) (gamma family) spending functions and (Alpha or Beta)(Time^C) spending functions are flexible spending functions that can be used to spend more alpha or beta early or late or evenly, depending on the choice of C.
Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data, Looks & Boundaries, Enter Boundaries, and Options tabs. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains most of the parameters and options for the general setup of the procedure.

Solve For

Find (Solve For)

Solve for either power, sample size, or enter the boundaries directly and solve for power and alpha. When solving for power or sample size, the look and boundary details are specified on the "Looks & Boundaries" tab and the "Enter Boundaries" tab is ignored. When entering the boundaries directly and solving for power and alpha, the boundaries are input on the "Enter Boundaries" tab and the "Looks & Boundaries" tab is ignored.

When solving for power or N1, the early-stopping boundaries are also calculated. High accuracy for early-stopping boundaries requires a very large number of simulations (Recommended 100,000 to 10,000,000).

The parameter selected here is the parameter displayed on the vertical axis of the plot.

Because this is a simulation based procedure, the search for the sample size may take several minutes or hours to complete. You may find it quicker and more informative to solve for Power for a range of sample sizes.

Error Rates

Power or Beta

Power is the probability of rejecting the null hypothesis when it is false. Power is equal to 1-Beta, so specifying power implicitly specifies beta.

Beta is the probability obtaining a false negative on the statistical test. That is, it is the probability of accepting a false null hypothesis.

In the context of simulated group sequential trials, the power is the proportion of the alternative hypothesis simulations that cross any one of the significance (efficacy) boundaries.

The valid range is between 0 to 1.

Different disciplines and protocols have different standards for setting power. A common choice is 0.90, but 0.80 is also popular.

You can enter a range of values such as 0.70 0.80 0.90 or 0.70 to 0.90 by 0.1.

Alpha (Significance Level)

Alpha is the probability of obtaining a false positive on the statistical test. That is, it is the probability of rejecting a true null hypothesis.

The null hypothesis is usually that the parameters (the means, proportions, etc.) are all equal.
In the context of simulated group sequential trials, alpha is the proportion of the null hypothesis simulations that cross any one of the significance (efficacy) boundaries.

Since Alpha is a probability, it is bounded by 0 and 1. Commonly, it is between 0.001 and 0.250. Alpha is often set to 0.05 for two-sided tests and to 0.025 for one-sided tests. You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.

**Sample Size**

**N1 (Sample Size Group 1)**
Enter a value for the sample size, N1. This is the number of subjects in the first group of the study at the final look.

You may enter a range such as 10 to 100 by 10 or a list of values separated by commas or blanks. You might try entering the same number two or three times to get an idea of the variability in your results. For example, you could enter "10 10 10".

**N2 (Sample Size Group 2)**
Enter a value (or range of values) for the sample size of group 2. This is the number of subjects in the second group of the study at the final look.

- **Use R**
  When Use R is entered here, N2 is calculated using the formula
  \[ N2 = \lceil R(N1) \rceil \]
  where \( R \) is the Sample Allocation Ratio and the operator \( \lceil \cdot \rceil \) is the first integer greater than or equal to \( Y \). For example, if you want \( N1 = N2 \), select Use R and set \( R = 1 \).

**R (Sample Allocation Ratio)**
Enter a value (or range of values) for \( R \), the allocation ratio between samples. This value is only used when \( N2 \) is set to Use R.

\[ R = \frac{N2}{N1} \]

When used, \( N2 \) is calculated from \( N1 \) using the formula: \( N2 = \lceil R(N1) \rceil \) where \( \lceil \cdot \rceil \) is the next integer greater than or equal to \( Y \). Note that setting \( R = 1.0 \) forces \( N2 = N1 \).

**Effect Size**

**Mean1 (Mean of Group 1, Control)**
Enter a value for the hypothesized mean of this group. The difference between the two means and the direction of the difference relative to the alternative hypothesis is of primary importance here.

For the simulations under the null and alternative hypothesis, the test statistics are calculated as mean of group 1 minus mean of group 2. For the purposes of simulation, if a two-sided alternative is selected, Mean1 should be larger than Mean2.

If you want to analyze the difference, enter the value of the difference for one of the means and zero for the other.

You may enter a range of values such as 10 20 30 or 0 to 100 by 25.
NIM (Non-Inferiority Margin)
This is the magnitude of the margin of non-inferiority. It must be entered as a positive number. When higher means are better, this value is the distance below the reference mean that is still considered non-inferior. When higher means are worse, this value is the distance above the reference mean that is still considered non-inferior. The % box may be checked to specify NIM as a percentage of Mean1.

% (NIM as a Percent of Mean1)
If this box is checked, the non-inferiority margin will be generated as a percent of Mean1.

Mean2 (Mean of Group 2, Treatment | H1)
Enter a value for the hypothesized mean of this group. The difference between the two means and the direction of the difference relative to the alternative hypothesis is of primary importance here. For the simulations under the null and alternative hypothesis, the test statistics are calculated as mean of group 1 minus mean of group 2. For the purposes of simulation, if a two-sided alternative is selected, Mean1 should be larger than Mean2. If you want to analyze the difference, enter the value of the difference for one of the means and zero for the other. You may enter a range of values such as 10 20 30 or 0 to 100 by 25.

Standard Deviation
Enter the desired value for the standard deviation of the population from which simulations will be generated. The standard deviation must be a positive number. This standard deviation is the standard deviation of each group. Press the "SD" button to obtain help on estimating the standard deviation. You can enter a range of values such as 1 2 3 or 1 to 10 by 1.

Test and Simulations

Higher Means Are
This option defines whether higher values of the response variable are to be considered better or worse. The choice here determines the direction of the non-inferiority test. If Higher Means Are Better the null hypothesis is Mean1 - Mean2 >= NIM and the alternative hypothesis is Mean1 - Mean2 < NIM. If Higher Means Are Worse the null hypothesis is Mean1 - Mean2 <= -NIM and the alternative hypothesis is Mean1 - Mean2 > -NIM.

Test Type
Specify which test statistic is to be reported on. The T-Test is the standard T-test based on the pooled variance and SS1 + SS2 - 2 degrees of freedom, where SS1 and SS2 are the sample sizes of groups 1 and 2 at the corresponding look.
The Mann-Whitney Test is the common non-parametric alternative and is based on the ranks of the observations. It is sometimes called the Mann-Whitney U, Mann-Whitney-Wilcoxon, Wilcoxon Rank-Sum, or Wilcoxon-Mann-Whitney test.

Simulations
This option specifies the number of iterations, $M$, used in the simulation. As the number of iterations is increased, the accuracy and running time of the simulation will be increased also.

The precision of the simulated power estimates are calculated from the binomial distribution. Thus, confidence intervals may be constructed for various power values. The following table gives an estimate of the precision that is achieved for various simulation sizes when the power is either 0.50 or 0.95. The table values are interpreted as follows: a 95% confidence interval of the true power is given by the power reported by the simulation plus and minus the ‘Precision’ amount given in the table.

<table>
<thead>
<tr>
<th>Simulation Size $M$</th>
<th>Precision when Power = 0.50</th>
<th>Precision when Power = 0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.100</td>
<td>0.044</td>
</tr>
<tr>
<td>500</td>
<td>0.045</td>
<td>0.019</td>
</tr>
<tr>
<td>1000</td>
<td>0.032</td>
<td>0.014</td>
</tr>
<tr>
<td>2000</td>
<td>0.022</td>
<td>0.010</td>
</tr>
<tr>
<td>5000</td>
<td>0.014</td>
<td>0.006</td>
</tr>
<tr>
<td>10000</td>
<td>0.010</td>
<td>0.004</td>
</tr>
<tr>
<td>50000</td>
<td>0.004</td>
<td>0.002</td>
</tr>
<tr>
<td>100000</td>
<td>0.003</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Notice that a simulation size of 1000 gives a precision of plus or minus 0.01 when the true power is 0.95. Also note that as the simulation size is increased beyond 5000, there is only a small amount of additional accuracy achieved.

However, when solving for Power or $N_1$, the simulations are also used to calculate the look boundaries. To obtain precise boundary estimates, the number of simulations needs to be high. This consideration competes with the length of time to complete the simulation. When solving for power, a large number of simulations (100,000 or 1,000,000) will finish in several minutes. When solving for $N_1$, perhaps 10,000 simulations can be run for each iteration. Then, a final run with the resulting $N_1$ solving for power can be run with more simulations.

Looks & Boundaries Tab
The Data tab contains most of the parameters and options for the general setup of the procedure.

Looks and Boundaries
Specification of Looks and Boundaries
Choose whether spending functions will be used to divide alpha and beta for each look (Simple Specification), or whether the percents of alpha and beta to be spent at each look will be specified directly (Custom Specification).

Under Simple Specification, the looks are automatically considered to be equally spaced. Under Custom Specification, the looks may be equally spaced or custom defined based on the percent of accumulated information.
Looks and Boundaries – Simple Specification

Number of Equally Spaced Looks
Select the total number of looks that will be used if the study is not stopped early for the crossing of a boundary.

Alpha Spending Function
Specify the type of alpha spending function to use.
The O’Brien-Fleming Analog spends very little alpha at the beginning and much more at the final looks. The Pocock Analog and (Alpha)(Time) spending functions spend alpha more evenly across the looks. The Hwang-Shih-DeCani (C) (sometimes called the gamma family) spending functions and (Alpha)(Time^C) spending functions are flexible spending functions that can be used to spend more alpha early or late or evenly, depending on the choice of C.

C (Alpha Spending)
C is used to define the Hwang-Shih-DeCani (C) or (Alpha)(Time^C) spending functions.
For the Hwang-Shih-DeCani (C) spending function, negative values of C spend more of alpha at later looks, values near 0 spend alpha evenly, and positive values of C spend more of alpha at earlier looks.
For the (Alpha)(Time^C) spending function, only positive values for C are permitted. Values of C near zero spend more of alpha at earlier looks, values near 1 spend alpha evenly, and larger values of C spend more of alpha at later looks.

Type of Futility Boundary
This option determines whether or not futility boundaries will be created, and if so, whether they are binding or non-binding.
Futility boundaries are boundaries such that, if crossed at a given look, stop the study in favor of H0.
Binding futility boundaries are computed in concert with significance boundaries. They are called binding because they require the stopping of a trial if they are crossed. If the trial is not stopped, the probability of a false positive will exceed alpha.
When Non-binding futility boundaries are computed, the significance boundaries are first computed, ignoring the futility boundaries. The futility boundaries are then computed. These futility boundaries are non-binding because continuing the trial after they are crossed will not affect the overall probability of a false positive declaration.

Number of Skipped Futility Looks
In some trials it may be desirable to wait a number of looks before examining the trial for futility. This option allows the beta to begin being spent after a specified number of looks.
The Number of Skipped Futility Looks should be less than the number of looks.

Beta Spending Function
Specify the type of beta spending function to use.
The O’Brien-Fleming Analog spends very little beta at the beginning and much more at the final looks. The Pocock Analog and (Beta)(Time) spending functions spend beta more evenly across the looks. The Hwang-Shih-DeCani (C) (sometimes called the gamma family) spending functions
and \((\text{Beta})(\text{Time}^C)\) spending functions are flexible spending functions that can be used to spend more beta early or late or evenly, depending on the choice of \(C\).

**C (Beta Spending)**

C is used to define the Hwang-Shih-DeCani (C) or \((\text{Beta})(\text{Time}^C)\) spending functions.

For the Hwang-Shih-DeCani (C) spending function, negative values of \(C\) spend more of beta at later looks, values near 0 spend beta evenly, and positive values of \(C\) spend more of beta at earlier looks.

For the \((\text{Beta})(\text{Time}^C)\) spending function, only positive values for \(C\) are permitted. Values of \(C\) near zero spend more of beta at earlier looks, values near 1 spend beta evenly, and larger values of \(C\) spend more of beta at later looks.

---

**Looks and Boundaries – Custom Specification**

**Number of Looks**

This is the total number of looks of either type (significance or futility or both).

**Equally Spaced**

If this box is checked, the Accumulated Information boxes are ignored and the accumulated information is evenly spaced.

**Type of Futility Boundary**

This option determines whether or not futility boundaries will be created, and if so, whether they are binding or non-binding.

Futility boundaries are boundaries such that, if crossed at a given look, stop the study in favor of H0.

Binding futility boundaries are computed in concert with significance boundaries. They are called binding because they require the stopping of a trial if they are crossed. If the trial is not stopped, the probability of a false positive will exceed alpha.

When Non-binding futility boundaries are computed, the significance boundaries are first computed, ignoring the futility boundaries. The futility boundaries are then computed. These futility boundaries are non-binding because continuing the trial after they are crossed will not affect the overall probability of a false positive declaration.

**Accumulated Information**

The accumulated information at each look defines the proportion or percent of the sample size that is used at that look.

These values are accumulated information values so they must be increasing.

Proportions, percents, or sample sizes may be entered. All proportions, percents, or sample sizes will be divided by the value at the final look to create an accumulated information proportion for each look.

**Percent of Alpha Spent**

This is the percent of the total alpha that is spent at the corresponding look. It is not the cumulative value.
Percents, proportions, or alphas may be entered here. Each of the values is divided by the sum of the values to obtain the proportion of alpha that is used at the corresponding look.

**Percent of Beta Spent**
This is the percent of the total beta (1-power) that is spent at the corresponding look. It is not the cumulative value.

Percents, proportions, or betas may be entered here. Each of the values is divided by the sum of the values to obtain the proportion of beta that is used at the corresponding look.

---

**Enter Boundaries Tab**
The Data tab contains most of the parameters and options for the general setup of the procedure.

---

**Looks and Boundaries**

**Number of Looks**
This is the total number of looks of either type (significance or futility or both).

**Equally Spaced**
If this box is checked, the Accumulated Information boxes are ignored and the accumulated information is evenly spaced.

**Types of Boundaries**
This option determines whether or not futility boundaries will be entered.

Futility boundaries are boundaries such that, if crossed at a given look, stop the study in favor of H0.

**Accumulated Information**
The accumulated information at each look defines the proportion or percent of the sample size that is used at that look.

These values are accumulated information values so they must be increasing.

Proportions, percents, or sample sizes may be entered. All proportions, percents, or sample sizes will be divided by the value at the final look to create an accumulated information proportion for each look.

**Significance Boundary**
Enter the value of the significance boundary corresponding to the chosen test statistic. These are sometimes called efficacy boundaries.

**Futility Boundary**
Enter the value of the futility boundary corresponding to the chosen test statistic.
Options Tab

The Options tab contains limits on the number of iterations and various options about individual tests.

Maximum Iterations

Maximum Iterations Before Search Termination
Specify the maximum N1 before the search for N1 is aborted.

Since simulations for large sample sizes are very computationally intensive and hence time-consuming, this value can be used to stop searches when N1 is larger than reasonable sample sizes for the study.

This applies only when "Find (Solve For)" is set to N1.

The procedure uses a binary search when searching for N1. If a value for N1 is tried that exceeds this value, and the power is not reached, a warning message will be shown on the output indicating the desired power was not reached.

We recommend a value of at least 20000.

Random Numbers

Random Number Pool Size
This is the size of the pool of random values from which the random samples will be drawn. Pools should be at least the maximum of 10,000 and twice the number of simulations. You can enter Automatic and an appropriate value will be calculated.

If you do not want to draw numbers from a pool, enter 0 here.

Matching Boundaries at Final Look

Beta Search Increment
For each simulation, when futility bounds are computed, the appropriate beta is found by searching from 0 to 1 by this increment. Smaller increments are more refined, but the search takes longer.

We recommend 0.001 or 0.0001.
Example 1 – Power and Output

A clinical trial is to be conducted over a two-year period to compare the mean response of a new treatment to that of the current treatment. The current response mean is 108. The researchers would like to determine if the new treatment is not inferior to the current treatment, as it has fewer side effects and is less expensive. The new treatment is considered non-inferior if it has a mean greater than 103. Although the researchers do not know the true mean of the new treatment, they would like to examine the power that is achieved if the mean of the new treatment is also 108. The standard deviation for both groups is assumed to be 20. The sample size at the final look is to be 200 per group. Testing will be done at the 0.05 significance level. A total of five tests are going to be performed on the data as they are obtained. The O’Brien-Fleming (Analog) boundaries will be used.

Find the power and test boundaries assuming equal sample sizes per arm.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Group-Sequential Non-Inferiority Tests for Two Means (Simulation) procedure window by expanding Means, then Two Independent Means, then clicking on Group-Sequential, and then clicking on Group-Sequential Non-Inferiority Tests for Two Means (Simulation). You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Find (Solve For)</td>
<td>Power</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>200</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Mean1 (Control)</td>
<td>108</td>
</tr>
<tr>
<td>NIM</td>
<td>5</td>
</tr>
<tr>
<td>Mean2 (Treatment</td>
<td>H1)</td>
</tr>
<tr>
<td>S</td>
<td>20</td>
</tr>
<tr>
<td>Higher Means Are</td>
<td>Better</td>
</tr>
<tr>
<td>Test Type</td>
<td>T-Test</td>
</tr>
<tr>
<td>Simulations</td>
<td>20000</td>
</tr>
</tbody>
</table>

Looks and Boundaries Tab

<table>
<thead>
<tr>
<th>Specification of Looks and Boundaries</th>
<th>Simple</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Equally Spaced Looks</td>
<td>5</td>
</tr>
<tr>
<td>Alpha Spending Function</td>
<td>...O’Brien-Fleming Analog...</td>
</tr>
</tbody>
</table>
Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results and Plots

Scenario 1 Numeric Results for Group Sequential Test of Non-Inferiority
Higher Means are Better
Hypotheses: H0: Mean1 - Mean2 = NI; H1: Mean1 - Mean2 < NI
Test Statistic: T-Test
Alpha-Spending Function: O'Brien-Fleming Analog
Beta-Spending Function: None
Futility Boundary Type: None
Number of Looks: 5
Simulations: 20000
Pool Size: 40000

Numeric Summary for Scenario 1

<table>
<thead>
<tr>
<th>Value</th>
<th>Power 95% LCL</th>
<th>Power 95% UCL</th>
<th>Alpha Target</th>
<th>Alpha Actual</th>
<th>Alpha 95% LCL</th>
<th>Alpha 95% UCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7938</td>
<td>0.7882</td>
<td>0.7994</td>
<td>0.0500</td>
<td>0.0500</td>
<td>0.0470</td>
<td>0.0530</td>
</tr>
</tbody>
</table>

----- Average Sample Size ----

<table>
<thead>
<tr>
<th>N1</th>
<th>N2</th>
<th>Grp1</th>
<th>Grp2</th>
<th>Non-Inf. Margin</th>
<th>Mean1</th>
<th>Mean2</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>200</td>
<td>198</td>
<td>198</td>
<td>5.0</td>
<td>108.0</td>
<td>108.0</td>
<td>20.0</td>
</tr>
</tbody>
</table>

Report Definitions

Power is the probability of rejecting a false null hypothesis at one of the looks. It is the total proportion of alternative hypothesis simulations that are outside the significance boundaries.
Power 95% LCL and UCL are the lower and upper confidence limits for the power estimate. The width of the interval is based on the number of simulations.
Target Alpha is the user-specified probability of rejecting a true null hypothesis. It is the total alpha spent.
Actual Alpha is the alpha level that was actually achieved by the experiment. It is the total proportion of the null hypothesis simulations that are outside the significance boundaries.
Alpha 95% LCL and UCL are the lower and upper confidence limits for the actual alpha estimate. The width of the interval is based on the number of simulations.
Beta is the probability of accepting a false null hypothesis. It is the total proportion of alternative hypothesis simulations that do not cross the significance boundaries.
N1 and N2 are the sample sizes of each group if the study reaches the final look.
Average Sample Size Given H0 Grp1 and Grp2 are the average or expected sample sizes of each group if H0 is true. These are based on the proportion of null hypothesis simulations that cross the significance or futility boundaries at each look.
Average Sample Size Given H1 Grp1 and Grp2 are the average or expected sample sizes of each group if H1 is true. These are based on the proportion of alternative hypothesis simulations that cross the significance or futility boundaries at each look.
Non-inferiority margin is the distance from the control mean that is still considered non-inferior.
Mean1, Mean2, and Std Dev are the parameters that were set by the user to define the null and alternative simulation distributions.

Summary Statements

Group sequential trials with group sample sizes of 200 and 200 at the final look achieve 78% power to detect a difference of 5.0 at the 0.0500 significance level (alpha) using a one-sided T-Test.
Group-Sequential Non-Inferiority Tests for Two Means (Simulation) 478-19

Accumulated Information Details for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>Percent</th>
<th>Accumulated Sample Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Group 1</td>
</tr>
<tr>
<td>1</td>
<td>20.0</td>
<td>40</td>
</tr>
<tr>
<td>2</td>
<td>40.0</td>
<td>80</td>
</tr>
<tr>
<td>3</td>
<td>60.0</td>
<td>120</td>
</tr>
<tr>
<td>4</td>
<td>80.0</td>
<td>160</td>
</tr>
<tr>
<td>5</td>
<td>100.0</td>
<td>200</td>
</tr>
</tbody>
</table>

Accumulated Information Details Definitions

Look is the number of the look.
Accumulated Information Percent is the percent of the sample size accumulated up to the corresponding look.
Accumulated Sample Size Group 1 is total number of individuals in group 1 at the corresponding look.
Accumulated Sample Size Group 2 is total number of individuals in group 2 at the corresponding look.
Accumulated Sample Size Total is total number of individuals in the study (group 1 + group 2) at the corresponding look.

Boundaries for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>T-Value Scale</th>
<th>P-Value Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-4.6117</td>
<td>0.0000</td>
</tr>
<tr>
<td>2</td>
<td>-2.9561</td>
<td>0.0018</td>
</tr>
<tr>
<td>3</td>
<td>-2.3362</td>
<td>0.0102</td>
</tr>
<tr>
<td>4</td>
<td>-1.9639</td>
<td>0.0252</td>
</tr>
<tr>
<td>5</td>
<td>-1.7773</td>
<td>0.0381</td>
</tr>
</tbody>
</table>

Boundaries Definitions

Look is the number of the look.
Significance Boundary T-Value Scale is the value such that statistics outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. They are sometimes called efficacy boundaries.
Significance Boundary P-Value Scale is the value such that P-Values outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. This P-Value corresponds to the T-Value Boundary and is sometimes called the nominal alpha.

Boundary Plot
Boundary Plot - P-Value

Significance Boundaries with 95% Simulation Confidence Intervals for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>T-Value Boundary Value</th>
<th>95% LCL</th>
<th>95% UCL</th>
<th>P-Value Boundary Value</th>
<th>95% LCL</th>
<th>95% UCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-4.6117</td>
<td>0.0000</td>
<td></td>
<td>0.0000</td>
<td>0.0007</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-2.9561</td>
<td>-3.1336</td>
<td>-2.8627</td>
<td>0.0018</td>
<td>0.0024</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>-2.3362</td>
<td>-2.4246</td>
<td>-2.2891</td>
<td>0.0012</td>
<td>0.0080</td>
<td>0.0115</td>
</tr>
<tr>
<td>4</td>
<td>-1.9639</td>
<td>-1.9969</td>
<td>-1.9248</td>
<td>0.0252</td>
<td>0.0233</td>
<td>0.0276</td>
</tr>
<tr>
<td>5</td>
<td>-1.7773</td>
<td>-1.8062</td>
<td>-1.7478</td>
<td>0.0381</td>
<td>0.0358</td>
<td>0.0406</td>
</tr>
</tbody>
</table>

Significance Boundary Confidence Limit Definitions

Look is the number of the look.
Look is the number of the look.
T-Value Boundary Value is the value such that statistics outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. They are sometimes called efficacy boundaries.
P-Value Boundary Value is the value such that P-Values outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. This P-Value corresponds to the T-Value Boundary and is sometimes called the nominal alpha.
95% LCL and UCL are the lower and upper confidence limits for the boundary at the given look. The width of the interval is based on the number of simulations.

Alpha-Spending and Null Hypothesis Simulation Details for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>Target</th>
<th>Actual</th>
<th>Proportion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Signif. Boundary</td>
<td>Spending Function</td>
<td>Cum. H1 Sims Outside Signif. Boundary</td>
</tr>
<tr>
<td></td>
<td>T-Value Scale</td>
<td>P-Value Scale</td>
<td>Alpha</td>
</tr>
<tr>
<td>1</td>
<td>-4.6117</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>2</td>
<td>-2.9561</td>
<td>0.0018</td>
<td>0.0019</td>
</tr>
<tr>
<td>3</td>
<td>-2.3362</td>
<td>0.0102</td>
<td>0.0095</td>
</tr>
<tr>
<td>4</td>
<td>-1.9639</td>
<td>0.0252</td>
<td>0.0170</td>
</tr>
<tr>
<td>5</td>
<td>-1.7773</td>
<td>0.0381</td>
<td>0.0216</td>
</tr>
</tbody>
</table>
Group-Sequential Non-Inferiority Tests for Two Means (Simulation) 478-21

Alpha-Spending Details Definitions

Look is the number of the look.
Significance Boundary T-Value Scale is the value such that statistics outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. They are sometimes called efficacy boundaries.
Significance Boundary P-Value Scale is the value such that P-Values outside this boundary at the corresponding look indicate termination of the study and rejection of the null hypothesis. This P-Value corresponds to the Significance T-Value Boundary and is sometimes called the nominal alpha.
Spending Function Alpha is the intended portion of alpha allocated to the particular look based on the alpha-spending function.
Cumulative Spending Function Alpha is the intended accumulated alpha allocated to the particular look. It is the sum of the Spending Function Alpha up to the corresponding look.
Alpha Spent is the proportion of the null hypothesis simulations resulting in statistics outside the Significance Boundary at this look.
Cumulative Alpha Spent is the proportion of the null hypothesis simulations resulting in Significance Boundary termination up to and including this look. It is the sum of the Alpha Spent up to the corresponding look.
Proportion H1 Sims Outside Significance Boundary is the proportion of the alternative hypothesis simulations resulting in statistics outside the Significance Boundary at this look. It may be thought of as the incremental power.
Cumulative H1 Sims Outside Significance Boundary is the proportion of the alternative hypothesis simulations resulting in Significance Boundary termination up to and including this look. It is the sum of the Proportion H1 Sims Outside Significance Boundary up to the corresponding look.

Run Time: 48.00 seconds.

The values obtained from any given run of this example will vary slightly due to the variation in simulations.

Example 2 – Power for One-Sided Test with Futility Boundaries

Suppose researchers would like to compare two treatments with a non-inferiority test at each look. Further, suppose they would like to terminate the study early when it can be deemed highly unlikely that the new treatment is non-inferior to the standard. Suppose the control group mean is 108. The researchers wish to know the power of the test if the treatment group mean is also 108. The sample size at the final look is to be 200 per group. Testing will be done at the 0.05 significance level. A total of five tests are going to be performed on the data as they are obtained. The O’Brien-Fleming (Analog) boundaries will be used for both significance and futility boundaries.

Find the power and test boundaries assuming equal sample sizes per arm.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Group-Sequential Non-Inferiority Tests for Two Means (Simulation) procedure window by expanding Means, then Two Independent Means, then clicking on Group-Sequential, and then clicking on Group-Sequential Non-Inferiority Tests for Two Means (Simulation). You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.
Option | Value
---|---
**Data Tab**
Find (Solve For) | Power
Power | Ignored since this is the Find setting
Alpha | 0.05
N1 (Sample Size Group 1) | 200
N2 (Sample Size Group 2) | Use R
R (Allocation Ratio) | 1.0
Mean1 (Control) | 108
NIM | 5
Mean2 (Treatment | H1) | 108
S | 20
Higher Means Are | Better
Test Type | T-Test
Simulations | 20000

**Looks and Boundaries Tab**
Specification of Looks and Boundaries | Simple
Number of Equally Spaced Looks | 5
Alpha Spending Function | O'Brien-Fleming Analog
Type of Futility Boundary | Non-Binding
Number of Skipped Futility Looks | 0
Beta Spending Function | O'Brien-Fleming Analog

**Output**
Click the Run button to perform the calculations and generate the following output.

**Numeric Results and Plots**

**Scenario 1 Numeric Results for Group Sequential Test of Non-Inferiority**
Higher Means are Better
Hypotheses: H0: Mean1 - Mean2 = NIM; H1: Mean1 - Mean2 < NIM
Test Statistic: T-Test
Alpha-Spending Function: O'Brien-Fleming Analog
Beta-Spending Function: O'Brien-Fleming Analog
Futility Boundary Type: Non-binding
Number of Looks: 5
Simulations: 20000
Pool Size: 40000

**Numeric Summary for Scenario 1**

<table>
<thead>
<tr>
<th>Value</th>
<th>95% LCL</th>
<th>95% UCL</th>
<th>Target</th>
<th>Actual</th>
<th>95% LCL</th>
<th>95% UCL</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7305</td>
<td>0.7244</td>
<td>0.7366</td>
<td>0.0500</td>
<td>0.0406</td>
<td>0.0379</td>
<td>0.0433</td>
<td>0.2695</td>
</tr>
</tbody>
</table>
### Average Sample Size

<table>
<thead>
<tr>
<th>N1</th>
<th>N2</th>
<th>Grp1</th>
<th>Grp2</th>
<th>Grp1 Margin</th>
<th>Mean1</th>
<th>Mean2</th>
<th>Std Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>200</td>
<td>101</td>
<td>101</td>
<td>5.0</td>
<td>108.0</td>
<td>108.0</td>
<td>20.0</td>
</tr>
</tbody>
</table>

#### Accumulated Information Details for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>Percent</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20.0</td>
<td>40</td>
<td>40</td>
<td>80</td>
</tr>
<tr>
<td>2</td>
<td>40.0</td>
<td>80</td>
<td>80</td>
<td>160</td>
</tr>
<tr>
<td>3</td>
<td>60.0</td>
<td>120</td>
<td>120</td>
<td>240</td>
</tr>
<tr>
<td>4</td>
<td>80.0</td>
<td>160</td>
<td>160</td>
<td>320</td>
</tr>
<tr>
<td>5</td>
<td>100.0</td>
<td>200</td>
<td>200</td>
<td>400</td>
</tr>
</tbody>
</table>

#### Boundaries for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>T-Value Scale</th>
<th>P-Value Scale</th>
<th>T-Value Scale</th>
<th>P-Value Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-3.7891</td>
<td>0.0001</td>
<td>1.0859</td>
<td>0.8596</td>
</tr>
<tr>
<td>2</td>
<td>-2.9619</td>
<td>0.0018</td>
<td>-0.1501</td>
<td>0.4404</td>
</tr>
<tr>
<td>3</td>
<td>-2.3396</td>
<td>0.0101</td>
<td>-0.8176</td>
<td>0.2072</td>
</tr>
<tr>
<td>4</td>
<td>-1.9647</td>
<td>0.0252</td>
<td>-1.2866</td>
<td>0.0996</td>
</tr>
<tr>
<td>5</td>
<td>-1.7319</td>
<td>0.0420</td>
<td>-1.7319</td>
<td>0.0420</td>
</tr>
</tbody>
</table>

#### Boundary Plot

[Boundary Plot Image]
Boundary Plot - P-Value

Significance Boundaries with 95% Simulation Confidence Intervals for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>T-Value</th>
<th>95% LCL</th>
<th>95% UCL</th>
<th>P-Value</th>
<th>95% LCL</th>
<th>95% UCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-3.7891</td>
<td>0.0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-2.9619</td>
<td>-3.1283</td>
<td>-2.8905</td>
<td>0.0018</td>
<td>0.0010</td>
<td>0.0022</td>
</tr>
<tr>
<td>3</td>
<td>-2.3396</td>
<td>-2.3970</td>
<td>-2.2897</td>
<td>0.0101</td>
<td>0.0087</td>
<td>0.0115</td>
</tr>
<tr>
<td>4</td>
<td>-1.9647</td>
<td>-2.0082</td>
<td>-1.9309</td>
<td>0.0252</td>
<td>0.0227</td>
<td>0.0272</td>
</tr>
<tr>
<td>5</td>
<td>-1.7319</td>
<td>-1.7635</td>
<td>-1.7055</td>
<td>0.0420</td>
<td>0.0393</td>
<td>0.0444</td>
</tr>
</tbody>
</table>

Futility Boundaries with 95% Simulation Confidence Intervals for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>T-Value</th>
<th>95% LCL</th>
<th>95% UCL</th>
<th>P-Value</th>
<th>95% LCL</th>
<th>95% UCL</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0859</td>
<td>1.0468</td>
<td>1.1278</td>
<td>0.8596</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>-0.1501</td>
<td>-0.1696</td>
<td>-0.1269</td>
<td>0.4404</td>
<td>0.4328</td>
<td>0.4496</td>
</tr>
<tr>
<td>3</td>
<td>-0.8176</td>
<td>-0.8403</td>
<td>-0.7962</td>
<td>0.2072</td>
<td>0.2008</td>
<td>0.2134</td>
</tr>
<tr>
<td>4</td>
<td>-1.2866</td>
<td>-1.3064</td>
<td>-1.2633</td>
<td>0.0996</td>
<td>0.0962</td>
<td>0.1037</td>
</tr>
<tr>
<td>5</td>
<td>-1.7319</td>
<td>-1.7582</td>
<td>-1.7175</td>
<td>0.0420</td>
<td>0.0397</td>
<td>0.0433</td>
</tr>
</tbody>
</table>

Alpha-Spending and Null Hypothesis Simulation Details for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>T-Value</th>
<th>95% LCL</th>
<th>95% UCL</th>
<th>Signif. Boundary Scale</th>
<th>P-Value Boundary Scale</th>
<th>Alpha Spending Function</th>
<th>Cum. Alpha Spending</th>
<th>Proportion H0 Sims</th>
<th>Cum. H0 Sims</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-3.7891</td>
<td>0.0001</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>2</td>
<td>-2.9619</td>
<td>0.0018</td>
<td>0.0019</td>
<td>0.0019</td>
<td>0.0020</td>
<td>0.0020</td>
<td>0.0000</td>
<td>0.1432</td>
<td>0.1432</td>
</tr>
<tr>
<td>3</td>
<td>-2.3396</td>
<td>0.0101</td>
<td>0.0114</td>
<td>0.0095</td>
<td>0.0114</td>
<td>0.0114</td>
<td>0.0095</td>
<td>0.4281</td>
<td>0.5713</td>
</tr>
<tr>
<td>4</td>
<td>-1.9647</td>
<td>0.0252</td>
<td>0.0170</td>
<td>0.0284</td>
<td>0.0165</td>
<td>0.0279</td>
<td>0.0165</td>
<td>0.2424</td>
<td>0.8137</td>
</tr>
<tr>
<td>5</td>
<td>-1.7319</td>
<td>0.0420</td>
<td>0.0216</td>
<td>0.0500</td>
<td>0.0128</td>
<td>0.0406</td>
<td>0.0128</td>
<td>0.1016</td>
<td>0.9153</td>
</tr>
</tbody>
</table>

---
### Example 3 – Enter Boundaries

With a set-up similar to Example 2, suppose we wish to investigate the properties of a set of significance (-3, -3, -3, -2, -1) and futility (2, 1, 0, 0, -1) boundaries.

#### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Group-Sequential Non-Inferiority Tests for Two Means (Simulation)** procedure window by expanding **Means**, then **Two Independent Means**, then clicking on **Group-Sequential**, and then clicking on **Group-Sequential Non-Inferiority Tests for Two Means (Simulation)**. You may then make the appropriate entries as listed below, or open **Example 3** by going to the **File** menu and choosing **Open Example Template**.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.0859</td>
<td>0.8596</td>
<td>0.0136</td>
<td>0.0136</td>
<td>0.0137</td>
<td>0.0137</td>
<td>0.0052</td>
<td>0.0052</td>
<td>0.0052</td>
</tr>
<tr>
<td>2</td>
<td>-0.1501</td>
<td>0.4404</td>
<td>0.0675</td>
<td>0.0811</td>
<td>0.0675</td>
<td>0.0812</td>
<td>0.0821</td>
<td>0.0873</td>
<td>0.3439</td>
</tr>
<tr>
<td>3</td>
<td>-0.8176</td>
<td>0.0996</td>
<td>0.0630</td>
<td>0.2175</td>
<td>0.0631</td>
<td>0.2175</td>
<td>0.2567</td>
<td>0.2674</td>
<td>0.6113</td>
</tr>
<tr>
<td>4</td>
<td>-1.2866</td>
<td>0.0420</td>
<td>0.0525</td>
<td>0.2700</td>
<td>0.0520</td>
<td>0.2700</td>
<td>0.2695</td>
<td>0.1193</td>
<td>0.7305</td>
</tr>
<tr>
<td>5</td>
<td>-1.7319</td>
<td>0.0420</td>
<td>0.0525</td>
<td>0.2700</td>
<td>0.0520</td>
<td>0.2700</td>
<td>0.2695</td>
<td>0.1193</td>
<td>0.7305</td>
</tr>
</tbody>
</table>

Run Time: 54.05 seconds.

The values obtained from any given run of this example will vary slightly due to the variation in simulations.
Enter Boundaries Tab
Number of Looks.......................... 5
Types of Boundaries ....................... Significance and Futility Boundaries
Significance Boundary ................. -3 -3 -3 -2 -1 (for looks 1 through 5)
Futility Boundary ........................ 2 1 0 0 -1 (for looks 1 through 5)

Output
Click the Run button to perform the calculations and generate the following output.

Numeric Results and Plots

| Scenario 1 Numeric Results for Group Sequential Test of Non-Inferiority |
|---|---|---|---|---|---|---|---|---|---|
| Higher Means are Better |
| Hypotheses: H0: Mean1 - Mean2 = NIM; H1: Mean1 - Mean2 < NIM |
| Test Statistic: T-Test |
| Type of Boundaries: Significance and Futility Boundaries |
| Number of Looks: 5 |
| Simulations: 20000 |
| Pool Size: 40000 |

Numeric Summary for Scenario 1

<table>
<thead>
<tr>
<th>Value</th>
<th>95% LCL</th>
<th>95% UCL</th>
<th>Value</th>
<th>95% LCL</th>
<th>95% UCL</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9244</td>
<td>0.9207</td>
<td>0.9281</td>
<td>0.1544</td>
<td>0.1493</td>
<td>0.1594</td>
<td>0.0756</td>
</tr>
</tbody>
</table>

----- Average Sample Size ----

<table>
<thead>
<tr>
<th>N1</th>
<th>N2</th>
<th>Grp1</th>
<th>Grp2</th>
<th>Grp1</th>
<th>Grp2</th>
<th>Margin</th>
<th>Mean1</th>
<th>Mean2</th>
<th>Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>200</td>
<td>147</td>
<td>147</td>
<td>162</td>
<td>162</td>
<td>5.0</td>
<td>108.0</td>
<td>108.0</td>
<td>20.0</td>
</tr>
</tbody>
</table>

Accumulated Information Details for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>Accumulated Information</th>
<th>Percent</th>
<th>Group 1</th>
<th>Group 2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>20.0</td>
<td>40</td>
<td>40</td>
<td>80</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>40.0</td>
<td>80</td>
<td>80</td>
<td>160</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>60.0</td>
<td>120</td>
<td>120</td>
<td>240</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>80.0</td>
<td>160</td>
<td>160</td>
<td>320</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>100.0</td>
<td>200</td>
<td>200</td>
<td>400</td>
<td></td>
</tr>
</tbody>
</table>

Boundaries for Scenario 1

<table>
<thead>
<tr>
<th>Look</th>
<th>-- Significance Boundary --</th>
<th>----- Futility Boundary -----</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T-Value</td>
<td>Scale</td>
</tr>
<tr>
<td>1</td>
<td>-3.0000</td>
<td>0.0018</td>
</tr>
<tr>
<td>2</td>
<td>-3.0000</td>
<td>0.0016</td>
</tr>
<tr>
<td>3</td>
<td>-3.0000</td>
<td>0.0015</td>
</tr>
<tr>
<td>4</td>
<td>-2.0000</td>
<td>0.0232</td>
</tr>
<tr>
<td>5</td>
<td>-1.0000</td>
<td>0.1590</td>
</tr>
</tbody>
</table>
Example 4 – Validation Using Simulation

With a set-up similar to Example 1, we examine the power and alpha generated by the set of two-sided significance boundaries (-4.6117, -2.9561, -2.3362, -1.9639, -1.7773).

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Group-Sequential Non-Inferiority Tests for Two Means (Simulation) procedure window by expanding Means, then Two Independent Means, then clicking on Group-Sequential, and then clicking on Group-Sequential Non-Inferiority Tests for Two Means (Simulation). You may then make the appropriate entries as listed below, or open Example 4 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Alpha and Power (Enter Boundaries)</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>N1 (Sample Size Group 1)</td>
<td>200</td>
</tr>
<tr>
<td>N2 (Sample Size Group 2)</td>
<td>Use R</td>
</tr>
<tr>
<td>R (Allocation Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>Mean1 (Control)</td>
<td>108</td>
</tr>
<tr>
<td>NIM</td>
<td>5</td>
</tr>
</tbody>
</table>
Mean2 (Treatment | H1) ......................... 108
S ......................................................... 20
Higher Means Are ......................... Better
Test Type ............................................. T-Test
Simulations ........................................... 20000

Enter Boundaries Tab
Number of Looks ......................... 5
Types of Boundaries ......................... Significance Boundaries
Significance Boundary .................... -4.6117, -2.9561, -2.3362, -1.9639, -1.7773

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results and Plots

<table>
<thead>
<tr>
<th>Scenario 1 Numeric Results for Group Sequential Test of Non-Inferiority</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher Means are Better</td>
</tr>
<tr>
<td>Hypotheses: H0: Mean1 - Mean2 = NIM; H1: Mean1 - Mean2 &lt; NIM</td>
</tr>
<tr>
<td>Test Statistic: T-Test</td>
</tr>
<tr>
<td>Type of Boundaries: Significance Boundaries Only</td>
</tr>
<tr>
<td>Number of Looks: 5</td>
</tr>
<tr>
<td>Simulations: 20000</td>
</tr>
<tr>
<td>Pool Size: 40000</td>
</tr>
</tbody>
</table>

### Numerical Summary for Scenario 1

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Power</td>
<td>Power</td>
<td>Power</td>
<td>Alpha</td>
<td>Alpha</td>
<td>Alpha</td>
<td>Beta</td>
</tr>
<tr>
<td>Value</td>
<td>95% LCL</td>
<td>95% UCL</td>
<td>Value</td>
<td>95% LCL</td>
<td>95% UCL</td>
<td>Value</td>
</tr>
<tr>
<td>0.7851</td>
<td>0.7794</td>
<td>0.7907</td>
<td>0.0458</td>
<td>0.0429</td>
<td>0.0487</td>
<td>0.2150</td>
</tr>
</tbody>
</table>

### Average Sample Size

<table>
<thead>
<tr>
<th>N1</th>
<th>N2</th>
<th>Grp1</th>
<th>Grp2</th>
<th>Grp1</th>
<th>Grp2</th>
<th>Margin</th>
<th>Mean1</th>
<th>Mean2</th>
<th>Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>200</td>
<td>200</td>
<td>198</td>
<td>198</td>
<td>157</td>
<td>157</td>
<td>5.0</td>
<td>108.0</td>
<td>108.0</td>
<td>20.0</td>
</tr>
</tbody>
</table>

Run Time: 44.14 seconds.

The values obtained from any given run of this example will vary slightly due to the variation in simulations. The power and alpha generated with these boundaries are very close to the values of Example 1.
478-30  Group-Sequential Non-Inferiority Tests for Two Means (Simulation)
Chapter 480

Inequality Tests for Two Means in a Cluster-Randomized Design

Introduction

Cluster Randomization refers to the situation in which the means of two groups, made up of $M$ clusters of $N$ individuals each, are to be tested using a modified $t$ test. In this case, the basic experimental unit is a cluster instead of an individual.

Technical Details

Our formulation comes from Donner and Klar (1996). Denote an observation by $X_{ijk}$ where $i = 1, 2$ is the group, $j = 1, 2, \ldots, M$ is a cluster in group $i$, and $k = 1, 2, \ldots, N$ is an individual in cluster $j$ of group $i$. Each cluster mean, $\bar{X}_{ij}$, has a population mean of $\mu_i$ and variance

$$Var(\bar{X}_i) = \left(\frac{\sigma^2}{N}\right)[1 + (N - 1)\rho]$$

where $\sigma^2$ is the variance of $X_{ijk}$ and $\rho$ is the intracluster correlation coefficient. This correlation made be thought of as the simple correlation between any two observations on the same individual. It may also be thought of as the proportion of total variance in the observations that can be attributed to difference between clusters.
The power for the two-sided, two-sample \( t \) test using the above formulation is calculated by

\[
Power = 1 - P(t \leq t_{\alpha/2}, df, \lambda) + P(t \leq -t_{\alpha/2}, df, \lambda)
\]

where

\[
df = 2(M-1)
\]

\[
\lambda = \frac{d}{\sqrt{\left[2\left(1 + (N-1)\rho\right) / (MN)\right]^{1/2}}}
\]

\[
d = \frac{|\mu_1 - \mu_2|}{\sigma}
\]

### Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

#### Data Tab

The Data tab contains most of the parameters and options that you will be concerned with.

**Solve For**

**Find (Solve For)**

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are \( D, S, M, N, \) Alpha, and Power and Beta.

Under most situations, you will select either Power and Beta to calculate power or \( N \) to calculate sample size.

Note that the value selected here always appears as the vertical axis on the charts.

The program is set up to evaluate power directly. For the other parameters, a search is made using an iterative procedure until an appropriate value is found.

#### Error Rates

**Power or Beta**

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.
If your only interest is in determining the appropriate sample size for a confidence interval, set beta to 0.5.

**Alpha (Significance Level)**
This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.

---

**Sample Size**

**M (Number of Clusters)**
Enter a value (or range of values) for the number of clusters, \( M \), per group.

You may enter a range of values such as 2, 4, 6 or 2 to 12 by 2.

**N (Individuals Per Cluster)**
Enter a value (or range of values) for the number of individuals, \( N \), per cluster.

You may enter a range of values such as 100, 200, 300 or 100 to 300 by 50.

---

**Effect Size – Mean Difference**

**D (Difference Between Means)**
This is the absolute value of the difference between the two group means. This value, divided by the standard deviation, becomes the effect size.

---

**Effect Size – Standard Deviation**

**S (Standard Deviation)**
Enter a value (or range of values) for the standard deviation. This value is only used as the divisor of the effect size. Hence, if you do not know the standard deviation, you can enter a one here and use effect size units for \( D \), the difference.

Remember, this is the standard deviation that occurs when the same individual is measured over and over.

---

**Effect Size – Intracluster Correlation**

**R (Intracluster Correlation)**
Enter a value (or range of values) for the intracluster correlation. This correlation may be thought of as the simple correlation between any two observations on the same individual. It may also be thought of as the proportion of total variance in the observations that can be attributed to difference between clusters.

Although the actual range for this value is from zero to one, typical values range from 0.002 to 0.010.
Test

Alternative Hypothesis
Specify whether the test is one-sided or two-sided. A two-sided hypothesis states that the values are not equal without specifying which is greater. If you do not have any special reason to do otherwise, you should use the two-sided option.

When a two-sided hypothesis is selected, the value of alpha is split in half. Everything else remains the same.

Iterations Tab
This tab sets an option used in the iterative procedures.

Maximum Iterations
Maximum Iterations Before Search Termination
Specify the maximum number of iterations allowed before the search for the criterion of interest is aborted. When the maximum number of iterations is reached without convergence, the criterion is left blank. A value of 500 is recommended.

Example 1 – Calculating Power
Suppose that a study is to be conducted in which $D = 0.2; S = 1.0; R = 0.01; M = 6; \alpha = 0.01, 0.05$; and $N = 50$ to $300$ by $50$ and beta is to be calculated.

Setup
This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means in a Cluster-Randomized Design procedure window by expanding Means, then Two Independent Means, then clicking on Test (Inequality), and then clicking on Tests for Two Means in a Cluster-Randomized Design. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find parameter</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.01 0.05</td>
</tr>
<tr>
<td>M (Number of Clusters)</td>
<td>6</td>
</tr>
<tr>
<td>N (Individuals Per Cluster)</td>
<td>50 to 300 by 50</td>
</tr>
<tr>
<td>D (Difference Between Means)</td>
<td>0.2</td>
</tr>
<tr>
<td>S (Standard Deviation)</td>
<td>1.0</td>
</tr>
<tr>
<td>R (Intracluster Correlation)</td>
<td>0.01</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Two-Sided</td>
</tr>
</tbody>
</table>
**Inequality Tests for Two Means in a Cluster-Randomized Design**

**Annotated Output**

Click the Run button to perform the calculations and generate the following output.

**Numeric Results**

<table>
<thead>
<tr>
<th>M (Clusters per Group)</th>
<th>N (Individuals per Cluster)</th>
<th>D (Difference)</th>
<th>R (Intracluster Correlation)</th>
<th>S (Standard Deviation)</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>50</td>
<td>0.200</td>
<td>0.01000</td>
<td>1.000</td>
<td>0.01000</td>
<td>0.81246</td>
</tr>
<tr>
<td>6</td>
<td>50</td>
<td>0.200</td>
<td>0.01000</td>
<td>1.000</td>
<td>0.05000</td>
<td>0.55800</td>
</tr>
<tr>
<td>6</td>
<td>100</td>
<td>0.200</td>
<td>0.01000</td>
<td>1.000</td>
<td>0.01000</td>
<td>0.69689</td>
</tr>
<tr>
<td>6</td>
<td>100</td>
<td>0.200</td>
<td>0.01000</td>
<td>1.000</td>
<td>0.05000</td>
<td>0.39872</td>
</tr>
<tr>
<td>6</td>
<td>150</td>
<td>0.200</td>
<td>0.01000</td>
<td>1.000</td>
<td>0.01000</td>
<td>0.62668</td>
</tr>
<tr>
<td>6</td>
<td>150</td>
<td>0.200</td>
<td>0.01000</td>
<td>1.000</td>
<td>0.05000</td>
<td>0.32088</td>
</tr>
<tr>
<td>6</td>
<td>200</td>
<td>0.200</td>
<td>0.01000</td>
<td>1.000</td>
<td>0.01000</td>
<td>0.58090</td>
</tr>
<tr>
<td>6</td>
<td>200</td>
<td>0.200</td>
<td>0.01000</td>
<td>1.000</td>
<td>0.05000</td>
<td>0.27611</td>
</tr>
<tr>
<td>6</td>
<td>250</td>
<td>0.200</td>
<td>0.01000</td>
<td>1.000</td>
<td>0.01000</td>
<td>0.54899</td>
</tr>
<tr>
<td>6</td>
<td>250</td>
<td>0.200</td>
<td>0.01000</td>
<td>1.000</td>
<td>0.05000</td>
<td>0.24741</td>
</tr>
<tr>
<td>6</td>
<td>300</td>
<td>0.200</td>
<td>0.01000</td>
<td>1.000</td>
<td>0.01000</td>
<td>0.52557</td>
</tr>
<tr>
<td>6</td>
<td>300</td>
<td>0.200</td>
<td>0.01000</td>
<td>1.000</td>
<td>0.05000</td>
<td>0.22758</td>
</tr>
</tbody>
</table>

**Report Definitions**

- **Power** is the probability of rejecting a false null hypothesis. It should be close to one.
- **M** is the number of clusters per group. There are two groups.
- **N** is the number of individuals per cluster.
- **D** is difference between the group means.
- **R** is intracluster correlation.
- **S** is standard deviation within an individual.
- **Alpha** is the probability of rejecting a true null hypothesis. It should be small.
- **Beta** is the probability of accepting a false null hypothesis. It should be small.

**Summary Statements**

A sample size of 6 clusters per group with 50 individuals per cluster achieves 19% power to detect a difference of 0.200 between the group means when the standard deviation is 1.000 and the intracluster correlation is 0.01000 using a two-sided T-test with a significance level of 0.01000.

This report shows the power for each of the scenarios.

**Plots Section**

This plot shows the power versus the cluster size for the two alpha values.
Example 2 – Validation using Donner and Klar

Donner and Klar (1996) page 436 provide a table in which several power values are calculated. When alpha is 0.05, $D$ is 0.2, $R$ is 0.001, $S$ is 1.0, and $M$ is 3, they calculate a power of 0.43 for an $N$ of 100, 0.79 for an $N$ of 300, and 0.91 for an $N$ of 500.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means in a Cluster-Randomized Design procedure window by expanding Means, then Two Independent Means, then clicking on Test (Inequality), and then clicking on Tests for Two Means in a Cluster-Randomized Design. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find parameter</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>M (Number of Clusters)</td>
<td>3</td>
</tr>
<tr>
<td>N (Individuals Per Cluster)</td>
<td>100 300 500</td>
</tr>
<tr>
<td>D (Difference Between Means)</td>
<td>0.2</td>
</tr>
<tr>
<td>S (Standard Deviation)</td>
<td>1.0</td>
</tr>
<tr>
<td>R (Intracluster Correlation)</td>
<td>0.001</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Two-Sided</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Numeric Results for Two-Sided Test</th>
<th>Power</th>
<th>Number of Clusters</th>
<th>Number of Individuals Per Clusters</th>
<th>D</th>
<th>Intracluster Correlation</th>
<th>R</th>
<th>Standard Deviation</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.43008</td>
<td>3</td>
<td>100</td>
<td>0.200</td>
<td>0.00100</td>
<td>1.00</td>
<td>0.05000</td>
<td>0.56992</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.79236</td>
<td>3</td>
<td>300</td>
<td>0.200</td>
<td>0.00100</td>
<td>1.00</td>
<td>0.05000</td>
<td>0.20764</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.90905</td>
<td>3</td>
<td>500</td>
<td>0.200</td>
<td>0.00100</td>
<td>1.00</td>
<td>0.05000</td>
<td>0.09095</td>
<td></td>
</tr>
</tbody>
</table>

As you can see, PASS has calculated the same power values as Donner and Klar (1996).
Chapter 485

Inequality Tests for Paired Means

Introduction

The one-sample paired $t$-test may be used to test whether the mean difference of two populations is greater than, less than, or not equal to a specific value. Because the $t$ distribution is used to calculate critical values for the test, this test is often called the paired $t$-test. If the standard deviation of the differences is known, the normal distribution is used instead of the $t$ distribution and the test is officially known as the $z$ test.

This module also calculates the power of the nonparametric analog of the $t$-test, the Wilcoxon test.

Test Procedure

1. **Find the critical value.** Assume that the true mean difference is $D_0$. Choose a value $T_a$ so that the probability of rejecting $H_0$ when $H_0$ is true is equal to a specified value called $\alpha$. Using the $t$ distribution, select $T_a$ so that $\Pr(t > T_a) = \alpha$. This value is found using a $t$ probability table or a computer program (like PASS).

2. **Select a sample of $n$ items from the population and compute the $t$ statistic.** Call this value $T$. If $T > T_a$ reject the null hypothesis that the mean equals $D_0$ in favor of an alternative hypothesis that the mean equals $D1$ where $D1 > D0$.

Following is a specific example. Suppose we want to test the hypothesis that a variable, $X$, which is made up of paired differences, has a mean of 0 versus the alternative hypothesis that the mean is greater than 0. Suppose that previous studies have shown that the standard deviation of the paired differences, $\sigma$, is 40. A random sample of 100 pairs is used.

We first compute the critical value, $T_a$. The value of $T_a$ that yields $\alpha = 0.05$ is 6.6. If the paired difference mean computed from a sample is greater than 6.6, reject the hypothesis that the mean is 0. Otherwise, do not reject the hypothesis. We call the region greater than 6.6 the Rejection Region and values less than or equal to 6.6 the Acceptance Region of the significance test.
Now suppose that you want to compute the power of this testing procedure. In order to compute the power, we must specify an alternative value for the mean. We decide to compute the power if the true mean difference were 10. Figure 2 shows how to compute the power in this case.

The *power* is the probability of rejecting $H_0$ when the true mean is 10. Since we reject $H_0$ when the calculated mean is greater than 6.6, the probability of a Type-II error (called $\beta$) is given by the dark, shaded area of the second graph. This value is 0.196. The power is equal to $1 - \beta$ or 0.804.

Note that there are six parameters that may be varied in this situation: two means, standard deviation, alpha, beta, and the sample size.
Assumptions

This section describes the assumptions that are made when you use one of these tests. The key assumption relates to normality or non-normality of the data. One of the reasons for the popularity of the t-test is its robustness in the face of assumption violation. However, if the assumptions are not met, the significance levels and the power of the t-test may be invalidated. Unfortunately, in practice it often happens that several assumptions are not met. Take the steps to check the assumptions before you make important decisions based on these tests.

Paired T-Test Assumptions

The assumptions of the paired t-test are:

1. The data are continuous (not discrete).
2. The data, i.e., the differences for the matched-pairs, follow a normal probability distribution.
3. The sample of pairs is a simple random sample from its population. Each individual in the population has an equal probability of being selected in the sample.

Wilcoxon Signed-Rank Test Assumptions

The assumptions of the Wilcoxon signed-rank test are as follows (note that the difference is between a data value and the hypothesized median or between the two data values of a pair):

1. The differences are continuous (not discrete).
2. The distribution of each difference is symmetric.
3. The differences are mutually independent.
4. The differences all have the same median.
5. The measurement scale is at least interval.

Limitations

There are few limitations when using these tests. Sample sizes may range from a few to several hundred. If your data are discrete with at least five unique values, you can often ignore the continuous variable assumption. Perhaps the greatest restriction is that your data come from a random sample of the population. If you do not have a random sample, your significance levels will probably be incorrect.
Technical Details

**Standard Deviation Known**

When the standard deviation is known, the power is calculated as follows for a directional alternative (one-tailed test) in which \( D_1 > D_0 \).

1. Find \( z_\alpha \) such that \( 1 - \Phi(z_\alpha) = \alpha \), where \( \Phi(x) \) is the area under the standardized normal curve to the left of \( x \).

2. Calculate: \( x_\alpha = D_0 + z_\alpha \frac{\sigma}{\sqrt{n}} \).

3. Calculate: \( z_\alpha = \frac{x_\alpha - M_1}{\frac{\sigma}{\sqrt{n}}} \).

4. Power = \( 1 - \Phi(z_\alpha) \).

**Standard Deviation Unknown**

When the standard deviation is unknown, the power is calculated as follows for a directional alternative (one-tailed test) in which \( D_1 > D_0 \).

1. Find \( t_\alpha \) such that \( 1 - T_{df}(t_\alpha) = \alpha \), where \( T_{df}(t_\alpha) \) is the area under a central-\( t \) curve to the left of \( x \) and \( df = n - 1 \).

2. Calculate: \( x_\alpha = D_0 + t_\alpha \frac{\sigma}{\sqrt{n}} \).

3. Calculate the noncentrality parameter: \( \lambda_\alpha = \frac{D_1 - D_0}{\frac{\sigma}{\sqrt{n}}} \).

4. Calculate: \( t_\alpha = \frac{x_\alpha - D_1}{\frac{\sigma}{\sqrt{n}}} + \lambda_\alpha \).

5. Calculate: Power = \( 1 - T_{df',\lambda}(t_\alpha) \), where \( T_{df',\lambda}(x) \) is the area under a noncentral-\( t \) curve with degrees of freedom \( df' \) and noncentrality parameter \( \lambda \) to the left of \( x \).
Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains most of the parameters and options that you will be concerned with.

Solve For

Find (Solve For)

This option specifies the parameter to be calculated from the values of the other parameters. Under most conditions, you would select either Power and Beta or N.

Select N when you want to determine the sample size needed to achieve a given power and alpha error level.

Select Power and Beta when you want to calculate the power of an experiment that has already been run.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.
Sample Size

N (Sample Size)
This option specifies one or more values of the sample size, the number of pairs in the study. This value must be an integer greater than one. Note that you may enter a list of values using the syntax 50 100 150 200 250 or 50 to 250 by 50.

Effect Size – Means

Mean of Paired Differences (Alternative)
Enter a value (or range of values) for the mean of paired differences under the alternative hypothesis. This value indicates the minimum detectible difference for the corresponding power and sample size.

Effect Size – Standard Deviation

S (Standard Deviation of Paired Differences)
Enter an estimate of the standard deviation of paired differences (must be positive).
Use results from a previous (or pilot) study or the range divided by 5.
You can enter a range of values such as 1 2 3 or 1 to 10 by 1.
Press the Estimator button to load the Standard Deviation Estimator window.

Known Standard Deviation
This option specifies whether the standard deviation (sigma) is known or unknown. In almost all experimental situations, the standard deviation is not known. However, great calculation efficiencies are obtained if the standard deviation is assumed to be known.

When this box is checked, the program performs its calculations assuming that the standard deviation is known. This results in the use of the normal distribution in all probability calculations. Calculations using this option will be much faster than for the unknown standard deviation case. The results for either case will be close when the sample size is over 30.

When this box is not checked, the program assumes that the standard deviation is not known and will be estimated from the data when the $t$-test is run. This results in probability calculations using the noncentral-$t$ distribution. This distribution requires a lot more calculations than does the normal distribution.

The calculation speed comes into play whenever the Find option is set to something besides Beta. In these cases, the program uses a special searching algorithm which requires numerous iterations. You will note a real difference in calculation speed depending on whether this option is checked.

A reasonable strategy would be to leave this option checked while you are experimenting with the parameters and then turn it off when you are ready for your final results.
**Test**

**Alternative Hypothesis**
This option specifies the alternative hypothesis. This implicitly specifies the direction of the hypothesis test. The null hypothesis is always $H_0: \text{Mean of Paired Differences} = 0$.

Note that the alternative hypothesis enters into power calculations by specifying the rejection region of the hypothesis test. Its accuracy is critical.

Possible selections are:

- **Ha: Mean of Paired Diffs $\neq 0$**
  This is the most common selection. It yields the *two-tailed* $t$-test. Use this option when you are testing whether the means are different but you do not want to specify beforehand which mean is larger. Many scientific journals require two-tailed tests.

- **Ha: Mean of Paired Diffs $< 0$**
  This option yields a *one-tailed* $t$-test.

- **Ha: Mean of Paired Diffs $> 0$**
  This option yields a *one-tailed* $t$-test.

**Nonparametric Adjustment**
This option makes appropriate sample size adjustments for the Wilcoxon test. Results by Al-Sunduqchi and Guenther (1990) indicate that power calculations for the Wilcoxon test may be made using the standard $t$-test formulations with a simple adjustment to the sample size. The size of the adjustment depends upon the actual distribution of the data. They give sample size adjustment factors for four distributions. These are 1 for the uniform distribution, $2/3$ for the double exponential distribution, $9/\pi^2$ for the logistic distribution, and $\pi/3$ for the normal distribution.

The options are as follows:

- **Ignore**
  Do not make a Wilcoxon adjustment. This indicates that you want to analyze a $t$-test, not the Wilcoxon test.

- **Uniform**
  Make the Wilcoxon sample size adjustment assuming the uniform distribution. Since the factor is one, this option performs the same as Ignore. It is included for completeness.

- **Double Exponential**
  Make the Wilcoxon sample size adjustment assuming that the data actually follow the double exponential distribution.

- **Logistic**
  Make the Wilcoxon sample size adjustment assuming that the data actually follow the logistic distribution.
• Normal
  Make the Wilcoxon sample size adjustment assuming that the data actually follow the normal distribution.

Population Size
This is the number of subjects in the population. Usually, you assume that samples are drawn from a very large (infinite) population. Occasionally, however, situations arise in which the population of interest is of limited size. In these cases, appropriate adjustments must be made.

When a finite population size is specified, the standard deviation is reduced according to the formula:

$$\sigma_1^2 = \left(1 - \frac{n}{N}\right) \sigma^2$$

where $n$ is the sample size, $N$ is the population size, $\sigma$ is the original standard deviation, and $\sigma_1$ is the new standard deviation.

The quantity $n/N$ is often called the sampling fraction. The quantity $\left(1 - \frac{n}{N}\right)$ is called the finite population correction factor.

Example 1 – Paired T-Test

Usually, a researcher designs a study to compare two or more groups of subjects, so the one sample case described in this chapter occurs infrequently. However, there is a popular research design that does lead to the single mean test: paired observations.

For example, suppose researchers want to study the impact of an exercise program on the individual’s weight. To do so they randomly select $N$ individuals, weigh them, put them through the exercise program, and weigh them again. The variable of interest is not their actual weight, but how much their weight changed.

In this design, the data are analyzed using a one-sample $t$-test on the differences between the paired observations. The null hypothesis is that the average difference is zero. The alternative hypothesis is that the average difference is some nonzero value.

To study the impact of an exercise program on weight loss, the researchers decide to conduct a study that will be analyzed using the paired $t$-test. A sample of individuals will be weighed before and after a specified exercise program that will last three months. The difference in their weights will be analyzed.

Past experiments of this type have had standard deviations in the range of 10 to 15 pounds. The researcher wants to detect a difference of 5 pounds or more. Alpha values of 0.01 and 0.05 will be tried. Beta is set to 0.20 so that the power is 80%. How large of a sample must the researchers take?
Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for One Mean procedure window by expanding Means, then One Mean, then clicking on Test (Inequality), and then clicking on Tests for One Mean. You may then make the appropriate entries as listed below, or open Example 4 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N</td>
</tr>
<tr>
<td>Power</td>
<td>0.80</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.01 0.05</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>Ignored since this is the Find setting.</td>
</tr>
<tr>
<td>Mean of Paired Diffs (Alternative)</td>
<td>-5</td>
</tr>
<tr>
<td>S (SD of Paired Differences)</td>
<td>10 12.5 15</td>
</tr>
<tr>
<td>Known Standard Deviation</td>
<td>Not checked</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Ha: Mean of Paired Diffs ≠ 0</td>
</tr>
<tr>
<td>Nonparametric Adjustment</td>
<td>Ignore</td>
</tr>
<tr>
<td>Population Size</td>
<td>Infinite</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results and Plots

<table>
<thead>
<tr>
<th>Numeric Results for Paired T-Test</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Null Hypothesis: Mean of Paired Differences = 0, Alternative Hypothesis: Mean of Paired Differences ≠ 0</td>
<td>Unknown standard deviation.</td>
</tr>
<tr>
<td>Power</td>
<td>0.80939 0.80778 0.80434 0.80252 0.80230</td>
</tr>
<tr>
<td>N</td>
<td>51 34 77 109 73</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.01000 0.05000 0.05000 0.01000 0.05000</td>
</tr>
<tr>
<td>Beta</td>
<td>0.19061 0.19222 0.19566 0.19748 0.19770</td>
</tr>
<tr>
<td>Mean of Paired Differences</td>
<td>-5.0 -5.0 -5.0 -5.0 -5.0</td>
</tr>
<tr>
<td>S</td>
<td>10.0 10.0 12.5 15.0 15.0</td>
</tr>
<tr>
<td>Effect Size</td>
<td>0.500 0.500 0.400 0.333 0.333</td>
</tr>
</tbody>
</table>
The report shows the values of each of the parameters, one scenario per row. We were solving for the sample size, \( N \).

Note that depending on our choice of assumptions, the sample size ranges from 34 to 109. Hence, the researchers have to make a careful determination of which standard deviation and significance level should be used.

### Example 2 – Validation using Zar

This procedure uses the same mechanics as the Tests for One Mean procedure. The validation for this procedure is obtained through Examples 6 and 7 of Chapter 400.
Chapter 490

Inequality Tests for Paired Means (Simulation)

Introduction

This procedure allows you to study the power and sample size of several statistical tests of the null hypothesis that the difference between two correlated means is equal to a specific value versus the alternative that it is greater than, less than, or not-equal to that value. The paired t-test is commonly used in this situation. Other tests have been developed for the case when the data are not normally distributed. These additional tests include the Wilcoxon signed-ranks test, the sign test, and the computer-intensive bootstrap test.

Paired data may occur because two measurements are made on the same subject or because measurements are made on two subjects that have been matched according to other, often demographic, variables. Hypothesis tests on paired data can be analyzed by considering the differences between the paired items. The distribution of differences is usually symmetric. In fact, the distribution must be symmetric if the individual distributions of the two items are identical. Hence, the paired t-test and the Wilcoxon signed-rank test are appropriate for paired data even when the distributions of the individual items are not normal.

The details of the power analysis of the paired t-test using analytic techniques are presented in another PASS chapter and they won’t be duplicated here. This chapter will only consider power analysis using computer simulation.

Technical Details

Computer simulation allows us to estimate the power and significance level that is actually achieved by a test procedure in situations that are not mathematically tractable. Computer simulation was once limited to mainframe computers. But, in recent years, as computer speeds have increased, simulation studies can be completed on desktop and laptop computers in a reasonable period of time.

The steps to a simulation study are

1. Specify the test procedure and the test statistic. This includes the significance level, sample size, and underlying data distributions.
2. Generate a random sample \( X_1, X_2, \ldots, X_n \) from the distribution specified by the alternative hypothesis. Calculate the test statistic from the simulated data and determine if the null hypothesis is accepted or rejected. These samples are used to calculate the power of the test. In the case of paired data, the individual values are simulated as the difference between two other random variables. These samples are constructed so that they exhibit a certain amount of correlation.

3. Generate a random sample \( Y_1, Y_2, \ldots, Y_n \) from the distribution specified by the null hypothesis. Calculate the test statistic from the simulated data and determine if the null hypothesis is accepted or rejected. These samples are used to calculate the significance level of the test. In the case of paired data, the individual values are simulated as the difference between two other random variables. These samples are constructed so that they exhibit a certain amount of correlation.

4. Repeat steps 2 and 3 several thousand times, tabulating the number of times the simulated data leads to a rejection of the null hypothesis. The power is the proportion of simulated samples in step 2 that lead to rejection. The significance level is the proportion of simulated samples in step 3 that lead to rejection.

---

**Simulating Paired Distributions**

Paired data occur when two observations are correlated. Examples of paired designs are pre–post designs, cross-over designs, and matched pair designs.

In order to simulate paired data, the simulation should mimic the actual data generation process as closely as possible. Since paired data are analyzed by creating the individual difference between each pair, the simulation should also create data as the difference between two variates. Paired data exhibit a correlation between the two variates. As this correlation between the variates increases, the variance of the difference decreases. Thus it is important not only to specify the distributions of the two variates that will be differenced, but to also specify their correlation.

Obtaining paired samples from arbitrary distributions with a set correlation is difficult because the joint, bivariate distribution must be specified and simulated. Rather than specify the bivariate distribution, *PASS* requires the specification of the two marginal distributions and the correlation between them.

Monte Carlo samples with given marginal distributions and correlation are generated using the method suggested by Gentle (1998). The method begins by generating a large population of random numbers from the two distributions. Each of these populations is evaluated to determine if their means are within a small relative tolerance (0.0001) of the target mean. If the actual mean is not within the tolerance of the target mean, individual members of the population are replaced with new random numbers if the new random number moves the mean towards its target. Only a few hundred such swaps are required to bring the actual mean to within tolerance of the target mean.

The next step is to obtain the target correlation. This is accomplished by permuting one of the populations until they have the desired correlation.

The above steps provide a large pool of random numbers that exhibit the desired characteristics. This pool is then sampled at random using the uniform distribution to obtain the random numbers used in the simulation.
This algorithm may be stated as follows.

1. Draw individual samples of size M from the two distributions where M is a large number, usually over 10,000. Adjust these samples so that they have the specified mean and standard deviation. Label these samples A and B. Create an index of the values of A and B according to the order in which they are generated. Thus, the first value of A and the first value of B are indexed as one, the second values of A and B are indexed as two, and so on up to the final set which is indexed as M.

2. Compute the correlation between the two generated variates.

3. If the computed correlation is within a small tolerance (usually less than 0.001) of the specified correlation, go to step 7.

4. Select two indices (I and J) at random using uniform random numbers.

5. Determine what will happen to the correlation if \( B_i \) is swapped with \( B_J \). If the swap will result in a correlation that is closer to the target value, swap the indices and proceed to step 6. Otherwise, go to step 4.

6. If the computed correlation is within the desired tolerance of the target correlation, go to step 7. Otherwise, go to step 4.

7. End with a population with the required marginal distributions and correlation.

Now, to complete the simulation, random samples of the designated size are drawn from this population.

**Test Statistics**

This section describes the test statistics that are available in this procedure. Note that these test statistics are computed on the differences. Thus, when the equation refers to an \( X \) value, this \( X \) value is assumed to be a difference between two individual variates.

**One-Sample t-Test**

The one-sample t-test assumes that the data are a simple random sample from a population of normally-distributed values that all have the same mean and variance. This assumption implies that the data are continuous and their distribution is symmetric. The calculation of the t-test proceeds as follows:

\[
t_{n-1} = \frac{\overline{X} - \mu_0}{s_{\overline{X}}}
\]

where

\[
\overline{X} = \frac{\sum_{i=1}^{n} X_i}{n},
\]

\[
s_{\overline{X}} = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n-1}}.
\]
and $M_0$ is the value of the difference hypothesized by the null hypothesis.

The significance of the test statistic is determined by computing the p-value. If this p-value is less than a specified level (usually 0.05), the hypothesis is rejected. Otherwise, no conclusion can be reached.

**Wilcoxon Signed-Rank Test**

The Wilcoxon signed-rank test is a popular, nonparametric substitute for the t-test. It assumes that the data follow a symmetric distribution. The test is computed using the following steps.

1. Subtract the hypothesized mean, $D_0$, from each data value. Rank the values according to their absolute values.
2. Compute the sum of the positive ranks $S_\text{p}$ and the sum of the negative ranks $S_\text{n}$. The test statistic, $W$, is the minimum of $S_\text{p}$ and $S_\text{n}$.
3. Compute the mean and standard deviation of $W$ using the formulas
   \[
   \mu_W = \frac{n(n+1)}{4} \quad \text{and} \quad \sigma_W = \sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{\sum i^2 - \sum t}{48}}
   \]
   where $t_i$ represents the number of times the $i^{th}$ value occurs.
4. Compute the $z$ value using
   \[
   z_W = \frac{W - \mu_W}{\sigma_W}
   \]

For cases when $n$ is less than 38, the significance level is found from a table of exact probabilities for the Wilcoxon test. When $n$ is greater than or equal to 38, the significance of the test statistic is determined by computing the p-value using the standard normal distribution. If this p-value is less than a specified level (usually 0.05), the null hypothesis is rejected in favor of the alternative hypothesis. Otherwise, no conclusion can be reached.

**Sign Test**

The sign test is popular because it is simple to compute. It assumes that the data follow the same distribution. The test is computed using the following steps.

1. Count the number of values strictly greater than $M_0$. Call this value $X$.
2. Count the number of values strictly less than $M_0$. Call this value $Y$.
3. Set $m = X + Y$.
4. Under the null hypothesis, $X$ is distributed as a binomial random variable with a proportion of 0.5 and sample size of $m$.

The significance of $X$ is calculated using binomial probabilities.

**Bootstrap Test**

The one-sample bootstrap procedure for testing whether the mean is equal to a specific value is given in Efron & Tibshirani (1993) pages 224-227. The bootstrap procedure is as follows.

1. Compute the mean of the sample. Call it $\overline{X}$. 
2. Compute the t-value using the standard t-test. The formula for this computation is

\[ t = \frac{X - M_0}{s_X} \]

3. Draw a random, with-replacement sample of size \( n \) from the original \( X \) values. Call this sample \( Y_1, Y_2, \ldots, Y_n \).

4. Compute the t-value of this bootstrap sample using the formula

\[ t = \frac{Y - X}{s_Y} \]

5. For a two-tailed test, if \( |t_Y| > |t| \) then add one to a counter variable \( A \).

6. Repeat steps 3 – 5 \( B \) times. \( B \) may be anywhere from 100 to 10,000.

7. Compute the \( p \)-value of the bootstrap test as \( \frac{A + 1}{B + 1} \)

8. Steps 1 – 7 complete one simulation iteration. Repeat these steps \( M \) times, where \( M \) is the number of simulations. The power and significance level is equal to the percent of the time the \( p \)-value is less than the nominal alpha of the test.

Note that the bootstrap test is a time-consuming test to run, especially if you set \( B \) to a value larger than 100.

The Problem of Differing Standard Deviations

Care must be used when either the null or alternative distribution is not normal. In these cases, the standard deviation is usually not specified directly. For example, you might use a gamma distribution with a shape parameter of 1.5 and a mean of 4 as the null distribution and a gamma distribution with the same shape parameter and a mean of 5 as the alternative distribution. This allows you to compare the two means. However, note that although the shape parameters are constant, the standard deviations are not. Thus the null and alternatives not only have different means, but different standard deviations!

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data and Options tabs. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains most of the parameters and options that will be of interest.

Solve For

Find (Solve For)

This option specifies whether you want to find Power or \( N \) from the simulation. Select Power when you want to estimate the power of a certain scenario. Select \( N \) when you want to determine
the sample size needed to achieve a given power and alpha error level. Finding $N$ is very computationally intensive, and so it may take a long time to complete.

**Error Rates**

**Power or Beta**

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of equal means when in fact the means are different.

Values must be between zero and one. Historically, the value of 0.80 ($\beta = 0.20$) was used for power. Now, 0.90 ($\beta = 0.10$) is also commonly used.

A single value may be entered here or a range of values such as $0.8 \text{ to } 0.95 \text{ by } 0.05$ may be entered.

**Alpha (Significance Level)**

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as $0.01 \text{ to } 0.10 \text{ by } 0.01$.

**Sample Size**

**N (Sample Size)**

This option specifies one or more values of the sample size, the number of subjects in the study. The paired design assumes that a pair of observations will be obtained from each subject. Thus there will be $2N$ observations simulated, resulting in $N$ differences.

This value must be an integer greater than one. You may enter a list of values using the syntax $50 \text{ 100 150 200 250}$ or $50 \text{ to } 250 \text{ by } 50$.

**Effect Size**

**Item A (and B) Distribution|H0**

These options specify the distributions of the two items making up the pair under the null hypothesis, H0. The difference between the means of these two distributions is the difference that is tested, $\text{Diff}_0$.

Usually, you will want $\text{Diff}_0 = 0$. This zero difference is specified by entering $M_0$ for the mean parameter in each of the distributions and then entering an appropriate value for the $M_0$ parameter below.

All of the distributions are parameterized so that the mean is entered first. For example, if you wanted to test whether the mean of a normal distributed variable is five, you could enter $N(5, S)$ or $N(M_0, S)$ here.
The parameters of each distribution are specified using numbers or letters. If letters are used, their values are specified in the boxes below. The value $M0$ is reserved for the value of the mean under the null hypothesis.

Following is a list of the distributions that are available and the syntax used to specify them:

- Beta = A($M0$, $A$, $B$, Minimum)
- Binomial = B($M0$, $N$)
- Cauchy = C($M0$, Scale)
- Constant = K(Value)
- Exponential = E($M0$)
- F = F($M0$, $DF1$)
- Gamma = G($M0$, $A$)
- Multinomial = M($P1$, $P2$, ..., $Pk$)
- Normal = N($M0$, SD)
- Poisson = P($M0$)
- Student’s T = T($M0$, D)
- Tukey’s Lambda = L($M0$, $S$, Skewness, Elongation)
- Uniform = U($M0$, Minimum)
- Weibull = W($M0$, $B$)

Details of writing mixture distributions, combined distributions, and compound distributions are found in the chapter on Data Simulation and are not repeated here.

**Finding the Value of the Mean of a Specified Distribution**

Except for the multinomial distribution, the distributions have been parameterized in terms of their means, since this is the parameter being tested. The mean of a distribution created as a linear combination of other distributions is found by applying the linear combination to the individual means. However, the mean of a distribution created by multiplying or dividing other distributions is not necessarily equal to applying the same function to the individual means. For example, the mean of $4N(4, 5) + 2N(5, 6)$ is $4*4 + 2*5 = 26$, but the mean of $4N(4, 5) * 2N(5, 6)$ is not exactly $4*4*2*5 = 160$ (although it is close).

**Item A (and B) Distribution | H1**

These options specify the distributions of the two items making up the pair under the alternative hypothesis, H1. The difference between the means of these two distributions is the difference that is assumed to be the true value of the difference. That is, this is the difference at which the power is computed.

Usually, the mean difference is specified by entering $M1$ for the mean parameter in the distribution expression for item A and $M0$ for the mean parameter in the distribution expression for item B. The mean difference under H1 then becomes the value of $M1 - M0$.

The parameters of each distribution are specified using numbers or letters. If letters are used, their values are specified in the boxes below. The value $M1$ is reserved for the value of the mean under the alternative hypothesis.

Following is a list of the distributions that are available and the syntax used to specify them:

- Beta = A($M1$, $A$, $B$, Minimum)
- Binomial = B($M1$, $N$)
- Cauchy = C($M1$, Scale)
- Constant = K(Value)
Exponential=E(M1)
F=F(M1,DF1)
Gamma=G(M1,A)
Multinomial=M(P1,P2,...,Pk)
Normal=N(M1,SD)
Poisson=P(M1)
Student's T=T(M1,D)
Tukey's Lambda=L(M1,S,Skewness,Elongation)
Uniform=U(M1,Minimum)
Weibull=W(M1,B)

Details of writing mixture distributions, combined distributions, and compound distributions are found in the chapter on Data Simulation and will not be repeated here.

**Effect Size – Distribution Parameters**

**M0 (Mean|H0)**
These values are substituted for the $M_0$ in the four distribution specifications given above. $M_0$ is intended to be the value of the mean hypothesized by the null hypothesis, $H_0$.
You can enter a list of values using the syntax 0 1 2 3 or 0 to 3 by 1.

**M1 (Mean|H1)**
These values are substituted for the $M_1$ in the four distribution specifications given above. $M_1$ is intended to be the value of the mean hypothesized by the alternative hypothesis, $H_1$.
You can enter a list of values using the syntax 0 1 2 3 or 0 to 3 by 1.

**Parameter Values (S, A, B)**
Enter the numeric value(s) of the parameters listed above. These values are substituted for the corresponding letter in all four distribution specifications.
You can enter a list of values using the syntax 0 1 2 3 or 0 to 3 by 1.
You can also change the letter that is used as the name of this parameter using the pull-down menu to the side.

**Effect Size – Distribution Parameters**

**R (Correlation of Items A & B)**
Specify the value of the correlation between items (variates) A and B of the pair.
Since this is a correlation, it must be between -1 and 1. However, some distributions (such as the multinomial distribution) have a maximum possible correlation that is far less than one.
Typical values are between 0 and 0.4.

**Test and Simulations**

**Alternative Hypothesis**
This option specifies the alternative hypothesis, $H_1$. This implicitly specifies the direction of the hypothesis test. The null hypothesis is always $H_0$: Diff = Diff$0$. 
Note that the alternative hypothesis enters into power calculations by specifying the rejection region of the hypothesis test. Its accuracy is critical.

Possible selections are:

- **Difference <> Diff0**
  This is the most common selection. It yields a *two-tailed test*. Use this option when you are testing whether the mean is different from a specified value Diff0, but you do not want to specify beforehand whether it is smaller or larger. Most scientific journals require two-tailed tests.

- **Difference < Diff0**
  This option yields a *one-tailed test*. Use it when you want to test whether the true mean is less than Diff0.

- **Difference > Diff0**
  This option yields a *one-tailed test*. Use it when you want to test whether the true mean is greater than Diff0. Note that this option could be used for a *non-inferiority test*.

**Simulations**

This option specifies the number of iterations, M, used in the simulation. As the number of iterations is increased, the accuracy and running time of the simulation will be increased also.

The precision of the simulated power estimates are calculated from the binomial distribution. Thus, confidence intervals may be constructed for various power values. The following table gives an estimate of the precision that is achieved for various simulation sizes when the power is either 0.50 or 0.95. The table values are interpreted as follows: a 95% confidence interval of the true power is given by the power reported by the simulation plus and minus the ‘Precision’ amount given in the table.

<table>
<thead>
<tr>
<th>Simulation Size M</th>
<th>Precision when Power = 0.50</th>
<th>Precision when Power = 0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.100</td>
<td>0.044</td>
</tr>
<tr>
<td>500</td>
<td>0.045</td>
<td>0.019</td>
</tr>
<tr>
<td>1000</td>
<td>0.032</td>
<td>0.014</td>
</tr>
<tr>
<td>2000</td>
<td>0.022</td>
<td>0.010</td>
</tr>
<tr>
<td>5000</td>
<td>0.014</td>
<td>0.006</td>
</tr>
<tr>
<td>10000</td>
<td>0.010</td>
<td>0.004</td>
</tr>
<tr>
<td>50000</td>
<td>0.004</td>
<td>0.002</td>
</tr>
<tr>
<td>100000</td>
<td>0.003</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Notice that a simulation size of 1000 gives a precision of plus or minus 0.01 when the true power is 0.95. Also note that as the simulation size is increased beyond 5000, there is only a small amount of additional accuracy achieved.

**Test Type**

Specify which test statistic (t-test, Wilcoxon test, sign test, or bootstrap test) is to be simulated. Although the t-test is the most commonly used test statistic, it is based on assumptions that may not be viable in many situations. For your data, you may find that one of the other tests is more accurate (actual alpha = target alpha) and more precise (better power).

Note that the bootstrap test is computationally intensive, so it can be very slow to calculate.
Options Tab

The Options tab contains limits on the number of iterations and various options about individual tests.

Maximum Iterations

Maximum Iterations Before Search Termination

Specify the maximum number of iterations before the search for the sample size, N, is aborted. When the maximum number of iterations is reached without convergence, the sample size is left blank. We recommend a value of at least 500.

Bootstrap Iterations

Bootstrap Iterations

Specify the number of iterations used in the bootstrap hypothesis test. This value is only used if the bootstrap test is displayed on the reports. The running time of the procedure depends heavily on the number of iterations specified here.

Recommendations by authors of books discussing the bootstrap are from 100 to 10,000. If you enter a large (greater than 500) value, the simulation may take several hours to run.

Random Numbers

Random Number Pool Size

This is the size of the pool of random values from which the random samples will be drawn. Populations of at least 10,000 should be used. Also, the value should be about twice the number of simulations. You can enter Automatic and an appropriate value will be calculated.

Note that values over 50,000 may take a long time to permute to achieve the target means and correlation.

Correlation

Maximum Switches

This option specifies the maximum number of index switches that can be made while searching for a permutation of item B that yields a correlation within the specified range. A value near 5,000,000 may be necessary when the correlation is near one.

Correlation Tolerance

Specify the amount above and below the target correlation that will still let a particular index-permutation to be selected for the population. For example, if you have selected a correlation of 0.3 and you set this tolerance to 0.001, then only populations with a correlation between 0.299 and 0.301 will be used. The recommended is 0.001 or smaller. Valid values are between 0 and 0.999.
Example 1 – Power at Various Sample Sizes

Researchers are planning a pre-post experiment to test whether the difference in response to a certain drug is different from zero. The researchers will use a paired t-test with an alpha level of 0.05. They want to compare the power at sample sizes of 50, 100, and 150 when the shift in the means is 0.6 from pre-test to post-test. They assume that the data are normally distributed with a standard deviation of 2 and that the correlation between the pre-test and post-test values is 0.20. Since this is an exploratory analysis, they set the number of simulation iterations to 2000.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Paired Means (Simulation) procedure window by expanding Means, then clicking on Paired Means, and then clicking on Tests for Paired Means (Simulation). You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>50 100 150</td>
</tr>
<tr>
<td>Item A Distribution</td>
<td>H0</td>
</tr>
<tr>
<td>Item B Distribution</td>
<td>H0</td>
</tr>
<tr>
<td>Item A Distribution</td>
<td>H1</td>
</tr>
<tr>
<td>Item B Distribution</td>
<td>H1</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H1)</td>
</tr>
<tr>
<td>S</td>
<td>2</td>
</tr>
<tr>
<td>R (Correlation of Items A &amp; B)</td>
<td>0.2</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>Diff ≠ Diff0</td>
</tr>
<tr>
<td>Simulations</td>
<td>2000</td>
</tr>
<tr>
<td>Test Type</td>
<td>T-Test</td>
</tr>
</tbody>
</table>
Inequality Tests for Paired Means (Simulation)

Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>Diff0</th>
<th>H0</th>
<th>Diff1</th>
<th>Corr</th>
<th>R</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>Beta</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.388</td>
<td>50</td>
<td>0.0</td>
<td>-0.6</td>
<td>0.200</td>
<td>0.050</td>
<td>0.040</td>
<td>0.612</td>
<td>0.050</td>
<td>0.040</td>
<td>0.6</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>(0.021)</td>
<td>[0.367, 0.409]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.009)</td>
<td></td>
<td></td>
<td>(0.009)</td>
<td>[0.031, 0.048]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.642</td>
<td>100</td>
<td>0.0</td>
<td>-0.6</td>
<td>0.200</td>
<td>0.050</td>
<td>0.041</td>
<td>0.359</td>
<td>0.050</td>
<td>0.041</td>
<td>0.6</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>(0.021)</td>
<td>[0.620, 0.663]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.009)</td>
<td></td>
<td></td>
<td>(0.009)</td>
<td>[0.032, 0.049]</td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.830</td>
<td>150</td>
<td>0.0</td>
<td>-0.6</td>
<td>0.200</td>
<td>0.050</td>
<td>0.042</td>
<td>0.170</td>
<td>0.050</td>
<td>0.042</td>
<td>0.6</td>
<td>2.0</td>
<td></td>
</tr>
<tr>
<td>(0.016)</td>
<td>[0.814, 0.846]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>(0.009)</td>
<td></td>
<td></td>
<td>(0.009)</td>
<td>[0.033, 0.050]</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:

Report Definitions
Power is the probability of rejecting a false null hypothesis.
N is the size of the sample drawn from the population.
Diff0 is the paired-difference mean (A-B) assuming the null hypothesis, H0. This is the value being tested.
Diff1 is the paired-difference mean (A-B) assuming the alternative hypothesis, H1. This is the true value.
R is the correlation between the paired items.
Target Alpha is the probability of rejecting a true null hypothesis. It is set by the user.
Actual Alpha is the alpha level that was actually achieved by the experiment.
Beta is the probability of accepting a false null hypothesis.
Second Row: (Power Inc.) [95% LCL and UCL Power] (Alpha Inc.) [95% LCL and UCL Alpha]

Summary Statements
A sample size of 50 achieves 39% power to detect a difference of -0.6 between the null hypothesis mean difference of 0.0 and the actual mean difference of -0.6 at the 0.050 significance level (alpha) using a two-sided Paired T-Test. These results are based on 2000 Monte Carlo samples from the null distribution: Normal(M0 S) - Normal(M0 S) and the alternative distribution: Normal(M0 S) - Normal(M1 S).

Plots Section
This report shows the estimated power for each scenario. The first row shows the parameter settings and the estimated power and significance level (Actual Alpha).
The second row shows two 95% confidence intervals in brackets: the first for the power and the second for the significance level. Half the width of each confidence interval is given in parentheses as a fundamental measure of the accuracy of the simulation. As the number of simulations is increased, the width of the confidence intervals will decrease.

## Example 2 – Finding the Sample Size

Continuing with Example 1, the researchers want to determine how large a sample is needed to obtain a power of 0.90?

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Paired Means (Simulation) procedure window by expanding Means, then clicking on Paired Means, and then clicking on Tests for Paired Means (Simulation). You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>(N)</td>
</tr>
<tr>
<td>Power</td>
<td>(0.90)</td>
</tr>
<tr>
<td>Alpha</td>
<td>(0.05)</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Item A Distribution</td>
<td>H0</td>
</tr>
<tr>
<td>Item B Distribution</td>
<td>H0</td>
</tr>
<tr>
<td>Item A Distribution</td>
<td>H1</td>
</tr>
<tr>
<td>Item B Distribution</td>
<td>H1</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H1)</td>
</tr>
<tr>
<td>S</td>
<td>(2)</td>
</tr>
<tr>
<td>R (Correlation of Items A &amp; B)</td>
<td>(0.2)</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>(\text{Diff} \neq \text{Diff0})</td>
</tr>
<tr>
<td>Simulations</td>
<td>(2000)</td>
</tr>
<tr>
<td>Test Type</td>
<td>T-Test</td>
</tr>
</tbody>
</table>
Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results of Search for N

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>Diff0</th>
<th>H0</th>
<th>H1</th>
<th>Corr</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>Beta</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.897</td>
<td>187</td>
<td>0.0</td>
<td>0.0</td>
<td>-0.6</td>
<td>0.200</td>
<td>0.050</td>
<td>0.053</td>
<td>0.0</td>
<td>0.0</td>
<td>0.6</td>
<td>2.0</td>
</tr>
</tbody>
</table>

Notes:
Number of Monte Carlo Samples: 2000. Simulation Run Time: 95.53 seconds.

The required sample size of 187 achieved a power of 0.897. The power of 0.897 is less than the target value of 0.900 because the sample size search algorithm re-simulates the power for the final sample size. Thus it is possible for the search algorithm to converge to a sample size which exhibits the desired power, but then on a succeeding simulation to achieve a power that is slightly less than the target. To achieve more accuracy, a reasonable strategy would be to run simulations to obtain the powers using N’s from 180 to 200 using a simulation size of 5000 or greater.

Example 3 – Comparative Results

Continuing with Example 2, the researchers want to study the characteristics of alternative test statistics.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Paired Means (Simulation) procedure window by expanding Means, then clicking on Paired Means, and then clicking on Tests for Paired Means (Simulation). You may then make the appropriate entries as listed below, or open Example 3 by going to the File menu and choosing Open Example Template.

Option | Value
---|---
Data Tab | Power
Find (Solve For) | Power
Power | Ignored since this is the Find setting
Alpha | 0.05
N (Sample Size) | 50 100 150 200
Item A Distribution|H0 | N(M0 S)
Item B Distribution|H0 | N(M0 S)
Item A Distribution|H1 | N(M0 S)
Item B Distribution|H1 | N(M1 S)
M0 (Mean|H0) | 0
M1 (Mean|H1) | 0.6
S | 2
R (Correlation of Items A & B) | 0.2
Data Tab (continued)

Alternative Hypothesis .............................................. \( \text{Diff} \neq \text{Diff}_0 \)
Simulations ................................................................. 2000
Test Type ................................................................. T-Test

Reports Tab
Show Comparative Reports ........................................... Checked
Show Comparative Plots ............................................... Checked

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

Power Comparison for Testing Mean Difference = \( \text{Diff}_0 \).  Hypotheses: \( H_0: \text{Diff}_1=\text{Diff}_0; H_1: \text{Diff}_1\neq\text{Diff}_0 \)
\( H_0 \) Dist'n: Normal(\( M_0 \) \( S \)) - Normal(\( M_0 \) \( S \))
\( H_1 \) Dist'n: Normal(\( M_0 \) \( S \)) - Normal(\( M_1 \) \( S \))

<table>
<thead>
<tr>
<th>N</th>
<th>( \text{H}_0 ) Diff (( \text{Diff}_0 ))</th>
<th>( \text{H}_1 ) Diff (( \text{Diff}_1 ))</th>
<th>Corr (R)</th>
<th>Target Power</th>
<th>T-Test Power</th>
<th>Wilcox Power</th>
<th>Sign Power</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.0</td>
<td>-0.6</td>
<td>0.200</td>
<td>0.050</td>
<td>0.355</td>
<td>0.335</td>
<td>0.187</td>
<td>0.0</td>
<td>0.6</td>
<td>2.0</td>
</tr>
<tr>
<td>100</td>
<td>0.0</td>
<td>-0.6</td>
<td>0.200</td>
<td>0.050</td>
<td>0.644</td>
<td>0.610</td>
<td>0.385</td>
<td>0.0</td>
<td>0.6</td>
<td>2.0</td>
</tr>
<tr>
<td>150</td>
<td>0.0</td>
<td>-0.6</td>
<td>0.200</td>
<td>0.050</td>
<td>0.843</td>
<td>0.818</td>
<td>0.617</td>
<td>0.0</td>
<td>0.6</td>
<td>2.0</td>
</tr>
<tr>
<td>200</td>
<td>0.0</td>
<td>-0.6</td>
<td>0.200</td>
<td>0.050</td>
<td>0.921</td>
<td>0.905</td>
<td>0.734</td>
<td>0.0</td>
<td>0.6</td>
<td>2.0</td>
</tr>
</tbody>
</table>

Number of Monte Carlo Iterations: 2000.  Simulation Run Time: 10.50 seconds.

Alpha Comparison for Testing Mean Difference = \( \text{Diff}_0 \).  Hypotheses: \( H_0: \text{Diff}_1=\text{Diff}_0; H_1: \text{Diff}_1\neq\text{Diff}_0 \)
\( H_0 \) Dist'n: Normal(\( M_0 \) \( S \)) - Normal(\( M_0 \) \( S \))
\( H_1 \) Dist'n: Normal(\( M_0 \) \( S \)) - Normal(\( M_1 \) \( S \))

<table>
<thead>
<tr>
<th>N</th>
<th>( \text{H}_0 ) Diff (( \text{Diff}_0 ))</th>
<th>( \text{H}_1 ) Diff (( \text{Diff}_1 ))</th>
<th>Corr (R)</th>
<th>Target Alpha</th>
<th>T-Test Alpha</th>
<th>Wilcox Alpha</th>
<th>Sign Alpha</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>50</td>
<td>0.0</td>
<td>-0.6</td>
<td>0.200</td>
<td>0.059</td>
<td>0.060</td>
<td>0.034</td>
<td>0.0</td>
<td>0.0</td>
<td>0.6</td>
<td>2.0</td>
</tr>
<tr>
<td>100</td>
<td>0.0</td>
<td>-0.6</td>
<td>0.200</td>
<td>0.050</td>
<td>0.052</td>
<td>0.033</td>
<td>0.0</td>
<td>0.0</td>
<td>0.6</td>
<td>2.0</td>
</tr>
<tr>
<td>150</td>
<td>0.0</td>
<td>-0.6</td>
<td>0.200</td>
<td>0.055</td>
<td>0.050</td>
<td>0.041</td>
<td>0.0</td>
<td>0.0</td>
<td>0.6</td>
<td>2.0</td>
</tr>
<tr>
<td>200</td>
<td>0.0</td>
<td>-0.6</td>
<td>0.200</td>
<td>0.058</td>
<td>0.058</td>
<td>0.050</td>
<td>0.0</td>
<td>0.0</td>
<td>0.6</td>
<td>2.0</td>
</tr>
</tbody>
</table>

Number of Monte Carlo Iterations: 2000.  Simulation Run Time: 10.50 seconds.

These results show that for paired data, the t-test and Wilcoxon test have very similar power and alpha values. The sign test is less accurate and less powerful.
**Example 4 – Validation**

We will validate this procedure by comparing its results to those of the regular one-sample t-test, a procedure that has already been validated. For this run, we will use the settings of Example 1: $M_0 = 0$, $M_1 = 0.6$, $\alpha = 0.05$, $N = 50$, $R = 0.2$, and $S = 2$.

Note that to run this example using the regular one-sample t-test procedure, the variance will have to be altered to account for the correlation of 0.20. The adjusted standard deviation is equal to $S$ times the square root of $2(1 - R)$, which, in this case, is 2.530. Running this through the regular One Mean procedure yields a power of 0.376.

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Paired Means (Simulation) procedure window by expanding Means, then clicking on Paired Means, and then clicking on Tests for Paired Means (Simulation). You may then make the appropriate entries as listed below, or open Example 4 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>$N$ (Sample Size)</td>
<td>50</td>
</tr>
<tr>
<td>Item A Distribution</td>
<td>H0</td>
</tr>
<tr>
<td>Item B Distribution</td>
<td>H0</td>
</tr>
<tr>
<td>Item A Distribution</td>
<td>H1</td>
</tr>
<tr>
<td>Item B Distribution</td>
<td>H1</td>
</tr>
<tr>
<td>$M_0$ (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>$M_1$ (Mean</td>
<td>H1)</td>
</tr>
<tr>
<td>$S$</td>
<td>2</td>
</tr>
<tr>
<td>$R$ (Correlation of Items A &amp; B)</td>
<td>0.2</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>$\text{Diff} \neq \text{Diff}_0$</td>
</tr>
<tr>
<td>Simulations</td>
<td>10000</td>
</tr>
<tr>
<td>Test Type</td>
<td>T-Test</td>
</tr>
</tbody>
</table>

**Options Tab**

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Random Number Pool Size</td>
<td>50000 (Increase to 5 times Simulations)</td>
</tr>
</tbody>
</table>
Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>Diff0</th>
<th>H0</th>
<th>Diff1</th>
<th>Corr</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>Beta</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.373</td>
<td>50</td>
<td>0.0</td>
<td>-0.6</td>
<td>0.200</td>
<td>0.050</td>
<td>0.049</td>
<td>0.049</td>
<td>0.627</td>
<td>0.0</td>
<td>0.6</td>
<td>2.0</td>
</tr>
</tbody>
</table>

Notes:
Number of Monte Carlo Samples: 10000. Simulation Run Time: 30.97 seconds.

The power matches the exact value of 0.376 quite well. We re-ran the procedure several times and obtained power values from 0.370 to 0.396.

Example 5 – Non-Inferiority Test

A non-inferiority test is appropriate when you want to show that a new treatment is no worse than the standard. For example, suppose that a standard diagnostic test has an average score of 70. Unfortunately, this diagnostic test is expensive. A promising new diagnostic test must be compared to the standard. Researchers want to show that it is no worse than the standard.

Because of many benefits from the new test, clinicians are willing to adopt it even if it is slightly less accurate than the current test. How much less can the score of the new treatment be and still be adopted? Should it be adopted if the difference is -1? -2? -5? -10? There is an amount below 0 at which the difference between the two treatments is no longer considered ignorable. After thoughtful discussion with several clinicians, the margin of equivalence is set to -5.

The developers decided to use a paired t-test. They must design an experiment to test the hypothesis that the average difference between the two tests is greater than -5. The statistical hypothesis to be tested is

\[ H_0: A - B \leq -5 \quad \text{versus} \quad H_1: A - B > -5 \]

where A represents the mean of the new test and B represents the mean of the standard test. Notice that when the null hypothesis is rejected, the conclusion is that the average difference is greater than -5.

Past experience has shown that the standard deviation is 5.0 and the correlation is 0.2. Following proper procedure, the researchers decide to use a significance level of 0.025 for this one-sided test to keep it comparable to the usual value of 0.05 for a two-sided test. They decide to look at the power for sample sizes of 5, 10, 15, 20, and 25 subjects. They decide to compute the power for the case when the two tests are actually equal.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Paired Means (Simulation) procedure window by expanding Means, then clicking on Paired Means, and then clicking on Tests for Paired Means (Simulation). You may then make the appropriate entries as listed below, or open Example 5 by going to the File menu and choosing Open Example Template.
Option | Value
--- | ---
Data Tab | Power
Find (Solve For) | Power
Ignored since this is the Find setting
Alpha | 0.025
N (Sample Size) | 5 10 15 20 25
Item A Distribution|H0 | N(M0 S)
Item B Distribution|H0 | N(M1 S)
Item A Distribution|H1 | N(M0 S)
Item B Distribution|H1 | N(M0 S)
M0 (Mean|H0) | 0
M1 (Mean|H1) | 5
S | 5
R (Correlation of Items A & B) | 0.2
Alternative Hypothesis | Diff > Diff0
Simulations | 2000
Test Type | T-Test

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

| Power | N | M0 | M1 | S | H0 | H1 | Diff0 | Diff1 | R | Corr | Target Alpha | Actual Alpha | Beta | M0 | M1 | S |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0.308 | 5 | 0.692 | 0.617 | 0.816 | 0.916 | 0.968 |
| 0.617 | 10 | 0.638 | 0.816 | 0.833 | 0.817 | 0.638 |
| 0.916 | 20 | 0.903 | 0.903 | 0.903 | 0.912 | 0.890 |
| 0.968 | 25 | 0.960 | 0.976 | 0.976 | 0.968 | 0.976 |
| 0.020 | 0.288 | 0.021 | 0.799 | 0.017 | 0.012 | 0.008 |
| 0.032 | 0.034 | 0.028 | 0.031 | 0.020 | 0.018 | 0.016 |
| 0.023 | 0.007 | 0.024 | 0.007 | 0.027 | 0.007 | 0.025 |
| 0.011 | 0.007 | 0.016 | 0.007 | 0.019 | 0.007 | 0.007 |
| Notes: Number of Monte Carlo Samples: 2000. Simulation Run Time: 13.34 seconds.

We see that a power of 0.8 is achieved at about 15 subjects, while a power of 0.9 requires about 20 subjects.
Chapter 495

Equivalence Tests for Paired Means (Simulation)

Introduction

This procedure allows you to study the power and sample size of tests of equivalence of means of two correlated variables. Schuirmann’s (1987) two one-sided tests (TOST) approach is used to test equivalence. The paired t-test is commonly used in this situation. Other tests have been developed for the case when the data are not normally distributed. These additional tests include the Wilcoxon signed-ranks test, the sign test, and the computer-intensive bootstrap test.

Paired data may occur because two measurements are made on the same subject or because measurements are made on two subjects that have been matched according to other, often demographic, variables. Hypothesis tests on paired data can be analyzed by considering the differences between the paired items. The distribution of differences is usually symmetric. In fact, the distribution must be symmetric if the individual distributions of the two items are identical. Hence, the paired t-test and the Wilcoxon signed-rank test are appropriate for paired data even when the distributions of the individual items are not normal.

The details of the power analysis of the paired t-test using analytic techniques are presented in another PASS chapter and they won’t be duplicated here. This chapter will only consider power analysis using computer simulation.

Technical Details

Computer simulation allows us to estimate the power and significance-level that is actually achieved by a test procedure in situations that are not mathematically tractable. Computer simulation was once limited to mainframe computers. But, in recent years, as computer speeds have increased, simulation studies can be completed on desktop and laptop computers in a reasonable period of time.

The steps to a simulation study are as follows.

1. Specify the test procedure and the test statistic. This includes the significance level, sample size, and underlying data distributions.
2. Generate a random sample \( X_1, X_2, \ldots, X_n \) from the distribution specified by the alternative hypothesis. Calculate the test statistic from the simulated data and determine if the null hypothesis is accepted or rejected. These samples are used to calculate the power of the test. In the case of paired data, the individual values are simulated as the difference between two other random variables. These samples are constructed so that they exhibit a certain amount of correlation.

3. Generate a random sample \( Y_1, Y_2, \ldots, Y_n \) from the distribution specified by the null hypothesis. Calculate the test statistic from the simulated data and determine if the null hypothesis is accepted or rejected. These samples are used to calculate the significance-level of the test. In the case of paired data, the individual values are simulated as the difference between two other random variables. These samples are constructed so that they exhibit a certain amount of correlation.

4. Repeat steps 2 and 3 several thousand times, tabulating the number of times the simulated data leads to a rejection of the null hypothesis. The power is the proportion of simulation samples in step 2 that lead to rejection. The significance-level is the proportion of simulated samples in step 3 that lead to rejection.

---

**Simulating Paired Distributions**

Paired data occur when two observations are correlated. Examples of paired designs are pre – post designs, cross-over designs, and matched pair designs.

In order to simulate paired data, the simulation should mimic the actual data generation process as closely as possible. Since paired data are analyzed by creating the individual difference between each pair, the simulation should also create data as the difference between two variates. Paired data exhibit a correlation between the two variates. As this correlation between the variates increases, the variance of the difference decreases. Thus it is important not only to specify the distributions of the two variates that will be differenced, but to also specify their correlation.

Obtaining paired samples from arbitrary distributions with a set correlation is difficult because the joint, bivariate distribution must be specified and simulated. Rather than specify the bivariate distribution, \texttt{PASS} requires the specification of the two marginal distributions and the correlation between them.

Monte Carlo samples with given marginal distributions and correlation are generated using the method suggested by Gentle (1998). The method begins by generating a large population of random numbers from the two distributions. Each of these populations is evaluated to determine if their means are within a small relative tolerance (0.0001) of the target mean. If the actual mean is not within the tolerance of the target mean, individual members of the population are replaced with new random numbers if the new random number moves the mean towards its target. Only a few hundred such swaps are required to bring the actual mean to within tolerance of the target mean.

The next step is to obtain the target correlation. This is accomplished by permuting one of the populations until they have the desired correlation.

The above steps provide a large pool of random numbers that exhibit the desired characteristics. This pool is then sampled at random using the uniform distribution to obtain the random numbers used in the simulation.
This algorithm may be stated as follows.

1. Draw individual samples of size $M$ from the two distributions where $M$ is a large number, usually over 10,000. Adjust these samples so that they have the specified mean and standard deviation. Label these samples A and B. Create an index of the values of A and B according to the order in which they are generated. Thus, the first value of A and the first value of B are indexed as one, the second values of A and B are indexed as two, and so on up to the final set which is indexed as $M$.

2. Compute the correlation between the two generated variates.

3. If the computed correlation is within a small tolerance (usually less than 0.001) of the specified correlation, go to step 7.

4. Select two indices (I and J) at random using uniform random numbers.

5. Determine what will happen to the correlation if $B_I$ is swapped with $B_J$. If the swap will result in a correlation that is closer to the target value, swap the indices and proceed to step 6. Otherwise, go to step 4.

6. If the computed correlation is within the desired tolerance of the target correlation, go to step 7. Otherwise, go to step 4.

7. End with a population with the required marginal distributions and correlation.

Now, to complete the simulation, random samples of the designated size are drawn from this population.

**Simulating Data for an Equivalence Test**

Simulating equivalence data is more complex than simulating data for a regular two-sided test. An equivalence test essentially reverses the roles of the null and alternative hypothesis. In so doing, the null hypothesis becomes

$$H_0: (\mu_1 - \mu_2) \leq -D \text{ or } (\mu_1 - \mu_2) \geq D$$

where $D$ is the margin of equivalence. Thus the null hypothesis is made up of two simple hypotheses:

$$H_{01}: (\mu_1 - \mu_2) \leq -D$$

$$H_{02}: (\mu_1 - \mu_2) \geq D$$

The additional complexity comes in deciding which of the two simple null hypotheses are used to simulate data for the null hypothesis situation. The choice becomes more problematic when asymmetric equivalence limits are chosen. In that case, you may want to try simulating using each simple null hypothesis in turn.

To generate data for the null hypotheses, you generate data for each group. The difference in the means of these two groups will become one of the equivalence limits. The other equivalence limit will be determined by symmetry and will always have a sign that is the negative of the first equivalence limit.
Test Statistics

This section describes the test statistics that are available in this procedure. Note that these test statistics are computed on the differences. Thus, when the equation refers to an X value, this X value is assumed to be a difference between two individual variates.

One-Sample t-Test

The one-sample t-test assumes that the data are a simple random sample from a population of normally-distributed values that all have the same mean and variance. This assumption implies that the data are continuous and their distribution is symmetric. The calculation of the t-test proceeds as follow

\[ t_{n-1} = \frac{\bar{X} - M0}{s_X} \]

where

\[ \bar{X} = \frac{\sum_{i=1}^{n} X_i}{n} , \]

\[ s_X = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}} , \]

and M0 is the value of the difference hypothesized by the null hypothesis.

The significance of the test statistic is determined by computing the p-value. If this p-value is less than a specified level (usually 0.05), the hypothesis is rejected. Otherwise, no conclusion can be reached.

Wilcoxon Signed-Rank Test

The Wilcoxon signed-rank test is a popular, nonparametric substitute for the t-test. It assumes that the data follow a symmetric distribution. The test is computed using the following steps.

1. Subtract the hypothesized mean, D0, from each data value. Rank the values according to their absolute values.
2. Compute the sum of the positive ranks \( Sp \) and the sum of the negative ranks \( Sn \). The test statistic, \( W \), is the minimum of \( Sp \) and \( Sn \).
3. Compute the mean and standard deviation of \( W \) using the formulas

\[ \mu_{w_s} = \frac{n(n+1)}{4} \quad \text{and} \quad \sigma_{w_s} = \sqrt{\frac{n(n+1)(2n+1)}{24} - \frac{\sum t_i^2 - \sum t}{48}} \]

where \( t_i \) represents the number of times the \( i^{th} \) value occurs.
4. Compute the \( z \) value using

\[ z_w = \frac{W - \mu_{w_s}}{\sigma_{w_s}} \]
For cases when \( n \) is less than 38, the significance level is found from a table of exact probabilities for the Wilcoxon test. When \( n \) is greater than or equal to 38, the significance of the test statistic is determined by computing the p-value using the standard normal distribution. If this p-value is less than a specified level (usually 0.05), the null hypothesis is rejected in favor of the alternative hypothesis. Otherwise, no conclusion can be reached.

**Sign Test**
The sign test is popular because it is simple to compute. It assumes that the data follow the same distribution. The test is computed using the following steps.

1. Count the number of values strictly greater than \( M_0 \). Call this value \( X \).
2. Count the number of values strictly less than \( M_0 \). Call this value \( Y \).
3. Set \( m = X + Y \).
4. Under the null hypothesis, \( X \) is distributed as a binomial random variable with a proportion of 0.5 and sample size of \( m \).

The significance of \( X \) is calculated using binomial probabilities.

**Bootstrap Test**
The one-sample bootstrap procedure for testing whether the mean is equal to a specific value is given in Efron & Tibshirani (1993) pages 224-227. The bootstrap procedure is as follows.

1. Compute the mean of the sample. Call it \( \bar{X} \).
2. Compute the t-value using the standard t-test. The formula for this computation is
   \[
   t_X = \frac{\bar{X} - M_0}{s_X}
   \]
3. Draw a random, with-replacement sample of size \( n \) from the original \( X \) values. Call this sample \( Y_1, Y_2, \ldots, Y_n \).
4. Compute the t-value of this bootstrap sample using the formula
   \[
   t_Y = \frac{\bar{Y} - \bar{X}}{s_T}
   \]
5. For a two-tailed test, if \( |t_Y| > |t_X| \) then add one to a counter variable \( A \).
6. Repeat steps 3 – 5 \( B \) times. \( B \) may be anywhere from 100 to 10,000.
7. Compute the p-value of the bootstrap test as \( (A + 1) / (B + 1) \)
8. Steps 1 – 7 complete one simulation iteration. Repeat these steps \( M \) times, where \( M \) is the number of simulations. The power and significance level is equal to the percent of the time the p-value is less than the nominal alpha of the test.

Note that the bootstrap test is a time-consuming test to run, especially if you set \( B \) to a value larger than 100.
The Problem of Differing Standard Deviations

Care must be used when either the null or alternative distribution is not normal. In these cases, the standard deviation is usually not specified directly. For example, you might use a gamma distribution with a shape parameter of 1.5 and a mean of 4 as the null distribution and a gamma distribution with the same shape parameter and a mean of 5 as the alternative distribution. This allows you to compare the two means. However, note that although the shape parameters are constant, the standard deviations are not. Thus the null and alternatives not only have different means, but different standard deviations!

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data and Options tabs. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains most of the parameters and options that you will be concerned with.

Solve For

Find (Solve For)

This option specifies whether you want to find Power or \( N \) from the simulation. Select Power when you want to estimate the power of a certain scenario. Select \( N \) when you want to determine the sample size needed to achieve a given power and alpha error level. Finding \( N \) is very computationally intensive, and so it may take a long time to complete.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.
You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.

**Sample Size**

**N (Sample Size)**
This option specifies one or more values of the sample size, the number of subjects in the study. The paired design assumes that a pair of observations will be obtained from each subject. Thus there will be 2N observations simulated, resulting in N differences.

This value must be an integer greater than one. You may enter a list of values using the syntax 50 100 150 200 250 or 50 to 250 by 50.

**Effect Size**

**Item A (and B) Distribution|H0**
These options specify the distributions of the two items making up the pair under the null hypothesis, H0. The difference between the means of these two distributions is the difference that is tested, Diff0.

Usually, you will want Diff0 = 0. This zero difference is specified by entering M0 for the mean parameter in each of the distributions and then entering an appropriate value for the M0 parameter below.

All of the distributions are parameterized so that the mean is entered first. For example, if you wanted to test whether the mean of a normal distributed variable is five, you could enter N(5, S) or N(M0, S) here.

The parameters of each distribution are specified using numbers or letters. If letters are used, their values are specified in the boxes below. The value M0 is reserved for the value of the mean under the null hypothesis.

Following is a list of the distributions that are available and the syntax used to specify them:

- Beta=A(M0, A, B, Minimum)
- Binomial=B(M0, N)
- Cauchy=C(M0, Scale)
- Constant=K(Value)
- Exponential=E(M0)
- F=F(M0, DF1)
- Gamma=G(M0, A)
- Multinomial=M(P1, P2, …, Pk)
- Normal=N(M0, SD)
- Poisson=P(M0)
- Student's T=T(M0, D)
- Tukey's Lambda=L(M0, S, Skewness, Elongation)
Uniform=U(M0,Minimum)
Weibull=W(M0,B)

Details of writing mixture distributions, combined distributions, and compound distributions are found in the chapter on Data Simulation and are not repeated here.

**Finding the Value of the Mean of a Specified Distribution**

Except for the multinomial distribution, the distributions have been parameterized in terms of their means, since this is the parameter being tested. The mean of a distribution created as a linear combination of other distributions is found by applying the linear combination to the individual means. However, the mean of a distribution created by multiplying or dividing other distributions is not necessarily equal to applying the same function to the individual means. For example, the mean of 4N(4, 5) + 2N(5, 6) is \(4 \times 4 + 2 \times 5 = 26\), but the mean of 4N(4, 5) * 2N(5, 6) is not exactly \(4 \times 4 \times 2 \times 5 = 160\) (although it is close).

**Item A (and B) Distribution|H1**

These options specify the distributions of the two items making up the pair under the alternative hypothesis, H1. The difference between the means of these two distributions is the difference that is assumed to be the true value of the difference. That is, this is the difference at which the power is computed.

Usually, the mean difference is specified by entering \(M1\) for the mean parameter in the distribution expression for item A and \(M0\) for the mean parameter in the distribution expression for item B. The mean difference under H1 then becomes the value of \(M1 - M0\).

The parameters of each distribution are specified using numbers or letters. If letters are used, their values are specified in the boxes below. The value \(M1\) is reserved for the value of the mean under the alternative hypothesis.

Following is a list of the distributions that are available and the syntax used to specify them:

- Beta=A(M1,A,B,Minimum)
- Binomial=B(M1,N)
- Cauchy=C(M1,Scale)
- Constant=K(Value)
- Exponential=E(M1)
- F=F(M1,DF1)
- Gamma=G(M1,A)
- Multinomial=M(P1,P2,…,Pk)
- Normal=N(M1,SD)
- Poisson=P(M1)
- Student's T=T(M1,D)
- Tukey's Lambda=L(M1,S,Skewness,Elongation)
- Uniform=U(M1,Minimum)
- Weibull=W(M1,B)

Details of writing mixture distributions, combined distributions, and compound distributions are found in the chapter on Data Simulation and will not be repeated here.
Effect Size – Distribution Parameters

**M0 (Mean|H0)**
These values are substituted for the $M0$ in the four distribution specifications given above. $M0$ is intended to be the value of the mean hypothesized by the null hypothesis, H0.

You can enter a list of values using the syntax $0 \ 1 \ 2 \ 3$ or $0 \ to \ 3 \ by \ 1$.

**M1 (Mean|H1)**
These values are substituted for the $M1$ in the four distribution specifications given above. $M1$ is intended to be the value of the mean hypothesized by the alternative hypothesis, H1.

You can enter a list of values using the syntax $0 \ 1 \ 2 \ 3$ or $0 \ to \ 3 \ by \ 1$.

**R (Correlation of Items A & B)**
Specify the value of the correlation between items (variates) A and B of the pair.

Since this is a correlation, it must be between -1 and 1. However, some distributions (such as the multinomial distribution) have a maximum possible correlation that is far less than one.

Typical values are between 0 and 0.4.

**Parameter Values (S, A, B)**
Enter the numeric value(s) of the parameters listed above. These values are substituted for the corresponding letter in all four distribution specifications.

You can enter a list of values using the syntax $0 \ 1 \ 2 \ 3$ or $0 \ to \ 3 \ by \ 1$.

You can also change the letter that is used as the name of this parameter using the pull-down menu to the side.

---

Test and Simulations

**Equivalence Limit**
*Equivalence limits* are defined as the positive and negative limits around zero that define a zone of equivalence. This zone of equivalence is a set of difference values that define a region in which the two means are ‘close enough’ so that they are considered to be the same for practical purposes.

Rather than define these limits explicitly, they are set implicitly. This is done as follows. One limit is found by subtracting the Item B mean | H0 from the Item A mean | H0. If the limits are symmetric, the other limit is this difference times -1. To obtain symmetric limits, enter ‘Symmetric’ here.

If asymmetric limits are desired, a numerical value is specified here. It will be given the sign (+ or -) that is opposite the difference in the means discussed above.

For example, if the mean of A under H0 is 5, the mean of B under H0 is 4, and ‘Symmetric’ is entered here, the equivalence limits will be $5 - 4 = 1$ and -1. However, if the value ‘1.25’ is entered here, the equivalence limits are 1 and -1.25.

If you do not have a specific value in mind for the equivalence limit, a common value for an equivalence limit is 20% or 25% of the Item A (reference) mean.
Test Type
Specify which test statistic (t-test, Wilcoxon test, sign test, or bootstrap test) is to be simulated. Although the t-test is the most commonly used test statistic, it is based on assumptions that may not be viable in many situations. For your data, you may find that one of the other tests is more accurate (actual alpha = target alpha) and more precise (better power).

Note that the bootstrap test is computationally intensive, so it can be very slow to calculate.

Simulations
This option specifies the number of iterations, M, used in the simulation. As the number of iterations is increased, the accuracy and running time of the simulation will be increased also.

The precision of the simulated power estimates are calculated from the binomial distribution. Thus, confidence intervals may be constructed for various power values. The following table gives an estimate of the precision that is achieved for various simulation sizes when the power is either 0.50 or 0.95. The table values are interpreted as follows: a 95% confidence interval of the true power is given by the power reported by the simulation plus and minus the ‘Precision’ amount given in the table.

<table>
<thead>
<tr>
<th>Simulation Size M</th>
<th>Precision when Power = 0.50</th>
<th>Precision when Power = 0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.100</td>
<td>0.044</td>
</tr>
<tr>
<td>500</td>
<td>0.045</td>
<td>0.019</td>
</tr>
<tr>
<td>1000</td>
<td>0.032</td>
<td>0.014</td>
</tr>
<tr>
<td>2000</td>
<td>0.022</td>
<td>0.010</td>
</tr>
<tr>
<td>5000</td>
<td>0.014</td>
<td>0.006</td>
</tr>
<tr>
<td>10000</td>
<td>0.010</td>
<td>0.004</td>
</tr>
<tr>
<td>50000</td>
<td>0.004</td>
<td>0.002</td>
</tr>
<tr>
<td>100000</td>
<td>0.003</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Notice that a simulation size of 1000 gives a precision of plus or minus 0.01 when the true power is 0.95. Also note that as the simulation size is increased beyond 5000, there is only a small amount of additional accuracy achieved.

Options Tab
The Options tab contains limits on the number of iterations and various options about individual tests.

Maximum Iterations

Maximum Iterations Before Search Termination
Specify the maximum number of iterations before the search for the sample size, N, is aborted. When the maximum number of iterations is reached without convergence, the sample size is left blank. We recommend a value of at least 500.
Bootstrap Iterations

Bootstrap Iterations
Specify the number of iterations used in the bootstrap hypothesis test. This value is only used if
the bootstrap test is displayed on the reports. The running time of the procedure depends heavily
on the number of iterations specified here.

Recommendations by authors of books discussing the bootstrap are from 100 to 10,000. If you
enter a large (greater than 500) value, the simulation may take several hours to run.

Random Numbers

Random Number Pool Size
This is the size of the pool of random values from which the random samples will be drawn.
Populations of at least 10,000 should be used. Also, the value should be about twice the number
of simulations. You can enter Automatic and an appropriate value will be calculated.

Note that values over 50,000 may take a long time to permute to achieve the target means and
correlation.

Correlation

Maximum Switches
This option specifies the maximum number of index switches that can be made while searching
for a permutation of item B that yields a correlation within the specified range. A value near
5,000,000 may be necessary when the correlation is near one.

Correlation Tolerance
Specify the amount above and below the target correlation that will still let a particular
permutation to be selected for the population. For example, if you have selected a correlation of
0.3 and you set this tolerance to 0.001, then only populations with a correlation between 0.299
and 0.301 will be used. The recommended is 0.001 or smaller. Valid values are between 0 and
0.999.

Example 1 – Power at Various Sample Sizes

Researchers are planning an experiment to determine if the response to a new drug is equivalent
to the response to the standard drug. The average response level to the standard drug is 63 with a
standard deviation of 5. The researchers decide that if the average response level to the new drug
is between 60 and 66, they will consider it to be equivalent to the standard drug.

The researchers decide to use a paired design so that each subject can serve as their own control.
The response level for the standard drug will be measured for each subject. Then, followed by an
appropriate wash-out period of two days, the response level to the new drug will be measured.
From previous studies, they know that the correlation between the two response levels will be
between 0.1 and 0.20.

The researchers will analyze the data using an equivalence test based on the paired t-test with an
alpha level of 0.05. They want to compare the power at sample sizes of 10, 30, 50, and 70. They
assume that the data are normally distributed and that the true difference between the response
level of the two drugs is zero. Since this is an exploratory analysis, they set the number of simulation iterations to 2000.

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for Paired Means (Simulation)** procedure window by expanding **Means**, then clicking on **Paired Means**, and then clicking on **Equivalence Tests for Paired Means (Simulation)**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Tab</strong></td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>10 30 50 70</td>
</tr>
<tr>
<td>Item A (Reference) Dist'n</td>
<td>H0</td>
</tr>
<tr>
<td>Item B (Treatment) Dist'n</td>
<td>H0</td>
</tr>
<tr>
<td>Item A (Reference) Dist'n</td>
<td>H1</td>
</tr>
<tr>
<td>Item B (Treatment) Dist'n</td>
<td>H1</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H1)</td>
</tr>
<tr>
<td>S</td>
<td>5</td>
</tr>
<tr>
<td>R (Correlation of Items A &amp; B)</td>
<td>0.1 0.2</td>
</tr>
<tr>
<td>Equivalence Limit</td>
<td>Symmetric</td>
</tr>
<tr>
<td>Test Type</td>
<td>T-Test</td>
</tr>
<tr>
<td>Simulations</td>
<td>2000</td>
</tr>
</tbody>
</table>

**Annotated Output**

Click the Run button to perform the calculations and generate the following output.

**Numeric Results**

| Numeric Results for Testing Mean Equivalence.  | Hypotheses: H0: Diff>=|Diff0|; H1: Diff<|Diff0| |
|-----------------------------------------------|------------------------|
| H0 Dist'n: Normal(M0 S) - Normal(M1 S)        |                         |
| H1 Dist'n: Normal(M0 S) - Normal(M0 S)        |                         |
| Test Statistic: Paired T-Test                 |                         |

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>H0 Diff1</th>
<th>Lower Equiv. Limit</th>
<th>Upper Equiv. Limit</th>
<th>Corr R</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.030</td>
<td>10</td>
<td>0.0</td>
<td>-3.0</td>
<td>3.0</td>
<td>0.100</td>
<td>0.050</td>
<td>0.009</td>
<td>63</td>
<td>66</td>
<td>5</td>
</tr>
<tr>
<td>(0.007)</td>
<td>[0.022 0.037]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.055</td>
<td>10</td>
<td>0.0</td>
<td>-3.0</td>
<td>3.0</td>
<td>0.200</td>
<td>0.050</td>
<td>0.019</td>
<td>63</td>
<td>66</td>
<td>5</td>
</tr>
<tr>
<td>(0.010)</td>
<td>[0.045 0.065]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.560</td>
<td>30</td>
<td>0.0</td>
<td>-3.0</td>
<td>3.0</td>
<td>0.100</td>
<td>0.050</td>
<td>0.050</td>
<td>63</td>
<td>66</td>
<td>5</td>
</tr>
<tr>
<td>(0.022)</td>
<td>[0.538 0.581]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Report Definitions
Power is the probability of rejecting a false null hypothesis.
N is the size of the sample drawn from the population.
Diff1 is the paired-difference mean (A-B) assuming the alternative hypothesis, H1. This is the true value.
Lower Equiv Limit is the lower limit on a difference (A-B) that is considered as equivalent.
Upper Equiv Limit is the upper limit on a difference (A-B) that is considered as equivalent.
Diff0 is the paired-difference mean (A-B) assuming the null hypothesis, H0. This is one of the equivalence limits.
R is the correlation between the paired items.
Target Alpha is the probability of rejecting a true null hypothesis. It is set by the user.
Actual Alpha is the alpha level that was actually achieved by the experiment.
Beta is the probability of accepting a false null hypothesis.
Second Row: (Power Inc.) [95% LCL and UCL Power] (Alpha Inc.) [95% LCL and UCL Alpha]

Summary Statements
A sample size of 10 pairs with a correlation of 0.100 achieves 3% power to detect equivalence when the margin of equivalence is from -3.0 to 3.0 and the actual mean difference is 0.0. The significance level (alpha) is 0.050 using two one-sided Paired T-Tests. These results are based on 2000 Monte Carlo samples from the null distribution: Normal(M0 S) - Normal(M1 S) and the alternative distribution: Normal(M0 S) - Normal(M0 S).

Chart Section

This report shows the estimated power for each scenario. The first row shows the parameter settings and the estimated power and significance level (Actual Alpha).
The second row shows two 95% confidence intervals in brackets: the first for the power and the second for the significance level. Half the width of each confidence interval is given in parentheses as a fundamental measure of the accuracy of the simulation. As the number of simulations is increased, the width of the confidence intervals will decrease.
We see that a sample size of about 50 is needed to obtain a reasonable power level.
Example 2 – Finding the Sample Size

Continuing with Example 1, the researchers want to determine how large a sample is needed to obtain a power of 0.90? They decide to use a correlation of 0.10, since that will result in a larger, more conservative, sample size.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Equivalence Tests for Paired Means (Simulation) procedure window by expanding Means, then clicking on Paired Means, and then clicking on Equivalence Tests for Paired Means (Simulation). You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N</td>
</tr>
<tr>
<td>Power</td>
<td>0.9</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Item A (Reference) Dist'n</td>
<td>H0</td>
</tr>
<tr>
<td>Item B (Treatment) Dist'n</td>
<td>H0</td>
</tr>
<tr>
<td>Item A (Reference) Dist'n</td>
<td>H1</td>
</tr>
<tr>
<td>Item B (Treatment) Dist'n</td>
<td>H1</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H1)</td>
</tr>
<tr>
<td>S</td>
<td>5</td>
</tr>
<tr>
<td>R (Correlation of Items A &amp; B)</td>
<td>0.1</td>
</tr>
<tr>
<td>Equivalence Limit</td>
<td>Symmetric</td>
</tr>
<tr>
<td>Test Type</td>
<td>T-Test</td>
</tr>
<tr>
<td>Simulations</td>
<td>2000</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>H1 Diff1</th>
<th>Lower Equiv. Limit</th>
<th>Upper Equiv. Limit</th>
<th>Corr R</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.899</td>
<td>54</td>
<td>0.0</td>
<td>-3.0</td>
<td>3.0</td>
<td>0.100</td>
<td>0.050</td>
<td>(0.009)</td>
<td>63.0</td>
<td>66.0</td>
<td>5.0</td>
</tr>
</tbody>
</table>

The required sample size was 54 which achieved a power of 0.899.
The power of 0.899 is slightly less than the target value of 0.900 because the sample size search algorithm re-simulates the power for the final sample size. Thus it is possible for the search algorithm to converge to a sample size which exhibits the desired power, but then on the second simulation, achieves a power that is slightly less than the target. To obtain more accuracy, a reasonable strategy would be to run simulations to obtain the powers using N’s from 50 to 60 using a simulation size of 5000.

**Example 3 – Comparing Test Statistics**

Continuing with Example 2, the researchers want to study the characteristics of alternative test statistics.

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the *Equivalence Tests for Paired Means (Simulation)* procedure window by expanding Means, then clicking on Paired Means, and then clicking on *Equivalence Tests for Paired Means (Simulation)*. You may then make the appropriate entries as listed below, or open Example 3 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Tab</strong></td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>10 30 50 70</td>
</tr>
<tr>
<td>Item A (Reference) Dist'n</td>
<td>H0</td>
</tr>
<tr>
<td>Item B (Treatment) Dist'n</td>
<td>H0</td>
</tr>
<tr>
<td>Item A (Reference) Dist'n</td>
<td>H1</td>
</tr>
<tr>
<td>Item B (Treatment) Dist'n</td>
<td>H1</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H1)</td>
</tr>
<tr>
<td>S</td>
<td>5</td>
</tr>
<tr>
<td>R (Correlation of Items A &amp; B)</td>
<td>0.1</td>
</tr>
<tr>
<td>Equivalence Limit</td>
<td>Symmetric</td>
</tr>
<tr>
<td>Test Type</td>
<td>T-Test</td>
</tr>
<tr>
<td>Simulations</td>
<td>2000</td>
</tr>
</tbody>
</table>

**Reports Tab**

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Show Comparative Reports</td>
<td>Checked</td>
</tr>
<tr>
<td>Show Comparative Plots</td>
<td>Checked</td>
</tr>
</tbody>
</table>
Output
Click the Run button to perform the calculations and generate the following output.

Numeric Results

Power Comparison for Testing Equivalence. Hypotheses: H0: Diff>=|Diff0|; H1: Diff<|Diff0|
H0 Dist’n: Normal(M0 S) - Normal(M1 S)
H1 Dist’n: Normal(M0 S) - Normal(M0 S)

<table>
<thead>
<tr>
<th>N</th>
<th>H1 Diff (Diff1)</th>
<th>Lower Equiv. Limit</th>
<th>Upper Equiv. Limit</th>
<th>Corr (R)</th>
<th>Target Alpha</th>
<th>T-Test Power</th>
<th>Wilcoxon Power</th>
<th>Sign Power</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.0</td>
<td>-3.0</td>
<td>3.0</td>
<td>0.100</td>
<td>0.050</td>
<td>0.034</td>
<td>0.024</td>
<td>0.002</td>
<td>63.0</td>
<td>66.0</td>
<td>5.0</td>
</tr>
<tr>
<td>30</td>
<td>0.0</td>
<td>-3.0</td>
<td>3.0</td>
<td>0.100</td>
<td>0.050</td>
<td>0.537</td>
<td>0.494</td>
<td>0.275</td>
<td>63.0</td>
<td>66.0</td>
<td>5.0</td>
</tr>
<tr>
<td>50</td>
<td>0.0</td>
<td>-3.0</td>
<td>3.0</td>
<td>0.100</td>
<td>0.050</td>
<td>0.870</td>
<td>0.855</td>
<td>0.500</td>
<td>63.0</td>
<td>66.0</td>
<td>5.0</td>
</tr>
<tr>
<td>70</td>
<td>0.0</td>
<td>-3.0</td>
<td>3.0</td>
<td>0.100</td>
<td>0.050</td>
<td>0.966</td>
<td>0.953</td>
<td>0.768</td>
<td>63.0</td>
<td>66.0</td>
<td>5.0</td>
</tr>
</tbody>
</table>

Alpha Comparison for Testing Equivalence. Hypotheses: H0: Diff>=|Diff0|; H1: Diff<|Diff0|
H0 Dist’n: Normal(M0 S) - Normal(M1 S)
H1 Dist’n: Normal(M0 S) - Normal(M0 S)

<table>
<thead>
<tr>
<th>N</th>
<th>H1 Diff (Diff1)</th>
<th>Lower Equiv. Limit</th>
<th>Upper Equiv. Limit</th>
<th>Corr (R)</th>
<th>Target Alpha</th>
<th>T-Test Alpha</th>
<th>Wilcoxon Alpha</th>
<th>Sign Alpha</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>0.0</td>
<td>-3.0</td>
<td>3.0</td>
<td>0.100</td>
<td>0.050</td>
<td>0.010</td>
<td>0.009</td>
<td>0.001</td>
<td>63.0</td>
<td>66.0</td>
<td>5.0</td>
</tr>
<tr>
<td>30</td>
<td>0.0</td>
<td>-3.0</td>
<td>3.0</td>
<td>0.100</td>
<td>0.050</td>
<td>0.057</td>
<td>0.056</td>
<td>0.052</td>
<td>63.0</td>
<td>66.0</td>
<td>5.0</td>
</tr>
<tr>
<td>50</td>
<td>0.0</td>
<td>-3.0</td>
<td>3.0</td>
<td>0.100</td>
<td>0.050</td>
<td>0.052</td>
<td>0.049</td>
<td>0.024</td>
<td>63.0</td>
<td>66.0</td>
<td>5.0</td>
</tr>
<tr>
<td>70</td>
<td>0.0</td>
<td>-3.0</td>
<td>3.0</td>
<td>0.100</td>
<td>0.050</td>
<td>0.044</td>
<td>0.045</td>
<td>0.041</td>
<td>63.0</td>
<td>66.0</td>
<td>5.0</td>
</tr>
</tbody>
</table>

These results show that for paired data, the t-test and Wilcoxon test have very similar power and alpha values. The sign test is less accurate and less powerful.
Example 4 – Validation using Chow et al.

We will validate this procedure by comparing its results to those of Chow et al. (2003) page 55 in which the parameter values are: M0 = 0, M1 = 0.05, alpha = 0.05, N = 35, R = 0.0, and S = 0.070711. For these parameters, the power is given as 0.800.

Note that they give the standard deviation of the differences as 0.1. Since the correlation is 0.0, the standard deviation of the individual data values is given by 0.1/Sqrt(2) = 0.070711.

In order to understand the accuracy of the simulation, we will re-run the analysis five times.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Equivalence Tests for Paired Means (Simulation) procedure window by expanding Means, then clicking on Paired Means, and then clicking on Equivalence Tests for Paired Means (Simulation). You may then make the appropriate entries as listed below, or open Example 4 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For) ...........</td>
<td>Power</td>
</tr>
<tr>
<td>Power ........................</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha .......................</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Sample Size) .............</td>
<td>35 35 35 35 35</td>
</tr>
<tr>
<td>Item A (Reference) Dist’n</td>
<td>H0  ..............</td>
</tr>
<tr>
<td>Item B (Treatment) Dist’n</td>
<td>H0  ..............</td>
</tr>
<tr>
<td>Item A (Reference) Dist’n</td>
<td>H1  ..............</td>
</tr>
<tr>
<td>Item B (Treatment) Dist’n</td>
<td>H1  ..............</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0) ......................</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H1) ......................</td>
</tr>
<tr>
<td>S ................................</td>
<td>0.070711</td>
</tr>
<tr>
<td>R (Correlation of Items A &amp; B) ........</td>
<td>0.0</td>
</tr>
<tr>
<td>Equivalence Limit ...........</td>
<td>Symmetric</td>
</tr>
<tr>
<td>Test Type ...................</td>
<td>T-Test</td>
</tr>
<tr>
<td>Simulations ..................</td>
<td>2000</td>
</tr>
</tbody>
</table>
Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>H1 Diff</th>
<th>Lower Equiv. Limit</th>
<th>Upper Equiv. Limit</th>
<th>Corr R</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.813</td>
<td>35</td>
<td>0.0</td>
<td>-0.1</td>
<td>0.1</td>
<td>0.000</td>
<td>0.050</td>
<td>0.050</td>
<td>0.0</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>(0.017)</td>
<td>[0.795</td>
<td>0.830]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.813</td>
<td>35</td>
<td>0.0</td>
<td>-0.1</td>
<td>0.1</td>
<td>0.000</td>
<td>0.050</td>
<td>0.045</td>
<td>0.0</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>(0.017)</td>
<td>[0.796</td>
<td>0.830]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.803</td>
<td>35</td>
<td>0.0</td>
<td>-0.1</td>
<td>0.1</td>
<td>0.000</td>
<td>0.050</td>
<td>0.051</td>
<td>0.0</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>(0.017)</td>
<td>[0.785</td>
<td>0.820]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.799</td>
<td>35</td>
<td>0.0</td>
<td>-0.1</td>
<td>0.1</td>
<td>0.000</td>
<td>0.050</td>
<td>0.045</td>
<td>0.0</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>(0.018)</td>
<td>[0.781</td>
<td>0.816]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.826</td>
<td>35</td>
<td>0.0</td>
<td>-0.1</td>
<td>0.1</td>
<td>0.000</td>
<td>0.050</td>
<td>0.051</td>
<td>0.0</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>(0.017)</td>
<td>[0.809</td>
<td>0.843]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:

The powers match the analytic value of 0.800 quite well. Note how informative the confidence intervals are.
Chapter 496

Confidence Intervals for Paired Means

Introduction

This routine calculates the sample size necessary to achieve a specified distance from the paired sample mean difference to the confidence limit(s) at a stated confidence level for a confidence interval about the mean difference when the underlying data distribution is normal.

Caution: This procedure assumes that the standard deviation of the future sample will be the same as the standard deviation that is specified. If the standard deviation to be used in the procedure is estimated from a previous paired sample or represents the population standard deviation, the Confidence Intervals for Paired Means with Tolerance Probability procedure should be considered. That procedure controls the probability that the distance from the mean paired difference to the confidence limits will be less than or equal to the value specified.

Technical Details

For a paired sample mean difference from a normal distribution with known variance, a two-sided, 100(1 – $\alpha$)% confidence interval is calculated by

$$\bar{X}_{\text{Diff}} \pm \frac{z_{1-\alpha/2} \sigma_{\text{diff}}}{\sqrt{n}}$$

where $\bar{X}_{\text{Diff}}$ is the mean of the paired differences of the sample, and $\sigma_{\text{diff}}$ is the known standard deviation of paired sample differences.

A one-sided 100(1 – $\alpha$)% upper confidence limit is calculated by

$$\bar{X}_{\text{Diff}} + \frac{z_{1-\alpha} \sigma_{\text{diff}}}{\sqrt{n}}$$
Similarly, the one-sided 100(1 – \(\alpha\))% lower confidence limit is

\[
\bar{X}_{\text{Diff}} - \frac{z_{1-\alpha} \hat{\sigma}_{\text{Diff}}}{\sqrt{n}}
\]

For a paired sample mean difference from a normal distribution with unknown variance, a two-sided, 100(1 – \(\alpha\))% confidence interval is calculated by

\[
\bar{X}_{\text{Diff}} \pm \frac{t_{1-\alpha/2, n-1} \hat{\sigma}_{\text{Diff}}}{\sqrt{n}}
\]

where \(\bar{X}_{\text{Diff}}\) is the mean of the paired differences of the sample, and \(\hat{\sigma}_{\text{Diff}}\) is the estimated standard deviation of paired sample differences.

A one-sided 100(1 – \(\alpha\))% upper confidence limit is calculated by

\[
\bar{X}_{\text{Diff}} + \frac{t_{1-\alpha, n-1} \hat{\sigma}_{\text{Diff}}}{\sqrt{n}}
\]

Similarly, the one-sided 100(1 – \(\alpha\))% lower confidence limit is

\[
\bar{X}_{\text{Diff}} - \frac{t_{1-\alpha, n-1} \hat{\sigma}_{\text{Diff}}}{\sqrt{n}}
\]

Each confidence interval is calculated using an estimate of the mean difference plus and/or minus a quantity that represents the distance from the mean difference to the edge of the interval. For two-sided confidence intervals, this distance is sometimes called the precision, margin of error, or half-width. We will label this distance, \(D\).

The basic equation for determining sample size when \(D\) has been specified is

\[
D = \frac{z_{1-\alpha/2} \sigma_{\text{diff}}}{\sqrt{n}}
\]

when the standard deviation is known, and

\[
D = \frac{t_{1-\alpha/2, n-1} \hat{\sigma}_{\text{Diff}}}{\sqrt{n}}
\]

when the standard deviation is unknown. These equations can be solved for any of the unknown quantities in terms of the others. The value \(\alpha / 2\) is replaced by \(\alpha\) when a one-sided interval is used.

### Finite Population Size

The above calculations assume that samples are being drawn from a large (infinite) population. When the population is of finite size \(N\), an adjustment must be made. The adjustment reduces the standard deviation as follows:

\[
\sigma_{\text{finite}} = \sigma \sqrt{\left(1 - \frac{n}{N}\right)}
\]

This new standard deviation replaces the regular standard deviation in the above formulas.
Confidence Intervals for Paired Means

Confidence Level

The confidence level, $1 - \alpha$, has the following interpretation. If thousands of samples of $n$ items are drawn from a population using simple random sampling and a confidence interval is calculated for each sample, the proportion of those intervals that will include the true population mean difference is $1 - \alpha$.

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains most of the parameters and options that you will be concerned with.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for from the other parameters.

Confidence

Confidence Level

The confidence level, $1 - \alpha$, has the following interpretation. If thousands of samples of $n$ items are drawn from a population using simple random sampling and a confidence interval is calculated for each sample, the proportion of those intervals that will include the true population mean difference is $1 - \alpha$.

Often, the values 0.95 or 0.99 are used. You can enter single values or a range of values such as 0.90, 0.95 or 0.90 to 0.99 by 0.01.

Sample Size (Number of Pairs)

N (Sample Size)

Enter one or more values for the sample size. This is the number of pairs selected at random from the population to be in the study.

You can enter a single value or a range of values.

Precision

Distance from Mean Difference to Limit(s)

This is the distance from the confidence limit(s) to the mean paired difference. For two-sided intervals, it is also known as the precision, half-width, or margin of error.

You can enter a single value or a list of values. The value(s) must be greater than zero.
Standard Deviation of Paired Differences

S (Standard Deviation)
Enter a value (or range of values) for the standard deviation. You can use the results of a pilot study, a previous study, or a ball park estimate based on the range (e.g., Range/4) to estimate this parameter.

Know Standard Deviation
Check this box when you want to base your results on the normal distribution. When the box is not checked, calculations are based on the t-distribution. The difference between the two distributions is negligible when the sample sizes are large (>50).

One-Sided or Two-Sided Interval

Interval Type
Specify whether the interval to be used will be a one-sided or a two-sided confidence interval.

Population

Population Size
This is the number of pairs in the population. Usually, you assume that samples are drawn from a very large (infinite) population. Occasionally, however, situations arise in which the population of interest is of limited size. In these cases, appropriate adjustments must be made. This option sets the population size.

Iterations Tab
This tab sets an option used in the iterative procedures.

Maximum Iterations

Maximum Iterations Before Search Termination
Specify the maximum number of iterations allowed before the search for the criterion of interest is aborted. When the maximum number of iterations is reached without convergence, the criterion is left blank. A value of 500 is recommended.
Example 1 – Calculating Sample Size

A researcher would like to estimate the mean difference in weight following a specific diet using a two-sided 95% confidence interval. The confidence level is set at 0.95, but 0.99 is included for comparative purposes. The standard deviation estimate, based on the range of paired differences, is 9.6 lbs. The researcher would like the interval to be no wider than 10 lbs. (half-width = 5 lbs.), but will examine half-widths of 3, 4, 5, 6, and 7 lbs.

The goal is to determine the necessary sample size.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Confidence Intervals for Paired Means procedure window by expanding Means, then clicking on Paired Means, and then clicking on Confidence Intervals for Paired Means. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N (Sample Size)</td>
</tr>
<tr>
<td>Confidence Level</td>
<td>0.95 0.99</td>
</tr>
<tr>
<td>N (Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Distance from Mean to Limit(s)</td>
<td>3 to 7 by 1</td>
</tr>
<tr>
<td>S (Standard Deviation)</td>
<td>9.6</td>
</tr>
<tr>
<td>Interval Type</td>
<td>Two-Sided</td>
</tr>
<tr>
<td>Population Size</td>
<td>Infinite</td>
</tr>
</tbody>
</table>

Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>Sample Size (N)</th>
<th>Target Dist from Mean Diff to Limits</th>
<th>Actual Dist from Mean Diff to Limits</th>
<th>Standard Deviation (S)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.95000</td>
<td>42</td>
<td>3.000</td>
<td>2.992</td>
<td>9.600</td>
</tr>
<tr>
<td>0.99000</td>
<td>72</td>
<td>3.000</td>
<td>2.995</td>
<td>9.600</td>
</tr>
<tr>
<td>0.95000</td>
<td>25</td>
<td>4.000</td>
<td>3.963</td>
<td>9.600</td>
</tr>
<tr>
<td>0.99000</td>
<td>43</td>
<td>4.000</td>
<td>3.950</td>
<td>9.600</td>
</tr>
<tr>
<td>0.95000</td>
<td>17</td>
<td>5.000</td>
<td>4.936</td>
<td>9.600</td>
</tr>
<tr>
<td>0.99000</td>
<td>29</td>
<td>5.000</td>
<td>4.926</td>
<td>9.600</td>
</tr>
<tr>
<td>0.95000</td>
<td>13</td>
<td>6.000</td>
<td>5.801</td>
<td>9.600</td>
</tr>
<tr>
<td>0.99000</td>
<td>21</td>
<td>6.000</td>
<td>5.961</td>
<td>9.600</td>
</tr>
<tr>
<td>0.95000</td>
<td>10</td>
<td>7.000</td>
<td>6.867</td>
<td>9.600</td>
</tr>
<tr>
<td>0.99000</td>
<td>17</td>
<td>7.000</td>
<td>6.801</td>
<td>9.600</td>
</tr>
</tbody>
</table>

References

Report Definitions
Confidence level is the proportion of confidence intervals (constructed with this same confidence level, sample size, etc.) that would contain the population mean difference.
N is the size of the sample (or number of pairs) drawn from the population.
Dist from Mean Diff to Limit is the distance from the confidence limit(s) to the mean paired difference. For two-sided intervals, it is also known as the precision, half-width, or margin of error.
Target Dist from Mean Diff to Limit is the value of the distance that is entered into the procedure.
Actual Dist from Mean Diff to Limit is the value of the distance that is obtained from the procedure.
The standard deviation (S) is the standard deviation of the paired differences.

Summary Statements
A sample size of 42 produces a two-sided 95% confidence interval with a distance from the mean paired difference to the limits that is equal to 2.992 when the estimated standard deviation of the paired differences is 9.600.

This report shows the calculated sample size for each of the scenarios.

Plots Section

This plot shows the sample size versus the precision for the two confidence limits.

Example 2 – Validation
This procedure uses the same mechanics as the Confidence Intervals for One Mean procedure. The validation of this procedure is given in Examples 2 and 3 of the Confidence Intervals for One Mean procedure.
Chapter 497

Confidence Intervals for Paired Means with Tolerance Probability

Introduction

This routine calculates the sample size necessary to achieve a specified distance from the paired sample mean difference to the confidence limit(s) with a given tolerance probability at a stated confidence level for a confidence interval about a single mean difference when the underlying data distribution is normal.

Technical Details

For a paired sample mean difference from a normal distribution with unknown variance, a two-sided, 100(1 – α)% confidence interval is calculated by

\[ \bar{X}_{\text{Diff}} \pm t_{1-\alpha/2,\nu-1} \hat{\sigma}_{\text{Diff}} / \sqrt{n} \]

where \( \bar{X}_{\text{Diff}} \) is the mean of the paired differences of the sample, and \( \sigma_{\text{Diff}} \) is the known standard deviation of paired sample differences.

A one-sided 100(1 – α)% upper confidence limit is calculated by

\[ \bar{X}_{\text{Diff}} + t_{1-\alpha,\nu-1} \hat{\sigma}_{\text{Diff}} / \sqrt{n} \]

Similarly, the one-sided 100(1 – α)% lower confidence limit is

\[ \bar{X}_{\text{Diff}} - t_{1-\alpha,\nu-1} \hat{\sigma}_{\text{Diff}} / \sqrt{n} \]

Each confidence interval is calculated using an estimate of the mean difference plus and/or minus a quantity that represents the distance from the mean difference to the edge of the interval. For
two-sided confidence intervals, this distance is sometimes called the precision, margin of error, or half-width. We will label this distance, $D$.

The basic equation for determining sample size when $D$ has been specified is

$$D = \frac{t_{\alpha/2, n-1} \hat{\sigma}_{\text{diff}}}{\sqrt{n}}$$

Solving for $n$, we obtain

$$n = \left(\frac{t_{\alpha/2, n-1} \hat{\sigma}_{\text{diff}}}{D}\right)^2$$

This equation can be solved for any of the unknown quantities in terms of the others. The value $\alpha/2$ is replaced by $\alpha$ when a one-sided interval is used.

There is an additional subtlety that arises when the standard deviation is to be chosen for estimating sample size. The sample sizes determined from the formula above produce confidence intervals with the specified widths only when the future sample has a sample standard deviation of differences that is no greater than the value specified.

As an example, suppose that 15 pairs of individuals are sampled in a pilot study, and a standard deviation estimate of 3.5 is obtained from the sample. The purpose of a later study is to estimate the mean difference within 10 units. Suppose further that the sample size needed is calculated to be 57 pairs using the formula above with 3.5 as the estimate for the standard deviation. The sample of size 57 pairs is then obtained from the population, but the standard deviation of the 57 paired differences turns out to be 3.9 rather than 3.5. The confidence interval is computed and the distance from the mean difference to the confidence limits is greater than 10 units.

This example illustrates the need for an adjustment to adjust the sample size such that the distance from the mean difference to the confidence limits will be below the specified value with known probability.

Such an adjustment for situations where a previous sample is used to estimate the standard deviation is derived by Harris, Horvitz, and Mood (1948) and discussed in Zar (1984) and Hahn and Meeker (1991). The adjustment is

$$n = \left(\frac{t_{\alpha/2, m-1} \hat{\sigma}_{\text{diff}}}{D}\right)^2 F_{1-\gamma,m-1}$$

where $1 - \gamma$ is the probability that the distance from the mean difference to the confidence limit(s) will be below the specified value, and $m$ is the sample size in the previous paired sample that was used to estimate the standard deviation.

The corresponding adjustment when no previous sample is available is discussed in Kupper and Hafner (1989) and Hahn and Meeker (1991). The adjustment in this case is

$$n = \left(\frac{t_{\alpha/2, n-1} \hat{\sigma}_{\text{diff}}}{D}\right)^2 \left(\frac{\chi^2_{1-\gamma,n-1}}{n-1}\right)$$

where, again, $1 - \gamma$ is the probability that the distance from the mean difference to the confidence limit(s) will be below the specified value.
Each of these adjustments accounts for the variability in a future estimate of the standard deviation. In the first adjustment formula (Harris, Horvitz, and Mood, 1948), the distribution of the standard deviation is based on the estimate from a previous paired sample. In the second adjustment formula, the distribution of the standard deviation is based on a specified value that is assumed to be the population standard deviation of differences.

Finite Population Size
The above calculations assume that samples are being drawn from a large (infinite) population. When the population is of finite size \( N \), an adjustment must be made. The adjustment reduces the standard deviation as follows:

\[
\sigma_{\text{finite}} = \sigma \sqrt{1 - \frac{n}{N}}
\]

This new standard deviation replaces the regular standard deviation in the above formulas.

Confidence Level
The confidence level, \( 1 - \alpha \), has the following interpretation. If thousands of samples of \( n \) items are drawn from a population using simple random sampling and a confidence interval is calculated for each sample, the proportion of those intervals that will include the true population mean difference is \( 1 - \alpha \).

Procedure Options
This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab
The Data tab contains most of the parameters and options that you will be concerned with.

Solve For
Find (Solve For)
This option specifies the parameter to be solved for from the other parameters.

Confidence and Tolerance
Confidence Level (1 – Alpha)
The confidence level, \( 1 - \alpha \), has the following interpretation. If thousands of samples of \( n \) items are drawn from a population using simple random sampling and a confidence interval is calculated for each sample, the proportion of those intervals that will include the true population mean difference is \( 1 - \alpha \).
Often, the values 0.95 or 0.99 are used. You can enter single values or a range of values such as 0.90 0.95 0.99 or 0.90 to 0.99 by 0.01.

**Tolerance Probability**

This is the probability that a future interval with sample size \( N \) and the specified confidence level will have a distance from the mean paired difference to the limit(s) that is less than or equal to the distance specified.

If a tolerance probability is not used, as in the 'Confidence Intervals for Paired Means' procedure, the sample size is calculated for the expected distance from the mean paired difference to the limit(s), which assumes that the future standard deviation will also be the one specified.

Using a tolerance probability implies that the standard deviation of the future sample will not be known in advance, and therefore, an adjustment is made to the sample size formula to account for the variability in the standard deviation. Use of a tolerance probability is similar to using an upper bound for the standard deviation in the 'Confidence Intervals for Paired Means' procedure.

Values between 0 and 1 can be entered. The choice of the tolerance probability depends upon how important it is that the distance from the interval limit(s) to the mean difference is at most the value specified.

You can enter a range of values such as 0.70 0.80 0.90 or 0.70 to 0.95 by 0.05.

**Sample Size (Number of Pairs)**

**N (Sample Size or Number of Pairs)**

Enter one or more values for the sample size. This is the number of pairs selected at random from the population to be in the study.

You can enter a single value or a range of values.

**Precision**

**Distance from Mean Difference to Limit(s)**

This is the distance from the confidence limit(s) to the mean paired difference. For two-sided intervals, it is also known as the precision, half-width, or margin of error.

You can enter a single value or a list of values. The value(s) must be greater than zero.
Standard Deviation of Paired Differences

Standard Deviation Source
This procedure permits two sources for estimates of the standard deviation of paired differences:

- **S is a Population Standard Deviation**
  This option should be selected if there is no previous sample that can be used to obtain an estimate of the standard deviation of the paired differences. In this case, the algorithm assumes that future samples obtained will be from a population with standard deviation S.

- **S from a Previous Sample**
  This option should be selected if the estimate of the standard deviation of the paired differences is obtained from a previous random sample from the same distribution as the one to be sampled. The sample size of the previous sample must also be entered under 'Sample Size of Previous Sample'.

Standard Deviation of Paired Differences—S is a Population Standard Deviation

S (Standard Deviation)
Enter an estimate of the standard deviation of paired differences (must be positive). In this case, the algorithm assumes that future samples obtained will be from a population with standard deviation S.

One common method for estimating the standard deviation is the range divided by 4, 5, or 6. You can enter a range of values such as 1 2 3 or 1 to 10 by 1.

Press the Standard Deviation Estimator button to load the Standard Deviation Estimator window.

Standard Deviation of Paired Differences—S from a Previous Sample

S (SD Estimated from a Previous Sample)
Enter an estimate of the standard deviation of paired differences from a previous (or pilot) study. This value must be positive.

A range of values may be entered.

Press the Standard Deviation Estimator button to load the Standard Deviation Estimator window.

Sample Size (# of Pairs) of Previous Sample
Enter the sample size (number of pairs) that was used to estimate the standard deviation entered in S (SD Estimated from a Previous Sample).

This value is entered only when 'Standard Deviation Source:' is set to 'S from a Previous Sample'.
One-Sided or Two-Sided Interval

Interval Type
Specify whether the interval to be used will be a one-sided or a two-sided confidence interval.

Population

Population Size
This is the number of pairs in the population. Usually, you assume that samples are drawn from a very large (infinite) population. Occasionally, however, situations arise in which the population of interest is of limited size. In these cases, appropriate adjustments must be made. This option sets the population size.

Iterations Tab
This tab sets an option used in the iterative procedures.

Maximum Iterations

Maximum Iterations Before Search Termination
Specify the maximum number of iterations allowed before the search for the criterion of interest is aborted. When the maximum number of iterations is reached without convergence, the criterion is left blank. A value of 500 is recommended.

Example 1 – Calculating Sample Size

A researcher would like to estimate the mean difference in weight following a specific diet with 95% confidence. It is very important that the mean difference is estimated within 5 lbs. Data available from a previous study are used to provide an estimate of the standard deviation. The estimate of the standard deviation of before/after differences is 16.7 lbs, from a sample of size 17 individuals.

The goal is to determine the sample size necessary to obtain a two-sided confidence interval such that the mean weight is estimated within 5 lbs. Tolerance probabilities of 0.70 to 0.95 will be examined.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Confidence Intervals for Paired Means with Tolerance Probability procedure window by expanding Means, then clicking on Paired Means, and then clicking on Confidence Intervals for Paired Means with Tolerance Probability. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.
Confidence Intervals for Paired Means with Tolerance Probability

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N (Sample Size)</td>
</tr>
<tr>
<td>Confidence Level</td>
<td>0.95</td>
</tr>
<tr>
<td>Tolerance Probability</td>
<td>0.70 to 0.95 by 0.05</td>
</tr>
<tr>
<td>N (Sample Size or Number of Pairs)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Distance from Mean Diff to Limit(s)</td>
<td>5</td>
</tr>
<tr>
<td>Standard Deviation Source</td>
<td>S from a Previous Sample</td>
</tr>
<tr>
<td>S</td>
<td>16.7</td>
</tr>
<tr>
<td>Sample Size of Previous Sample</td>
<td>17</td>
</tr>
<tr>
<td>Interval Type</td>
<td>Two-Sided</td>
</tr>
<tr>
<td>Population Size</td>
<td>Infinite</td>
</tr>
</tbody>
</table>

Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Confidence Level</th>
<th>Sample Size</th>
<th>Target Dist from Mean Diff to Limits</th>
<th>Actual Dist from Mean Diff to Limits</th>
<th>Standard Deviation (S)</th>
<th>Tolerance Probability</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.95</td>
<td>58</td>
<td>5.000</td>
<td>4.970</td>
<td>16.700</td>
<td>0.70</td>
</tr>
<tr>
<td>0.95</td>
<td>61</td>
<td>5.000</td>
<td>4.996</td>
<td>16.700</td>
<td>0.75</td>
</tr>
<tr>
<td>0.95</td>
<td>66</td>
<td>5.000</td>
<td>4.967</td>
<td>16.700</td>
<td>0.80</td>
</tr>
<tr>
<td>0.95</td>
<td>71</td>
<td>5.000</td>
<td>4.985</td>
<td>16.700</td>
<td>0.85</td>
</tr>
<tr>
<td>0.95</td>
<td>79</td>
<td>5.000</td>
<td>4.973</td>
<td>16.700</td>
<td>0.90</td>
</tr>
<tr>
<td>0.95</td>
<td>92</td>
<td>5.000</td>
<td>4.981</td>
<td>16.700</td>
<td>0.95</td>
</tr>
</tbody>
</table>

Sample size for estimate of S from previous paired sample = 17.

References


Report Definitions

Confidence level is the proportion of confidence intervals (constructed with this same confidence level, sample size, etc.) that would contain the population mean difference.
N is the size of the sample (or number of pairs) drawn from the population.
Dist from Mean Diff to Limit is the distance from the confidence limit(s) to the mean paired difference. For two-sided intervals, it is also know as the precision, half-width, or margin of error.
Target Dist from Mean Diff to Limit is the value of the distance that is entered into the procedure.
Actual Dist from Mean Diff to Limit is the value of the distance that is obtained from the procedure.
The standard deviation (S) is the standard deviation of the paired differences.
Tolerance Probability is the probability that a future interval with sample size N and corresponding confidence level will have a distance from the mean difference to the limit(s) that is less than or equal to the specified distance.

Summary Statements

The probability is 0.70 that a sample size of 58 will produce a two-sided 95% confidence interval with a distance from the mean paired difference to the limits that is less than or equal to 4.970 if the population standard deviation is estimated to be 16.700 by a previous paired sample of size 17.
This report shows the calculated sample size for each of the scenarios.

**Plots Section**

This plot shows the sample size versus the tolerance probability.

**Example 2 – Validation**

This procedure uses the same mechanics as the Confidence Intervals for One Mean with Tolerance Probability procedure. The validation of this procedure is given in Examples 2, 3, and 4 of the Confidence Intervals for One Mean with Tolerance Probability procedure.
Chapter 500

Inequality Tests for Two Means in a 2x2 Cross-Over Design using Differences

Introduction

Senn (2002) defines a cross-over design as one in which each subject receives all treatments and the objective is to study differences among the treatments. The name cross-over comes from the most common case in which there are only two treatments. In this case, each subject crosses over from one treatment to the other. It is assumed that there is a washout period between treatments during which the response returns back to its baseline value. If this does not occur, there is said to be a carry-over effect.

A 2x2 cross-over design contains two sequences (treatment orderings) and two time periods (occasions). One sequence receives treatment A followed by treatment B. The other sequence receives B and then A. The design includes a washout period between responses to make certain that the effects of the first drug do no carry over to the second. Thus, the groups in this design are defined by the sequence in which the drugs are administered, not by the treatments they receive. Indeed, higher-order cross-over designs have been used in which the same treatment is used at both occasions.

Cross-over designs are employed because, if the no-carryover assumption is met, treatment differences are measured within a subject rather than between subjects—making a more precise measurement. Examples of the situations that might use a cross-over design are the comparison of anti-inflammatory drugs in arthritis and the comparison of hypotensive agents in essential hypertension. In both of these cases, symptoms are expected to return to their usual baseline level shortly after the treatment is stopped.
Advantages of Cross-Over Designs
A comparison of treatments on the same subject is expected to be more precise. The increased precision often translates into a smaller sample size. Also, patient enrollment into the study may be easier because each patient will receive both treatments. Finally, it is often more difficult to obtain a subject than to obtain a measurement.

Disadvantages of Cross-Over Designs
The statistical analysis of a cross-over experiment is more complex than a parallel-group experiment and requires additional assumptions. It may be difficult to separate the treatment effect from the period effect, the carry-over effect of the previous treatment, and the interaction between period and treatment.

The design cannot be used when the treatment (or the measurement of the response) alters the subject permanently. Hence, it should not be used to compare treatments that are intended to provide a cure.

Because subjects must be measured at least twice, it is often more difficult to keep patients enrolled in the study. It is arguably simpler to measure a subject once than to obtain their measurement twice. This is particularly true when the measurement process is painful, uncomfortable, embarrassing, or time consuming.

Technical Details
The 2x2 crossover design may be described as follows. Randomly assign the subjects to one of two sequence groups so that there are $N_1$ subjects in sequence one and $N_2$ subjects in sequence two. In order to achieve design balance, the sample sizes $N_1$ and $N_2$ are assumed to be equal so that $N_1 = N_2 = N / 2$.

Sequence one is given treatment A followed by treatment B. Sequence two is given treatment B followed by treatment A. The sequence is replicated $m$ times. So, if $m = 3$, the sequences are ABABAB and BABABA.

The usual method of analysis is the analysis of variance. However, the power and sample size formulas that follow are based on the t-test, not the F-test. This is done because, in the balanced case, the t-test and the analysis of variance F-test are equivalent. Also, the F-test is limited to a two-sided hypothesis, while the t-test allows both one-sided and two-sided hypotheses. This is important because one-sided hypotheses are used for non-inferiority and equivalence testing.

Cross-Over Analysis
The following discussion summarizes the presentation of Chow and Liu (1999). The general linear model for the standard 2x2 cross-over design is

$$Y_{ijkl} = \mu + S_{ik} + P_j + \mu_{(j,k)} + C_{(j-1,k)} + e_{ijkl}$$

where $i$ represents a subject (1 to $N_k$), $j$ represents the period (1 or 2), $k$ represents the sequence (1 or 2), and $l$ represents the replicate. The $S_{ik}$ represent the random effects of the subjects. The
Inequality Tests for Two Means in a 2x2 Cross-Over Design using Differences

$P_j$ represent the effects of the two periods. The $\mu_{(j,k)}$ represent the means of the two treatments. In the case of the 2x2 cross-over design

$$\mu_{(j,k)} = \begin{cases} 
\mu_1 & \text{if } k = j \\
\mu_2 & \text{if } k \neq j
\end{cases}$$

where the subscripts 1 and 2 represent treatments A and B, respectively.

The $C_{(j-1,k)}$ represent the carry-over effects. In the case of the 2x2 cross-over design

$$C_{(j-1,k)} = \begin{cases} 
C_1 & \text{if } j = 2, k = 1 \\
C_2 & \text{if } j = 2, k = 2 \\
0 & \text{otherwise}
\end{cases}$$

where the subscripts 1 and 2 represent treatments A and B, respectively.

Assuming that the average effect of the subjects is zero, the four means from the 2x2 cross-over design can be summarized using the following table.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Period 1</th>
<th>Period 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1 (AB)</td>
<td>$\mu_{11} = \mu + P_1 + \mu_1$</td>
<td>$\mu_{12} = \mu + P_1 + \mu_2 + C_1$</td>
</tr>
<tr>
<td>2 (BA)</td>
<td>$\mu_{21} = \mu + P_2 + \mu_1$</td>
<td>$\mu_{22} = \mu + P_2 + \mu_2 + C_2$</td>
</tr>
</tbody>
</table>

where $P_1 + P_2 = 0$ and $C_1 + C_2 = 0$.

**Test Statistic**

The presence of a treatment effect can be studied by testing whether $\mu_1 - \mu_2 = \delta$ using a $t$-test or an $F$-test. If the $F$-test is used, only a two-sided test is possible. The $t$ statistic is calculated as follows

$$t_d = \frac{(\bar{x}_T - \bar{x}_R) - \delta}{\hat{\sigma}_w \sqrt{\frac{2}{N}}}$$

where $\hat{\sigma}_w^2$ is the within mean square error from the appropriate ANOVA table.

The two-sided null hypothesis is rejected at the $\alpha$ significance level if $|t_d| > t_{\alpha/2,N-2}$. Similar results are available for a one-sided hypothesis test.

The $F$-test is calculated using a standard repeated-measures analysis of variance table in which the between factor is the sequence and the within factor is the treatment. The within mean square error provides an estimate of the within-subject variance $\sigma_w^2$. If prior studies used a $t$-test rather than an ANOVA to analyze the data, you may not have a direct estimate of $\sigma_w^2$. Instead, you will have an estimate of the variance of the period differences from the $t$-test, $\hat{\sigma}_d^2$. The two variances, $\sigma_d^2$ and $\sigma_w^2$, are functionally related by $\sigma_w^2 = 2\sigma_d^2$. Either variance can be entered.
Computing the Power

The power is calculated as follows for a directional alternative (one-sided test).

1. Find $t_\alpha$ such that $1 - T_{df}(t_\alpha) = \alpha$, where $T_{df}(x)$ is the area left of $x$ under a central-$t$ curve and $df = N - 2$.

2. Calculate the noncentrality parameter: $\lambda = \frac{\delta \sqrt{N}}{\sigma_w \sqrt{2}}$.

3. Calculate: Power = $1 - T'_{df,\lambda}(t_\alpha)$, where $T'_{df,\lambda}(x)$ is the area to the left of $x$ under a noncentral-$t$ curve with degrees of freedom $df$ and noncentrality parameter $\lambda$.

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains most of the parameters and options that you will be concerned with.

Solve For

Find (Solve For)

This option specifies the parameter to be calculated from the values of the other parameters. Under most conditions, you would select either Power and Beta or $N$.

Select $N$ when you want to determine the sample size needed to achieve a given power and alpha level.

Select Power and Beta when you want to calculate the power of an experiment that has already been run.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of equal means when in fact the means are different.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.
Alpha (Significance Level)
This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when you reject the null hypothesis of equal means when in fact the means are equal.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.

Sample Size
N (Total Sample Size)
This option specifies one or more values of the sample size, the number of individuals in the study (total subjects in both sequences). This value must be an integer greater than one.

Note that you may enter a list of values using the syntax 50,100,150,200,250 or 50 to 250 by 50.

Effect Size – Mean Differences
Diff0 (Mean Difference|H0)
Enter the difference between the treatment means under the null (H0) hypothesis. This is the value that is to be rejected when the t-test is significant. This value is commonly set to zero.

You may enter a range of values such as 10 20 30 or 0 to 100 by 25.

Diff1 (Mean Difference|H1)
Enter the difference between the population means under the alternative (H1) hypothesis. This is the value of the difference at which the power is calculated.

You may enter a range of values such as 10 20 30 or 0 to 100 by 25.

Effect Size – Standard Deviation
Specify S as Sw, SdPeriod, or SdPaired
Specify the form of the standard deviation that is entered in the box below.

- Sw
  Specify S as the square root of the within mean square error from a repeated measures ANOVA. This is the most common method since cross-over designs are usually analyzed using ANOVA.

- SdPeriod
  Specify the standard deviation S as the the standard deviation of the period differences for each subject within each sequence. Note SdPeriod^2 = var((Yi2k - Yi1k)/2) = Sw^2 / 2.

- SdPaired
  Specify the standard deviation S as the the standard deviation of the paired differences. Note SdPaired^2 = var(Yi2k - Yi1k) = 2 * Sw^2.
S (Value of Sw, SdPeriod, or SdPaired)
Specify the value(s) of the standard deviation S. The interpretation of this value depends on the entry in "Specify S as Sw, SdPeriod or SdPaired" above.

If S=Sw is selected, this is the value of Sw which is sqrt(WMSE), where WMSE is the within mean square error from the ANOVA table used to analyze the Cross-Over design. Note Sw^2 = var(Yijk).

If S=SdPeriod is selected, this is the value of SdPeriod, which is the standard deviation of the period differences for each subject within each sequence. Note SdPeriod^2 = var((Yi2k - Yi1k)/2) = Sw^2 / 2.

If S=SdPaired is selected, this is the value of Sd which is the standard deviation of the paired differences. Note SdPaired^2 = var(Yi2k - Yi1k) = 2 * Sw^2.

These values must be positive. A list of values may be entered.

Test

Alternative Hypothesis
This option specifies the alternative hypothesis. This implicitly specifies the direction of the hypothesis test. The null hypothesis is always \( H_0 : \text{Diff0} = \text{Diff1} \).

Note that the alternative hypothesis enters into power calculations by specifying the rejection region of the hypothesis test. Its accuracy is critical.

Possible selections are:

- **H1: DIFF0 <> DIFF1**
  This is the most common selection. It yields the two-sided t-test. Use this option when you are testing whether the means are different but you do not want to specify beforehand which mean is larger. Many scientific journals require two-sided tests.

- **H1: DIFF0 > DIFF1**
  This option yields a one-sided t-test. Use it when you are only interested in the case in which the actual difference is less than Diff0.

- **H1: DIFF0 < DIFF1**
  This option yields a one-sided t-test. Use it when you are only interested in the case in which actual difference is greater than Diff0.
Example 1 – Power Analysis

Suppose you want to consider the power of a balanced cross-over design that will be analyzed using the two-sided t-test approach. The difference between the treatment means under H0 is 0. Similar experiments have had a standard deviation of the differences (Sd) of 10. Compute the power when the true differences are 5 and 10 at sample sizes between 5 and 50. The significance level is 0.05.

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means in a 2x2 Cross-Over Design [Differences] procedure window by expanding Means, then 2x2 Cross-Over Design, then clicking on Test (Inequality), and then clicking on Tests for Two Means in a 2x2 Cross-Over Design [Differences]. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td></td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>5 10 15 20 30 40 50</td>
</tr>
<tr>
<td>Diff0 (Mean Difference</td>
<td>H0)</td>
</tr>
<tr>
<td>Diff1 (Mean Difference</td>
<td>H1)</td>
</tr>
<tr>
<td>Specify S as</td>
<td>SdPeriod</td>
</tr>
<tr>
<td>S</td>
<td>10</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>H1: Diff0 ≠ Diff1</td>
</tr>
</tbody>
</table>

**Annotated Output**

Click the Run button to perform the calculations and generate the following output.

**Numeric Results**

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>Diff0</th>
<th>Diff1</th>
<th>Alpha</th>
<th>Beta</th>
<th>Sd</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0691</td>
<td>5</td>
<td>0.000</td>
<td>5.000</td>
<td>0.050</td>
<td>0.9309</td>
<td>10.000</td>
<td>0.500</td>
</tr>
<tr>
<td>0.1077</td>
<td>10</td>
<td>0.000</td>
<td>5.000</td>
<td>0.050</td>
<td>0.8923</td>
<td>10.000</td>
<td>0.500</td>
</tr>
<tr>
<td>0.1463</td>
<td>15</td>
<td>0.000</td>
<td>5.000</td>
<td>0.050</td>
<td>0.8537</td>
<td>10.000</td>
<td>0.500</td>
</tr>
<tr>
<td>0.1851</td>
<td>20</td>
<td>0.000</td>
<td>5.000</td>
<td>0.050</td>
<td>0.8149</td>
<td>10.000</td>
<td>0.500</td>
</tr>
<tr>
<td>0.2624</td>
<td>30</td>
<td>0.000</td>
<td>5.000</td>
<td>0.050</td>
<td>0.7376</td>
<td>10.000</td>
<td>0.500</td>
</tr>
<tr>
<td>0.3379</td>
<td>40</td>
<td>0.000</td>
<td>5.000</td>
<td>0.050</td>
<td>0.6621</td>
<td>10.000</td>
<td>0.500</td>
</tr>
<tr>
<td>0.4101</td>
<td>50</td>
<td>0.000</td>
<td>5.000</td>
<td>0.050</td>
<td>0.5899</td>
<td>10.000</td>
<td>0.500</td>
</tr>
<tr>
<td>0.1266</td>
<td>5</td>
<td>0.000</td>
<td>10.000</td>
<td>0.050</td>
<td>0.8734</td>
<td>10.000</td>
<td>1.000</td>
</tr>
<tr>
<td>0.2863</td>
<td>10</td>
<td>0.000</td>
<td>10.000</td>
<td>0.050</td>
<td>0.7137</td>
<td>10.000</td>
<td>1.000</td>
</tr>
<tr>
<td>0.4339</td>
<td>15</td>
<td>0.000</td>
<td>10.000</td>
<td>0.050</td>
<td>0.5861</td>
<td>10.000</td>
<td>1.000</td>
</tr>
<tr>
<td>0.5620</td>
<td>20</td>
<td>0.000</td>
<td>10.000</td>
<td>0.050</td>
<td>0.4380</td>
<td>10.000</td>
<td>1.000</td>
</tr>
<tr>
<td>0.7529</td>
<td>30</td>
<td>0.000</td>
<td>10.000</td>
<td>0.050</td>
<td>0.2471</td>
<td>10.000</td>
<td>1.000</td>
</tr>
<tr>
<td>0.8690</td>
<td>40</td>
<td>0.000</td>
<td>10.000</td>
<td>0.050</td>
<td>0.1310</td>
<td>10.000</td>
<td>1.000</td>
</tr>
<tr>
<td>0.9337</td>
<td>50</td>
<td>0.000</td>
<td>10.000</td>
<td>0.050</td>
<td>0.0663</td>
<td>10.000</td>
<td>1.000</td>
</tr>
</tbody>
</table>
500-8 Inequality Tests for Two Means in a 2x2 Cross-Over Design using Differences

Report Definitions
Power is the probability of rejecting a false null hypothesis. It should be close to one.
N is the total sample size drawn from all sequences. The sample is divided equally among sequences.
Alpha is the probability of a false positive.
Beta is the probability of a false negative.
Diff0 is the mean difference under the null hypothesis, H0.
Diff1 is the mean difference under the alternative hypothesis, H1.
Sd is the standard deviation of the difference.
Effect Size, |Diff0-Diff1|/Sd, is the relative magnitude of the effect under the alternative.

Summary Statements
A two-sided t-test achieves 7% power to infer that the mean difference is not 0.000 when the total sample size of a 2x2 cross-over design is 5, the actual mean difference is 5.000, the standard deviation of the differences is 10.000, and the significance level is 0.0500.

Plots Sections

This report shows the values of each of the parameters, one scenario per row. This plot shows the relationship between sample size and power. We see that a sample size of about 46 is needed when Diff1 = 10 for 90% power, while Diff1 = 5 never reaches 90% power in this range of sample sizes.
Example 2 – Finding the Sample Size

Continuing with Example 1, suppose the researchers want to find the exact sample size necessary to achieve 90% power for both values of Diff1.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means in a 2x2 Cross-Over Design [Differences] procedure window by expanding Means, then 2x2 Cross-Over Design, then clicking on Test (Inequality), and then clicking on Tests for Two Means in a 2x2 Cross-Over Design [Differences]. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>( N ) (Sample Size)</td>
</tr>
<tr>
<td>Power</td>
<td>0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Diff0 (Mean Difference</td>
<td>H0)</td>
</tr>
<tr>
<td>Diff1 (Mean Difference</td>
<td>H1)</td>
</tr>
<tr>
<td>Specify S as</td>
<td>SdPeriod</td>
</tr>
<tr>
<td>S</td>
<td>10</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>( H_1: \text{Diff0} \neq \text{Diff1} )</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Numeric Results for Cross-Over Design</th>
<th>Null Hypothesis: ( \text{Diff0}=\text{Diff1} )</th>
<th>Alternative Hypothesis: ( \text{Diff0}&lt;&gt;\text{Diff1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power</td>
<td>( \text{N} )</td>
<td>( \text{Diff0} )</td>
</tr>
<tr>
<td>0.9032</td>
<td>172</td>
<td>0.000</td>
</tr>
<tr>
<td>0.9125</td>
<td>46</td>
<td>0.000</td>
</tr>
</tbody>
</table>

This report shows the exact sample size necessary for each scenario.

Note that the search for \( N \) is conducted across only even values of \( N \) since the design is assumed to be balanced.
Example 3 – Validation using Julious

Julious (2004) page 1933 presents an example in which Diff0 = 0.0, Diff1 = 10, Sw = 20, alpha = 0.05, and beta = 0.10. Julious obtains a sample size of 86.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means in a 2x2 Cross-Over Design [Differences] procedure window by expanding Means, then 2x2 Cross-Over Design, then clicking on Test (Inequality), and then clicking on Tests for Two Means in a 2x2 Cross-Over Design [Differences]. You may then make the appropriate entries as listed below, or open Example 3 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N (Sample Size)</td>
</tr>
<tr>
<td>Power</td>
<td>0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Diff0 (Mean Difference</td>
<td>H0)</td>
</tr>
<tr>
<td>Diff1 (Mean Difference</td>
<td>H1)</td>
</tr>
<tr>
<td>Specify S as</td>
<td>Sw</td>
</tr>
<tr>
<td>S</td>
<td>20</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>H1: Diff0 ≠ Diff1</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>Diff0</th>
<th>Diff1</th>
<th>Alpha</th>
<th>Beta</th>
<th>Sw</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.906483</td>
<td>88</td>
<td>0.000</td>
<td>10.000</td>
<td>0.050000</td>
<td>0.093435</td>
<td>20.000</td>
<td>0.500</td>
</tr>
</tbody>
</table>

PASS obtained a sample size of 88, two higher than that obtained by Julious (2004). However, if you look at the power achieved by an N of 86, you will find that it is 0.899997—slightly less than the goal of 0.90.
Chapter 505

Inequality Tests for Two Means in a 2x2 Cross-Over Design using Ratios

Introduction

This procedure calculates power and sample size for a 2x2 cross-over design in which the logarithm of the outcome is a continuous normal random variable. This routine deals with the case in which the statistical hypotheses are expressed in terms of ratios of means instead of differences of means.

The details of testing two treatments using data from a 2x2 cross-over design are given in another chapter and they will not be repeated here. If the logarithms of the responses can be assumed to follow a normal distribution, hypotheses stated in terms of the ratio can be transformed into hypotheses about the difference. The details of this analysis are given in Julious (2004). They will only be summarized here.

Testing Using Ratios

It will be convenient to adopt the following specialized notation for the discussion of these tests.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PASS Input/Output</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu_T )</td>
<td>Not used</td>
<td>Treatment mean. This is the treatment (group 2) mean.</td>
</tr>
<tr>
<td>( \mu_R )</td>
<td>Not used</td>
<td>Reference mean. This is the reference (group 1) mean.</td>
</tr>
<tr>
<td>( \phi )</td>
<td>R1</td>
<td>True ratio. This is the value of ( \phi = \mu_T / \mu_R ) at which the power is calculated.</td>
</tr>
</tbody>
</table>
Note that the actual values of $\mu_T$ and $\mu_R$ are not needed. Only the ratio of these values is needed for power and sample size calculations.

The null hypothesis is

$$H_0: \phi = \phi_0$$

and the alternative hypothesis is

$$H_1: \phi \neq \phi_0$$

**Log Transformation**

In many cases, hypotheses stated in terms of ratios are more convenient than hypotheses stated in terms of differences. This is because ratios can be interpreted as percentages, but differences must be interpreted as actual amounts in their original scale. Hence, it has become a common practice to take the following steps in hypothesis testing.

1. State the statistical hypotheses in terms of ratios.
2. Transform these into hypotheses about differences by taking logarithms.
3. Analyze the logged data—that is, do the analysis in terms of the difference.
4. Draw the conclusion in terms of the ratio.

The details of step 2 for the null hypothesis are as follows.

$$\phi = \phi_0$$

$$\Rightarrow \phi = \left\{ \frac{\mu_T}{\mu_R} \right\}$$

$$\Rightarrow \ln(\phi) \neq \{\ln(\mu_T) - \ln(\mu_R)\}$$

Thus, a hypothesis about the ratio of the means on the original scale can be translated into a hypothesis about the difference of two means on the logged scale.

**Coefficient of Variation**

The coefficient of variation (COV) is the ratio of the standard deviation to the mean. This parameter can be used to represent the variation in the data because of a unique relationship that it has in the case of log-normal data.

Suppose the variable $X$ is the logarithm of the original variable $Y$. That is, $X = \ln(Y)$ and $Y = \exp(X)$. Label the mean and variance of $X$ as $\mu_x$ and $\sigma_x^2$, respectively. Similarly, label the mean and variance of $Y$ as $\mu_y$ and $\sigma_y^2$, respectively. If $X$ is normally distributed, then $Y$ is log-normally distributed. Julious (2004) presents the following well-known relationships between these two variables.
\[ \mu_y = \left( e^{\frac{\mu_x + \sigma_x^2}{2}} \right) \]

\[ \sigma_y^2 = \mu_y^2 \left( e^{\sigma_x^2} - 1 \right) \]

From this relationship, the coefficient of variation of \( Y \) can be found to be

\[ COV_Y = \sqrt{\frac{\mu_y^2}{\mu_y}} \cdot \left( e^{\sigma_x^2} - 1 \right) \]

\[ = \sqrt{e^{\sigma_x^2} - 1} \]

\[ = \sqrt{e^{\sigma_w^2} - 1} \]

where \( \sigma_w^2 \) is the within mean square error from the analysis of variance of the logged data.

Solving this relationship for \( \sigma_x^2 \), the standard deviation of \( X \) can be stated in terms of the coefficient of variation of \( Y \). This equation is

\[ \sigma_x = \sqrt{\ln \left( COV_Y^2 + 1 \right)} \]

Similarly, the mean of \( X \) is

\[ \mu_x = \frac{\mu_y}{\ln \left( COV_Y^2 + 1 \right)} \]

Thus, the hypotheses can be stated in the original (\( Y \)) scale and then the power can be analyzed in the transformed (\( X \)) scale.

**Power Calculation**

As is shown above, the hypotheses can be stated in the original (\( Y \)) scale using ratios or the logged (\( X \)) scale using differences. In either case, the power and sample size calculations are made using the formulas for testing the difference in two means. These formulas are presented in another chapter and are not duplicated here.

**Procedure Options**

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

**Data Tab**

The Data tab contains the parameters associated with this test such as the means, sample sizes, alpha, and power.
Solve For

Find (Solve For)
This option specifies the parameter to be solved for from the other parameters. In most situations, you will select either Power and Beta for a power analysis or \( N_1 \) for sample size determination.

Error Rates

Power or Beta
This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of equal means when in fact the means are different.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

Alpha (Significance Level)
This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when you reject the null hypothesis of equal means when in fact the means are equal.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.

Sample Size

N (Total Sample Size)
This option specifies one or more values of the total sample size, the number of individuals in the study (total subjects in both sequences). This value must be an integer greater than one.

When \( N \) is even, it is split evenly between the two sequences. When \( N \) is odd, the first sequence has one more subject than the second sequence.

Note that you may enter a list of values using the syntax 50 100 150 200 250 or 50 to 250 by 50.

Effect Size – Ratios

R0 (Ratio Under H0)
This is the value of the ratio of the two means assumed by the null hypothesis, H0. Usually, \( R_0 = 1.0 \) which implies that the two means are equal. However, you may test other values of \( R_0 \) as well. Strictly speaking, any positive number is valid, but, usually, 1.0 is used.

Warning: you cannot use the same value for both \( R_0 \) and \( R_1 \).
**R1 (True Ratio)**
This is the value of the ratio of the two means at which the power is to be calculated. Often, a range of values will be tried. For example, you might try the four values:
1.05 1.10 1.15 1.20
Strictly speaking, any positive number is valid. However, numbers between 0.50 and 2.00 are usually used.
Warning: you cannot use the same value for both R0 and R1.

---

**Effect Size – Coefficient of Variation**

**COV (Coefficient of Variation)**
The coefficient of variation is used to specify the variability (standard deviation). It is important to realize that this is the COV defined on the original (not log) scale. This value must be determined from past experience or from a pilot study. It is most easily calculated from the within mean-square error of the analysis of variance of the logged data using the relationship

\[ COV_Y = \sqrt{e^{\sigma_w^2}} - 1. \]

If prior studies used a t-test to analyze the logged data, you will not have a direct estimate of \( \hat{\sigma}_w^2 \). However, the two variances, \( \sigma_d^2 \) and \( \sigma_w^2 \), are functionally related by \( \sigma_d^2 = 2\sigma_w^2 \).

---

**Test**

**Alternative Hypothesis**
This option specifies the alternative hypothesis. This implicitly specifies the direction of the hypothesis test. Possible selections are:

- **H1: R1 <> R0**
  This is the most common selection. It yields the _two-tailed t-test_. Use this option when you are testing whether the means are different, but you do not want to specify beforehand which mean is larger.

- **H1: R1 < R0**
  This option yields a _one-tailed t-test_. Use it when you are only interested in the case in which Mean1 is greater than Mean2.

- **H1: R1 > R0**
  This option yields a _one-tailed t-test_. Use it when you are only interested in the case in which Mean1 is less than Mean2.
Example 1 – Finding Power

A company has developed a generic drug for treating rheumatism and wants to show that it is better than the standard drug. Responses for either treatment are assumed to follow a lognormal distribution. A 2x2 cross-over design will be used and the logged data will be analyzed using an appropriate analysis of variance. Note that using an analysis of variance instead of a t-test to analyze the data forces the researchers to use two-sided tests.

Past experience leads the researchers to set the COV to 0.50. The significance level is 0.05. The power will be computed for R1 equal to 1.10 and 1.20. Sample sizes between 20 and 220 will be included in the initial analysis.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means in a 2x2 Cross-Over Design [Ratios] procedure window by expanding Means, then 2x2 Cross-Over Design, then clicking on Test (Inequality), and then clicking on Tests for Two Means in a 2x2 Cross-Over Design [Ratios]. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>20 to 220 by 40</td>
</tr>
<tr>
<td>R0 (Ratio Under H0)</td>
<td>1.0</td>
</tr>
<tr>
<td>R1 (True Ratio)</td>
<td>1.1 1.2</td>
</tr>
<tr>
<td>COV (Coefficient of Variation)</td>
<td>0.50</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>R1 ≠ R0 (Two-Sided)</td>
</tr>
</tbody>
</table>
Inequality Tests for Two Means in a 2x2 Cross-Over Design using Ratios

Annotated Output

Click the Run button to perform the calculations and generate the following output.

**Numeric Results**

<table>
<thead>
<tr>
<th>Power</th>
<th>Total Sample Size</th>
<th>Mean Ratio Under H0</th>
<th>Mean Ratio Under H1</th>
<th>Effect Size (ES)</th>
<th>Coefficient of Variation (COV)</th>
<th>Significance Level (Alpha)</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0928</td>
<td>20</td>
<td>1.000</td>
<td>1.100</td>
<td>0.143</td>
<td>0.500</td>
<td>0.0500</td>
<td>0.9072</td>
</tr>
<tr>
<td>0.1925</td>
<td>60</td>
<td>1.000</td>
<td>1.100</td>
<td>0.143</td>
<td>0.500</td>
<td>0.0500</td>
<td>0.8075</td>
</tr>
<tr>
<td>0.2925</td>
<td>100</td>
<td>1.000</td>
<td>1.100</td>
<td>0.143</td>
<td>0.500</td>
<td>0.0500</td>
<td>0.7075</td>
</tr>
<tr>
<td>0.3885</td>
<td>140</td>
<td>1.000</td>
<td>1.100</td>
<td>0.143</td>
<td>0.500</td>
<td>0.0500</td>
<td>0.6115</td>
</tr>
<tr>
<td>0.4777</td>
<td>180</td>
<td>1.000</td>
<td>1.100</td>
<td>0.143</td>
<td>0.500</td>
<td>0.0500</td>
<td>0.5223</td>
</tr>
<tr>
<td>0.5627</td>
<td>220</td>
<td>1.000</td>
<td>1.100</td>
<td>0.143</td>
<td>0.500</td>
<td>0.0500</td>
<td>0.4373</td>
</tr>
<tr>
<td>0.6178</td>
<td>260</td>
<td>1.000</td>
<td>1.100</td>
<td>0.143</td>
<td>0.500</td>
<td>0.0500</td>
<td>0.4373</td>
</tr>
<tr>
<td>0.5474</td>
<td>60</td>
<td>1.000</td>
<td>1.200</td>
<td>0.273</td>
<td>0.500</td>
<td>0.0500</td>
<td>0.7884</td>
</tr>
<tr>
<td>0.7711</td>
<td>100</td>
<td>1.000</td>
<td>1.200</td>
<td>0.273</td>
<td>0.500</td>
<td>0.0500</td>
<td>0.2289</td>
</tr>
<tr>
<td>0.8937</td>
<td>140</td>
<td>1.000</td>
<td>1.200</td>
<td>0.273</td>
<td>0.500</td>
<td>0.0500</td>
<td>0.1063</td>
</tr>
<tr>
<td>0.9537</td>
<td>180</td>
<td>1.000</td>
<td>1.200</td>
<td>0.273</td>
<td>0.500</td>
<td>0.0500</td>
<td>0.0463</td>
</tr>
<tr>
<td>0.9808</td>
<td>220</td>
<td>1.000</td>
<td>1.200</td>
<td>0.273</td>
<td>0.500</td>
<td>0.0500</td>
<td>0.0192</td>
</tr>
</tbody>
</table>

Report Definitions

Power is the probability of rejecting a false null hypothesis. It should be close to one.
N is the total sample size drawn from all sequences. The sample is divided equally among sequences.
R0 is the ratio of the means (Mean2/Mean1) under the null hypothesis, H0.
R1 is the ratio of the means (Mean2/Mean1) at which the power is calculated.
ES is the effect size which is |Ln(R0)-Ln(R1)| / (sigma).
COV is the coefficient of variation on the original scale. The value of sigma is calculated from this.
Alpha is the probability of a false positive H0.
Beta is the probability of a false negative H0.

Summary Statements

A two-sided t-test achieves 9% power to infer that the mean ratio is not 1.000 when the total sample size of a 2x2 cross-over design is 20, the actual mean ratio is 1.100, the coefficient of variation is 0.500, and the significance level is 0.0500.

This report shows the power for the indicated scenarios.

Plots Section

This plot shows the power versus the sample size.
Example 2 – Validation

We will validate this procedure by showing that it gives the identical results to the regular test on differences—a procedure that has been validated. We will use the same settings as those given in Example 1. Since the output for this example is shown above, all that we need is the output from the procedure that uses differences.

To run the power analysis on differences, we need the values of Diff1 (which correspond to R1) and Sw. The value of Diff0 will be zero.

\[ Sw = \sqrt{\ln(COV^2 + 1)} \]
\[ = \sqrt{\ln(0.5^2 + 1)} \]
\[ = 0.472381 \]

\[ Diff1 = \ln(R1) \]
\[ = \ln(1.10) \]
\[ = 0.095310 \]

\[ Diff1 = \ln(R1) \]
\[ = \ln(1.20) \]
\[ = 0.182322 \]

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means in a 2x2 Cross-Over Design [Differences] procedure window by expanding Means, then 2x2 Cross-Over Design, then clicking on Test (Inequality), and then clicking on Tests for Two Means in a 2x2 Cross-Over Design [Differences]. You may then make the appropriate entries as listed below, or open Example 1d by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>20 to 220 by 40</td>
</tr>
<tr>
<td>Diff0 (Mean Difference</td>
<td>H0)</td>
</tr>
<tr>
<td>Diff1 (Mean Difference</td>
<td>H1)</td>
</tr>
<tr>
<td>Specify S as Sw or Sd</td>
<td>Sw</td>
</tr>
<tr>
<td>S (Value of Sw or Sd)</td>
<td>0.472381</td>
</tr>
<tr>
<td>Alternative Hypothesis</td>
<td>H1: DIFF0 ≠ Diff1</td>
</tr>
</tbody>
</table>
Output
Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>Diff0</th>
<th>Diff1</th>
<th>Alpha</th>
<th>Beta</th>
<th>Sw</th>
<th>Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0928</td>
<td>20</td>
<td>0.000</td>
<td>0.095</td>
<td>0.050</td>
<td>0.9072</td>
<td>0.472</td>
<td>0.202</td>
</tr>
<tr>
<td>0.1925</td>
<td>60</td>
<td>0.000</td>
<td>0.095</td>
<td>0.050</td>
<td>0.8075</td>
<td>0.472</td>
<td>0.202</td>
</tr>
<tr>
<td>0.2925</td>
<td>100</td>
<td>0.000</td>
<td>0.095</td>
<td>0.050</td>
<td>0.7075</td>
<td>0.472</td>
<td>0.202</td>
</tr>
<tr>
<td>0.3885</td>
<td>140</td>
<td>0.000</td>
<td>0.095</td>
<td>0.050</td>
<td>0.6115</td>
<td>0.472</td>
<td>0.202</td>
</tr>
<tr>
<td>0.4777</td>
<td>180</td>
<td>0.000</td>
<td>0.095</td>
<td>0.050</td>
<td>0.5223</td>
<td>0.472</td>
<td>0.202</td>
</tr>
<tr>
<td>0.5627</td>
<td>220</td>
<td>0.000</td>
<td>0.095</td>
<td>0.050</td>
<td>0.4373</td>
<td>0.472</td>
<td>0.202</td>
</tr>
<tr>
<td>0.2116</td>
<td>20</td>
<td>0.000</td>
<td>0.182</td>
<td>0.050</td>
<td>0.7884</td>
<td>0.472</td>
<td>0.386</td>
</tr>
<tr>
<td>0.5474</td>
<td>60</td>
<td>0.000</td>
<td>0.182</td>
<td>0.050</td>
<td>0.4526</td>
<td>0.472</td>
<td>0.386</td>
</tr>
<tr>
<td>0.7711</td>
<td>100</td>
<td>0.000</td>
<td>0.182</td>
<td>0.050</td>
<td>0.2289</td>
<td>0.472</td>
<td>0.386</td>
</tr>
<tr>
<td>0.8937</td>
<td>140</td>
<td>0.000</td>
<td>0.182</td>
<td>0.050</td>
<td>0.1063</td>
<td>0.472</td>
<td>0.386</td>
</tr>
<tr>
<td>0.9537</td>
<td>180</td>
<td>0.000</td>
<td>0.182</td>
<td>0.050</td>
<td>0.0463</td>
<td>0.472</td>
<td>0.386</td>
</tr>
<tr>
<td>0.9808</td>
<td>220</td>
<td>0.000</td>
<td>0.182</td>
<td>0.050</td>
<td>0.0192</td>
<td>0.472</td>
<td>0.386</td>
</tr>
</tbody>
</table>

You can compare these power values with those shown above in Example 1 to validate the procedure. You will find that the power values are identical.
505-10  Inequality Tests for Two Means in a 2x2 Cross-Over Design using Ratios
Chapter 508

Tests for Two Means in a 2x2 Cross-Over Design with Non-Zero Null using Differences

Introduction

This procedure computes power and sample size for non-zero null tests in 2x2 cross-over designs in which the outcome is a continuous normal random variable. The details of sample size calculation for the 2x2 cross-over design are presented in the 2x2 Cross-Over Designs chapter and they will not be duplicated here. This chapter only discusses those changes necessary for non-zero null tests. Sample size formulas for non-zero null tests of cross-over designs are presented in Chow et al. (2003) pages 63-68.

Cross-Over Designs

Senn (2002) defines a cross-over design as one in which each subject receives all treatments and the objective is to study differences among the treatments. The name cross-over comes from the most common case in which there are only two treatments. In this case, each subject crosses over from one treatment to the other. It is assumed that there is a washout period between treatments during which the response returns back to its baseline value. If this does not occur, there is said to be a carry-over effect.
Two Means in a 2x2 Cross-Over Design with Non-Zero Null using Differences

A 2x2 cross-over design refers to two treatments (periods) and two sequences (treatment orderings). One sequence receives treatment A followed by treatment B. The other sequence receives B and then A. The design includes a washout period between responses to make certain that the effects of the first drug do no carry-over to the second. Thus, the groups in this design are defined by the sequence in which the two drugs are administered, not by the treatments they receive.

Cross-over designs are employed because, if the no-carryover assumption is met, treatment differences are measured within a subject rather than between subjects—making a more precise measurement. Examples of the situations that might use a cross-over design are the comparison of anti-inflammatory drugs in arthritis and the comparison of hypotensive agents in essential hypertension. In both of these cases, symptoms are expected to return to their usual baseline level shortly after the treatment is stopped.

The Statistical Hypotheses

Both non-inferiority and superiority tests are examples of directional (one-sided) tests and their power and sample size can be calculated using the 2x2 Cross-Over Design procedure. However, at the urging of our users, we have developed this module which provides the input and output in formats that are convenient for these types of tests. This section reviews the specifics of non-inferiority and superiority testing.

Remember that in the usual t-test setting, the null (H0) and alternative (H1) hypotheses for one-sided tests are defined as

\[ H_0: \mu_X \leq A \quad \text{versus} \quad H_1: \mu_X > A \]

Rejecting H0 implies that the mean is larger than the value A. This test is called an upper-tailed test because it is rejected in samples in which the difference in sample means is larger than A.

Following is an example of a lower-tailed test.

\[ H_0: \mu_X \geq A \quad \text{versus} \quad H_1: \mu_X < A \]

Non-inferiority and non-zero null tests are special cases of the above directional tests. It will be convenient to adopt the following specialize notation for the discussion of these tests.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PASS Input/Output</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu_T )</td>
<td>Not used</td>
<td>Treatment mean. This is the treatment mean.</td>
</tr>
<tr>
<td>( \mu_R )</td>
<td>Not used</td>
<td>Reference mean. This is the mean of a reference population.</td>
</tr>
<tr>
<td>( M_s )</td>
<td>SM</td>
<td>Margin of superiority. This is a tolerance value that defines the magnitude of difference that is required for practical importance. This may be thought of as the smallest difference from the reference that is considered to be different.</td>
</tr>
<tr>
<td>( \delta )</td>
<td>D</td>
<td>True difference. This is the value of ( \mu_T - \mu_R ), the difference between the treatment and reference means. This is the value at which the power is calculated.</td>
</tr>
</tbody>
</table>
Note that the actual values of $\mu_T$ and $\mu_R$ are not needed. Only their difference is needed for power and sample size calculations.

**Non-Zero Null Tests**

A non-zero null test tests that the treatment mean is better than the reference mean by more than the superiority margin. The actual direction of the hypothesis depends on the response variable being studied.

**Case 1: High Values Good**

In this case, higher values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is greater than the reference mean by at least the margin of superiority. The value of $\delta$ must be greater than $|M_S|$. The following are equivalent sets of hypotheses.

$$H_0 : \mu_1 \leq \mu_2 + |M_S| \quad \text{versus} \quad H_1 : \mu_1 > \mu_2 + |M_S|$$

$$H_0 : \mu_1 - \mu_2 \leq |M_S| \quad \text{versus} \quad H_1 : \mu_1 - \mu_2 > |M_S|$$

$$H_0 : \delta \leq |M_S| \quad \text{versus} \quad H_1 : \delta > |M_S|$$

**Case 2: High Values Bad**

In this case, lower values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is less than the reference mean by at least the margin of superiority. The value of $\delta$ must be less than $-|M_S|$. The following are equivalent sets of hypotheses.

$$H_0 : \mu_1 \geq \mu_2 - |M_S| \quad \text{versus} \quad H_1 : \mu_1 < \mu_2 - |M_S|$$

$$H_0 : \mu_1 - \mu_2 \geq -|M_S| \quad \text{versus} \quad H_1 : \mu_1 - \mu_2 < -|M_S|$$

$$H_0 : \delta \geq -|M_S| \quad \text{versus} \quad H_1 : \delta < -|M_S|$$

**Test Statistics**

This section describes the test statistic that is used to perform the hypothesis test.

**T-Test**

A t-test is used to analyze the data. When the data are balanced between sequences, the two-sided t-test is equivalent to an analysis of variance F-test. The test assumes that the data are a simple random sample from a population of normally-distributed values that have the same variance. This assumption implies that the differences are continuous and normal. The calculation of the t-statistic proceeds as follow
Two Means in a 2x2 Cross-Over Design with Non-Zero Null using Differences

\[ t_d = \frac{(\bar{x}_d - \bar{x}_s) - \varepsilon}{\hat{\sigma}_w \sqrt{\frac{2}{N}}} \]

where \( \hat{\sigma}_w^2 \) is the within mean square error from the appropriate ANOVA table.

The significance of the test statistic is determined by computing the p-value. If this p-value is less than a specified level (usually 0.05), the hypothesis is rejected. That is, the one-sided null hypothesis is rejected at the \( \alpha \) significance level if \( t_d > t_{\alpha,N-2} \). Otherwise, no conclusion can be reached.

If prior studies used a t-test rather than an ANOVA to analyze the data, you may not have a direct estimate of \( \sigma_w^2 \). Instead, you will have an estimate of the variance of the period differences from the t-test, \( \hat{\sigma}_d^2 \). These variances are functionally related by \( \sigma_w^2 = 2\sigma_d^2 \). Either variance can be entered.

**Computing the Power**

The power is calculated as follows.

1. Find \( t_\alpha \) such that \( 1 - T_{df}(t_\alpha) = \alpha \), where \( T_{df}(x) \) is the area under a central-t curve to the left of \( x \) and \( df = N - 2 \).

2. Calculate the noncentrality parameter: 
   \[ \lambda = \frac{(\delta - \varepsilon)\sqrt{N}}{\sigma_w \sqrt{2}} \]

3. Calculate: Power = \( 1 - T'_{df,\lambda}(t_\alpha) \), where \( T'_{df,\lambda}(x) \) is the area under a noncentral-t curve with degrees of freedom \( df \) and noncentrality parameter \( \lambda \) to the left of \( x \).

**Procedure Options**

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

**Data Tab**

The Data tab contains most of the parameters and options that you will be concerned with.

**Solve For**

**Find (Solve For)**

This option specifies the parameter to be calculated from the values of the other parameters. Under most conditions, you would select either *Power and Beta* or *N*. 
Select $N$ when you want to determine the sample size needed to achieve a given power and alpha level.

Select Power and Beta when you want to calculate the power of an experiment that has already been run.

---

**Error Rates**

**Power or Beta**

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of inferiority when in fact the null hypothesis should be rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

**Alpha (Significance Level)**

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when you reject the null hypothesis when in fact the treatment is not superior.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.

---

**Sample Size**

**N (Total Sample Size)**

This option specifies one or more values of the sample size, the number of individuals in the study. This value must be an integer greater than one. Note that you may enter a list of values using the syntax 50,100,150,200,250 or 50 to 250 by 50.

---

**Effect Size – Mean Difference**

**SM (Superiority Margin)**

This is the magnitude of the margin of superiority. It must be entered as a positive number.

When higher means are better, this value is the distance above the reference mean that is required to be considered superior. When higher means are worse, this value is the distance below the reference mean that is required to be considered superior.

**D (True Difference)**

This is the actual difference between the treatment mean and the reference mean at which the power is calculated.
When higher means are better, this value should be greater than SM. When higher means are worse, this value should be negative and greater in magnitude than SM.

**Effect Size – Standard Deviation**

Specify S as Sw, SdPeriod, or SdPaired

Specify the form of the standard deviation that is entered in the box below.

- **Sw**
  Specify S as the square root of the within mean square error from a repeated measures ANOVA. This is the most common method since cross-over designs are usually analyzed using ANOVA.

- **SdPeriod**
  Specify the standard deviation S as the the standard deviation of the period differences for each subject within each sequence. Note SdPeriod^2 = var((Yi2k - Yi1k)/2) = Sw^2 / 2.

- **SdPaired**
  Specify the standard deviation S as the the standard deviation of the paired differences. Note SdPaired^2 = var(Yi2k - Yi1k) = 2 * Sw^2.

**S (Value of Sw, SdPeriod, or SdPaired)**

Specify the value(s) of the standard deviation S. The interpretation of this value depends on the entry in "Specify S as Sw, SdPeriod or SdPaired" above.

If S=Sw is selected, this is the value of Sw which is sqrt(WMSE), where WMSE is the within mean square error from the ANOVA table used to analyze the Cross-Over design. Note Sw^2 = var(Yijk).

If S=SdPeriod is selected, this is the value of SdPeriod, which is the standard deviation of the period differences for each subject within each sequence. Note SdPeriod^2 = var((Yi2k - Yi1k)/2) = Sw^2 / 2.

If S=SdPaired is selected, this is the value of Sd which is the standard deviation of the paired differences. Note SdPaired^2 = var(Yi2k - Yi1k) = 2 * Sw^2.

These values must be positive. A list of values may be entered.

**Test**

**Higher Means Are**

This option defines whether higher values of the response variable are to be considered better or worse.

The choice here determines the direction of the test.

If Higher Means Are Better the null hypothesis is Diff <= SM and the alternative hypothesis is Diff > SM. If Higher Means Are Worse the null hypothesis is Diff >= -SM and the alternative hypothesis is Diff < -SM.
Example 1 – Power Analysis

Suppose you want to consider the power of a balanced, cross-over design that will be analyzed using the t-test approach. You want to compute the power when the margin of superiority is either 5 or 10 at several sample sizes between 5 and 50. The true difference between the means under H0 is assumed to be 15. Similar experiments have had a value for $S_w$ of 10. The significance level is 0.025.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means in a 2x2 Cross-Over Design (Non-Zero Null) [Differences] procedure window by expanding Means, then 2x2 Cross-Over Design, then clicking on Test (Non-Zero Null), and then clicking on Tests for Two Means in a 2x2 Cross-Over Design (Non-Zero Null) [Differences]. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.025</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>5 10 15 20 30 40 50</td>
</tr>
<tr>
<td>SM (Superiority Margin)</td>
<td>5 10</td>
</tr>
<tr>
<td>D (True Difference)</td>
<td>15</td>
</tr>
<tr>
<td>Specify S as Sw or Sd</td>
<td>Sw</td>
</tr>
<tr>
<td>S (Value of Sw or Sd)</td>
<td>10</td>
</tr>
<tr>
<td>Higher Means Are</td>
<td>Better</td>
</tr>
</tbody>
</table>
Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results and Plots

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>Superiority Margin</th>
<th>Actual Difference</th>
<th>Significance Level</th>
<th>Standard Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(SM)</td>
<td>(D)</td>
<td>(Alpha)</td>
<td>(Sw)</td>
</tr>
<tr>
<td>0.20131</td>
<td>5</td>
<td>5.000</td>
<td>15.000</td>
<td>0.02500</td>
<td>0.79869</td>
</tr>
<tr>
<td>0.50245</td>
<td>10</td>
<td>5.000</td>
<td>15.000</td>
<td>0.02500</td>
<td>0.49755</td>
</tr>
<tr>
<td>0.71650</td>
<td>15</td>
<td>5.000</td>
<td>15.000</td>
<td>0.02500</td>
<td>0.28350</td>
</tr>
<tr>
<td>0.84845</td>
<td>20</td>
<td>5.000</td>
<td>15.000</td>
<td>0.02500</td>
<td>0.15155</td>
</tr>
<tr>
<td>0.96222</td>
<td>30</td>
<td>5.000</td>
<td>15.000</td>
<td>0.02500</td>
<td>0.03778</td>
</tr>
<tr>
<td>0.99173</td>
<td>40</td>
<td>5.000</td>
<td>15.000</td>
<td>0.02500</td>
<td>0.00827</td>
</tr>
<tr>
<td>0.99835</td>
<td>50</td>
<td>5.000</td>
<td>15.000</td>
<td>0.02500</td>
<td>0.00165</td>
</tr>
<tr>
<td>0.08310</td>
<td>5</td>
<td>10.000</td>
<td>15.000</td>
<td>0.02500</td>
<td>0.91690</td>
</tr>
<tr>
<td>0.16563</td>
<td>10</td>
<td>10.000</td>
<td>15.000</td>
<td>0.02500</td>
<td>0.83437</td>
</tr>
<tr>
<td>0.24493</td>
<td>15</td>
<td>10.000</td>
<td>15.000</td>
<td>0.02500</td>
<td>0.75507</td>
</tr>
<tr>
<td>0.32175</td>
<td>20</td>
<td>10.000</td>
<td>15.000</td>
<td>0.02500</td>
<td>0.67825</td>
</tr>
<tr>
<td>0.46414</td>
<td>30</td>
<td>10.000</td>
<td>15.000</td>
<td>0.02500</td>
<td>0.53586</td>
</tr>
<tr>
<td>0.58682</td>
<td>40</td>
<td>10.000</td>
<td>15.000</td>
<td>0.02500</td>
<td>0.41318</td>
</tr>
<tr>
<td>0.68785</td>
<td>50</td>
<td>10.000</td>
<td>15.000</td>
<td>0.02500</td>
<td>0.31215</td>
</tr>
</tbody>
</table>

Report Definitions

H0 (null hypothesis) is Diff <= SM, where D = Treatment Mean - Reference Mean.
H1 (alternative hypothesis) is Diff > SM.
Power is the probability of rejecting H0 when it is false.
N is the total sample size drawn from all sequences. The sample is divided equally among sequences.
SM is the magnitude of the margin of superiority. Since higher means are better, this value is positive and is
the distance above the reference mean that is required to be considered superior.
D is the mean difference at which the power is computed. D = Mean1 - Mean2 = treatment mean - reference mean.
Alpha is the probability of a false positive.
Beta is the probability of a false negative.
Sw is the square root of the within mean square error from the ANOVA table.

Summary Statements

A total sample size of 5 achieves 20% power to detect superiority using a one-sided t-test when
the margin of superiority is 5.000, the true mean difference is 15.000, the significance level
is 0.02500, the square root of the within mean square error is 10.000, and A 2x2 cross-over
design with an equal number in each sequence is used.
Two Means in a 2x2 Cross-Over Design with Non-Zero Null using Differences

This report shows the values of each of the parameters, one scenario per row. The plot shows the relationship between sample size and power. We see that a sample size of about 20 is needed to achieve 80% power when SM = 5.

Example 2 – Finding the Sample Size

Continuing with Example 1, suppose the researchers want to find the exact sample size necessary to achieve 90% power for both values of D.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means in a 2x2 Cross-Over Design (Non-Zero Null) [Differences] procedure window by expanding Means, then 2x2 Cross-Over Design, then clicking on Test (Non-Zero Null), and then clicking on Tests for Two Means in a 2x2 Cross-Over Design (Non-Zero Null) [Differences]. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N (Sample Size)</td>
</tr>
<tr>
<td>Power</td>
<td>0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.025</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>SM (Superiority Margin)</td>
<td>5 10</td>
</tr>
<tr>
<td>D (True Difference)</td>
<td>15</td>
</tr>
<tr>
<td>Specify S as Sw or Sd</td>
<td>Sw</td>
</tr>
<tr>
<td>S (Value of Sw or Sd)</td>
<td>10</td>
</tr>
<tr>
<td>Higher Means Are</td>
<td>Better</td>
</tr>
</tbody>
</table>
Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>Superiority Margin (SM)</th>
<th>Actual Difference (D)</th>
<th>Significance Level (Alpha)</th>
<th>Standard Deviation (Sw)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.91139</td>
<td>24</td>
<td>5.000</td>
<td>15.000</td>
<td>0.02500</td>
<td>10.000</td>
</tr>
<tr>
<td>0.90648</td>
<td>88</td>
<td>10.000</td>
<td>15.000</td>
<td>0.02500</td>
<td>10.000</td>
</tr>
</tbody>
</table>

This report shows the exact sample size necessary for each scenario.

Note that the search for N is conducted across only even values of N since the design is assumed to be balanced.

Example 3 – Validation using Julious

This procedure uses the same mechanics as the Non-Inferiority Tests for Two Means in a 2x2 Cross-Over Design using Differences procedure. We refer the user to Example 3 of Chapter 510 for the validation.
Chapter 510

Non-Inferiority Tests for Two Means in a 2x2 Cross-Over Design using Differences

Introduction

This procedure computes power and sample size for non-inferiority tests in 2x2 cross-over designs in which the outcome is a continuous normal random variable. The details of sample size calculation for the 2x2 cross-over design are presented in the 2x2 Cross-Over Designs chapter and they will not be duplicated here. This chapter only discusses those changes necessary for non-inferiority tests. Sample size formulas for non-inferiority tests of cross-over designs are presented in Chow et al. (2003) pages 63-68.

Cross-Over Designs

Senn (2002) defines a cross-over design as one in which each subject receives all treatments and the objective is to study differences among the treatments. The name cross-over comes from the most common case in which there are only two treatments. In this case, each subject crosses over from one treatment to the other. It is assumed that there is a washout period between treatments during which the response returns back to its baseline value. If this does not occur, there is said to be a carry-over effect.
A 2x2 cross-over design refers to two treatments (periods) and two sequences (treatment orderings). One sequence receives treatment A followed by treatment B. The other sequence receives B and then A. The design includes a washout period between responses to make certain that the effects of the first drug do no carry-over to the second. Thus, the groups in this design are defined by the sequence in which the two drugs are administered, not by the treatments they receive.

Cross-over designs are employed because, if the no-carryover assumption is met, treatment differences are measured within a subject rather than between subjects—making a more precise measurement. Examples of the situations that might use a cross-over design are the comparison of anti-inflammatory drugs in arthritis and the comparison of hypotensive agents in essential hypertension. In both of these cases, symptoms are expected to return to their usual baseline level shortly after the treatment is stopped.

### The Statistical Hypotheses

Both non-inferiority and superiority tests are examples of directional (one-sided) tests and their power and sample size can be calculated using the 2x2 Cross-Over Design procedure. However, at the urging of our users, we have developed this module which provides the input and output in formats that are convenient for these types of tests. This section reviews the specifics of non-inferiority and superiority testing.

Remember that in the usual t-test setting, the null (H0) and alternative (H1) hypotheses for one-sided tests are defined as

\[
H_0: \mu_X \leq A \quad \text{versus} \quad H_1: \mu_X > A
\]

Rejecting H0 implies that the mean is larger than the value A. This test is called an upper-tailed test because it is rejected in samples in which the difference in sample means is larger than A.

Following is an example of a lower-tailed test.

\[
H_0: \mu_X \geq A \quad \text{versus} \quad H_1: \mu_X < A
\]

Non-inferiority and superiority tests are special cases of the above directional tests. It will be convenient to adopt the following specialize notation for the discussion of these tests.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PASS Input/Output</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu_T )</td>
<td>Not used</td>
<td>Treatment mean. This is the treatment mean.</td>
</tr>
<tr>
<td>( \mu_R )</td>
<td>Not used</td>
<td>Reference mean. This is the mean of a reference population.</td>
</tr>
</tbody>
</table>
Non-Inferiority Tests for Two Means in a 2x2 Cross-Over Design using Differences

\(M_{NI}\) NIM

Margin of non-inferiority. This is a tolerance value that defines the magnitude of the amount that is not of practical importance. This may be thought of as the largest change from the baseline that is considered to be trivial. The absolute value is shown to emphasize that this is a magnitude. The sign of the value will be determined by the specific design that is being used.

\(\delta\) D

True difference. This is the value of \(\mu_T - \mu_R\), the difference between the treatment and reference means. This is the value at which the power is calculated.

Note that the actual values of \(\mu_T\) and \(\mu_R\) are not needed. Only their difference is needed for power and sample size calculations.

Non-Inferiority Tests

A non-inferiority test tests that the treatment mean is not worse than the reference mean by more than the equivalence margin. The actual direction of the hypothesis depends on the response variable being studied.

Case 1: High Values Good, Non-Inferiority Test

In this case, higher values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is no less than a small amount below the reference mean. The value of \(\delta\) is often set to zero. The following are equivalent sets of hypotheses.

\[H_0 : \mu_1 \leq \mu_2 - |M_{NI}| \quad \text{versus} \quad H_1 : \mu_1 > \mu_2 - |M_{NI}|\]

\[H_0 : \mu_1 - \mu_2 \leq -|M_{NI}| \quad \text{versus} \quad H_1 : \mu_1 - \mu_2 > -|M_{NI}|\]

\[H_0 : \delta \leq -|M_{NI}| \quad \text{versus} \quad H_1 : \delta > -|M_{NI}|\]

Case 2: High Values Bad, Non-Inferiority Test

In this case, lower values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is no more than a small amount above the reference mean. The value of \(\delta\) is often set to zero. The following are equivalent sets of hypotheses.

\[H_0 : \mu_1 \geq \mu_2 + |M_{NI}| \quad \text{versus} \quad H_1 : \mu_1 < \mu_2 + |M_{NI}|\]

\[H_0 : \mu_1 - \mu_2 \geq |M_{NI}| \quad \text{versus} \quad H_1 : \mu_1 - \mu_2 < |M_{NI}|\]

\[H_0 : \delta \geq |M_{NI}| \quad \text{versus} \quad H_1 : \delta < |M_{NI}|\]
Test Statistics

This section describes the test statistic that is used to perform the hypothesis test.

T-Test

A t-test is used to analyze the data. When the data are balanced between sequences, the two-sided t-test is equivalent to an analysis of variance F-test. The test assumes that the data are a simple random sample from a population of normally-distributed values that have the same variance. This assumption implies that the differences are continuous and normal. The calculation of the t-statistic proceeds as follow

\[
t_d = \frac{(\bar{x}_T - \bar{x}_R) - \varepsilon}{\hat{\sigma}_w \sqrt{\frac{2}{N}}}
\]

where \(\hat{\sigma}_w^2\) is the within mean square error from the appropriate ANOVA table.

The significance of the test statistic is determined by computing the p-value. If this p-value is less than a specified level (usually 0.05), the hypothesis is rejected. That is, the one-sided null hypothesis is rejected at the \(\alpha\) significance level if \(t_d > t_{\alpha, N-2}\). Otherwise, no conclusion can be reached.

If prior studies used a t-test rather than an ANOVA to analyze the data, you may not have a direct estimate of \(\sigma_w^2\). Instead, you will have an estimate of the variance of the period differences from the t-test, \(\hat{\sigma}_d^2\). These variances are functionally related by \(\sigma_w^2 = 2\sigma_d^2\). Either variance can be entered.

Computing the Power

The power is calculated as follows.

1. Find \(t_{\alpha}\) such that \(1 - T_{df}(t_{\alpha}) = \alpha\), where \(T_{df}(x)\) is the area under a central-t curve to the left of \(x\) and \(df = N - 2\).

2. Calculate the noncentrality parameter: \(\lambda = \frac{(\delta - \varepsilon)\sqrt{N}}{\sigma_w \sqrt{2}}\).

3. Calculate: Power = \(1 - T_{df, \lambda}(t_{\alpha})\), where \(T_{df, \lambda}(x)\) is the area under a noncentral-t curve with degrees of freedom \(df\) and noncentrality parameter \(\lambda\) to the left of \(x\).
Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains most of the parameters and options that you will be concerned with.

Solve For

Find (Solve For)

This option specifies the parameter to be calculated from the values of the other parameters. Under most conditions, you would select either Power and Beta or N.

Select N when you want to determine the sample size needed to achieve a given power and alpha level.

Select Power and Beta when you want to calculate the power of an experiment that has already been run.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of inferiority when in fact the treatment mean is non-inferior.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when you reject the null hypothesis of different means when in fact the means are different.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.
Sample Size

N (Total Sample Size)
This option specifies one or more values of the sample size, the number of individuals in the study. This value must be an integer greater than one. Note that you may enter a list of values using the syntax 50,100,150,200,250 or 50 to 250 by 50.

Effect Size – Mean Difference

NIM (Non-Inferiority Margin)
This is the magnitude of the margin of non-inferiority. It must be entered as a positive number.

When higher means are better, this value is the distance below the reference mean that is still considered non-inferior. When higher means are worse, this value is the distance above the reference mean that is still considered non-inferior.

D (True Difference)
This is the actual difference between the treatment mean and the reference mean at which power is calculated.

For non-inferiority tests, this value is often set to zero. When this value is non-zero, care should be taken that this value is consistent with whether higher means are better or worse.

Effect Size – Standard Deviation

Specify S as Sw, SdPeriod, or SdPaired
Specify the form of the standard deviation that is entered in the box below.

- **Sw**
  Specify S as the square root of the within mean square error from a repeated measures ANOVA. This is the most common method since cross-over designs are usually analyzed using ANOVA.

- **SdPeriod**
  Specify the standard deviation S as the the standard deviation of the period differences for each subject within each sequence. Note SdPeriod^2 = var((Yi2k - Yi1k)/2) = Sw^2 / 2.

- **SdPaired**
  Specify the standard deviation S as the the standard deviation of the paired differences. Note SdPaired^2 = var(Yi2k - Yi1k) = 2 * Sw^2.

S (Value of Sw, SdPeriod, or SdPaired)
Specify the value(s) of the standard deviation S. The interpretation of this value depends on the entry in "Specify S as Sw, SdPeriod or SdPaired" above.

If S=Sw is selected, this is the value of Sw which is sqrt(WMSE), where WMSE is the within mean square error from the ANOVA table used to analyze the Cross-Over design. Note Sw^2 = var(Yijk).
IF $S=S_{\text{period}}$ is selected, this is the value of $S_{\text{period}}$, which is the standard deviation of the period differences for each subject within each sequence. Note $S_{\text{period}}^2 = \text{var}((Y_{i2k} - Y_{i1k})/2) = Sw^2 / 2$.

IF $S=S_{\text{paired}}$ is selected, this is the value of $S_d$ which is the standard deviation of the paired differences. Note $S_{\text{paired}}^2 = \text{var}(Y_{i2k} - Y_{i1k}) = 2 * Sw^2$.

These values must be positive. A list of values may be entered.

**Test**

**Higher Means Are**

This option defines whether higher values of the response variable are to be considered better or worse.

The choice here determines the direction of the non-inferiority test.

If Higher Means Are Better the null hypothesis is $\text{Diff} \leq -\text{NIM}$ and the alternative hypothesis is $\text{Diff} > -\text{NIM}$. If Higher Means Are Worse the null hypothesis is $\text{Diff} \geq \text{NIM}$ and the alternative hypothesis is $\text{Diff} < \text{NIM}$.

**Example 1 – Power Analysis**

Suppose you want to consider the power of a balanced, cross-over design that will be analyzed using the $t$-test approach. You want to compute the power when the margin of equivalence is either 5 or 10 at several sample sizes between 5 and 50. The true difference between the means under $H_0$ is assumed to be 0. Similar experiments have had an $Sw$ of 10. The significance level is 0.025.

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the *Non-Inferiority Tests for Two Means in a 2x2 Cross-Over Design [Differences]* procedure window by expanding *Means*, then *2x2 Cross-Over Design*, then clicking on *Non-Inferiority*, and then clicking on *Non-Inferiority Tests for Two Means in a 2x2 Cross-Over Design [Differences]*. You may then make the appropriate entries as listed below, or open Example 1 by going to the *File* menu and choosing *Open Example Template*.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.025</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>5 10 15 20 30 40 50</td>
</tr>
<tr>
<td>NIM (Non-Inferiority Margin)</td>
<td>5 10</td>
</tr>
<tr>
<td>D (True Difference)</td>
<td>0</td>
</tr>
<tr>
<td>Specify S as</td>
<td>Sw</td>
</tr>
<tr>
<td>S</td>
<td>10</td>
</tr>
<tr>
<td>Higher Means Are</td>
<td>Better</td>
</tr>
</tbody>
</table>
Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results and Plots

**Numeric Results for Non-Inferiority T-Test (H0: Diff <= -NIM; H1: Diff > -NIM)**

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>Non-Inf. Margin (-NIM)</th>
<th>Actual Difference (D)</th>
<th>Significance Level (Alpha)</th>
<th>Beta</th>
<th>Standard Deviation (Sw)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.08310</td>
<td>5</td>
<td>-5.000</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.91690</td>
<td>10.000</td>
</tr>
<tr>
<td>0.16563</td>
<td>10</td>
<td>-5.000</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.83437</td>
<td>10.000</td>
</tr>
<tr>
<td>0.24493</td>
<td>15</td>
<td>-5.000</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.75507</td>
<td>10.000</td>
</tr>
<tr>
<td>0.32175</td>
<td>20</td>
<td>-5.000</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.67825</td>
<td>10.000</td>
</tr>
<tr>
<td>0.46414</td>
<td>30</td>
<td>-5.000</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.53586</td>
<td>10.000</td>
</tr>
<tr>
<td>0.58662</td>
<td>40</td>
<td>-5.000</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.41318</td>
<td>10.000</td>
</tr>
<tr>
<td>0.68785</td>
<td>50</td>
<td>-5.000</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.31215</td>
<td>10.000</td>
</tr>
<tr>
<td>0.70131</td>
<td>5</td>
<td>-10.000</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.79869</td>
<td>10.000</td>
</tr>
<tr>
<td>0.50245</td>
<td>10</td>
<td>-10.000</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.49755</td>
<td>10.000</td>
</tr>
<tr>
<td>0.71650</td>
<td>15</td>
<td>-10.000</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.28350</td>
<td>10.000</td>
</tr>
<tr>
<td>0.84845</td>
<td>20</td>
<td>-10.000</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.15155</td>
<td>10.000</td>
</tr>
<tr>
<td>0.96222</td>
<td>30</td>
<td>-10.000</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.03778</td>
<td>10.000</td>
</tr>
<tr>
<td>0.99173</td>
<td>40</td>
<td>-10.000</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.00827</td>
<td>10.000</td>
</tr>
<tr>
<td>0.99835</td>
<td>50</td>
<td>-10.000</td>
<td>0.000</td>
<td>0.02500</td>
<td>0.00165</td>
<td>10.000</td>
</tr>
</tbody>
</table>

**Report Definitions**

H0 (null hypothesis) is that Diff <= -NIM, where Diff = Treatment Mean - Reference Mean.
H1 (alternative hypothesis) is that Diff > -NIM.

Power is the probability of rejecting H0 when it is false.
N is the total sample size drawn from all sequences. The sample is divided equally among sequences.
-NIM is the magnitude and direction of the margin of non-inferiority. Since higher means are better, this value is negative and is the distance below the reference mean that is still considered non-inferior.
D is the actual difference between the treatment and reference means that is used in the power calculations.
Alpha is the probability of a false positive.
Beta is the probability of a false negative.
Sw is the square root of the within mean square error from the ANOVA table.

**Summary Statements**

A total sample size of 5 achieves 8% power to detect non-inferiority using a one-sided t-test when the margin of non-inferiority is -5.000, the true mean difference is 0.000, the significance level is 0.02500, and the square root of the within mean square error is 10.000. A 2x2 cross-over design with an equal number in each sequence is used.
This report shows the values of each of the parameters, one scenario per row. The plot shows the relationship between sample size and power. We see that a sample size of about 20 is needed to achieve 80% power when NIM = -10.

**Example 2 – Finding the Sample Size**

Continuing with Example 1, suppose the researchers want to find the exact sample size necessary to achieve 90% power for both values of D.

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the *Non-Inferiority Tests for Two Means in a 2x2 Cross-Over Design [Differences]* procedure window by expanding *Means*, then *2x2 Cross-Over Design*, then clicking on *Non-Inferiority*, and then clicking on *Non-Inferiority Tests for Two Means in a 2x2 Cross-Over Design [Differences]*. You may then make the appropriate entries as listed below, or open Example 2 by going to the *File* menu and choosing *Open Example Template*.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For).............</td>
<td>N (Sample Size).............</td>
</tr>
<tr>
<td>Power............</td>
<td>0.90</td>
</tr>
<tr>
<td>Alpha..................</td>
<td>0.025</td>
</tr>
<tr>
<td>N (Total Sample Size)........</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>NIM (Non-Inferiority Margin)</td>
<td>5 10</td>
</tr>
<tr>
<td>D (True Difference)..........</td>
<td>0</td>
</tr>
<tr>
<td>Specify S as ...............</td>
<td>Sw</td>
</tr>
<tr>
<td>S...........................</td>
<td>10</td>
</tr>
<tr>
<td>Higher Means Are ...........</td>
<td>Better</td>
</tr>
</tbody>
</table>

**Output**

Click the Run button to perform the calculations and generate the following output.

**Numeric Results**

<table>
<thead>
<tr>
<th>Non-Inferiority Margin (-NIM)</th>
<th>Actual Difference (D)</th>
<th>Significance Level (Alpha)</th>
<th>Beta</th>
<th>Standard Deviation (Sw)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.90648</td>
<td>88</td>
<td>-5.000</td>
<td>0.000</td>
<td>0.02500</td>
</tr>
<tr>
<td>0.91139</td>
<td>24</td>
<td>-10.000</td>
<td>0.000</td>
<td>0.02500</td>
</tr>
</tbody>
</table>

This report shows the exact sample size necessary for each scenario.

Note that the search for N is conducted across only even values of N since the design is assumed to be balanced.
Example 3 – Validation using Julious

Julious (2004) page 1953 presents an example in which $D = 0.0$, $E = 10$, $Sw = 20.00$, $alpha = 0.025$, and $beta = 0.10$. Julious obtains a sample size of 86.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Non-Inferiority Tests for Two Means in a 2x2 Cross-Over Design [Differences] procedure window by expanding Means, then 2x2 Cross-Over Design, then clicking on Non-Inferiority, and then clicking on Non-Inferiority Tests for Two Means in a 2x2 Cross-Over Design [Differences]. You may then make the appropriate entries as listed below, or open Example 3 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>$N$ (Sample Size)</td>
</tr>
<tr>
<td>Power</td>
<td>0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.025</td>
</tr>
<tr>
<td>$N$ (Total Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>NIM (Non-Inferiority Margin)</td>
<td>10</td>
</tr>
<tr>
<td>D (True Difference)</td>
<td>0</td>
</tr>
<tr>
<td>Specify S as</td>
<td>$Sw$</td>
</tr>
<tr>
<td>S</td>
<td>20</td>
</tr>
<tr>
<td>Higher Means Are</td>
<td>Better</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Numeric Results for Non-Inferiority T-Test (H0: Diff &lt;= -NIM; H1: Diff &gt; -NIM) Higher Means are Better</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>0.90648</td>
</tr>
</tbody>
</table>

PASS obtained a sample size of 88, two higher than that obtained by Julious (2004). However, if you look at the power achieved by an $N$ of 86, you will find that it is 0.899997—slightly less than the goal of 0.90.
Chapter 513

Tests for Two Means in a 2x2 Cross-Over Design with Non-Unity Null using Ratios

Introduction

This procedure calculates power and sample size of statistical tests for non-unity null tests from a 2x2 cross-over design. This routine deals with the case in which the statistical hypotheses are expressed in terms mean ratios rather than mean differences.

The details of testing the non-unity null of two treatments using data from a 2x2 cross-over design are given in another chapter and they will not be repeated here. If the logarithms of the responses can be assumed to follow the normal distribution, hypotheses about non-unity null hypotheses stated in terms of the ratio can be transformed into hypotheses about the difference. The details of this analysis are given in Julious (2004). They will only be summarized here.
Non-Inferiority Testing Using Ratios

It will be convenient to adopt the following specialized notation for the discussion of these tests.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PASS Input/Output</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu_T )</td>
<td>Not used</td>
<td>Treatment mean. This is the treatment mean.</td>
</tr>
<tr>
<td>( \mu_R )</td>
<td>Not used</td>
<td>Reference mean. This is the mean of a reference population.</td>
</tr>
<tr>
<td>( M_s )</td>
<td>SM</td>
<td>Margin of superiority. This is a tolerance value that defines the magnitude of difference that is required for practical importance. This may be thought of as the smallest difference from the reference that is considered to be different.</td>
</tr>
<tr>
<td>( \phi )</td>
<td>R1</td>
<td>True ratio. This is the value of ( \phi = \frac{\mu_T}{\mu_R} ) at which the power is calculated.</td>
</tr>
</tbody>
</table>

Note that the actual values of \( \mu_T \) and \( \mu_R \) are not needed. Only their ratio is needed for power and sample size calculations.

The null hypothesis of non-superiority is

\[ H_0 : \phi \leq \phi_L \text{ where } \phi_L > 1. \]

and the alternative hypothesis of superiority is

\[ H_1 : \phi > \phi_L. \]

Log Transformation

In many cases, hypotheses stated in terms of ratios are more convenient than hypotheses stated in terms of differences. This is because ratios can be interpreted as scale-less percentages, but differences must be interpreted as actual amounts in their original scale. Hence, it has become a common practice to take the following steps in hypothesis testing.

1. State the statistical hypotheses in terms of ratios.
2. Transform these into hypotheses about differences by taking logarithms.
3. Analyze the logged data—that is, do the analysis in terms of the difference.
4. Draw the conclusion in terms of the ratio.

The details of step 2 for the null hypothesis are as follows.

\[ \phi_L \leq \phi \]
\[ \Rightarrow \phi_L \leq \left\{ \frac{\mu_T}{\mu_R} \right\} \]
\[ \Rightarrow \ln(\phi_L) \leq \left\{ \ln(\mu_T) - \ln(\mu_R) \right\} \]

Thus, a hypothesis about the ratio of the means on the original scale can be translated into a hypothesis about the difference of two means on the logged scale.
Coefficient of Variation

The coefficient of variation (COV) is the ratio of the standard deviation to the mean. This parameter is used to represent the variation in the data because of a unique relationship that it has in the case of log-normal data.

Suppose the variable $X$ is the logarithm of the original variable $Y$. That is, $X = \ln(Y)$ and $Y = \exp(X)$. Label the mean and variance of $X$ as $\mu_X$ and $\sigma_X^2$, respectively. Similarly, label the mean and variance of $Y$ as $\mu_Y$ and $\sigma_Y^2$, respectively. If $X$ is normally distributed, then $Y$ is log-normally distributed. Julious (2004) presents the following well-known relationships between these two variables

$$
\mu_Y = \left(e^{\mu_X} + \frac{\sigma_X^2}{2}\right)
$$

$$
\sigma_Y^2 = \mu_Y \left(e^{\sigma_X^2} - 1\right)
$$

From this relationship, the coefficient of variation of $Y$ can be expressed as

$$
COV_Y = \frac{\sqrt{\mu_Y^2 \left(e^{\sigma_X^2} - 1\right)}}{\mu_Y}
$$

$$
= \sqrt{e^{\sigma_X^2} - 1}
$$

$$
= \sqrt{e^{\sigma_Y^2} - 1}
$$

where $\sigma_w^2$ is the within mean square error from the analysis of variance of the logged data.

Solving this relationship for $\sigma_X^2$, the standard deviation of $X$ can be stated in terms of the coefficient of variation of $Y$. This equation is

$$
\sigma_X = \sqrt{\ln\left(COV_Y^2 + 1\right)}
$$

Similarly, the mean of $X$ is

$$
\mu_X = \frac{\mu_Y}{\ln\left(COV_Y^2 + 1\right)}
$$

Thus, the hypotheses can be stated in the original (Y) scale and then power can be analyzed in the transformed (X) scale.

Power Calculation

As is shown above, the hypotheses can be stated in the original (Y) scale using ratios or the logged (X) scale using differences. Either way, the power and sample size calculations are made using the formulas for testing the equivalence of the difference in two means. These formulas are presented in another chapter and are not duplicated here.
Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains the parameters associated with this test such as the means, sample sizes, alpha, and power.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for from the other parameters. Under most situations, you will select either Power and Beta for a power analysis or N for sample size determination.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of non-superiority when in fact the treatment mean is superior.

Values must be between zero and one. Historically, the value of 0.20 was often used for beta. Recently, the standard has shifted to 0.10.

Power is defined as one minus beta. Power is equal to the probability of rejecting a false null hypothesis. Hence, specifying the beta error level also specifies the power level. For example, if you specify beta values of 0.05, 0.10, and 0.20, you are specifying the corresponding power values of 0.95, 0.90, and 0.80, respectively.

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when rejecting the null hypothesis of non-superiority when in fact the treatment group is not superior to the reference group.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.
Sample Size

N (Total Sample Size)
This option specifies one or more values of the sample size, the number of individuals in the study (total subjects in both sequences). This value must be an integer greater than one.

When N is even, it is split evenly between the two sequences. When N is odd, the first sequence has one more subject than the second sequence.

Note that you may enter a list of values using the syntax 50,100,150,200,250 or 50 to 250 by 50.

Effect Size – Ratios

SM (Superiority Margin)
This is the magnitude of the margin of superiority. It must be entered as a positive number.

When higher means are better, this value is the distance above one that is required for the mean ratio (Treatment Mean / Reference Mean) to be considered superior. When higher means are worse, this value is the distance below one that is required for the mean ratio (Treatment Mean / Reference Mean) to be considered superior.

R1 (True Ratio)
This is the value of the ratio of the two means (Treatment Mean / Reference Mean) at which the power is to be calculated.

When higher means are better, this value should be greater than 1+SM. When higher means are worse, this value should be less than 1-SM.

Effect Size – Coefficient of Variation

COV (Coefficient of Variation)
The coefficient of variation is the ratio of the standard deviation and the mean (SD/Mean). It is used to specify the variability (standard deviation). Note that this COV is defined on the original (not logarithmic) scale. This value must be determined from past experience or from a pilot study.

To be clear, consider the following definition. Suppose data on a response variable Y are collected. This procedure assumes that the values of X = Ln(Y) are analyzed using an appropriate ANOVA procedure. Thus, there are two sets of means and standard deviations: those of X labelled MX and SX and those of Y labelled MY and SY. The COV entered here is the COV of Y = SY/MY. For log-normal data, the following relationship exists: COV(Y) = SQR(Exp(SX*SX)-1) where SX is the square root of the within mean square error in the ANOVA table of the log-transformed values.

Test

Higher Means Are
This option defines whether higher values of the response variable are to be considered better or worse.

The choice here determines the direction of the test.
If Higher Means Are Better the null hypothesis is $R \leq 1 + SM$ and the alternative hypothesis is $R > 1 + SM$. If Higher Means Are Worse the null hypothesis is $R \geq 1 - SM$ and the alternative hypothesis is $R < 1 - SM$.

### Example 1 – Finding Power

A company has developed a generic drug for treating rheumatism and wants to show that it is superior to the standard drug by a certain amount. A 2x2 cross-over design will be used to test the superiority of the treatment drug to the reference drug.

Researchers have decided to set the margin of superiority to 0.20. Past experience leads the researchers to set the COV to 1.50. The significance level is 0.05. The power will be computed assuming that the true ratio is 1.40. Sample sizes between 50 and 550 will be included in the analysis.

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means in a 2x2 Cross-Over Design (Non-Unity Null) [Ratios] procedure window by expanding Means, then 2x2 Cross-Over Design, then clicking on Test (Non-Zero Null), and then clicking on Tests for Two Means in a 2x2 Cross-Over Design (Non-Unity Null) [Ratios]. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>50 to 550 by 100</td>
</tr>
<tr>
<td>SM (Superiority Margin)</td>
<td>0.20</td>
</tr>
<tr>
<td>R1 (True Ratio)</td>
<td>1.40</td>
</tr>
<tr>
<td>COV (Coefficient of Variation)</td>
<td>1.50</td>
</tr>
<tr>
<td>Higher Means Are</td>
<td>Better</td>
</tr>
</tbody>
</table>
Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>Superiority Margin (SM)</th>
<th>Superiority Bound (SB)</th>
<th>Actual Ratio (R1)</th>
<th>Significance Level (Alpha)</th>
<th>Beta</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1724</td>
<td>50</td>
<td>0.2000</td>
<td>1.2000</td>
<td>1.4000</td>
<td>0.0500</td>
<td>0.8276</td>
<td>1.5000</td>
</tr>
<tr>
<td>0.3369</td>
<td>150</td>
<td>0.2000</td>
<td>1.2000</td>
<td>1.4000</td>
<td>0.0500</td>
<td>0.6631</td>
<td>1.5000</td>
</tr>
<tr>
<td>0.4754</td>
<td>250</td>
<td>0.2000</td>
<td>1.2000</td>
<td>1.4000</td>
<td>0.0500</td>
<td>0.5246</td>
<td>1.5000</td>
</tr>
<tr>
<td>0.5909</td>
<td>350</td>
<td>0.2000</td>
<td>1.2000</td>
<td>1.4000</td>
<td>0.0500</td>
<td>0.4091</td>
<td>1.5000</td>
</tr>
<tr>
<td>0.6850</td>
<td>450</td>
<td>0.2000</td>
<td>1.2000</td>
<td>1.4000</td>
<td>0.0500</td>
<td>0.3150</td>
<td>1.5000</td>
</tr>
<tr>
<td>0.7602</td>
<td>550</td>
<td>0.2000</td>
<td>1.2000</td>
<td>1.4000</td>
<td>0.0500</td>
<td>0.2398</td>
<td>1.5000</td>
</tr>
</tbody>
</table>

Report Definitions

H0 (null hypothesis) is \( R \leq 1+SM \), where \( R = \frac{\text{Treatment Mean}}{\text{Reference Mean}} \).
H1 (alternative hypothesis) is \( R > 1+SM \).
Power is the probability of rejecting H0 when it is false.
N is the total sample size drawn from all sequences. The sample is divided equally among sequences.
SM is the magnitude of the margin of superiority. Since higher means are better, this value is positive and is the distance above one that is required to be considered superior.
SB is the corresponding bound to the superiority margin, and equals \( 1 + SM \).
R1 is the mean ratio (treatment/reference) at which the power is computed.
Alpha is the probability of falsely rejecting H0.
Beta is the probability of not rejecting H0 when it is false.
COV is the coefficient of variation on the original scale.

Summary Statements

A total sample size of 50 achieves 17% power to detect superiority using a one-sided t-test when the margin of superiority is 0.2000, the true mean ratio is 1.4000, the significance level is 0.0500, and the coefficient of variation on the original, unlogged scale is 1.5000. A 2x2 cross-over design with an equal number in each sequence is used.

This report shows the power for the indicated scenarios. Note that if they want 80% power, they will require a sample of more than 450 subjects.
Plot Section

This plot shows the power versus the sample size.

Example 2 – Validation

This procedure uses the same mechanics as the Non-Inferiority Tests for Two Means in a 2x2 Cross-Over Design using Ratios procedure. We refer the user to Example 2 of Chapter 515 for the validation.
Chapter 515

Non-Inferiority Tests for Two Means in a 2x2 Cross-Over Design using Ratios

Introduction

This procedure calculates power and sample size of statistical tests for non-inferiority tests from a 2x2 cross-over design. This routine deals with the case in which the statistical hypotheses are expressed in terms mean ratios rather than mean differences.

The details of testing the non-inferiority of two treatments using data from a 2x2 cross-over design are given in another chapter and they will not be repeated here. If the logarithms of the responses can be assumed to follow the normal distribution, hypotheses about non-inferiority stated in terms of the ratio can be transformed into hypotheses about the difference. The details of this analysis are given in Julious (2004). They will only be summarized here.
Non-Inferiority Testing Using Ratios

It will be convenient to adopt the following specialized notation for the discussion of these tests.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PASS Input/Output</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_T$</td>
<td>Not used</td>
<td>Treatment mean. This is the treatment mean.</td>
</tr>
<tr>
<td>$\mu_R$</td>
<td>Not used</td>
<td>Reference mean. This is the mean of a reference population.</td>
</tr>
<tr>
<td>$M_{NI}$</td>
<td>NIM</td>
<td>Margin of non-inferiority. This is a tolerance value that defines the maximum amount that is not of practical importance. This is the largest change in the mean ratio from the baseline value (usually one) that is still considered to be trivial.</td>
</tr>
<tr>
<td>$\phi$</td>
<td>RI</td>
<td>True ratio. This is the value of $\phi = \frac{\mu_T}{\mu_R}$ at which the power is calculated.</td>
</tr>
</tbody>
</table>

Note that the actual values of $\mu_T$ and $\mu_R$ are not needed. Only their ratio is needed for power and sample size calculations.

The null hypothesis of inferiority is

$$H_0: \phi \leq \phi_L \quad \text{where} \quad \phi_L < 1.$$  

and the alternative hypothesis of non-inferiority is

$$H_1: \phi > \phi_L$$

Log Transformation

In many cases, hypotheses stated in terms of ratios are more convenient than hypotheses stated in terms of differences. This is because ratios can be interpreted as scale-less percentages, but differences must be interpreted as actual amounts in their original scale. Hence, it has become a common practice to take the following steps in hypothesis testing.

1. State the statistical hypotheses in terms of ratios.
2. Transform these into hypotheses about differences by taking logarithms.
3. Analyze the logged data—that is, do the analysis in terms of the difference.
4. Draw the conclusion in terms of the ratio.

The details of step 2 for the null hypothesis are as follows.

$$\phi_L \leq \phi$$

$$\Rightarrow \phi_L \leq \left\{ \frac{\mu_T}{\mu_R} \right\}$$

$$\Rightarrow \ln(\phi_L) \leq \left\{ \ln(\mu_T) - \ln(\mu_R) \right\}$$

Thus, a hypothesis about the ratio of the means on the original scale can be translated into a hypothesis about the difference of two means on the logged scale.
Coefficient of Variation

The coefficient of variation (COV) is the ratio of the standard deviation to the mean. This parameter is used to represent the variation in the data because of a unique relationship that it has in the case of log-normal data.

Suppose the variable $X$ is the logarithm of the original variable $Y$. That is, $X = \ln(Y)$ and $Y = \exp(X)$. Label the mean and variance of $X$ as $\mu_X$ and $\sigma_X^2$, respectively. Similarly, label the mean and variance of $Y$ as $\mu_Y$ and $\sigma_Y^2$, respectively. If $X$ is normally distributed, then $Y$ is log-normally distributed. Julious (2004) presents the following well-known relationships between these two variables

$$\mu_Y = \left( e^{\mu_X + \frac{\sigma_X^2}{2}} \right)$$

$$\sigma_Y^2 = \mu_Y^2 \left( e^{\sigma_X^2} - 1 \right)$$

From this relationship, the coefficient of variation of $Y$ can be expressed as

$$COV_Y = \frac{\sqrt{\mu_Y^2 \left( e^{\sigma_X^2} - 1 \right)}}{\mu_Y}$$

$$= \sqrt{e^{\sigma_X^2} - 1}$$

$$= \sqrt{e^2 - 1}$$

where $\sigma_w^2$ is the within mean square error from the analysis of variance of the logged data.

Solving this relationship for $\sigma_X^2$, the standard deviation of $X$ can be stated in terms of the coefficient of variation of $Y$. This equation is

$$\sigma_X = \sqrt{\ln \left( COV_Y^2 + 1 \right)}$$

Similarly, the mean of $X$ is

$$\mu_X = \frac{\mu_Y}{\ln \left( COV_Y^2 + 1 \right)}$$

Thus, the hypotheses can be stated in the original $(Y)$ scale and then power can be analyzed in the transformed $(X)$ scale.

Power Calculation

As is shown above, the hypotheses can be stated in the original $(Y)$ scale using ratios or the logged $(X)$ scale using differences. Either way, the power and sample size calculations are made using the formulas for testing the equivalence of the difference in two means. These formulas are presented in another chapter and are not duplicated here.
Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains the parameters associated with this test such as the means, sample sizes, alpha, and power.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for from the other parameters. Under most situations, you will select either Power and Beta for a power analysis or N for sample size determination.

Error Rates

Power or Beta

This option specifies one or more values for the probability of a type-II error (beta). A type-II error occurs when you fail to reject the null hypothesis of inferiority when in fact the treatment mean is non-inferior.

Values must be between zero and one. Historically, the value of 0.20 was often used for beta. Recently, the standard has shifted to 0.10.

Power is defined as one minus beta. Power is equal to the probability of rejecting a false null hypothesis. Hence, specifying the beta error level also specifies the power level. For example, if you specify beta values of 0.05, 0.10, and 0.20, you are specifying the corresponding power values of 0.95, 0.90, and 0.80, respectively.

Alpha (Significance Level)

Specify one or more values of alpha. Alpha is the probably of a type-I error. A type-I error occurs when you reject the null hypothesis of inferiority when in fact the treatment group is not inferior to the reference group.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.

Sample Size

N (Total Sample Size)

This option specifies one or more values of the sample size, the number of individuals in the study (total subjects in both sequences). This value must be an integer greater than one.
When N is even, it is split evenly between the two sequences. When N is odd, the first sequence has one more subject than the second sequence.

Note that you may enter a list of values using the syntax 50,100,150,200,250 or 50 to 250 by 50.

---

**Effect Size – Ratios**

**NIM (Non-Inferiority Margin)**
This is the magnitude of the margin of non-inferiority. It must be entered as a positive number.

When higher means are better, this value is the distance below one for which the mean ratio (Treatment Mean / Reference Mean) still indicates non-inferiority of the treatment mean. E.g., a value of 0.2 here specifies that mean ratios greater than 0.8 indicate non-inferiority of the treatment mean.

When higher means are worse, this value is the distance above one for which the mean ratio (Treatment Mean / Reference Mean) still indicates non-inferiority of the treatment mean. E.g., a value of 0.2 here specifies that mean ratios less than 1.2 indicate non-inferiority of the treatment mean.

Recommended values: 0.20 is a common value for the parameter.

**R1 (True Ratio)**
This is the value of the ratio of the two means (Treatment Mean / Reference Mean) at which the power is to be calculated.

Often, the ratio will be set to one. However, some authors recommend using a ratio slightly different than one, such as 0.95 (when higher means are "better") or 1.05 (when higher means are "worse"), since this will require a larger sample size.

---

**Effect Size – Coefficient of Variation**

**COV (Coefficient of Variation)**
The coefficient of variation is the ratio of the standard deviation and the mean (SD/Mean). It is used to specify the variability (standard deviation). Note that this COV is defined on the original (not logarithmic) scale. This value must be determined from past experience or from a pilot study.

To be clear, consider the following definition. Suppose data on a response variable Y are collected. This procedure assumes that the values of X = Ln(Y) are analyzed using an appropriate ANOVA procedure. Thus, there are two sets of means and standard deviations: those of X labelled MX and SX and those of Y labelled MY and SY. The COV entered here is the COV of Y = SY/MY. For log-normal data, the following relationship exists: COV(Y) = SQR(Exp(SX*SX)-1) where SX is the square root of the within mean square error in the ANOVA table of the log-transformed values.

---

**Test**

**Higher Means Are**
This option defines whether higher values of the response variable are to be considered better or worse.

The choice here determines the direction of the non-inferiority test.
If Higher Means Are Better the null hypothesis is \( R \leq 1 \cdot \text{NIM} \) and the alternative hypothesis is \( R > 1 \cdot \text{NIM} \). If Higher Means Are Worse the null hypothesis is \( R \geq 1 + \text{NIM} \) and the alternative hypothesis is \( R < 1 + \text{NIM} \).

### Example 1 – Finding Power

A company has developed a generic drug for treating rheumatism and wants to show that it is not inferior to standard drug. A 2x2 cross-over design will be used to test the non-inferiority of the treatment drug to the reference drug.

Researchers have decided to set the margin of non-inferiority to 0.20. Past experience leads the researchers to set the COV to 1.50. The significance level is 0.05. The power will be computed assuming that the true ratio is one. Sample sizes between 50 and 550 will be included in the analysis.

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Non-Inferiority Tests for Two Means in a 2x2 Cross-Over Design [Ratios] procedure window by expanding Means, then 2x2 Cross-Over Design, then clicking on Non-Inferiority, and then clicking on Non-Inferiority Tests for Two Means in a 2x2 Cross-Over Design [Ratios]. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>50 to 550 by 100</td>
</tr>
<tr>
<td>NIM (Non-Inferiority Margin)</td>
<td>0.20</td>
</tr>
<tr>
<td>R1 (True Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>COV (Coefficient of Variation)</td>
<td>1.50</td>
</tr>
<tr>
<td>Higher Means Are</td>
<td>Better</td>
</tr>
</tbody>
</table>
Non-Inferiority Tests for Two Means in a 2x2 Cross-Over Design using Ratios

Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>Margin (NIM)</th>
<th>Bound (NIB)</th>
<th>Actual Ratio (R1)</th>
<th>Significance Level (Alpha)</th>
<th>Beta</th>
<th>COV</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2638</td>
<td>50</td>
<td>0.2000</td>
<td>0.8000</td>
<td>1.0000</td>
<td>0.0500</td>
<td>0.7362</td>
</tr>
<tr>
<td>0.5505</td>
<td>150</td>
<td>0.2000</td>
<td>0.8000</td>
<td>1.0000</td>
<td>0.0500</td>
<td>0.4495</td>
</tr>
<tr>
<td>0.7411</td>
<td>250</td>
<td>0.2000</td>
<td>0.8000</td>
<td>1.0000</td>
<td>0.0500</td>
<td>0.2589</td>
</tr>
<tr>
<td>0.8574</td>
<td>350</td>
<td>0.2000</td>
<td>0.8000</td>
<td>1.0000</td>
<td>0.0500</td>
<td>0.1426</td>
</tr>
<tr>
<td>0.9241</td>
<td>450</td>
<td>0.2000</td>
<td>0.8000</td>
<td>1.0000</td>
<td>0.0500</td>
<td>0.0759</td>
</tr>
<tr>
<td>0.9607</td>
<td>550</td>
<td>0.2000</td>
<td>0.8000</td>
<td>1.0000</td>
<td>0.0500</td>
<td>0.0393</td>
</tr>
</tbody>
</table>

Report Definitions

H0 (null hypothesis) is R <= 1-NIM, where R = Treatment Mean / Reference Mean.

H1 (alternative hypothesis) is R > 1-NIM.

Power is the probability of rejecting H0 when it is false.

N is the total sample size drawn from all sequences. The sample is divided equally among sequences.

-NIM is the magnitude and direction of the margin of non-inferiority. Since higher means are better, this value is negative and is the distance below one that is still considered non-inferior.

NIB is the corresponding bound to the non-inferiority margin, and equals 1 - NIM.

R1 is the mean ratio (treatment/reference) at which the power is computed.

Alpha is the probability of falsely rejecting H0 when it is false.

Beta is the probability of not rejecting H0 when it is false.

COV is the coefficient of variation on the original scale.

Summary Statements

A total sample size of 50 achieves 26% power to detect non-inferiority using a one-sided t-test when the margin of non-inferiority is -0.2000, the true mean ratio is 1.0000, the significance level is 0.0500, and the coefficient of variation on the original, unlogged scale is 1.5000. A 2x2 cross-over design with an equal number in each sequence is used.

This report shows the power for the indicated scenarios. Note that if they want 90% power, they will require a sample of around 450 subjects.

Plot Section

This plot shows the power versus the sample size.
## Example 2 – Validation

We could not find a validation example for this procedure in the statistical literature. Therefore, we will show that this procedure gives the same results as the non-inferiority test on differences—a procedure that has been validated. We will use the same settings as those given in Example 1. Since the output for this example is shown above, only the output from the procedure that uses differences is shown below.

To run the inferiority test on differences, we need the values of $|E|$ and $Sw$.

$$Sw = \sqrt{\ln(COV^2 + 1)}$$

$$= \sqrt{\ln(1.5^2 + 1)}$$

$$= 1.085659$$

$$E = \sqrt{\ln(1 - E)}$$

$$= \sqrt{\ln(0.8)}$$

$$= 0.223144$$

## Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Non-Inferiority Tests for Two Means in a 2x2 Cross-Over Design [Differences] procedure window by expanding Means, then 2x2 Cross-Over Design, then clicking on Non-Inferiority, and then clicking on Non-Inferiority Tests for Two Means in a 2x2 Cross-Over Design [Differences]. You may then make the appropriate entries as listed below, or open Example 1a by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>50 to 550 by 100</td>
</tr>
<tr>
<td>NIM (Non-Inferiority Margin)</td>
<td>0.223144</td>
</tr>
<tr>
<td>D (True Difference)</td>
<td>0</td>
</tr>
<tr>
<td>Specify S as</td>
<td>Sw</td>
</tr>
<tr>
<td>S</td>
<td>1.085659</td>
</tr>
<tr>
<td>Higher Means Are</td>
<td>Better</td>
</tr>
</tbody>
</table>
Output
Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>N</th>
<th>Non-Infer. Margin (-NIM)</th>
<th>Actual Difference (D)</th>
<th>Significance Level (Alpha)</th>
<th>Beta</th>
<th>Standard Deviation (Sw)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2638</td>
<td>50</td>
<td>-0.223</td>
<td>0.000</td>
<td>0.0500</td>
<td>0.7362</td>
<td>1.086</td>
</tr>
<tr>
<td>0.5505</td>
<td>150</td>
<td>-0.223</td>
<td>0.000</td>
<td>0.0500</td>
<td>0.4495</td>
<td>1.086</td>
</tr>
<tr>
<td>0.7411</td>
<td>250</td>
<td>-0.223</td>
<td>0.000</td>
<td>0.0500</td>
<td>0.2589</td>
<td>1.086</td>
</tr>
<tr>
<td>0.8574</td>
<td>350</td>
<td>-0.223</td>
<td>0.000</td>
<td>0.0500</td>
<td>0.1426</td>
<td>1.086</td>
</tr>
<tr>
<td>0.9242</td>
<td>450</td>
<td>-0.223</td>
<td>0.000</td>
<td>0.0500</td>
<td>0.0758</td>
<td>1.086</td>
</tr>
<tr>
<td>0.9607</td>
<td>550</td>
<td>-0.223</td>
<td>0.000</td>
<td>0.0500</td>
<td>0.0393</td>
<td>1.086</td>
</tr>
</tbody>
</table>

You can compare these power values with those shown above in Example 1 to validate the procedure. You will find that the power values are identical.
515-10 Non-Inferiority Tests for Two Means in a 2x2 Cross-Over Design using Ratios
Chapter 520

Equivalence Tests for Two Means in a 2x2 Cross-Over Design using Differences

Introduction

This procedure calculates power and sample size of statistical tests of equivalence of the means of a 2x2 cross-over design which is analyzed with a t-test. Schuirmann’s (1987) two one-sided tests (TOST) approach is used to test equivalence. Only a brief introduction to the subject will be given here. For a comprehensive discussion on the subject, refer to Chow and Liu (1999) and Julious (2004).

Measurements are made on individuals that have been randomly assigned to one of two sequences. The first sequence receives the treatment followed by the reference (AB). The second sequence receives the reference followed by the treatment (BA). This cross-over design may be analyzed by a TOST equivalence test to show that the two means do not differ by more than a small amount, called the margin of equivalence.

The definition of equivalence has been refined in recent years using the concepts of prescribability and switchability. Prescribability refers to ability of a physician to prescribe either of two drugs at the beginning of the treatment. However, once prescribed, no other drug can be substituted for it. Switchability refers to the ability of a patient to switch from one drug to another during treatment without adverse effects. Prescribability is associated with equivalence of location and variability. Switchability is associated with the concept of individual equivalence. This procedure analyzes average equivalence. Thus, it partially analyzes prescribability. It does not address equivalence of variability or switchability.
Cross-Over Designs

Senn (2002) defines a *cross-over* design as one in which each subject receives all treatments and the objective is to study differences among the treatments. The name *cross-over* comes from the most common case in which there are only two treatments. In this case, each subject *crosses over* from one treatment to the other. It is assumed that there is a *washout* period between treatments during which the response returns back to its baseline value. If this does not occur, there is said to be a *carryover* effect.

A 2x2 cross-over design contains two *sequences* (treatment orderings) and two time periods (occasions). One sequence receives treatment A followed by treatment B. The other sequence receives B and then A. The design includes a washout period between responses to make certain that the effects of the first drug do no carry over to the second. Thus, the groups in this design are defined by the sequence in which the drugs are administered, not by the treatments they receive. Indeed, higher-order cross-over designs have been used in which the same treatment is used at both occasions.

Cross-over designs are employed because, if the no-carryover assumption is met, treatment differences are measured within a subject rather than between subjects—making a more precise measurement. Examples of the situations that might use a cross-over design are the comparison of anti-inflammatory drugs in arthritis and the comparison of hypotensive agents in essential hypertension. In both of these cases, symptoms are expected to return to their usual baseline level shortly after the treatment is stopped.

Advantages of Cross-Over Designs

A comparison of treatments on the same subject is expected to be more precise. The increased precision often translates into a smaller sample size. Also, patient enrollment into the study may be easier because each patient will receive both treatments. Finally, it is often more difficult to obtain a subject than to obtain a measurement.

Disadvantages of Cross-Over Designs

The statistical analysis of a cross-over experiment is more complex than a parallel-group experiment and requires additional assumptions. It may be difficult to separate the treatment effect from the period effect, the carry-over effect of the previous treatment, and the interaction between period and treatment.

The design cannot be used when the treatment (or the measurement of the response) alters the subject permanently. Hence, it should not be used to compare treatments that are intended to provide a cure.

Because subjects must be measured at least twice, it is often more difficult to keep patients enrolled in the study. It is arguably simpler to measure a subject once than to obtain their measurement twice. This is particularly true when the measurement process is painful, uncomfortable, embarrassing, or time consuming.
Outline of an Equivalence Test

PASS follows the two one-sided tests approach described by Schuirmann (1987) and Phillips (1990). It will be convenient to adopt the following specialize notation for the discussion of these tests.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PASS Input/Output</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu_T )</td>
<td>Not used</td>
<td>Treatment mean. This is the treatment mean.</td>
</tr>
<tr>
<td>( \mu_R )</td>
<td>Not used</td>
<td>Reference mean. This is the mean of a reference population.</td>
</tr>
<tr>
<td>( \epsilon )</td>
<td>(</td>
<td>E</td>
</tr>
<tr>
<td>( \delta )</td>
<td>D</td>
<td>True difference. This is the value of ( \mu_T - \mu_R ), the difference between the treatment and reference means. This is the value at which the power is calculated.</td>
</tr>
</tbody>
</table>

Note that the actual values of \( \mu_T \) and \( \mu_R \) are not needed. Only their difference is needed for power and sample size calculations.

The null hypothesis of non-equivalence is

\[
H_0: \delta \leq \epsilon_L \text{ or } \delta \geq \epsilon_U, \text{ where } \epsilon_L < 0, \epsilon_U > 0.
\]

and the alternative hypothesis of equivalence is

\[
H_1: \epsilon_L < \delta < \epsilon_U.
\]

Test Statistics

This section describes the test statistic that is used to perform the hypothesis test.

T-Test

A t-test is used to analyze the data. The test assumes that the data are a simple random sample from a population of normally-distributed values that have the same variance. This assumption implies that the differences are continuous and normal. The calculation of the two, one-sided t-tests proceeds as follow

\[
T_L = \frac{(\bar{x}_T - \bar{x}_R) - \epsilon_L}{\hat{\sigma}_w \sqrt{\frac{2}{N}}}, \quad T_U = \frac{(\bar{x}_T - \bar{x}_R) - \epsilon_U}{\hat{\sigma}_w \sqrt{\frac{2}{N}}}
\]

where \( \hat{\sigma}_w^2 \) is the within mean square error from the appropriate ANOVA table.
The significance of each test statistic is determined by computing the p-value. If this p-value is less than a specified level (usually 0.05), the null hypothesis is rejected.

If prior studies used a t-test rather than an ANOVA to analyze the data, you may not have a direct estimate of $\sigma_w^2$. Instead, you will have an estimate of the variance of the period differences from the t-test, $\hat{\sigma}_d^2$. These variances are functionally related by $\sigma_w^2 = 2\sigma_d^2$. Either variance can be entered.

### Power Calculation

The power of this test is given by

$$\Pr(T_L \geq t_{1-\alpha,N-2} \text{ and } T_U \leq -t_{1-\alpha,N-2})$$

where $T_L$ and $T_U$ are distributed as the bivariate, noncentral $t$ distribution with noncentrality parameters $\Delta_L$ and $\Delta_U$ given by

$$\Delta_L = \frac{\delta - e_L}{\sigma_w \sqrt{\frac{2}{N}}}$$

$$\Delta_U = \frac{\delta - e_U}{\sigma_w \sqrt{\frac{2}{N}}}$$

### Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

### Data Tab

The Data tab contains the parameters associated with this test such as the means, sample sizes, alpha, and beta.

### Solve For

**Find (Solve For)**

This option specifies the parameter to be solved for from the other parameters. Under most situations, you will select either Power and Beta for a power analysis or $N$ for sample size determination.

Select $N$ when you want to calculate the sample size needed to achieve a given power and alpha level.
Select *Power and Beta* when you want to calculate the power of an experiment that has already been run.

### Error Rates

#### Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of unequal means when in fact the means are equal.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as *0.8 to 0.95 by 0.05* may be entered.

#### Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when you reject the null hypothesis of unequal means when in fact the means are unequal.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as *0.01 0.05 0.10* or *0.01 to 0.10 by 0.01*.

### Sample Size

#### N (Total Sample Size)

This option specifies one or more values of the sample size, the number of individuals in the study (total subjects in both sequences). This value must be an integer greater than one.

When N is even, it is split evenly between the two sequences. When N is odd, the first sequence has one more subject than the second sequence.

Note that you may enter a list of values using the syntax *50,100,150,200,250* or *50 to 250 by 50*.

### Effect Size – Equivalence Limits

#### |EU| (Upper Equivalence Limit)

This value gives upper limit on equivalence. Differences outside EL and EU are not considered equivalent. Differences between them are considered equivalent.

Note that EL<0 and EU>0. Also, you must have EL<0<EU.

#### -|EL| (Lower Equivalence Limit)

This value gives lower limit on equivalence. Differences outside EL and EU are not considered equivalent. Differences between them are.
If you want symmetric limits, enter -UPPER LIMIT for EL to force $EL = -|EU|$. Note that $EL < 0$ and $EU > 0$. Also, you must have $EL < D < EU$. Finally, the scale of these numbers must match the scale of $S$.

**Effect Size – True Mean Difference**

**D (True Difference)**
This is the true difference between the two means at which the power is to be computed. Often this value is set to zero, but it can be non-zero as long as it is between the equivalence limits $EL$ and $EU$.

**Effect Size – Standard Deviation**

**Specify S as Sw or Sd**
Specify the form of the standard deviation that is entered in the box below.

- **Sw**
  Specify $S$ as the square root of the within mean square error from a repeated measures ANOVA. This is the most common method since cross-over designs are usually analyzed using ANOVA.

- **Sd**
  Specify $S$ as the standard deviation of the individual treatment differences computed for each subject. This option is used when you have previous studies that produced this value.

**S (Value of Sw or Sd)**
Specify the value(s) of the standard deviation $S$. The interpretation of this value depends on the entry in **Specify S as Sw or Sd** above. If $S = Sw$ is selected, this is the value of $Sw$ which is $\text{SQR(WMSE)}$ where WMSE is the within mean square error from the ANOVA table used to analyze the Cross-Over design. If $S = Sd$ is selected, this is the value of $Sd$ which is the standard deviation of the period differences—pooled from both sequences.

These values must be positive. A list of values may be entered.

You can press the SD button to load the Standard Deviation Estimator window.
Example 1 – Finding Power

A cross-over design is to be used to compare the impact of two drugs on diastolic blood pressure. The average diastolic blood pressure after administration of the reference drug is known to be 96 mmHg. Researchers believe this average may drop to 92 mmHg with the use of a new drug. The within mean square error of similar studies is 324. Its square root is 18.

Following FDA guidelines, the researchers want to show that the diastolic blood pressure with the new drug is within 20% of the diastolic blood pressure with the reference drug. Thus, the equivalence limits of the mean difference of the two drugs are -19.2 and 19.2. They decide to calculate the power for a range of sample sizes between 6 and 100. The significance level is 0.05.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Equivalence Tests for Two Means in a 2x2 Cross-Over Design [Differences] procedure window by expanding Means, then 2x2 Cross-Over Design, then clicking on Equivalence, and then clicking on Equivalence Tests for Two Means in a 2x2 Cross-Over Design [Differences]. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>6 10 16 20 40 60 80 100</td>
</tr>
<tr>
<td></td>
<td>EU</td>
</tr>
<tr>
<td>-</td>
<td>EL</td>
</tr>
<tr>
<td>D (True Difference)</td>
<td>-4</td>
</tr>
<tr>
<td>Specify S as</td>
<td>Sw</td>
</tr>
<tr>
<td>S</td>
<td>18</td>
</tr>
</tbody>
</table>
Annotated Output
Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1470</td>
<td>6</td>
<td>-19.20</td>
<td>19.20</td>
<td>-4.00</td>
<td>18.00</td>
<td>0.0500</td>
<td>0.8530</td>
</tr>
<tr>
<td>0.3873</td>
<td>10</td>
<td>-19.20</td>
<td>19.20</td>
<td>-4.00</td>
<td>18.00</td>
<td>0.0500</td>
<td>0.6127</td>
</tr>
<tr>
<td>0.6997</td>
<td>16</td>
<td>-19.20</td>
<td>19.20</td>
<td>-4.00</td>
<td>18.00</td>
<td>0.0500</td>
<td>0.3003</td>
</tr>
<tr>
<td>0.8104</td>
<td>20</td>
<td>-19.20</td>
<td>19.20</td>
<td>-4.00</td>
<td>18.00</td>
<td>0.0500</td>
<td>0.1896</td>
</tr>
<tr>
<td>0.9804</td>
<td>40</td>
<td>-19.20</td>
<td>19.20</td>
<td>-4.00</td>
<td>18.00</td>
<td>0.0500</td>
<td>0.0196</td>
</tr>
<tr>
<td>0.9983</td>
<td>60</td>
<td>-19.20</td>
<td>19.20</td>
<td>-4.00</td>
<td>18.00</td>
<td>0.0500</td>
<td>0.0017</td>
</tr>
<tr>
<td>0.9999</td>
<td>80</td>
<td>-19.20</td>
<td>19.20</td>
<td>-4.00</td>
<td>18.00</td>
<td>0.0500</td>
<td>0.0001</td>
</tr>
<tr>
<td>1.0000</td>
<td>100</td>
<td>-19.20</td>
<td>19.20</td>
<td>-4.00</td>
<td>18.00</td>
<td>0.0500</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Report Definitions
Power is the probability of rejecting non-equivalence when the means are equivalent.
N is the total number of subjects split between both sequences.
EU & EL are the maximum allowable differences that still result in equivalence.
D is the difference between the means at which the power is computed.
Sw is the square root of the within mean square error from the ANOVA table.
Alpha is the probability of rejecting non-equivalence when the means are non-equivalent.
Beta is the probability of accepting non-equivalence when the means are equivalent.

Summary Statements
In an equivalence test of means using two one-sided tests on data from a two-period cross-over design, a total sample size of 6 achieves 15% power at a 5% significance level when the true difference between the means is -4.00, the square root of the within mean square error is 18.00, and the equivalence limits are -19.20 and 19.20.

This report shows the power for the indicated scenarios. Note that if they want 90% power, they will require a sample of around 30 subjects.

Plots Section
This plot shows the power versus the sample size.
Example 2 – Finding Sample Size

Continuing with Example 1, the researchers want to find the exact sample size needed to achieve both 80% power and 90% power.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Equivalence Tests for Two Means in a 2x2 Cross-Over Design [Differences] procedure window by expanding Means, then 2x2 Cross-Over Design, then clicking on Equivalence, and then clicking on Equivalence Tests for Two Means in a 2x2 Cross-Over Design [Differences]. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N</td>
</tr>
<tr>
<td>Power</td>
<td>0.8 0.9</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td></td>
<td>EU</td>
</tr>
<tr>
<td>-</td>
<td>EL</td>
</tr>
<tr>
<td>D (True Difference)</td>
<td>-4</td>
</tr>
<tr>
<td>Specify S as</td>
<td>Sw</td>
</tr>
<tr>
<td>S</td>
<td>18</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>Total Sample Size</th>
<th>Lower Equiv Limit</th>
<th>Upper Equiv Limit</th>
<th>True Difference</th>
<th>Standard Deviation</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9032</td>
<td>26</td>
<td>-19.20</td>
<td>19.20</td>
<td>-4.00</td>
<td>18.00</td>
<td>0.0500</td>
<td>0.0968</td>
</tr>
<tr>
<td>0.8104</td>
<td>20</td>
<td>-19.20</td>
<td>19.20</td>
<td>-4.00</td>
<td>18.00</td>
<td>0.0500</td>
<td>0.1896</td>
</tr>
</tbody>
</table>

We note that 20 subjects are needed to achieve 80% power and 26 subjects are needed to achieve 90% power.
Example 3 – Validation using Phillips

Phillips (1990) page 142 presents a table of sample sizes for various parameter values. In this table, the treatment mean, standard deviation, and equivalence limits are all specified as percentages of the reference mean. We will reproduce the second line of the table in which the square root of the within mean square error is 20%; the equivalence limits are 20%; the treatment mean is 100%, 95%, 90%, and 85%; the power is 70%; and the significance level is 0.05. Phillips reports total sample size as 16, 20, 40, and 152 corresponding to the four treatment mean percentages. We will now setup this example in PASS.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Equivalence Tests for Two Means in a 2x2 Cross-Over Design [Differences] procedure window by expanding Means, then 2x2 Cross-Over Design, then clicking on Equivalence, and then clicking on Equivalence Tests for Two Means in a 2x2 Cross-Over Design [Differences]. You may then make the appropriate entries as listed below, or open Example 3 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For) ..................</td>
<td>N</td>
</tr>
<tr>
<td>Power</td>
<td>0.7</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Total Sample Size) ................</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td></td>
<td>EU</td>
</tr>
<tr>
<td>-</td>
<td>EL</td>
</tr>
<tr>
<td>D (True Difference) ................</td>
<td>0 -5 -10 -15</td>
</tr>
<tr>
<td>Specify S as .........................</td>
<td>Sw</td>
</tr>
<tr>
<td>S ..................................................</td>
<td>20</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>Total Sample Size (N)</th>
<th>Lower Equiv. Limit</th>
<th>Upper Equiv. Limit</th>
<th>True Difference</th>
<th>Standard Deviation</th>
<th>Sw</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.7001</td>
<td>152</td>
<td>-20.00</td>
<td>20.00</td>
<td>-15.00</td>
<td>20.00</td>
<td>0.0500</td>
<td>0.2999</td>
<td></td>
</tr>
<tr>
<td>0.7092</td>
<td>40</td>
<td>-20.00</td>
<td>20.00</td>
<td>-10.00</td>
<td>20.00</td>
<td>0.0500</td>
<td>0.2908</td>
<td></td>
</tr>
<tr>
<td>0.7221</td>
<td>20</td>
<td>-20.00</td>
<td>20.00</td>
<td>-5.00</td>
<td>20.00</td>
<td>0.0500</td>
<td>0.2779</td>
<td></td>
</tr>
<tr>
<td>0.7031</td>
<td>16</td>
<td>-20.00</td>
<td>20.00</td>
<td>0.00</td>
<td>20.00</td>
<td>0.0500</td>
<td>0.2969</td>
<td></td>
</tr>
</tbody>
</table>

Note that PASS has obtained the same samples sizes as Phillips (1990).
Example 4 – Validation using Machin

Machin et al. (1997) page 107 present an example of determining the sample size for a cross-over design in which the reference mean is 35.03, the treatment mean is 35.03, the standard deviation, entered as the square root of the within mean square error, is 40% of the reference mean, the limits are plus or minus 20% of the reference mean, the power is 80%, and the significance level is 0.10. Machin et al. calculate the total sample size to be 54.

When the parameters are given as percentages of the reference mean, it is easy enough to calculate the exact amounts by applying those percentages. However, the percentages can all be entered directly as long as all parameters (EU, EL, D, and Sw) are specified as percentages.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Equivalence Tests for Two Means in a 2x2 Cross-Over Design [Differences] procedure window by expanding Means, then 2x2 Cross-Over Design, then clicking on Equivalence, and then clicking on Equivalence Tests for Two Means in a 2x2 Cross-Over Design [Differences]. You may then make the appropriate entries as listed below, or open Example 4 by going to the File menu and choosing Open Example Template.

Option | Value
--- | ---
Data Tab | 
Find (Solve For) | N
Power | 0.8
Alpha | 0.10
N (Total Sample Size) | Ignored since this is the Find setting
|EU| (Upper Equivalence Limit) | 20
|-|EL| (Lower Equivalence Limit) | -Upper Limit
D (True Difference) | 0
Specify S as | Sw
S | 40

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Total Sample Size</th>
<th>Lower Equiv. Limit</th>
<th>Upper Equiv. Limit</th>
<th>True Difference</th>
<th>Standard Deviation</th>
<th>Power</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>54</td>
<td>-20.00</td>
<td>20.00</td>
<td>0.00</td>
<td>40.00</td>
<td>0.8050</td>
<td>0.1000</td>
<td>0.1950</td>
</tr>
</tbody>
</table>

Note that PASS also has obtained a sample size of 54.
Example 5 – Validation using Chow and Liu

Chow and Liu (1999) page 153 present an example of determining the sample size for a cross-over design in which the reference mean is 82.559, the treatment mean is 82.559, the standard deviation, entered as the square root of the within mean square error, is 15.66%, the limits are plus or minus 20%, the power is 80%, and the significance level is 0.05. They calculate a sample size of 12. PASS calculates a sample size of 13. To see why PASS has increased the sample size by one, we will evaluate the power at sample sizes of 10, 12, 13, 14, and 16.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Equivalence Tests for Two Means in a 2x2 Cross-Over Design [Differences] procedure window by expanding Means, then 2x2 Cross-Over Design, then clicking on Equivalence, and then clicking on Equivalence Tests for Two Means in a 2x2 Cross-Over Design [Differences]. You may then make the appropriate entries as listed below, or open Example 5 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>10 12 13 14 16</td>
</tr>
<tr>
<td></td>
<td>20</td>
</tr>
<tr>
<td></td>
<td>Upper Limit</td>
</tr>
<tr>
<td></td>
<td>-Upper Limit</td>
</tr>
<tr>
<td>D (True Difference)</td>
<td>0</td>
</tr>
<tr>
<td>Specify S as</td>
<td>Sw</td>
</tr>
<tr>
<td>S</td>
<td>15.66</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>Total Sample Size (N)</th>
<th>Lower Limit</th>
<th>Upper Limit</th>
<th>True Difference</th>
<th>Standard Deviation Sw</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.6643</td>
<td>10</td>
<td>-20.00</td>
<td>20.00</td>
<td>0.00</td>
<td>15.66</td>
<td>0.0500</td>
<td>0.3357</td>
</tr>
<tr>
<td>0.7932</td>
<td>12</td>
<td>-20.00</td>
<td>20.00</td>
<td>0.00</td>
<td>15.66</td>
<td>0.0500</td>
<td>0.2068</td>
</tr>
<tr>
<td>0.8363</td>
<td>13</td>
<td>-20.00</td>
<td>20.00</td>
<td>0.00</td>
<td>15.66</td>
<td>0.0500</td>
<td>0.1637</td>
</tr>
<tr>
<td>0.8752</td>
<td>14</td>
<td>-20.00</td>
<td>20.00</td>
<td>0.00</td>
<td>15.66</td>
<td>0.0500</td>
<td>0.1248</td>
</tr>
<tr>
<td>0.9258</td>
<td>16</td>
<td>-20.00</td>
<td>20.00</td>
<td>0.00</td>
<td>15.66</td>
<td>0.0500</td>
<td>0.0742</td>
</tr>
</tbody>
</table>

The power for $N = 12$ is 0.7932. The power for $N = 13$ is 0.8363. Hence, to achieve better than 80% power, a sample size of 13 is necessary. However, 0.7932 is sufficiently close to 0.800 to make $N = 12$ a reasonable choice (as Chow and Liu did).
Example 6 – Validation using Senn

Senn (1993) page 217 presents an example of determining the sample size for a cross-over design in which the reference mean is equal to the treatment mean, the standard deviation, entered as the square root of the within mean square error, is 45, the equivalence limits are plus or minus 30, the power is 80%, and the significance level is 0.05. He calculates a sample size of 40.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for Two Means in a 2x2 Cross-Over Design [Differences]** procedure window by expanding **Means**, then **2x2 Cross-Over Design**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for Two Means in a 2x2 Cross-Over Design [Differences]**. You may then make the appropriate entries as listed below, or open **Example 6** by going to the **File** menu and choosing **Open Example Template**.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N</td>
</tr>
<tr>
<td>Power</td>
<td>0.8</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td></td>
<td>EU</td>
</tr>
<tr>
<td></td>
<td>EL</td>
</tr>
<tr>
<td>D (True Difference)</td>
<td>0</td>
</tr>
<tr>
<td>Specify S as</td>
<td>Sw</td>
</tr>
<tr>
<td>S</td>
<td>45</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

**Numeric Results**

<table>
<thead>
<tr>
<th>Power</th>
<th>Total Sample Size</th>
<th>Lower Equiv. Limit</th>
<th>Upper Equiv. Limit</th>
<th>True Difference</th>
<th>Standard Deviation</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.8004</td>
<td>40</td>
<td>-30.00</td>
<td>30.00</td>
<td>0.00</td>
<td>45.00</td>
<td>0.0500</td>
<td>0.1996</td>
</tr>
</tbody>
</table>

PASS also calculates a sample size of 40.
520-14 Equivalence Tests for Two Means in a 2x2 Cross-Over Design using Differences
Chapter 525

Equivalence Tests for Two Means in a 2x2 Cross-Over Design using Ratios

Introduction

This procedure calculates power and sample size of statistical tests of equivalence of the means from a 2x2 cross-over design which is analyzed with a t-test. This routine deals with the case in which the statistical hypotheses are expressed in terms mean of ratios rather than mean differences.

The details of testing the equivalence of two treatments using data from a 2x2 cross-over design are given in another chapter and will not be repeated here. If the logarithms of the responses can be assumed to follow the normal distribution, hypotheses about the equivalence of two means stated in terms of the ratio can be transformed into hypotheses about the difference. The details of this analysis are given in Julious (2004). They will only be summarized here.

Equivalence Testing Using Ratios

PASS follows the two one-sided tests approach described by Schuirmann (1987) and Phillips (1990). It will be convenient to adopt the following specialized notation for the discussion of these tests.
525-2 Equivalence Tests for Two Means in a 2x2 Cross-Over Design using Ratios

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PASS Input/Output</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu_T )</td>
<td>Not used</td>
<td>Treatment mean. This is the treatment mean.</td>
</tr>
<tr>
<td>( \mu_R )</td>
<td>Not used</td>
<td>Reference mean. This is the mean of a reference population.</td>
</tr>
<tr>
<td>( \phi_L, \phi_U )</td>
<td>RL, RU</td>
<td>Margin of equivalence. These limits that define an interval of the ratio of the means in which their difference is so small that it may be ignored.</td>
</tr>
<tr>
<td>( \phi )</td>
<td>R1</td>
<td>True ratio. This is the value of ( \phi = \frac{\mu_T}{\mu_R} ) at which the power is calculated.</td>
</tr>
</tbody>
</table>

Note that the actual values of \( \mu_T \) and \( \mu_R \) are not needed. Only their ratio is needed for power and sample size calculations.

The null hypothesis of non-equivalence is

\[
H_0: \phi \leq \phi_L \text{ or } \phi \geq \phi_U \text{ where } \phi_L < 1, \phi_U > 1.
\]

and the alternative hypothesis of equivalence is

\[
H_1: \phi_L < \phi < \phi_U
\]

Log-Transformation

In many cases, hypotheses stated in terms of ratios are more convenient than hypotheses stated in terms of differences. This is because ratios can be interpreted as scale-less percentages, but differences must be interpreted as actual amounts in their original scale. Hence, it has become a common practice to take the following steps in hypothesis testing.

1. State the statistical hypotheses in terms of ratios.
2. Transform these into hypotheses about differences by taking logarithms.
3. Analyze the logged data—that is, do the analysis in terms of the difference.
4. Draw the conclusion in terms of the ratio.

The details of step 2 for the null hypothesis are as follows.

\[
\phi_L \leq \phi \leq \phi_U
\]

\[
\Rightarrow \phi_L \leq \frac{\mu_T}{\mu_R} \leq \phi_U
\]

\[
\Rightarrow \ln(\phi_L) \leq \{\ln(\mu_T) - \ln(\mu_R)\} \leq \ln(\phi_U)
\]

Thus, a hypothesis about the ratio of the means on the original scale can be translated into a hypothesis about the difference of two means on the logged scale.

When performing an equivalence test on the difference between means, the usual procedure is to set the equivalence limits symmetrically above and below zero. Thus the equivalence limits will be plus or minus an appropriate amount. The common practice is to do the same when the data are being analyzed on the log scale. However, when symmetric limits are set on the log scale,
they do not translate to symmetric limits on the original scale. Instead, they translate to limits that are the inverses of each other.

Perhaps these concepts can best be understood by considering an example. Suppose the researchers have determined that the lower equivalence limit should be 80% on the original scale. Since they are planning to use a log scale for their analysis, they transform this limit to the log scale by taking the logarithm of 0.80. The result is -0.223144. Wanting symmetric limits, they set the upper equivalence limit to 0.223144. Exponentiating this value, they find that \( \exp(0.223144) = 1.25 \). Note that \( 1/(0.80) = 1.25 \). Thus, the limits on the original scale are 80% and 125%, not 80% and 120%.

Using this procedure, appropriate equivalence limits for the ratio of two means can be easily determined. Here are a few sets of equivalence limits.

<table>
<thead>
<tr>
<th>Specified Percent Change</th>
<th>Lower Limit Original Scale</th>
<th>Upper Limit Original Scale</th>
<th>Lower Limit Log Scale</th>
<th>Upper Limit Log Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>-25%</td>
<td>75.0%</td>
<td>133.3%</td>
<td>-0.287682</td>
<td>0.287682</td>
</tr>
<tr>
<td>+25%</td>
<td>80.0%</td>
<td>125.0%</td>
<td>-0.223144</td>
<td>0.223144</td>
</tr>
<tr>
<td>-20%</td>
<td>80.0%</td>
<td>125.0%</td>
<td>-0.223144</td>
<td>0.223144</td>
</tr>
<tr>
<td>+20%</td>
<td>83.3%</td>
<td>120.0%</td>
<td>-0.182322</td>
<td>0.182322</td>
</tr>
<tr>
<td>-10%</td>
<td>90.0%</td>
<td>111.1%</td>
<td>-0.105361</td>
<td>0.105361</td>
</tr>
<tr>
<td>+10%</td>
<td>90.9%</td>
<td>110.0%</td>
<td>-0.095310</td>
<td>0.095310</td>
</tr>
</tbody>
</table>

Note that negative percent-change values specify the lower limit first, while positive percent-change values specify the upper limit first. After the first limit is found, the other limit is calculated as its inverse.

### Coefficient of Variation

The coefficient of variation (COV) is the ratio of the standard deviation to the mean. This parameter can be used to represent the variation in the data because of a unique relationship that it has in the case of log-normal data.

Suppose the variable \( X \) is the logarithm of the original variable \( Y \). That is, \( X = \ln(Y) \) and \( Y = \exp(X) \). Label the mean and variance of \( X \) as \( \mu_X \) and \( \sigma_X^2 \), respectively. Similarly, label the mean and variance of \( Y \) as \( \mu_Y \) and \( \sigma_Y^2 \), respectively. If \( X \) is normally distributed, then \( Y \) is log-normally distributed. Julious (2004) presents the following well-known relationships between these two variables

\[
\begin{align*}
\mu_Y &= \left( e^{\mu_X + \frac{\sigma_X^2}{2}} \right) \\
\sigma_Y^2 &= \mu_Y \left( e^{\sigma_X^2} - 1 \right)
\end{align*}
\]

From this relationship, the coefficient of variation of \( Y \) can be expressed as
Equivalence Tests for Two Means in a 2x2 Cross-Over Design using Ratios

\[
COV_Y = \sqrt{\frac{\mu^2_y (e^{\sigma^2_Y} - 1)}{\mu_Y}}
\]

\[
= \sqrt{e^{\sigma^2_Y} - 1}
\]

Solving this relationship for \( \sigma^2_Y \), the standard deviation of \( X \) can be stated in terms of the coefficient of variation of \( Y \). This equation is

\[
\sigma_X = \sqrt{\ln(COV^2_Y + 1)}
\]

Similarly, the mean of \( X \) is

\[
\mu_X = \frac{\mu_y}{\ln(COV^2_Y + 1)}
\]

Thus, the hypotheses can be stated in the original (\( Y \)) scale and then the power can be analyzed in the transformed (\( X \)) scale.

**Power Calculation**

As is shown above, the hypotheses can be stated in the original (\( Y \)) scale using ratios or the logged (\( X \)) scale using differences. Either way, the power and sample size calculations are made using the formulas for testing the equivalence of the difference in two means. These formulas are presented in another chapter and are not duplicated here.

**Procedure Options**

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

**Data Tab**

The Data tab contains the parameters associated with this test such as the means, sample sizes, alpha, and power.

**Solve For**

**Find (Solve For)**

This option specifies the parameter to be solved for from the other parameters. Under most situations, you will select either Power and Beta for a power analysis or \( N \) for sample size determination.
Error Rates

Power or Beta
This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of unequal means when in fact the means are equal.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

Alpha (Significance Level)
This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when you reject the null hypothesis of unequal means when in fact the means are unequal.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.

Sample Size

N (Total Sample Size)
This option specifies one or more values of the sample size, the number of individuals in the study (total subjects in both sequences). This value must be an integer greater than one.

When N is even, it is split evenly between the two sequences. When N is odd, the first sequence has one more subject than the second sequence.

Note that you may enter a list of values using the syntax 50,100,150,200,250 or 50 to 250 by 50.

Effect Size – Equivalence Limits

RU (Upper Equivalence Limit)
Enter the upper equivalence limit for the ratio of the two means. When the ratio of the means is between this value and RL, the two means are said to be equivalent. The value must be greater than one. A popular choice is 1.25. Note that this value is not a percentage.

If you enter 1/RL, then 1/RL will be calculated and used here. This choice is commonly used because RL and 1/RL give limits that are of equal magnitude on the log scale.
RL (Lower Equivalence Limit)
Enter the lower equivalence limit for the ratio of the two means. When the ratio of the means is
between this value and RU, the two means are said to be equivalent. The value must be less than
one. A popular choice is 0.80. Note that this value is not a percentage.

If you enter $1/RU$, then $1/RU$ will be calculated and used here. This choice is commonly used
because RU and $1/RU$ give limits that are of equal magnitude on the log scale.

Effect Size – True Ratio

R1 (True Ratio)
This is the value of the ratio of the two means at which the power is to be calculated. Usually, the
ratio will be assumed to be one. However, some authors recommend calculating the power using
a ratio of 1.05 since this will require a larger sample size.

Effect Size – Coefficient of Variation

COV (Coefficient of Variation)
The coefficient of variation is used to specify the variability (standard deviation). It is important
to realize that this is the COV defined on the original (not logged) scale. This value must be
determined from past experience or from a pilot study. It is most easily calculated from the within
mean-square error of the analysis of variance using the logged data using the relationship

$$COV_y = \sqrt{e^{\sigma_d^2} - 1}.$$  

If prior studies used a t-test to analyze the logged data, you will not have a direct estimate of $\hat{\sigma}_w^2$.
However, the two variances, $\sigma_d^2$ and $\sigma_w^2$, are functionally related. The relationship between these
quantities is $\sigma_d^2 = 2\sigma_w^2$. 

Example 1 – Finding Power

A company has opened a new manufacturing plant and wants to show that the drug produced in the new plant is equivalent to that produced in an older plant. A cross-over design will be used to test the equivalence of drugs produced at the two plants.

Researchers have decided to set the equivalence limits for the ratio at 0.90 and 1.111 (note that 1.111 = 1/0.90). Past experience leads the researchers to set the COV to 0.50. The significance level is 0.05. The power will be computed assuming that the true ratio is one. Sample sizes between 50 and 550 will be included in the analysis.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Equivalence Tests for Two Means in a 2x2 Cross-Over Design [Ratios] procedure window by expanding Means, then 2x2 Cross-Over Design, then clicking on Equivalence, and then clicking on Equivalence Tests for Two Means in a 2x2 Cross-Over Design [Ratios]. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>50 to 550 by 100</td>
</tr>
<tr>
<td>RU (Upper Equivalence Limit)</td>
<td>1/RL</td>
</tr>
<tr>
<td>RL (Lower Equivalence Limit)</td>
<td>0.90</td>
</tr>
<tr>
<td>R1 (True Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>COV (Coefficient of Variation)</td>
<td>0.50</td>
</tr>
</tbody>
</table>
Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0000</td>
<td>50</td>
<td>0.9000</td>
<td>1.1111</td>
<td>1.1111</td>
<td>1.0000</td>
<td>0.5000</td>
<td>0.0500</td>
<td>1.0000</td>
</tr>
<tr>
<td>0.2190</td>
<td>150</td>
<td>0.9000</td>
<td>1.1111</td>
<td>1.1111</td>
<td>1.0000</td>
<td>0.5000</td>
<td>0.0500</td>
<td>0.7810</td>
</tr>
<tr>
<td>0.6002</td>
<td>250</td>
<td>0.9000</td>
<td>1.1111</td>
<td>1.1111</td>
<td>1.0000</td>
<td>0.5000</td>
<td>0.0500</td>
<td>0.3998</td>
</tr>
<tr>
<td>0.8064</td>
<td>350</td>
<td>0.9000</td>
<td>1.1111</td>
<td>1.1111</td>
<td>1.0000</td>
<td>0.5000</td>
<td>0.0500</td>
<td>0.1936</td>
</tr>
<tr>
<td>0.9101</td>
<td>450</td>
<td>0.9000</td>
<td>1.1111</td>
<td>1.1111</td>
<td>1.0000</td>
<td>0.5000</td>
<td>0.0500</td>
<td>0.0899</td>
</tr>
<tr>
<td>0.9596</td>
<td>550</td>
<td>0.9000</td>
<td>1.1111</td>
<td>1.1111</td>
<td>1.0000</td>
<td>0.5000</td>
<td>0.0500</td>
<td>0.0404</td>
</tr>
</tbody>
</table>

Report Definitions

Power is the probability of rejecting non-equivalence when the means are equivalent.  
N is the total number of subjects split between both sequences.  
RU & RL are the upper and lower equivalence limits. Ratios between these limits are equivalent.  
R1 is the ratio of the means at which the power is computed.  
COV is the coefficient of variation on the original scale.  
Alpha is the probability of rejecting non-equivalence when the means are non-equivalent.  
Beta is the probability of accepting non-equivalence when the means are equivalent.

Summary Statements

In an equivalence test of means using two one-sided tests on data from a two-period cross-over design, a total sample size of 50 achieves 0% power at a 5% significance level when the true ratio of the means is 1.0000, the coefficient of variation on the original, unlogged scale is 0.5000, and the equivalence limits of the mean ratio are 0.9000 and 1.1111.

This report shows the power for the indicated scenarios. Note that if they want 90% power, they will require a sample of around 450 subjects.

Plots Section

This plot shows the power versus the sample size.
Example 2 – Validation using Julious

Julious (2004) page 1963 presents a table of sample sizes for various parameter values. The power is 0.90 and the significance level is 0.05. The COV is set to 0.25, the ‘level of bioequivalence’ is set to 10%, 15%, 20%, and 25%, and the true ratio is set to 1.00, the necessary sample sizes are 120, 52, 28, and 18. Note that the level of bioequivalence as defined in Julious (2004) is equal to 1 – RL.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Equivalence Tests for Two Means in a 2x2 Cross-Over Design [Ratios] procedure window by expanding Means, then 2x2 Cross-Over Design, then clicking on Equivalence, and then clicking on Equivalence Tests for Two Means in a 2x2 Cross-Over Design [Ratios]. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N</td>
</tr>
<tr>
<td>Power</td>
<td>0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>RU (Upper Equivalence Limit)</td>
<td>1/RL</td>
</tr>
<tr>
<td>RL (Lower Equivalence Limit)</td>
<td>0.90 0.85 0.80 0.75</td>
</tr>
<tr>
<td>R1 (True Ratio)</td>
<td>1.00</td>
</tr>
<tr>
<td>COV (Coefficient of Variation)</td>
<td>0.25</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>Total Sample Size (N)</th>
<th>Lower Equiv. Limit of Ratio (RL)</th>
<th>Upper Equiv. Limit of Ratio (RU)</th>
<th>True Ratio (R1)</th>
<th>Coefficient of Variation (COV)</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9121</td>
<td>18</td>
<td>0.7500</td>
<td>1.3333</td>
<td>1.0000</td>
<td>0.2500</td>
<td>0.0500</td>
<td>0.0879</td>
</tr>
<tr>
<td>0.9023</td>
<td>28</td>
<td>0.8000</td>
<td>1.2500</td>
<td>1.0000</td>
<td>0.2500</td>
<td>0.0500</td>
<td>0.0977</td>
</tr>
<tr>
<td>0.9060</td>
<td>52</td>
<td>0.8500</td>
<td>1.1765</td>
<td>1.0000</td>
<td>0.2500</td>
<td>0.0500</td>
<td>0.0940</td>
</tr>
<tr>
<td>0.9012</td>
<td>120</td>
<td>0.9000</td>
<td>1.1111</td>
<td>1.0000</td>
<td>0.2500</td>
<td>0.0500</td>
<td>0.0988</td>
</tr>
</tbody>
</table>

Note that PASS obtains the same samples sizes as Julious (2004).
525-10 Equivalence Tests for Two Means in a 2x2 Cross-Over Design using Ratios
Introduction

This procedure calculates power and sample size for non-zero null tests which use the difference in the means of a higher-order cross-over design. Measurements are made on individuals that have been randomly assigned to one of several treatment sequences. Only a brief introduction to the subject will be given here. For a comprehensive discussion on the subject, refer to Chen et al. (1997) and Chow et al. (2003).

Cross-Over Designs

Senn (2002) defines a cross-over design as one in which each subject receives all treatments at least once and the object is to study differences among the treatments. The name cross-over comes from the most common case in which there are only two treatments. In this case, each subject crosses over from one treatment to another. It is assumed that there is a washout period between treatments during which the response returns to its baseline value. If this does not occur, there is said to be a carryover effect.

A 2x2 cross-over design refers to two treatments (periods) and two sequences (treatment orderings). One sequence of treatments is treatment A followed by treatment B. The other
sequence is B and then A. The design includes a washout period between responses to make certain that the effects of the first drug do no carryover to the second. Thus, the groups of subjects in this design are defined by the sequence in which the two treatments are administered, not by the treatments they receive.

Higher-Order Cross-Over Designs

Chen et al. (1997) present the results for four cross-over designs that are more complicated than the 2x2 design. Assume that the two treatments are labeled A and B. The available designs are defined by the order and number of times the two treatments are administered.

Balaam’s Design

Balaam’s design has four sequences with two treatments each. It is popular because it allows the intrasubject variabilities to be estimated. The design is

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Period 1</th>
<th>Period 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>A</td>
</tr>
</tbody>
</table>

Two-Sequence Dual Design

This design has two sequences with three periods each. It is popular because it allows the intrasubject variabilities to be estimated. The design is

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>

Four-Period Design with Two Sequences

This design has two sequences of four periods each. The design is

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
<th>Period 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
</tbody>
</table>

Four-Period Design with Four Sequences

This design has four sequences of four periods each. The design is

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
<th>Period 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>1</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
</tbody>
</table>

Advantages of Cross-Over Designs

A comparison of treatments on the same subject is expected to be more precise. The increased precision often translates into a smaller sample size. Also, patient enrollment may be easier to obtain because each patient will receive both treatments.
Disadvantages of Cross-Over Designs

The statistical analysis of a cross-over experiment is more complex than a parallel-group experiment and requires additional assumptions. In a cross-over experiment, it may be difficult to separate the treatment effect from the time effect and the carry-over effect of the previous treatment.

These cross-over designs cannot be used when the treatment (or the measurement of the response) alters the subject permanently. Hence, it cannot be used to compare treatments that are intended to provide a cure.

Because subjects must be measured at least twice, it may be more difficult to keep patients enrolled in the study. This is particularly true when the measurement process is painful, uncomfortable, embarrassing, or time consuming.

The Statistical Hypotheses

Both non-inferiority and superiority tests are examples of directional (one-sided) tests. Remember that in the usual t-test setting, the null (H0) and alternative (H1) hypotheses for one-sided tests are defined as

\[ H_0 : \delta \leq A \text{ versus } H_1 : \delta > A \]

Rejecting H0 implies that the mean is larger than the value A. This test is called an *upper-tailed test* because H0 is rejected only in samples in which the difference in sample means is larger than A.

Following is an example of a *lower-tailed test*.

\[ H_0 : \delta \geq A \text{ versus } H_1 : \delta < A \]

*Non-inferiority* and *superiority* tests are special cases of the above directional tests. It will be convenient to adopt the following specialize notation for the discussion of these tests.
Two Means in a Higher-Order Cross-Over Design with Non-Zero Null (Differences)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PASS Input/Output</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\mu_T$</td>
<td>Not used</td>
<td>Treatment mean. This is the treatment mean.</td>
</tr>
<tr>
<td>$\mu_R$</td>
<td>Not used</td>
<td>Reference mean. This is the mean of a reference population.</td>
</tr>
<tr>
<td>$M_s$</td>
<td>SM</td>
<td>Margin of superiority. This is a tolerance value that defines the magnitude of difference that is required for practical importance. This may be thought of as the smallest difference from the reference that is considered to be different.</td>
</tr>
<tr>
<td>$\delta$</td>
<td>D</td>
<td>True difference. This is the value of $\mu_T - \mu_R$, the difference between the treatment and reference means. This is the value at which the power is calculated.</td>
</tr>
</tbody>
</table>

Note that the actual values of $\mu_T$ and $\mu_R$ are not needed. Only their difference is needed for power and sample size calculations.

**Non-Zero Null Tests**

A non-zero null test tests that the treatment mean is better than the reference mean by more than the superiority margin. The actual direction of the hypothesis depends on the response variable being studied.

**Case 1: High Values Good**

In this case, higher values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is greater than the reference mean by at least the margin of superiority. The value of $\delta$ must be greater than $M_s$. The following are equivalent sets of hypotheses.

\[
H_0 : \mu_1 \leq \mu_2 + M_s \quad \text{versus} \quad H_1 : \mu_1 > \mu_2 + M_s
\]

\[
H_0 : \mu_1 - \mu_2 \leq M_s \quad \text{versus} \quad H_1 : \mu_1 - \mu_2 > M_s
\]

\[
H_0 : \delta \leq M_s \quad \text{versus} \quad H_1 : \delta > M_s
\]

**Case 2: High Values Bad**

In this case, lower values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is less than the reference mean by at least the margin of superiority. The value of $\delta$ must be less than $-M_s$. The following are equivalent sets of hypotheses.

\[
H_0 : \mu_1 \geq \mu_2 - M_s \quad \text{versus} \quad H_1 : \mu_1 < \mu_2 - M_s
\]

\[
H_0 : \mu_1 - \mu_2 \geq -M_s \quad \text{versus} \quad H_1 : \mu_1 - \mu_2 < -M_s
\]

\[
H_0 : \delta \geq -M_s \quad \text{versus} \quad H_1 : \delta < -M_s
\]
Test Statistics

The analysis for assessing non-inferiority and non-zero null tests using higher-order cross-over designs is discussed in detail in Chapter 9 of Chow and Liu (2000). Unfortunately, their presentation is too lengthy to give here. Their method involves the computation of an analysis of variance to estimate the error variance. It also describes the construction of confidence limits for appropriate contrasts. One-sided confidence limits can be used for non-inferiority tests. Details of this approach are given in Chapter 3 of Chow et al. (2003). We refer you to these books for details.

Power Calculation

The power of the non-inferiority and superiority tests for the case in which higher values are better is given by

\[
Power = T_v \left( \frac{\delta - \varepsilon}{\sigma_w \sqrt{b \cdot n}} \right) - t_{v,1-\alpha}
\]

where \(T\) represents the cumulative \(t\) distribution, \(V\) and \(b\) depend on the design, \(\sigma_w\) is the square root of the within mean square error from the ANOVA table used to analyze the cross-over design, and \(n\) is the average number of subjects per sequence. Note that the constants \(V\) and \(b\) depend on the design as follows.

The power of the non-inferiority and superiority tests for the case in which higher values are worse is given by

\[
Power = 1 - T_v \left( t_{v,1-\alpha} - \frac{\varepsilon - \delta}{\sigma_w \sqrt{b \cdot n}} \right)
\]

The constants \(V\) and \(b\) depend on the design as follows:

<table>
<thead>
<tr>
<th>Design Type</th>
<th>Parameters ((V, b))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balaam’s Design</td>
<td>(V = 4n - 3, \ b = 2).</td>
</tr>
<tr>
<td>Two-Sequence Dual Design</td>
<td>(V = 4n - 4, \ b = 3/4).</td>
</tr>
<tr>
<td>Four-Period Design with Two Sequences</td>
<td>(V = 6n - 5, \ b = 11/20).</td>
</tr>
<tr>
<td>Four-Period Design with Four Sequences</td>
<td>(V = 12n - 5, \ b = 1/4).</td>
</tr>
</tbody>
</table>

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains the parameters associated with this test such as the means, sample sizes, alpha, and power.
Solve For

Find (Solve For)
This option specifies the parameter to be solved for from the other parameters. Under most situations, you will select either Power and Beta for a power analysis or N for sample size determination.

Select N when you want to calculate the sample size needed to achieve a given power and alpha level. Note that there are two choices for finding N. Select N (Equal Per Sequence) when you want the design to have an equal number of subjects per sequence. Select N (Exact) when you want to find the exact sample size even though the number of subjects cannot be dividing equally among the sequences.

Select Power and Beta when you want to calculate the power of an experiment.

Error Rates

Power or Beta
This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

Alpha (Significance Level)
This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.

Sample Size

N (Total Sample Size)
This option specifies one or more values of the sample size, the number of individuals in the study (total subjects in all sequences). This value must be an integer greater than one.

Effect Size – Mean Difference

SM (Superiority Margin)
This is the magnitude of the margin of superiority. It must be entered as a positive number.
Two Means in a Higher-Order Cross-Over Design with Non-Zero Null (Differences) 528-7

When higher means are better, this value is the distance above the reference mean that is required to be considered superior. When higher means are worse, this value is the distance below the reference mean that is required to be considered superior.

**D (True Difference)**
This is the actual difference between the treatment mean and the reference mean at which the power is calculated.

When higher means are better, this value should be greater than SM. When higher means are worse, this value should be negative and greater in magnitude than SM.

**Effect Size – Standard Deviation**

**Sw (Within Standard Error)**
Specify one or more values of Sw, which is SQR(WMSE) where WMSE is the within mean square error from the ANOVA table used to analyze the cross-over design. These values must be positive.

You can press the Standard Deviation Estimator button to load the Standard Deviation Estimator window.

**Test**

**Design Type**
Specify the type of cross-over design that you are analyzing. Note that all of these designs assume that you are primarily interested in the overall difference between the two treatment means.

**Higher Means Are**
This option defines whether higher values of the response variable are to be considered better or worse. The choice here determines the direction of the test.

If Higher Means Are Better the null hypothesis is Diff <= SM and the alternative hypothesis is Diff > SM. If Higher Means Are Worse the null hypothesis is Diff >= -SM and the alternative hypothesis is Diff < -SM.
Example 1 – Finding Power

Researchers want to calculate the power of a non-zero null test using data from a two-sequence, dual cross-over design. The margin of superiority is either 5 or 10 at several sample sizes between 6 and 66. The true difference between the means under is assumed to be 15. Similar experiments have had a standard deviation (Sw) of 10. The significance level is 0.025.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means in a Higher-Order Cross-Over Design (Non-Zero Null) [Differences] procedure window by expanding Means, then MxK Cross-Over Design, then clicking on Test (Non-Zero Null), and then clicking on Tests for Two Means in a Higher-Order Cross-Over Design (Non-Zero Null) [Differences]. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.025</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>6 to 66 by 10</td>
</tr>
<tr>
<td>SM (Superiority Margin)</td>
<td>5 10</td>
</tr>
<tr>
<td>D (True Difference)</td>
<td>15</td>
</tr>
<tr>
<td>Sw (Within Standard Error)</td>
<td>10</td>
</tr>
<tr>
<td>Design Type</td>
<td>2x3 (Two-Sequence Dual)</td>
</tr>
<tr>
<td>Higher Means Are</td>
<td>Better</td>
</tr>
</tbody>
</table>

Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

Numeric Results for Testing Non-Inferiority Using the Difference

<table>
<thead>
<tr>
<th>Higher Means Are Better</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design: Two-Sequence Dual Cross-Over. Hypotheses: H0: Diff &lt;= SM; H1: Diff &gt; SM.</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Power</th>
<th>Total Sample Size (N)</th>
<th>Sequences and Periods (SxP)</th>
<th>Superiority Margin (SM)</th>
<th>Difference for Power (D)</th>
<th>Standard Error of Diff. (Sw)</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.3837</td>
<td>6</td>
<td>2x3</td>
<td>5.00</td>
<td>15.00</td>
<td>10.00</td>
<td>0.0250</td>
<td>0.6163</td>
</tr>
<tr>
<td>0.8832</td>
<td>16</td>
<td>2x3</td>
<td>5.00</td>
<td>15.00</td>
<td>10.00</td>
<td>0.0250</td>
<td>0.1168</td>
</tr>
<tr>
<td>0.9818</td>
<td>26</td>
<td>2x3</td>
<td>5.00</td>
<td>15.00</td>
<td>10.00</td>
<td>0.0250</td>
<td>0.0182</td>
</tr>
<tr>
<td>0.9975</td>
<td>36</td>
<td>2x3</td>
<td>5.00</td>
<td>15.00</td>
<td>10.00</td>
<td>0.0250</td>
<td>0.0025</td>
</tr>
<tr>
<td>0.9997</td>
<td>46</td>
<td>2x3</td>
<td>5.00</td>
<td>15.00</td>
<td>10.00</td>
<td>0.0250</td>
<td>0.0003</td>
</tr>
<tr>
<td>1.0000</td>
<td>56</td>
<td>2x3</td>
<td>5.00</td>
<td>15.00</td>
<td>10.00</td>
<td>0.0250</td>
<td>0.0000</td>
</tr>
<tr>
<td>1.0000</td>
<td>66</td>
<td>2x3</td>
<td>5.00</td>
<td>15.00</td>
<td>10.00</td>
<td>0.0250</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Report continues…
Two Means in a Higher-Order Cross-Over Design with Non-Zero Null (Differences)  528-9

References

Report Definitions
Power is the probability of rejecting H0 (concluding non-inferiority) when H0 is false.
N is the total number of subjects. They are divided evenly among all sequences.
S is the number of sequences.
P is the number of periods per sequence.
SM is the magnitude of the margin of superiority. Since higher means are better, this value is positive and is the distance above the reference mean that is required to be considered superior.
D is the mean difference at which the power is computed. D = Mean1 - Mean2 = treatment mean - reference mean.
Sw is the square root of the within mean square error from the ANOVA table.
Alpha is the probability of falsely rejecting H0 (falsely concluding non-inferiority).
Beta is the probability of not rejecting H0 when it is false.
Two-Sequence Dual Cross-Over Design with pattern: ABB; BAA

Summary Statements
In a superiority test on data for which higher values are better drawn from a two-sequence dual cross-over design, a total sample size of 6 achieves 38% power at a 3% significance level when the true difference between the means is 15.00, the square root of the within mean square error is 10.00, and the superiority margin is 5.00.

This report shows the power for the indicated scenarios.

Plots Section

This plot shows the power versus the sample size.
Example 2 – Finding Sample Size

Continuing with Example 1, the researchers want to find the exact sample size needed to achieve both 80% power and 90% power.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means in a Higher-Order Cross-Over Design (Non-Zero Null) [Differences] procedure window by expanding Means, then MxK Cross-Over Design, then clicking on Test (Non-Zero Null), and then clicking on Tests for Two Means in a Higher-Order Cross-Over Design (Non-Zero Null) [Differences]. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N (Equal Per Sequence)</td>
</tr>
<tr>
<td>Power</td>
<td>0.80 0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.025</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>SM (Superiority Margin)</td>
<td>5 10</td>
</tr>
<tr>
<td>D (True Difference)</td>
<td>15</td>
</tr>
<tr>
<td>Sw (Within Standard Error)</td>
<td>10</td>
</tr>
<tr>
<td>Design Type</td>
<td>2x3 (Two-Sequence Dual)</td>
</tr>
<tr>
<td>Higher Means Are</td>
<td>Better</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Numeric Results for Testing Non-Inferiority Using the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher Means are Better</td>
</tr>
<tr>
<td>Design: Two-Sequence Dual Cross-Over. Hypotheses: H0: Diff &lt;= SM; H1: Diff &gt; SM.</td>
</tr>
<tr>
<td>Power</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>0.9184</td>
</tr>
<tr>
<td>0.8343</td>
</tr>
<tr>
<td>0.9084</td>
</tr>
<tr>
<td>0.8153</td>
</tr>
</tbody>
</table>

When the superiority margin is set to 10, 66 subjects are needed to achieve 90% power and 50 subjects are needed to achieve at least 80% power.
Example 3 – Validation

This procedure uses the same mechanics as the Non-Inferiority Tests for Two Means in a Higher-Order Cross-Over Design using Differences procedure. We refer the user to Example 3 of Chapter 530 for the validation.
Two Means in a Higher-Order Cross-Over Design with Non-Zero Null (Differences)
Chapter 530

Non-Inferiority Tests for Two Means in a Higher-Order Cross-Over Design using Differences

Introduction

This procedure calculates power and sample size for non-inferiority tests which use the difference in the means of a higher-order cross-over design. Measurements are made on individuals that have been randomly assigned to one of several treatment sequences. Only a brief introduction to the subject will be given here. For a comprehensive discussion on the subject, refer to Chen et al. (1997) and Chow et al. (2003).

Cross-Over Designs

Senn (2002) defines a cross-over design as one in which each subject receives all treatments at least once and the object is to study differences among the treatments. The name cross-over comes from the most common case in which there are only two treatments. In this case, each subject crosses over from one treatment to another. It is assumed that there is a washout period between treatments during which the response returns to its baseline value. If this does not occur, there is said to be a carryover effect.

A 2x2 cross-over design refers to two treatments (periods) and two sequences (treatment orderings). One sequence of treatments is treatment A followed by treatment B. The other
sequence is B and then A. The design includes a washout period between responses to make certain that the effects of the first drug do no carryover to the second. Thus, the groups of subjects in this design are defined by the sequence in which the two treatments are administered, not by the treatments they receive.

Higher-Order Cross-Over Designs
Chen et al. (1997) present the results for four cross-over designs that are more complicated than the 2x2 design. Assume that the two treatments are labeled A and B. The available designs are defined by the order and number of times the two treatments are administered.

Balaam’s Design
Balaam’s design has four sequences with two treatments each. It is popular because it allows the intrasubject variabilities to be estimated. The design is

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Period 1</th>
<th>Period 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>A</td>
</tr>
</tbody>
</table>

Two-Sequence Dual Design
This design has two sequences with three periods each. It is popular because it allows the intrasubject variabilities to be estimated. The design is

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>

Four-Period Design with Two Sequences
This design has two sequences of four periods each. The design is

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
<th>Period 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
</tbody>
</table>

Four-Period Design with Four Sequences
This design has four sequences of four periods each. The design is

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
<th>Period 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
</tbody>
</table>

Advantages of Cross-Over Designs
A comparison of treatments on the same subject is expected to be more precise. The increased precision often translates into a smaller sample size. Also, patient enrollment may be easier to obtain because each patient will receive both treatments.
Disadvantages of Cross-Over Designs

The statistical analysis of a cross-over experiment is more complex than a parallel-group experiment and requires additional assumptions. In a cross-over experiment, it may be difficult to separate the treatment effect from the time effect and the carry-over effect of the previous treatment.

These cross-over designs cannot be used when the treatment (or the measurement of the response) alters the subject permanently. Hence, it cannot be used to compare treatments that are intended to provide a cure.

Because subjects must be measured at least twice, it may be more difficult to keep patients enrolled in the study. This is particularly true when the measurement process is painful, uncomfortable, embarrassing, or time consuming.

The Statistical Hypotheses

Both non-inferiority and superiority tests are examples of directional (one-sided) tests. Remember that in the usual t-test setting, the null (H0) and alternative (H1) hypotheses for one-sided tests are defined as

\[ H_0: \delta \leq A \text{ versus } H_1: \delta > A \]

Rejecting H0 implies that the mean is larger than the value A. This test is called an upper-tailed test because H0 is rejected only in samples in which the difference in sample means is larger than A.

Following is an example of a lower-tailed test.

\[ H_0: \delta \geq A \text{ versus } H_1: \delta < A \]

Non-inferiority and superiority tests are special cases of the above directional tests. It will be convenient to adopt the following specialize notation for the discussion of these tests.
Non-Inferiority of Two Means in a Higher-Order Cross-Over Design (Differences)

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PASS Input/Output</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu_T )</td>
<td>Not used</td>
<td>Treatment mean. This is the treatment mean.</td>
</tr>
<tr>
<td>( \mu_R )</td>
<td>Not used</td>
<td>Reference mean. This is the mean of a reference population.</td>
</tr>
<tr>
<td>( M_{NI} )</td>
<td>NIM</td>
<td>Margin of non-inferiority. This is a tolerance value that defines the magnitude of the amount that is not of practical importance. This may be thought of as the largest change from the baseline that is considered to be trivial. The absolute value is shown to emphasize that this is a magnitude. The sign of the value will be determined by the specific design that is being used.</td>
</tr>
<tr>
<td>( \delta )</td>
<td>D</td>
<td>True difference. This is the value of ( \mu_T - \mu_R ), the difference between the treatment and reference means. This is the value at which the power is calculated.</td>
</tr>
</tbody>
</table>

Note that the actual values of \( \mu_T \) and \( \mu_R \) are not needed. Only their difference is needed for power and sample size calculations.

Non-Inferiority Tests

A non-inferiority test tests that the treatment mean is not worse than the reference mean by more than the equivalence margin. The actual direction of the hypothesis depends on the response variable being studied.

**Case 1: High Values Good, Non-Inferiority Test**

In this case, higher values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is no less than a small amount below the reference mean. The value of \( \delta \) is often set to zero. The following are equivalent sets of hypotheses.

\[
H_0 : \mu_1 \leq \mu_2 - |M_{NI}| \quad \text{versus} \quad H_1 : \mu_1 > \mu_2 - |M_{NI}|
\]

\[
H_0 : \mu_1 - \mu_2 \leq -|M_{NI}| \quad \text{versus} \quad H_1 : \mu_1 - \mu_2 > -|M_{NI}|
\]

\[
H_0 : \delta \leq -|M_{NI}| \quad \text{versus} \quad H_1 : \delta > -|M_{NI}|
\]

**Case 2: High Values Bad, Non-Inferiority Test**

In this case, lower values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is no more than a small amount above the reference mean. The value of \( \delta \) is often set to zero. The following are equivalent sets of hypotheses.

\[
H_0 : \mu_1 \geq \mu_2 + |M_{NI}| \quad \text{versus} \quad H_1 : \mu_1 < \mu_2 + |M_{NI}|
\]

\[
H_0 : \mu_1 - \mu_2 \geq |M_{NI}| \quad \text{versus} \quad H_1 : \mu_1 - \mu_2 < |M_{NI}|
\]

\[
H_0 : \delta \geq |M_{NI}| \quad \text{versus} \quad H_1 : \delta < |M_{NI}|
\]
Test Statistics

The analysis for assessing equivalence (and thus non-inferiority) using higher-order cross-over designs is discussed in detail in Chapter 9 of Chow and Liu (2000). Unfortunately, their presentation is too lengthy to give here. Their method involves the computation of an analysis of variance to estimate the error variance. It also describes the construction of confidence limits for appropriate contrasts. One-sided confidence limits can be used for non-inferiority tests. Details of this approach are given in Chapter 3 of Chow et al. (2003). We refer you to these books for details.

Power Calculation

The power of the non-inferiority and superiority tests for the case in which higher values are better is given by

$$Power = T_{V}(\frac{\delta - \epsilon}{\sigma_{w} \sqrt{b / n}}) - t_{v,1-\alpha}$$

where $T$ represents the cumulative $t$ distribution, $V$ and $b$ depend on the design, $\sigma_{w}$ is the square root of the within mean square error from the ANOVA table used to analyze the cross-over design, and $n$ is the average number of subjects per sequence. Note that the constants $V$ and $b$ depend on the design as follows.

The power of the non-inferiority and superiority tests for the case in which higher values are worse is given by

$$Power = 1 - T_{V}(t_{v,1-\alpha} - \frac{\epsilon - \delta}{\sigma_{w} \sqrt{b / n}})$$

The constants $V$ and $b$ depend on the design as follows:

<table>
<thead>
<tr>
<th>Design Type</th>
<th>Parameters ($V, b$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balaam’s Design</td>
<td>$V = 4n - 3, b = 2.$</td>
</tr>
<tr>
<td>Two-Sequence Dual Design</td>
<td>$V = 4n - 4, b = 3/4.$</td>
</tr>
<tr>
<td>Four-Period Design with Two Sequences</td>
<td>$V = 6n - 5, b = 11/20.$</td>
</tr>
<tr>
<td>Four-Period Design with Four Sequences</td>
<td>$V = 12n - 5, b = 1/4.$</td>
</tr>
</tbody>
</table>

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains the parameters associated with this test such as the means, sample sizes, alpha, and power.
Solve For

Find (Solve For)
This option specifies the parameter to be solved for from the other parameters. Under most situations, you will select either Power and Beta for a power analysis or N for sample size determination.

Select N when you want to calculate the sample size needed to achieve a given power and alpha level. Note that there are two choices for finding N. Select N (Equal Per Sequence) when you want the design to have an equal number of subjects per sequence. Select N (Exact) when you want to find the exact sample size even though the number of subjects cannot be dividing equally among the sequences.

Select Power and Beta when you want to calculate the power of an experiment.

Error Rates

Power or Beta
This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

Alpha (Significance Level)
This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.

Sample Size

N (Total Sample Size)
This option specifies one or more values of the sample size, the number of individuals in the study (total subjects in all sequences). This value must be an integer greater than one.

Effect Size – Mean Difference

NIM (Non-Inferiority Margin)
This is the magnitude of the margin of non-inferiority. It must be entered as a positive number.
When higher means are better, this value is the distance below the reference mean that is still considered non-inferior. When higher means are worse, this value is the distance above the reference mean that is still considered non-inferior.

**D (True Difference)**

This is the actual difference between the treatment mean and the reference mean at which power is calculated.

For non-inferiority tests, this value is often set to zero. When this value is non-zero, care should be taken that this value is consistent with whether higher means are better or worse.

---

**Effect Size – Standard Deviation**

**Sw (Within Standard Error)**

Specify the value(s) of the standard deviation Sw. This is the value of Sw which is $\text{SQR}(\text{WMSE})$ where WMSE is the within mean square error from the ANOVA table used to analyze the Cross-Over design.

These values must be positive. A list of values may be entered.

---

**Test**

**Design Type**

Specify the type of cross-over design that you are analyzing. Note that all of these designs assume that you are primarily interested in the overall difference between the two treatment means.

**Higher Means Are**

This option defines whether higher values of the response variable are to be considered better or worse.

The choice here determines the direction of the non-inferiority test.

If Higher Means Are Better the null hypothesis is $\text{Diff} \leq -\text{NIM}$ and the alternative hypothesis is $\text{Diff} > -\text{NIM}$. If Higher Means Are Worse the null hypothesis is $\text{Diff} \geq \text{NIM}$ and the alternative hypothesis is $\text{Diff} < \text{NIM}$. 
Example 1 – Finding Power

Researchers want to calculate the power of a non-inferiority test using data from a two-sequence, dual cross-over design. The margin of equivalence is either 5 or 10 at several sample sizes between 6 and 66. The true difference between the means under is assumed to be 0. Similar experiments have had a standard deviation (Sw) of 10. The significance level is 0.025.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Non-Inferiority Tests for Two Means in a Higher-Order Cross-Over Design [Differences] procedure window by expanding Means, then MxK Cross-Over Design, then clicking on Non-Inferiority, and then clicking on Non-Inferiority Tests for Two Means in a Higher-Order Cross-Over Design [Differences]. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Power</td>
<td>0.025</td>
</tr>
<tr>
<td>Alpha</td>
<td>6 to 66 by 10</td>
</tr>
<tr>
<td>NIM (Non-Inferiority Margin)</td>
<td>5 10</td>
</tr>
<tr>
<td>D (True Difference)</td>
<td>0</td>
</tr>
<tr>
<td>Sw (Within Standard Error)</td>
<td>10</td>
</tr>
<tr>
<td>Design Type</td>
<td>2x3 (Two-Sequence Dual)</td>
</tr>
<tr>
<td>Higher Means Are</td>
<td>Better</td>
</tr>
</tbody>
</table>

Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>Total Sample Size (N)</th>
<th>Sequences and Periods (SxP)</th>
<th>Non-Inferiority Margin (-NIM)</th>
<th>Difference for Power (D)</th>
<th>Standard Error of Diff. (Sw)</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.1139</td>
<td>6</td>
<td>2x3</td>
<td>5.00</td>
<td>0.00</td>
<td>10.00</td>
<td>0.025</td>
<td>0.8861</td>
</tr>
<tr>
<td>0.3405</td>
<td>16</td>
<td>2x3</td>
<td>5.00</td>
<td>0.00</td>
<td>10.00</td>
<td>0.025</td>
<td>0.6595</td>
</tr>
<tr>
<td>0.5282</td>
<td>26</td>
<td>2x3</td>
<td>5.00</td>
<td>0.00</td>
<td>10.00</td>
<td>0.025</td>
<td>0.4718</td>
</tr>
<tr>
<td>0.6744</td>
<td>36</td>
<td>2x3</td>
<td>5.00</td>
<td>0.00</td>
<td>10.00</td>
<td>0.025</td>
<td>0.3256</td>
</tr>
<tr>
<td>0.7817</td>
<td>46</td>
<td>2x3</td>
<td>5.00</td>
<td>0.00</td>
<td>10.00</td>
<td>0.025</td>
<td>0.2183</td>
</tr>
<tr>
<td>0.8571</td>
<td>56</td>
<td>2x3</td>
<td>5.00</td>
<td>0.00</td>
<td>10.00</td>
<td>0.025</td>
<td>0.1429</td>
</tr>
<tr>
<td>0.9084</td>
<td>66</td>
<td>2x3</td>
<td>5.00</td>
<td>0.00</td>
<td>10.00</td>
<td>0.025</td>
<td>0.0916</td>
</tr>
</tbody>
</table>
Report continues...
References

Report Definitions
Power is the probability of rejecting H0 (concluding non-inferiority) when H0 is false.
N is the total number of subjects. They are divided evenly among all sequences.
P is the number of periods per sequence.
-NIM is the magnitude and direction of the margin of non-inferiority. Since higher means are better, this value is negative and is the distance below the reference mean that is still considered non-inferior.
D is the actual difference between the treatment and reference means that is used in the power calculations.
Sw is the square root of the within mean square error from the ANOVA table.
Alpha is the probability of falsely rejecting H0 (falsely concluding non-inferiority).
Beta is the probability of not rejecting H0 when it is false.

Two-Sequence Dual Cross-Over Design with pattern: ABB; BAA

Summary Statements
In a non-inferiority test on data for which higher values are better drawn from a two-sequence dual cross-over design, a total sample size of 6 achieves 11% power at a 3% significance level when the true difference between the means is 0.00, the square root of the within mean square error is 10.00, and the non-inferiority margin is -5.00.

This report shows the power for the indicated scenarios.

Plots Section

This plot shows the power versus the sample size.
Example 2 – Finding Sample Size

Continuing with Example 1, the researchers want to find the exact sample size needed to achieve both 80% power and 90% power.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Non-Inferiority Tests for Two Means in a Higher-Order Cross-Over Design [Differences] procedure window by expanding Means, then MxK Cross-Over Design, then clicking on Non-Inferiority, and then clicking on Non-Inferiority Tests for Two Means in a Higher-Order Cross-Over Design [Differences]. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N (Equal Per Sequence)</td>
</tr>
<tr>
<td>Power</td>
<td>0.80 0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.025</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>NIM (Non-Inferiority Margin)</td>
<td>5 10</td>
</tr>
<tr>
<td>D (True Difference)</td>
<td>0</td>
</tr>
<tr>
<td>Sw (Within Standard Error)</td>
<td>10</td>
</tr>
<tr>
<td>Design Type</td>
<td>2x3 (Two-Sequence Dual)</td>
</tr>
<tr>
<td>Higher Means Are</td>
<td>Better</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Numeric Results for Testing Non-Inferiority Using the Difference</th>
<th>Higher Means are Better</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design: Two-Sequence Dual Cross-Over. Hypotheses: H0: Diff &lt;= -NIM; H1: Diff &gt; -NIM.</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Power</th>
<th>Total Sample Size (N)</th>
<th>Sequences and Periods (SxP)</th>
<th>Non-Inferiority Margin (-NIM)</th>
<th>Difference for Power (D)</th>
<th>Standard Error of Diff. (Sw)</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9084</td>
<td>66</td>
<td>2x3</td>
<td>5.00</td>
<td>0.00</td>
<td>10.00</td>
<td>0.0250</td>
<td>0.0916</td>
</tr>
<tr>
<td>0.8153</td>
<td>50</td>
<td>2x3</td>
<td>5.00</td>
<td>0.00</td>
<td>10.00</td>
<td>0.0250</td>
<td>0.1847</td>
</tr>
<tr>
<td>0.9184</td>
<td>18</td>
<td>2x3</td>
<td>10.00</td>
<td>0.00</td>
<td>10.00</td>
<td>0.0250</td>
<td>0.0816</td>
</tr>
<tr>
<td>0.8343</td>
<td>14</td>
<td>2x3</td>
<td>10.00</td>
<td>0.00</td>
<td>10.00</td>
<td>0.0250</td>
<td>0.1657</td>
</tr>
</tbody>
</table>

When the non-inferiority margin is set to 5, 66 subjects are needed to achieve 90% power and 50 subjects are needed to achieve at least 80% power.
Example 3 – Validation

We could not find a validation example for this procedure in the statistical literature, so we will have to generate a validated example from within PASS. To do this, we use the Higher-Order, Cross-Over Equivalence using Differences procedure which was validated. By setting the upper equivalence limit to a large value (we used 22), we obtain results for a non-inferiority test.

Suppose the square root of the within mean square error is 0.10, the equivalence limit is 0.20, the difference between the means is 0.05, the power is 90%, and the significance level is 0.05 (see the Example 4 template). PASS calculates a sample size of 16.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Non-Inferiority Tests for Two Means in a Higher-Order Cross-Over Design [Differences] procedure window by expanding Means, then MxK Cross-Over Design, then clicking on Non-Inferiority, and then clicking on Non-Inferiority Tests for Two Means in a Higher-Order Cross-Over Design [Differences]. You may then make the appropriate entries as listed below, or open Example 3 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td>Find (Solve For) ......................................N (Equal Per Sequence)</td>
</tr>
<tr>
<td></td>
<td>Power ......................................................0.90</td>
</tr>
<tr>
<td></td>
<td>Alpha .......................................................0.05</td>
</tr>
<tr>
<td></td>
<td>N (Total Sample Size) ..................................Ignored since this is the Find setting</td>
</tr>
<tr>
<td></td>
<td>NIM (Non-Inferiority Margin) ..........................0.2</td>
</tr>
<tr>
<td></td>
<td>D (True Difference) .....................................0.05</td>
</tr>
<tr>
<td></td>
<td>Sw (Within Standard Error) ............................0.10</td>
</tr>
<tr>
<td></td>
<td>Design Type ..................................................4x2 (Balaam)</td>
</tr>
<tr>
<td></td>
<td>Higher Means Are .........................................Better</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Numeric Results for Testing Non-Inferiority Using the Difference</th>
</tr>
</thead>
<tbody>
<tr>
<td>Higher Means are Better</td>
</tr>
<tr>
<td>Design: Balaam’s Cross-Over. Hypotheses: H0: Diff &lt;= -NIM; H1: Diff &gt; -NIM.</td>
</tr>
<tr>
<td>Power</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>0.9495</td>
</tr>
</tbody>
</table>

PASS has also obtained a sample size of 16 using the non-inferiority procedure.
530-12 Non-Inferiority of Two Means in a Higher-Order Cross-Over Design (Differences)
Chapter 533

Tests for Two Means in a Higher-Order Cross-Over Design with Non-Unity Null using Ratios

Introduction

This procedure calculates power and sample size for non-zero null tests which use the ratio of the two means of a higher-order cross-over design. Measurements are made on individuals that have been randomly assigned to one of several treatment sequences. Only a brief introduction to the subject will be given here. For a comprehensive discussion on the subject, refer to Chen et al. (1997) and Chow et al. (2003).

Cross-Over Designs

Senn (2002) defines a cross-over design as one in which each subject receives all treatments at least once and the object is to study differences among the treatments. The name cross-over comes from the most common case in which there are only two treatments. In this case, each subject crosses over from one treatment to another. It is assumed that there is a washout period between treatments during which the response returns to its baseline value. If this does not occur, there is said to be a carryover effect.

A 2x2 cross-over design refers to two treatments (periods) and two sequences (treatment orderings). One sequence of treatments is treatment A followed by treatment B. The other
sequence is B and then A. The design includes a washout period between responses to make certain that the effects of the first drug do no carryover to the second. Thus, the groups of subjects in this design are defined by the sequence in which the two treatments are administered, not by the treatments they receive.

**Higher-Order Cross-Over Designs**

Chen et al. (1997) present the results for four cross-over designs that are more complicated than the 2x2 design. Assume that the two treatments are labeled A and B. The available designs are defined by the order and number of times the two treatments are administered.

**Balaam’s Design**

Balaam’s design has four sequences with two treatments each. It is popular because it allows the intrasubject variabilities to be estimated. The design is

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Period 1</th>
<th>Period 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>A</td>
</tr>
</tbody>
</table>

**Two-Sequence Dual Design**

This design has two sequences with three periods each. It is popular because it allows the intrasubject variabilities to be estimated. The design is

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>

**Four-Period Design with Two Sequences**

This design has two sequences of four periods each. The design is

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
<th>Period 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
</tbody>
</table>

**Four-Period Design with Four Sequences**

This design has four sequences of four periods each. The design is

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
<th>Period 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>1</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
</tbody>
</table>

**Advantages of Cross-Over Designs**

A comparison of treatments on the same subject is expected to be more precise. The increased precision often translates into a smaller sample size. Also, patient enrollment may be easier to obtain because each patient will receive both treatments.
Two Means in a Higher-Order Cross-Over Design with Non-Unity Null using Ratios 533-3

Disadvantages of Cross-Over Designs

The statistical analysis of a cross-over experiment is more complex than a parallel-group experiment and requires additional assumptions. In a cross-over experiment, it may be difficult to separate the treatment effect from the time effect and the carry-over effect of the previous treatment.

These cross-over designs cannot be used when the treatment (or the measurement of the response) alters the subject permanently. Hence, it cannot be used to compare treatments that are intended to provide a cure.

Because subjects must be measured at least twice, it may be more difficult to keep patients enrolled in the study. This is particularly true when the measurement process is painful, uncomfortable, embarrassing, or time consuming.

The Statistical Hypotheses

Both non-inferiority and non-zero null tests are examples of directional (one-sided) tests. Remember that in the usual t-test setting, the null (H0) and alternative (H1) hypotheses for one-sided tests are defined as

\[ H_0; \phi \leq A \text{ versus } H_1; \phi > A \]

Rejecting H0 implies that the ratio of the mean is larger than the value A. This test is called an upper-tailed test because H0 is rejected only in samples in which the ratio of the sample means is larger than A.

Following is an example of a lower-tailed test.

\[ H_0; \phi \geq A \text{ versus } H_1; \phi < A \]

Non-inferiority and non-zero null tests are special cases of the above directional tests. It will be convenient to adopt the following specialize notation for the discussion of these tests.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PASS Input/Output</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu_T )</td>
<td>Not used</td>
<td>Treatment mean. This is the treatment mean.</td>
</tr>
<tr>
<td>( \mu_R )</td>
<td>Not used</td>
<td>Reference mean. This is the mean of a reference population.</td>
</tr>
<tr>
<td>( M_s )</td>
<td>SM</td>
<td>Margin of superiority. This is a tolerance value that defines the magnitude of difference that is required for practical importance. This may be thought of as the smallest difference from the reference that is considered to be different.</td>
</tr>
<tr>
<td>( \phi )</td>
<td>R1</td>
<td>True ratio. This is the value of ( \phi = \frac{\mu_T}{\mu_R} ) at which the power is calculated.</td>
</tr>
</tbody>
</table>

Note that the actual values of \( \mu_T \) and \( \mu_R \) are not needed. Only their ratio is needed for power and sample size calculations.

The null hypothesis of non-superiority is
H₀ : φ ≤ φ₁ where φ₁ > 1.

and the alternative hypothesis of superiority is

H₁ : φ > φ₁

**Log-Transformation**

In many cases, hypotheses stated in terms of ratios are more convenient than hypotheses stated in terms of differences. This is because ratios can be interpreted as scale-less percentages, but differences must be interpreted as actual amounts in their original scale. Hence, it has become a common practice to take the following steps in hypothesis testing.

1. State the statistical hypotheses in terms of ratios.
2. Transform these into hypotheses about differences by taking logarithms.
3. Analyze the logged data—that is, do the analysis in terms of the difference.
4. Draw the conclusion in terms of the ratio.

The details of step 2 for the alternative hypothesis are as follows.

\[ \phi₁ ≤ \phi \]

\[ \Rightarrow \phi₁ ≤ \left\{ \frac{\mu_r}{\mu_R} \right\} \]

\[ \Rightarrow \ln(\phi₁) ≤ \{\ln(\mu_r) - \ln(\mu_R)\} \]

Thus, a hypothesis about the ratio of the means on the original scale can be translated into a hypothesis about the difference of two means on the logged scale.

**Coefficient of Variation**

The coefficient of variation (COV) is the ratio of the standard deviation to the mean. This parameter is used to represent the variation in the data because of a unique relationship that it has in the case of log-normal data.

Suppose the variable \(X\) is the logarithm of the original variable \(Y\). That is, \(X = \ln(Y)\) and \(Y = \exp(X)\). Label the mean and variance of \(X\) as \(\mu_X\) and \(\sigma_X^2\), respectively. Similarly, label the mean and variance of \(Y\) as \(\mu_Y\) and \(\sigma_Y^2\), respectively. If \(X\) is normally distributed, then \(Y\) is log-normally distributed. Julious (2004) presents the following well-known relationships between these two variables

\[ \mu_Y = e^{\mu_X + \frac{\sigma_X^2}{2}} \]

\[ \sigma_Y^2 = \mu_Y^2 \left(e^{\sigma_X^2} - 1\right) \]

From this relationship, the coefficient of variation of \(Y\) can be found to be
Two Means in a Higher-Order Cross-Over Design with Non-Unity Null using Ratios

\[ COV_Y = \frac{\mu_Y}{\mu_Y^2 (e^{\sigma_Y^2} - 1)} \]
 \[ = \sqrt{e^{\sigma_Y^2} - 1} \]
 \[ = \sqrt{e^{\sigma_X^2} - 1} \]

where \( \sigma_w^2 \) is the within mean square error from the analysis of variance of the logged data.

Solving this relationship for \( \sigma_X^2 \), the standard deviation of \( X \) can be stated in terms of the coefficient of variation of \( Y \). This equation is

\[ \sigma_X = \sqrt{\ln(COV_Y^2 + 1)} \]

Similarly, the mean of \( X \) is

\[ \mu_X = \frac{\mu_Y}{\ln(COV_Y^2 + 1)} \]

Thus, the hypotheses can be stated in the original (Y) scale and then power analyzed in the transformed (X) scale.

**Non-Inferiority and Superiority Tests**

A *non-zero null test* tests that the treatment mean is better than the reference mean by more than a small superiority margin. The actual direction of the hypothesis depends on the response variable being studied.

**Case 3: High Values Good, Superiority Test**

In this case, higher values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is greater than the reference mean by at least the margin of superiority. The null and alternative hypotheses are

\[ H_0: \frac{\mu_T}{\mu_R} \leq (1 + \varepsilon) \quad \text{versus} \quad H_1: \frac{\mu_T}{\mu_R} > (1 + \varepsilon) \]

\[ H_0: \ln(\mu_T) - \ln(\mu_R) \leq \ln(1 + \varepsilon) \quad \text{versus} \quad H_1: \ln(\mu_T) - \ln(\mu_R) > \ln(1 + \varepsilon) \]

**Case 4: High Values Bad, Superiority Test**

In this case, lower values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is less than the reference mean by at least the margin of superiority. The null and alternative hypotheses are

\[ H_0: \frac{\mu_T}{\mu_R} \geq (1 - \varepsilon) \quad \text{versus} \quad H_1: \frac{\mu_T}{\mu_R} < (1 - \varepsilon) \]

\[ H_0: \ln(\mu_T) - \ln(\mu_R) \geq \ln(1 - \varepsilon) \quad \text{versus} \quad H_1: \ln(\mu_T) - \ln(\mu_R) < \ln(1 - \varepsilon) \]
Test Statistics

The analysis for assessing non-inferiority and non-zero null hypotheses using higher-order cross-over designs is discussed in detail in Chapter 9 of Chow and Liu (2000). Unfortunately, their presentation is too lengthy to give here. Their method involves the computation of an analysis of variance to estimate the error variance. It also describes the construction of confidence limits for appropriate contrasts. One-sided confidence limits can be used for non-inferiority tests. Details of this approach are given in Chapter 3 of Chow et al. (2003). We refer you to these books for details.

Power Calculation

The power of the non-inferiority and superiority tests for the case in which higher values are better is given by

\[
Power = T_{t} \left( \frac{\ln(1 - \varepsilon)}{\sigma \sqrt{b/n}} - t_{V,1-\alpha} \right)
\]

where \( T \) represents the cumulative \( t \) distribution, \( V \) and \( b \) depend on the design, \( n \) is the average number of subjects per sequence, and

\[
\sigma = \sqrt{\ln(COV^2 + 1)}
\]

The power of the non-inferiority and superiority tests for the case in which higher values are worse is given by

\[
Power = 1 - T_{t} \left( t_{V,1-\alpha} - \frac{-\ln(1 + \varepsilon)}{\sigma \sqrt{b/n}} \right)
\]

The constants \( V \) and \( b \) depend on the design as follows:

<table>
<thead>
<tr>
<th>Design Type</th>
<th>Parameters ((V,b))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balaam’s Design</td>
<td>( V = 4n - 3, b = 2 ).</td>
</tr>
<tr>
<td>Two-Sequence Dual Design</td>
<td>( V = 4n - 4, b = 3/4 ).</td>
</tr>
<tr>
<td>Four-Period Design with Two Sequences</td>
<td>( V = 6n - 5, b = 11/20 ).</td>
</tr>
<tr>
<td>Four-Period Design with Four Sequences</td>
<td>( V = 12n - 5, b = 1/4 ).</td>
</tr>
</tbody>
</table>
Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains the parameters associated with this test such as the means, sample sizes, alpha, and power.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for from the other parameters. Under most situations, you will select either Power and Beta for a power analysis or N for sample size determination.

Select N when you want to calculate the sample size needed to achieve a given power and alpha level. Note that there are two choices for finding N. Select N (Equal Per Sequence) when you want the design to have an equal number of subjects per sequence. Select N (Exact) when you want to find the exact sample size even though the number of subjects cannot be dividing equally among the sequences.

Select Power and Beta when you want to calculate the power of an experiment.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.
Sample Size

N (Total Sample Size)
This option specifies one or more values of the sample size, the number of individuals in the study (total subjects in all sequences). These values must be integers greater than one.

Effect Size – Ratios

SM (Superiority Margin)
This is the magnitude of the margin of superiority. It must be entered as a positive number.

When higher means are better, this value is the distance above one that is required for the mean ratio (Treatment Mean / Reference Mean) to be considered superior. When higher means are worse, this value is the distance below one that is required for the mean ratio (Treatment Mean / Reference Mean) to be considered superior.

R1 (True Ratio)
This is the value of the ratio of the two means (Treatment Mean / Reference Mean) at which the power is to be calculated.

When higher means are better, this value should be greater than 1+SM. When higher means are worse, this value should be less than 1-SM.

Effect Size – Coefficient of Variation

COV (Coefficient of Variation)
The coefficient of variation is the ratio of the standard deviation and the mean (SD/Mean). It is used to specify the variability (standard deviation). Note that this COV is defined on the original (not logarithmic) scale. This value must be determined from past experience or from a pilot study.

To be clear, consider the following definition. Suppose data on a response variable Y are collected. This procedure assumes that the values of X = Ln(Y) are analyzed using an appropriate ANOVA procedure. Thus, there are two sets of means and standard deviations: those of X labelled MX and SX and those of Y labelled MY and SY. The COV entered here is the COV of Y = SY/MY. For log-normal data, the following relationship exists: COV(Y) = SQR(Exp(SX*SX)-1) where SX is the square root of the within mean square error in the ANOVA table of the log-transformed values.

Test

Design Type
Specify the type of cross-over design that you are analyzing. Note that all of these designs assume that you are primarily interested in the overall difference between the two treatment means.

Higher Means Are
This option defines whether higher values of the response variable are to be considered better or worse. The choice here determines the direction of the test.

If Higher Means Are Better the null hypothesis is R <= 1+SM and the alternative hypothesis is R > 1+SM. If Higher Means Are Worse the null hypothesis is R >= 1-SM and the alternative hypothesis is R < 1-SM.
Example 1 – Finding Power

A company has developed a generic drug for treating rheumatism and wants to show that it is superior to a standard drug by a small amount. Balaam’s cross-over design will be used.

Researchers have decided to set the margin of superiority at 0.20. Past experience leads the researchers to set the COV to 0.40. The significance level is 0.05. The power will be computed assuming that the true ratio is 1.40. Sample sizes between 50 and 550 will be included in the analysis. Note that several of these sample size values are not divisible by 4. This is not a problem here because are main goal is to get an overview of power versus sample size. When searching for the sample size, we can request that only designs divisible by 4 be considered.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means in a Higher-Order Cross-Over Design (Non-Unity Null) [Ratios] procedure window by expanding Means, then MxK Cross-Over Design, then clicking on Test (Non-Zero Null), and then clicking on Tests for Two Means in a Higher-Order Cross-Over Design (Non-Unity Null) [Ratios]. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>50 to 550 by 100</td>
</tr>
<tr>
<td>SM (Superiority Margin)</td>
<td>0.20</td>
</tr>
<tr>
<td>R1 (True Ratio)</td>
<td>1.40</td>
</tr>
<tr>
<td>COV (Coefficient of Variation)</td>
<td>0.40</td>
</tr>
<tr>
<td>Design Type</td>
<td>4x2 (Balaam)</td>
</tr>
<tr>
<td>Higher Means Are</td>
<td>Better</td>
</tr>
</tbody>
</table>


Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>Total Sample Size (N)</th>
<th>Sequences and Periods (SxP)</th>
<th>Superiority Margin (SM)</th>
<th>Mean Ratio for Power (R1)</th>
<th>Coef. of Variation (COV)</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.2507</td>
<td>50</td>
<td>4x2</td>
<td>0.20</td>
<td>1.40</td>
<td>0.40</td>
<td>0.0500</td>
<td>0.7493</td>
</tr>
<tr>
<td>0.5308</td>
<td>150</td>
<td>4x2</td>
<td>0.20</td>
<td>1.40</td>
<td>0.40</td>
<td>0.0500</td>
<td>0.4692</td>
</tr>
<tr>
<td>0.7207</td>
<td>250</td>
<td>4x2</td>
<td>0.20</td>
<td>1.40</td>
<td>0.40</td>
<td>0.0500</td>
<td>0.2793</td>
</tr>
<tr>
<td>0.8404</td>
<td>350</td>
<td>4x2</td>
<td>0.20</td>
<td>1.40</td>
<td>0.40</td>
<td>0.0500</td>
<td>0.1596</td>
</tr>
<tr>
<td>0.9116</td>
<td>450</td>
<td>4x2</td>
<td>0.20</td>
<td>1.40</td>
<td>0.40</td>
<td>0.0500</td>
<td>0.0884</td>
</tr>
<tr>
<td>0.9523</td>
<td>550</td>
<td>4x2</td>
<td>0.20</td>
<td>1.40</td>
<td>0.40</td>
<td>0.0500</td>
<td>0.0477</td>
</tr>
</tbody>
</table>

Report Definitions

H0 (null hypothesis) is \( R \leq 1 + SM \), where \( R = \text{Treatment Mean} / \text{Reference Mean} \).
H1 (alternative hypothesis) is \( R > 1 + SM \).
Power is the probability of rejecting H0 when H0 is false.
N is the total number of subjects. They are divided evenly among all sequences.
SM is the magnitude of the margin of superiority. Since higher means are better, this value is positive and is the distance above one that is required to be considered superior.
R1 is the ratio of the means at which the power is computed.
COV is the coefficient of variation on the original scale.
Alpha is the probability of falsely rejecting H0 (falsely concluding superiority).
Beta is the probability of not rejecting H0 when it is false.
Balaam's Cross-Over Design with pattern: AA; BB; AB; BA

Summary Statements

In a superiority test on data for which higher values are better, drawn from Balaam's cross-over design, a total sample size of 50 achieves 25% power at a 5% significance level when the true ratio of the means is 1.40, the coefficient of variation is 0.40, and the superiority margin is 0.20.

This report shows the power for the indicated scenarios.

Plots Section

This plot shows the power versus the sample size.
Example 2 – Finding Sample Size

Continuing with Example 1, the researchers want to find the exact sample size needed to achieve both 80% and 90% power.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Tests for Two Means in a Higher-Order Cross-Over Design (Non-Unity Null) [Ratios] procedure window by expanding Means, then MxK Cross-Over Design, then clicking on Test (Non-Zero Null), and then clicking on Tests for Two Means in a Higher-Order Cross-Over Design (Non-Unity Null) [Ratios]. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N (Equal Per Sequence)</td>
</tr>
<tr>
<td>Power</td>
<td>0.8 0.9</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>SM (Superiority Margin)</td>
<td>0.20</td>
</tr>
<tr>
<td>R1 (True Ratio)</td>
<td>1.40</td>
</tr>
<tr>
<td>COV (Coefficient of Variation)</td>
<td>0.40</td>
</tr>
<tr>
<td>Design Type</td>
<td>4x2 (Balaam)</td>
</tr>
<tr>
<td>Higher Means Are</td>
<td>Better</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th></th>
<th>Power</th>
<th>Total Sample Size</th>
<th>Sequences and Periods</th>
<th>Superiority Margin</th>
<th>Mean Ratio for Power</th>
<th>Coef. of Variation</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Numeric Results for Testing Non-Inferiority Using the Mean Ratio</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Higher Means are Better</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Design: Balaam’s Cross-Over. Hypotheses: H0: R &lt;= 1+SM; H1: R &gt; 1+SM.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.9015</td>
<td>432</td>
<td>4x2</td>
<td>0.20</td>
<td>1.40</td>
<td>0.40</td>
<td>0.0500</td>
<td>0.0985</td>
</tr>
<tr>
<td></td>
<td>0.8017</td>
<td>312</td>
<td>4x2</td>
<td>0.20</td>
<td>1.40</td>
<td>0.40</td>
<td>0.0500</td>
<td>0.1983</td>
</tr>
</tbody>
</table>

When the equivalence margin is set to 0.20, we note that 432 subjects are needed to achieve 90% power and 312 subjects are needed to achieve at least 80% power.
Example 3 – Validation

This procedure uses the same mechanics as the Non-Inferiority Tests for Two Means in a Higher-Order Cross-Over Design using Ratios procedure. We refer the user to Example 3 of Chapter 535 for the validation.
Chapter 535

Non-Inferiority Tests for Two Means in a Higher-Order Cross-Over Design using Ratios

Introduction

This procedure calculates power and sample size for non-inferiority tests which use the ratio of the two means of a higher-order cross-over design. Measurements are made on individuals that have been randomly assigned to one of several treatment sequences. Only a brief introduction to the subject will be given here. For a comprehensive discussion on the subject, refer to Chen et al. (1997) and Chow et al. (2003).

Cross-Over Designs

Senn (2002) defines a cross-over design as one in which each subject receives all treatments at least once and the object is to study differences among the treatments. The name cross-over comes from the most common case in which there are only two treatments. In this case, each subject crosses over from one treatment to another. It is assumed that there is a washout period between treatments during which the response returns to its baseline value. If this does not occur, there is said to be a carryover effect.

A 2x2 cross-over design refers to two treatments (periods) and two sequences (treatment orderings). One sequence of treatments is treatment A followed by treatment B. The other
sequence is B and then A. The design includes a washout period between responses to make certain that the effects of the first drug do no carryover to the second. Thus, the groups of subjects in this design are defined by the sequence in which the two treatments are administered, not by the treatments they receive.

Higher-Order Cross-Over Designs

Chen et al. (1997) present the results for four cross-over designs that are more complicated than the 2x2 design. Assume that the two treatments are labeled A and B. The available designs are defined by the order and number of times the two treatments are administered.

Balaam’s Design

Balaam’s design has four sequences with two treatments each. It is popular because it allows the intrasubject variabilities to be estimated. The design is

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Period 1</th>
<th>Period 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>A</td>
</tr>
</tbody>
</table>

Two-Sequence Dual Design

This design has two sequences with three periods each. It is popular because it allows the intrasubject variabilities to be estimated. The design is

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>

Four-Period Design with Two Sequences

This design has two sequences of four periods each. The design is

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
<th>Period 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
</tbody>
</table>

Four-Period Design with Four Sequences

This design has four sequences of four periods each. The design is

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
<th>Period 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>1</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
</tbody>
</table>

Advantages of Cross-Over Designs

A comparison of treatments on the same subject is expected to be more precise. The increased precision often translates into a smaller sample size. Also, patient enrollment may be easier to obtain because each patient will receive both treatments.
Disadvantages of Cross-Over Designs

The statistical analysis of a cross-over experiment is more complex than a parallel-group experiment and requires additional assumptions. In a cross-over experiment, it may be difficult to separate the treatment effect from the time effect and the carry-over effect of the previous treatment.

These cross-over designs cannot be used when the treatment (or the measurement of the response) alters the subject permanently. Hence, it cannot be used to compare treatments that are intended to provide a cure.

Because subjects must be measured at least twice, it may be more difficult to keep patients enrolled in the study. This is particularly true when the measurement process is painful, uncomfortable, embarrassing, or time consuming.

The Statistical Hypotheses

Both non-inferiority and non-zero null tests are examples of directional (one-sided) tests. Remember that in the usual t-test setting, the null (H0) and alternative (H1) hypotheses for one-sided tests are defined as

\[ H_0: \phi \leq A \quad \text{versus} \quad H_1: \phi > A \]

Rejecting \( H_0 \) implies that the ratio of the mean is larger than the value \( A \). This test is called an upper-tailed test because \( H_0 \) is rejected only in samples in which the ratio of the sample means is larger than \( A \).

Following is an example of a lower-tailed test.

\[ H_0: \phi \geq A \quad \text{versus} \quad H_1: \phi < A \]

Non-inferiority and non-zero null tests are special cases of the above directional tests. It will be convenient to adopt the following specialize notation for the discussion of these tests.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PASS Input/Output</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu_T )</td>
<td>Not used</td>
<td>Treatment mean. This is the treatment mean.</td>
</tr>
<tr>
<td>( \mu_R )</td>
<td>Not used</td>
<td>Reference mean. This is the mean of a reference population.</td>
</tr>
<tr>
<td>( M_{NI} )</td>
<td>NIM</td>
<td>Margin of non-inferiority. This is a tolerance value that defines the maximum amount that is not of practical importance. This is the largest change in the mean ratio from the baseline value (usually one) that is still considered to be trivial.</td>
</tr>
<tr>
<td>( \phi )</td>
<td>R1</td>
<td>True ratio. This is the value of ( \phi = \frac{\mu_T}{\mu_R} ) at which the power is calculated.</td>
</tr>
</tbody>
</table>

Note that the actual values of \( \mu_T \) and \( \mu_R \) are not needed. Only their ratio is needed for power and sample size calculations.

The null hypothesis of inferiority is

\[ H_0: \phi \leq \phi_L \quad \text{where} \quad \phi_L < 1. \]
and the alternative hypothesis of non-inferiority is

\[ H_1: \phi > \phi_L \]

### Log-Transformation

In many cases, hypotheses stated in terms of ratios are more convenient than hypotheses stated in terms of differences. This is because ratios can be interpreted as scale-less percentages, but differences must be interpreted as actual amounts in their original scale. Hence, it has become a common practice to take the following steps in hypothesis testing.

1. State the statistical hypotheses in terms of ratios.
2. Transform these into hypotheses about differences by taking logarithms.
3. Analyze the logged data—that is, do the analysis in terms of the difference.
4. Draw the conclusion in terms of the ratio.

The details of step 2 for the alternative hypothesis are as follows.

\[
\phi_L \leq \phi \\
\Rightarrow \phi_L \leq \left\{ \frac{\mu_L}{\mu_R} \right\} \\
\Rightarrow \ln(\phi_L) \leq \left\{ \ln(\mu_L) - \ln(\mu_R) \right\}
\]

Thus, a hypothesis about the ratio of the means on the original scale can be translated into a hypothesis about the difference of two means on the logged scale.

### Coefficient of Variation

The coefficient of variation (COV) is the ratio of the standard deviation to the mean. This parameter is used to represent the variation in the data because of a unique relationship that it has in the case of log-normal data.

Suppose the variable \( X \) is the logarithm of the original variable \( Y \). That is, \( X = \ln(Y) \) and \( Y = \exp(X) \). Label the mean and variance of \( X \) as \( \mu_X \) and \( \sigma^2_X \), respectively. Similarly, label the mean and variance of \( Y \) as \( \mu_Y \) and \( \sigma^2_Y \), respectively. If \( X \) is normally distributed, then \( Y \) is log-normally distributed. Julious (2004) presents the following well-known relationships between these two variables

\[
\mu_Y = \left( e^{\mu_X + \frac{\sigma^2_X}{2}} \right) \\
\sigma^2_Y = \mu_Y^2 \left( e^{\sigma^2_X} - 1 \right)
\]

From this relationship, the coefficient of variation of \( Y \) can be found to be
where \( \sigma_w^2 \) is the within mean square error from the analysis of variance of the logged data. Solving this relationship for \( \sigma_X^2 \), the standard deviation of \( X \) can be stated in terms of the coefficient of variation of \( Y \). This equation is

\[
\sigma_X = \sqrt{\ln(COV_y^2 + 1)}
\]

Similarly, the mean of \( X \) is

\[
\mu_X = \frac{\mu_Y}{\ln(COV_y^2 + 1)}
\]

Thus, the hypotheses can be stated in the original (\( Y \)) scale and then power analyzed in the transformed (\( X \)) scale.

Non-Inferiority Tests

A non-inferiority test tests that the treatment mean is not worse than the reference mean by more than a small non-inferiority margin. The actual direction of the hypothesis depends on the response variable being studied.

**Case 1: High Values Good, Non-Inferiority Test**

In this case, higher values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is no less than a small amount below the reference mean. The null and alternative hypotheses are

\[
H_0: \frac{\mu_T}{\mu_R} \leq (1 - \varepsilon) \quad \text{versus} \quad H_1: \frac{\mu_T}{\mu_R} > (1 - \varepsilon)
\]

\[
H_0: \ln(\mu_T) - \ln(\mu_R) \leq \ln(1 - \varepsilon) \quad \text{versus} \quad H_1: \ln(\mu_T) - \ln(\mu_R) > \ln(1 - \varepsilon)
\]

**Case 2: High Values Bad, Non-Inferiority Test**

In this case, lower values are better. The hypotheses are arranged so that rejecting the null hypothesis implies that the treatment mean is no more than a small amount above the reference mean. The null and alternative hypotheses are

\[
H_0: \frac{\mu_T}{\mu_R} \geq (1 + \varepsilon) \quad \text{versus} \quad H_1: \frac{\mu_T}{\mu_R} < (1 + \varepsilon)
\]

\[
H_0: \ln(\mu_T) - \ln(\mu_R) \geq \ln(1 + \varepsilon) \quad \text{versus} \quad H_1: \ln(\mu_T) - \ln(\mu_R) < \ln(1 + \varepsilon)
\]
Test Statistics

The analysis for assessing non-inferiority using higher-order cross-over designs is discussed in detail in Chapter 9 of Chow and Liu (2000). Unfortunately, their presentation is too lengthy to give here. Their method involves the computation of an analysis of variance to estimate the error variance. It also describes the construction of confidence limits for appropriate contrasts. One-sided confidence limits can be used for non-inferiority tests. Details of this approach are given in Chapter 3 of Chow et al. (2003). We refer you to these books for details.

Power Calculation

The power of the non-inferiority and superiority tests for the case in which higher values are better is given by

$$Power = T_p \left( \frac{\ln(1 - \varepsilon)}{\sigma_w \sqrt{b / n}} - t_{n-1, \alpha} \right)$$

where $T$ represents the cumulative $t$ distribution, $V$ and $b$ depend on the design, $n$ is the average number of subjects per sequence, and

$$\sigma_w = \sqrt{\ln(\text{COV}_y^2 + 1)}$$

The power of the non-inferiority and superiority tests for the case in which higher values are worse is given by

$$Power = 1 - T_p \left( t_{n-1, \alpha} - \frac{-\ln(1 + \varepsilon)}{\sigma_w \sqrt{b / n}} \right)$$

The constants $V$ and $b$ depend on the design as follows:

<table>
<thead>
<tr>
<th>Design Type</th>
<th>Parameters $(V, b)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balaam’s Design</td>
<td>$V = 4n - 3, b = 2$.</td>
</tr>
<tr>
<td>Two-Sequence Dual Design</td>
<td>$V = 4n - 4, b = 3/4$.</td>
</tr>
<tr>
<td>Four-Period Design with Two Sequences</td>
<td>$V = 6n - 5, b = 11/20$.</td>
</tr>
<tr>
<td>Four-Period Design with Four Sequences</td>
<td>$V = 12n - 5, b = 1/4$.</td>
</tr>
</tbody>
</table>
Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains the parameters associated with this test such as the means, sample sizes, alpha, and power.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for from the other parameters. Under most situations, you will select either Power and Beta for a power analysis or N for sample size determination.

Select N when you want to calculate the sample size needed to achieve a given power and alpha level. Note that there are two choices for finding N. Select N (Equal Per Sequence) when you want the design to have an equal number of subjects per sequence. Select N (Exact) when you want to find the exact sample size even though the number of subjects cannot be dividing equally among the sequences.

Select Power and Beta when you want to calculate the power of an experiment.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.
Sample Size

N (Total Sample Size)
This option specifies one or more values of the sample size, the number of individuals in the study (total subjects in all sequences). These values must be integers greater than one.

Effect Size – Ratios

NIM (Non-Inferiority Margin)
This is the magnitude of the margin of non-inferiority. It must be entered as a positive number.

When higher means are better, this value is the distance below one for which the mean ratio (Treatment Mean / Reference Mean) still indicates non-inferiority of the treatment mean. E.g., a value of 0.2 here specifies that mean ratios greater than 0.8 indicate non-inferiority of the treatment mean.

When higher means are worse, this value is the distance above one for which the mean ratio (Treatment Mean / Reference Mean) still indicates non-inferiority of the treatment mean. E.g., a value of 0.2 here specifies that mean ratios less than 1.2 indicate non-inferiority of the treatment mean.

Recommended values: 0.20 is a common value for the parameter.

R1 (True Ratio)
This is the value of the ratio of the two means (Treatment Mean / Reference Mean) at which the power is to be calculated.

Often, the ratio will be set to one. However, some authors recommend using a ratio slightly different than one, such as 0.95 (when higher means are "better") or 1.05 (when higher means are "worse"), since this will require a larger sample size.

Effect Size – Coefficient of Variation

COV (Coefficient of Variation)
The coefficient of variation is the ratio of the standard deviation and the mean (SD/Mean). It is used to specify the variability (standard deviation). Note that this COV is defined on the original (not logarithmic) scale. This value must be determined from past experience or from a pilot study.

To be clear, consider the following definition. Suppose data on a response variable Y are collected. This procedure assumes that the values of X = Ln(Y) are analyzed using an appropriate ANOVA procedure. Thus, there are two sets of means and standard deviations: those of X labelled MX and SX and those of Y labelled MY and SY. The COV entered here is the COV of Y = SY/MY. For log-normal data, the following relationship exists: COV(Y) = SQR(Exp(SX*SX)-1) where SX is the square root of the within mean square error in the ANOVA table of the log-transformed values.

Test

Design Type
Specify the type of cross-over design that you are analyzing. Note that all of these designs assume that you are primarily interested in the overall difference between the two treatment means.
Higher Means Are
This option defines whether higher values of the response variable are to be considered better or worse. The choice here determines the direction of the non-inferiority test.

If Higher Means Are Better the null hypothesis is $R \leq 1-NIM$ and the alternative hypothesis is $R > 1-NIM$. If Higher Means Are Worse the null hypothesis is $R \geq 1+NIM$ and the alternative hypothesis is $R < 1+NIM$.

Example 1 – Finding Power
A company has developed a generic drug for treating rheumatism and wants to show that it is not inferior to standard drug. Balaam’s cross-over design will be used.

Researchers have decided to set the margin of equivalence at 0.20. Past experience leads the researchers to set the COV to 0.40. The significance level is 0.05. The power will be computed assuming that the true ratio is one. Sample sizes between 50 and 550 will be included in the analysis. Note that several of these sample size values are not divisible by 4. This is not a problem here because the main goal is to get an overview of power versus sample size. When searching for the sample size, we can request that only designs divisible by 4 be considered.

Setup
This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Non-Inferiority Tests for Two Means in a Higher-Order Cross-Over Design [Ratios] procedure window by expanding Means, then MxK Cross-Over Design, then clicking on Non-Inferiority, and then clicking on Non-Inferiority Tests for Two Means in a Higher-Order Cross-Over Design [Ratios]. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td></td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Power</td>
<td></td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>50 to 550 by 100</td>
</tr>
<tr>
<td>NIM (Non-Inferiority Margin)</td>
<td>0.20</td>
</tr>
<tr>
<td>R1 (True Ratio)</td>
<td>1</td>
</tr>
<tr>
<td>COV (Coefficient of Variation)</td>
<td>0.40</td>
</tr>
<tr>
<td>Design Type</td>
<td>4x2 (Balaam)</td>
</tr>
<tr>
<td>Higher Means Are</td>
<td>Better</td>
</tr>
</tbody>
</table>
Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>Total Sample Size (N)</th>
<th>Sequences and Periods (SxP)</th>
<th>Non-Inferiority Margin (-NIM)</th>
<th>Mean Ratio for Power (R1)</th>
<th>Coef. of Variation (COV)</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.4096</td>
<td>50</td>
<td>4x2</td>
<td>0.20</td>
<td>1.00</td>
<td>0.40</td>
<td>0.0500</td>
<td>0.5904</td>
</tr>
<tr>
<td>0.8024</td>
<td>150</td>
<td>4x2</td>
<td>0.20</td>
<td>1.00</td>
<td>0.40</td>
<td>0.0500</td>
<td>0.1976</td>
</tr>
<tr>
<td>0.9853</td>
<td>250</td>
<td>4x2</td>
<td>0.20</td>
<td>1.00</td>
<td>0.40</td>
<td>0.0500</td>
<td>0.0562</td>
</tr>
<tr>
<td>0.9964</td>
<td>350</td>
<td>4x2</td>
<td>0.20</td>
<td>1.00</td>
<td>0.40</td>
<td>0.0500</td>
<td>0.0147</td>
</tr>
<tr>
<td>0.9992</td>
<td>450</td>
<td>4x2</td>
<td>0.20</td>
<td>1.00</td>
<td>0.40</td>
<td>0.0500</td>
<td>0.0036</td>
</tr>
</tbody>
</table>

Report Definitions

H0 (null hypothesis) is $R \leq 1-NIM$, where $R$ = Treatment Mean / Reference Mean.
H1 (alternative hypothesis) is $R > 1-NIM$.
Power is the probability of rejecting H0 (concluding non-inferiority) when H0 is false.
N is the total number of subjects. They are divided evenly among all sequences.
-NIM is the magnitude and direction of the margin of non-inferiority. Since higher means are better, this value is negative and is the distance below one that is still considered non-inferior.
R1 is the ratio of the means at which the power is computed.
COV is the coefficient of variation on the original scale.
Alpha is the probability of falsely rejecting H0 (falsely concluding non-inferiority).
Beta is the probability of not rejecting H0 when it is false.
Balaam's Cross-Over Design with pattern: AA; BB; AB; BA

Summary Statements

In a non-inferiority test on data for which higher values are better, drawn from Balaam's cross-over design, a total sample size of 50 achieves 41% power at a 5% significance level when the true ratio of the means is 1.00, the coefficient of variation is 0.40, and the non-inferiority margin is -0.20.

This report shows the power for the indicated scenarios.

Plots Section

This plot shows the power versus the sample size.
Example 2 – Finding Sample Size

Continuing with Example 1, the researchers want to find the exact sample size needed to achieve both 80% and 90% power.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Non-Inferiority Tests for Two Means in a Higher-Order Cross-Over Design [Ratios] procedure window by expanding Means, then MxK Cross-Over Design, then clicking on Non-Inferiority, and then clicking on Non-Inferiority Tests for Two Means in a Higher-Order Cross-Over Design [Ratios]. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N (Equal Per Sequence)</td>
</tr>
<tr>
<td>Power</td>
<td>0.8 0.9</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>NIM (Non-Inferiority Margin)</td>
<td>0.20</td>
</tr>
<tr>
<td>R1 (True Ratio)</td>
<td>1</td>
</tr>
<tr>
<td>COV (Coefficient of Variation)</td>
<td>0.40</td>
</tr>
<tr>
<td>Design Type</td>
<td>4x2 (Balaam)</td>
</tr>
<tr>
<td>Higher Means Are</td>
<td>Better</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>Total Sample Size (N)</th>
<th>Sequences and Periods (SxP)</th>
<th>Non-Inferiority Margin (-NIM)</th>
<th>Mean Ratio for Power (R1)</th>
<th>Coef. of Variation (COV)</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9027</td>
<td>208 4x2</td>
<td></td>
<td>0.20</td>
<td>1.00</td>
<td>0.40</td>
<td>0.0500</td>
<td>0.0973</td>
</tr>
<tr>
<td>0.8070</td>
<td>152 4x2</td>
<td></td>
<td>0.20</td>
<td>1.00</td>
<td>0.40</td>
<td>0.0500</td>
<td>0.1930</td>
</tr>
</tbody>
</table>

When the non-inferiority margin is set to 0.20, we note that 208 subjects are needed to achieve 90% power and 152 subjects are needed to achieve at least 80% power.
Example 3 – Validation

We could not find a validation example for this procedure in the statistical literature, so we will have to generate a validated example from within PASS. To do this, we use the Higher-Order, Cross-Over Equivalence using Ratios procedure which was validated. By setting the upper equivalence limit to a large value (we used 11), we obtain results for a non-inferiority test that can be used to validate this procedure.

In the other procedure, suppose the coefficient of variation is 0.40, the equivalence limits are 0.80 and 11.0, the true ratio of the means is 1, the power is 90%, and the significance level is 0.05. These settings are stored as Example4 in that procedure. PASS calculates a sample size of 208.

We will now setup this example in PASS. The only difference is that now we set E to 0.2 instead of RL to 0.8.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Non-Inferiority Tests for Two Means in a Higher-Order Cross-Over Design [Ratios] procedure window by expanding Means, then MxK Cross-Over Design, then clicking on Non-Inferiority, and then clicking on Non-Inferiority Tests for Two Means in a Higher-Order Cross-Over Design [Ratios]. You may then make the appropriate entries as listed below, or open Example 3 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N (Equal Per Sequence)</td>
</tr>
<tr>
<td>Power</td>
<td>0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>NIM (Non-Inferiority Margin)</td>
<td>0.2</td>
</tr>
<tr>
<td>R1 (True Ratio)</td>
<td>1.0</td>
</tr>
<tr>
<td>COV (Coefficient of Variation)</td>
<td>0.40</td>
</tr>
<tr>
<td>Design Type</td>
<td>4x2 (Balaam)</td>
</tr>
<tr>
<td>Higher Means Are</td>
<td>Better</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

<table>
<thead>
<tr>
<th>Power</th>
<th>Total Sample Size</th>
<th>Sequences and Periods</th>
<th>Equivalence Margin</th>
<th>Mean Ratio for Power</th>
<th>Coef. of Variation</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9027</td>
<td>208</td>
<td>4x2</td>
<td>0.20</td>
<td>1.00</td>
<td>0.40</td>
<td>0.0500</td>
<td>0.0973</td>
</tr>
</tbody>
</table>

PASS has also obtained the sample size of 208.
Chapter 540

Equivalence Tests for Two Means in a Higher-Order Cross-Over Design using Differences

Introduction

This procedure calculates power and sample size of statistical tests of equivalence of two means of higher-order cross-over designs when the analysis uses a t-test or equivalent. The parameter of interest is the difference of the two means. Schuirmann’s (1987) two one-sided tests (TOST) approach is used to test equivalence. Only a brief introduction to the subject will be given here. For a comprehensive discussion on the subject, refer to Chen, Chow, and Li (1997). The designs covered in this chapter are analyzed using what is called the ‘additive model’ in Chen et al (1997). The ‘multiplicative model’ is covered in the procedure that uses ratios.

Measurements are made on individuals that have been randomly assigned to one of several treatment sequences. This cross-over design may be analyzed by a TOST equivalence test to show that the two means do not differ by more than a small amount, called the margin of equivalence.
Cross-Over Designs

Senn (2002) defines a cross-over design as one in which each subject receives all treatments at least once and the object is to study differences among the treatments. The name cross-over comes from the most common case in which there are only two treatments. In this case, each subject crosses over from one treatment to another. It is assumed that there is a washout period between treatments during which the response returns to its baseline value. If this does not occur, there is said to be a carryover effect.

A 2x2 cross-over design refers to two treatments (periods) and two sequences (treatment orderings). One sequence of treatments is treatment A followed by treatment B. The other sequence is B and then A. The design includes a washout period between responses to make certain that the effects of the first drug do no carryover to the second. Thus, the groups of subjects in this design are defined by the sequence in which the two treatments are administered, not by the treatments they receive.

Higher-Order Cross-Over Designs

Chen et al. (1997) present the results for four cross-over designs that are more complicated than the 2x2 design. Assume that the two treatments are labeled A and B. The available designs are defined by the order and number of times the two treatments are administered.

Balaam’s Design

Balaam’s design has four sequences with two treatments each. It is popular because it allows the intrasubject variabilities to be estimated. The design is

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Period 1</th>
<th>Period 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>A</td>
</tr>
</tbody>
</table>

Two-Sequence Dual Design

This design has two sequences with three periods each. It is popular because it allows the intrasubject variabilities to be estimated. The design is

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>

Four-Period Design with Two Sequences

This design has two sequences of four periods each. The design is

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
<th>Period 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
</tbody>
</table>
Four-Period Design with Four Sequences

This design has four sequences of four periods each. The design is

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
<th>Period 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>1</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
</tbody>
</table>

Advantages of Cross-Over Designs

A comparison of treatments on the same subject is expected to be more precise. The increased precision often translates into a smaller sample size. Also, patient enrollment may be easier to obtain because each patient will receive both treatments.

Disadvantages of Cross-Over Designs

The statistical analysis of a cross-over experiment is more complex than a parallel-group experiment and requires additional assumptions. In a cross-over experiment, it may be difficult to separate the treatment effect from the time effect and the carry-over effect of the previous treatment.

These cross-over designs cannot be used when the treatment (or the measurement of the response) alters the subject permanently. Hence, it cannot be used to compare treatments that are intended to provide a cure.

Because subjects must be measured at least twice, it may be more difficult to keep patients enrolled in the study. This is particularly true when the measurement process is painful, uncomfortable, embarrassing, or time consuming.
Outline of an Equivalence Test

**PASS** follows the *two one-sided tests* approach described by Schuirmann (1987) and Phillips (1990). It will be convenient to adopt the following specialized notation for the discussion of these tests.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PASS Input/Output</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu_T )</td>
<td>Not used</td>
<td><em>Treatment mean</em>. This is the treatment mean.</td>
</tr>
<tr>
<td>( \mu_R )</td>
<td>Not used</td>
<td><em>Reference mean</em>. This is the mean of a reference population.</td>
</tr>
<tr>
<td>( \varepsilon )</td>
<td>[E]</td>
<td><em>Margin of equivalence</em>. This is a tolerance value that defines the maximum difference that is not of practical importance. This may be thought of as the largest change from the baseline that is considered to be trivial. The absolute value is shown to emphasize that this is a magnitude. The sign of the value will be determined by the specific design that is being used.</td>
</tr>
<tr>
<td>( \delta )</td>
<td>D</td>
<td><em>True difference</em>. This is the value of ( \mu_T - \mu_R ), the difference between the treatment and reference means. This is the value at which the power is calculated.</td>
</tr>
</tbody>
</table>

Note that the actual values of \( \mu_T \) and \( \mu_R \) are not needed. Only their difference is needed for power and sample size calculations.

The null hypothesis of non-equivalence is

\[
H_0: \delta \leq \varepsilon_L \text{ or } \delta \geq \varepsilon_U \text{ where } \varepsilon_L < 0, \varepsilon_U > 0.
\]

The alternative hypothesis of equivalence is

\[
H_1: \varepsilon_L < \delta < \varepsilon_U
\]

Test Statistics

The analysis for assessing equivalence using higher-order cross-over designs is discussed in detail in Chapter 9 of Chow and Liu (2000). Unfortunately, their presentation is too lengthy to give here. Their method involves the computation of an analysis of variance to estimate the error variance. It also describes the construction of confidence limits for appropriate contrasts. These confidence limits can then be compared to the equivalence limits to test for equivalence. We refer you to their book for details.
Power Calculation

The power is given by

$$\text{Power}(\delta) = T_{V} \left( \frac{\varepsilon_{U} - \delta - t_{V,1-\alpha}}{\sigma_{W} \sqrt{b/n}} \right) - T_{V} \left( t_{V,1-\alpha} - \frac{\delta - \varepsilon_{L}}{\sigma_{W} \sqrt{b/n}} \right)$$

where $T$ represents the cumulative $t$ distribution, $V$ and $b$ depend on the design, $\sigma_{W}$ is the square root of the within mean square error from the ANOVA table using data in the original scale used to analyze the cross-over design, and $n$ is the average number of subjects per sequence. The constants $V$ and $b$ depend on the design as follows.

<table>
<thead>
<tr>
<th>Design Type</th>
<th>Parameters $(V,b)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balaam’s Design</td>
<td>$V = 4n - 3, b = 2.$</td>
</tr>
<tr>
<td>Two-Sequence Dual Design</td>
<td>$V = 4n - 4, b = 3/4.$</td>
</tr>
<tr>
<td>Four-Period Design with Two Sequences</td>
<td>$V = 6n - 5, b = 11/20.$</td>
</tr>
<tr>
<td>Four-Period Design with Four Sequences</td>
<td>$V = 12n - 5, b = 1/4.$</td>
</tr>
</tbody>
</table>

The presentation of Chen et al (1997) uses the following, different parameterization.

$$\text{Power}(\theta) = T_{V} \left( \frac{\nabla_{U} - \theta - t_{V,1-\alpha}}{CV \sqrt{b/n}} \right) - T_{V} \left( t_{V,1-\alpha} - \frac{\theta - \nabla_{L}}{CV \sqrt{b/n}} \right)$$

where $\theta = \frac{\mu_{T} - \mu_{R}}{\mu_{R}}, CV = \frac{\sigma_{W}}{\mu_{R}}, \nabla_{L} = \frac{\varepsilon_{L}}{\mu_{R}},$ and $\nabla_{U} = \frac{\varepsilon_{U}}{\mu_{R}}.$ This parameterization has the advantage that the variables are scaled by the reference mean, so all you need to know is their relative magnitudes rather than their absolute values. It turns out that you can use either parameterization as input, as long as you are consistent.
Procedure Options
This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab
The Data tab contains the parameters associated with this test such as the means, sample sizes, alpha, and beta.

Solve For
Find (Solve For)
This option specifies the parameter to be solved for from the other parameters. Under most situations, you will select either Power and Beta for a power analysis or N for sample size determination.

Select N when you want to calculate the sample size needed to achieve a given power and alpha level. Note that there are two choices for finding N. Select N (Equal Per Sequence) when you want the design to have an equal number of subjects per sequence. Select N (Exact) when you want to find the exact sample size even though the number of subjects cannot be dividing equally among the sequences.

Select Power and Beta when you want to calculate the power of an experiment.

Error Rates
Power or Beta
This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

Alpha (Significance Level)
This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.
Sample Size

N (Total Sample Size)
This option specifies one or more values of the sample size, the number of individuals in the study (total subjects in all sequences). This value must be an integer greater than one.
You may enter a list of values using the syntax 50,100,150,200,250 or 50 to 250 by 50.

Effect Size – Equivalence Limits

|EU| (Upper Equivalence Limit)
This value gives the upper limit of equivalence. Differences outside EL and EU are not considered equivalent, while differences between them are.
Note that EL must be less than zero and EU must be greater than zero. Also, D, EL, and EU must satisfy EL<D<EU. Finally, the scale of these numbers must match the scale of Sw.
Alternatively, you can enter a value for \( \nabla_U = \frac{E_U}{\mu_R} \) here as long as you enter similarly scaled values for the other effect size parameters.

-|EL| (Lower Equivalence Limit)
This value gives lower limit on equivalence. Differences outside EL and EU are not considered equivalent, while differences between them are.
If you want symmetric limits, enter -UPPER LIMIT for EL to force EL = -|EU|.
Note that EL must be less than zero and EU must be greater than zero. Also, D, EL, and EU must satisfy EL<D<EU. Finally, the scale of these numbers must match the scale of Sw.
Alternatively, you can enter a value for \( \nabla_L = \frac{E_L}{\mu_R} \) here as long as you enter similarly scaled values for the other effect size parameters.

Effect Size – True Mean Difference

D (True Difference)
This is the true difference, \( \tilde{\delta} = \mu_T - \mu_R \), between the two means at which the power is to be computed. Often this value is set to zero, but it can be non-zero as long as it is between the equivalence limits, EL and EU. That is, D, EL, and EU must satisfy EL<D<EU. The scale of these numbers must match the scale of Sw.
Alternatively, you can enter \( \theta = \frac{\mu_T - \mu_R}{\mu_R} \) as long as you enter similarly scaled values for the other effect size parameters.
Effect Size – Standard Deviation

Sw (Within Standard Error)

Specify one or more values of Sw ($\sigma_w$), which is estimated by $\sqrt{\text{MSE}}$ from the ANOVA table calculated on the original (not logged) scale. These values must be positive. Alternatively, you can enter $CV = \frac{\sigma_w}{\mu}$ here, as long as you enter similarly scaled values in the other effect size parameters.

You can press the SD button to load the Standard Deviation Estimator window.

Test

Design Type

Specify the type of cross-over design that you are analyzing. Note that all of these designs assume that you are primarily interested in the overall difference between the two treatment means.
Example 1 – Finding Power

A two-sequence, dual cross-over design is to be used to compare the impact of two drugs on diastolic blood pressure. The average diastolic blood pressure after administration of the reference drug is 96 mmHg. Researchers believe this average may drop to 92 mmHg with the use of a new drug. The within mean square error found from similar studies is 324. Its square root is 18.

Following FDA guidelines, the researchers want to show that the diastolic blood pressure is within 20% of the diastolic blood pressure of the reference drug. Thus, the equivalence limits of the mean difference of the two drugs are -19.2 and 19.2. They decide to calculate the power for a range of sample sizes between 4 and 40. The significance level is 0.05.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Equivalence Tests for Two Means in a Higher-Order Cross-Over Design [Differences] procedure window by expanding Means, then MxK Cross-Over Design, then clicking on Equivalence, and then clicking on Equivalence Tests for Two Means in a Higher-Order Cross-Over Design [Differences]. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power</td>
</tr>
<tr>
<td></td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Power</td>
<td>0.05</td>
</tr>
<tr>
<td>Alpha</td>
<td></td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>4 6 8 10 12 14 16 18 20 30 40</td>
</tr>
<tr>
<td></td>
<td>EU</td>
</tr>
<tr>
<td>-</td>
<td>EL</td>
</tr>
<tr>
<td>D (True Difference)</td>
<td>-4</td>
</tr>
<tr>
<td>Sw (Within Standard Error)</td>
<td>18</td>
</tr>
<tr>
<td>Design Type</td>
<td>2x3 (Two-Sequence Dual)</td>
</tr>
</tbody>
</table>
Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>Total Sample Size (N)</th>
<th>Sequences and Periods (SxP)</th>
<th>Lower Equiv. Limit (EL)</th>
<th>Upper Equiv. Limit (EU)</th>
<th>Diff. for Power (D)</th>
<th>Standard Error of Diff. (Sw)</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0000</td>
<td>4</td>
<td>2x3</td>
<td>-19.20</td>
<td>19.20</td>
<td>-4.00</td>
<td>18.00</td>
<td>0.0500</td>
<td>1.0000</td>
</tr>
<tr>
<td>0.1878</td>
<td>6</td>
<td>2x3</td>
<td>-19.20</td>
<td>19.20</td>
<td>-4.00</td>
<td>18.00</td>
<td>0.0500</td>
<td>0.8122</td>
</tr>
<tr>
<td>0.4375</td>
<td>8</td>
<td>2x3</td>
<td>-19.20</td>
<td>19.20</td>
<td>-4.00</td>
<td>18.00</td>
<td>0.0500</td>
<td>0.5625</td>
</tr>
<tr>
<td>0.5985</td>
<td>10</td>
<td>2x3</td>
<td>-19.20</td>
<td>19.20</td>
<td>-4.00</td>
<td>18.00</td>
<td>0.0500</td>
<td>0.4015</td>
</tr>
</tbody>
</table>

Report Definitions

Power is the probability of rejecting non-equivalence when the means are equivalent.
N is the total number of subjects. They are divided evenly among all sequences.
S is the number of sequences.
P is the number of periods per sequence.
EU & EL are the upper & lower limits of the maximum allowable difference that results in equivalence.
D is the difference between the means at which the power is computed.
Sw is the square root of the within mean square error from the ANOVA table.
Alpha is the probability of rejecting non-equivalence when they are non-equivalent.
Beta is the probability of accepting non-equivalence when they are equivalent.
Two-Sequence Dual Cross-Over Design with pattern: ABB; BAA

Summary Statements

In an equivalence test of means using two one-sided tests on data from a two-sequence dual cross-over design, a total sample size of 4 achieves 0% power at a 5% significance level when the true difference between the means is -4.00, the square root of the within mean square error is 18.00, and the equivalence limits are -19.20 and 19.20.

This report shows the power for the indicated scenarios.

Plots Section

This plot shows the power versus the sample size.
**Example 2 – Finding Sample Size**

Continuing with Example 1, the researchers want to find the exact sample size needed to achieve both 80% and 90% power.

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Equivalence Tests for Two Means in a Higher-Order Cross-Over Design [Differences]** procedure window by expanding **Means**, then **MxK Cross-Over Design**, then clicking on **Equivalence**, and then clicking on **Equivalence Tests for Two Means in a Higher-Order Cross-Over Design [Differences]**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

**Option** | **Value**
---|---
Data Tab | 
Find (Solve For) | N (Equal Per Sequence)
Power | 0.80 0.90
Alpha | 0.05
N (Total Sample Size) | Ignored since this is the Find setting
|EU| (Upper Equivalence Limit) | 19.2
-|EL| (Lower Equivalence Limit) | -Upper Limit
D (True Difference) | -4
Sw (Within Standard Error) | 18
Design Type | 2x3 (Two-Sequence Dual)

### Output

Click the Run button to perform the calculations and generate the following output.

**Numeric Results**

<table>
<thead>
<tr>
<th>Power</th>
<th>Total Sample Size (N)</th>
<th>Sequences and Periods ($sxP$)</th>
<th>Lower Equiv. Limit (EL)</th>
<th>Upper Equiv. Limit (EU)</th>
<th>Diff. for Power (D)</th>
<th>Standard Error of Diff. (Sw)</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9119</td>
<td>20</td>
<td>2\times3</td>
<td>-19.20</td>
<td>19.20</td>
<td>-4.00</td>
<td>18.00</td>
<td>0.0500</td>
<td>0.0881</td>
</tr>
<tr>
<td>0.8411</td>
<td>16</td>
<td>2\times3</td>
<td>-19.20</td>
<td>19.20</td>
<td>-4.00</td>
<td>18.00</td>
<td>0.0500</td>
<td>0.1589</td>
</tr>
</tbody>
</table>

Twenty subjects are needed to achieve at least 90% power and sixteen subjects are needed to achieve at least 80% power.
Example 3 – Validation using Chen

Chen et al. (1997) page 757 present a table of sample sizes for various parameter values. In this table, the treatment mean, standard deviation, and equivalence limits are all specified as percentages of the reference mean. We will reproduce the seventeenth line of the table in which the square root of the within mean square error is 10%, the equivalence limits are 20%, the difference between the means is 0%, 5%, 10%, and 15%, the power is 90%, and the significance level is 0.05. Chen reports total sample sizes of 24, 36, 72, and 276. We will now setup this example in PASS.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Equivalence Tests for Two Means in a Higher-Order Cross-Over Design [Differences] procedure window by expanding Means, then MxK Cross-Over Design, then clicking on Equivalence, and then clicking on Equivalence Tests for Two Means in a Higher-Order Cross-Over Design [Differences]. You may then make the appropriate entries as listed below, or open Example 3 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N (Equal Per Sequence)</td>
</tr>
<tr>
<td>Power</td>
<td>0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td></td>
<td>EU</td>
</tr>
<tr>
<td></td>
<td>-EL</td>
</tr>
<tr>
<td>D (True Difference)</td>
<td>0 0.05 0.10 0.15</td>
</tr>
<tr>
<td>Sw (Within Standard Error)</td>
<td>0.1</td>
</tr>
<tr>
<td>Design Type</td>
<td>4x2 (Balaam)</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Numeric Results for Testing the Equivalence of Two Means</th>
</tr>
</thead>
<tbody>
<tr>
<td>Design: Balaam's Cross-Over</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9041</td>
<td>24</td>
<td>4x2</td>
<td>-0.20</td>
<td>0.20</td>
<td>0.00</td>
<td>0.10</td>
<td>0.0500</td>
<td>0.0959</td>
</tr>
<tr>
<td>0.9266</td>
<td>36</td>
<td>4x2</td>
<td>-0.20</td>
<td>0.20</td>
<td>0.05</td>
<td>0.10</td>
<td>0.0500</td>
<td>0.0734</td>
</tr>
<tr>
<td>0.9065</td>
<td>72</td>
<td>4x2</td>
<td>-0.20</td>
<td>0.20</td>
<td>0.10</td>
<td>0.10</td>
<td>0.0500</td>
<td>0.0935</td>
</tr>
<tr>
<td>0.9003</td>
<td>276</td>
<td>4x2</td>
<td>-0.20</td>
<td>0.20</td>
<td>0.15</td>
<td>0.10</td>
<td>0.0500</td>
<td>0.0997</td>
</tr>
</tbody>
</table>

PASS obtains the same samples sizes as Chen et al. (1997).
Chapter 545

Equivalence Tests for Two Means in a Higher-Order Cross-Over Design using Ratios

Introduction

This procedure calculates power and sample size of statistical tests of equivalence of two means of higher-order cross-over designs when the analysis uses a t-test or equivalent. The parameter of interest is the ratio of the two means. Schuirmann’s (1987) two one-sided tests (TOST) approach is used to test equivalence. Only a brief introduction to the subject will be given here. For a comprehensive discussion on the subject, refer to Chen, Chow, and Li (1997). The designs covered in this chapter are analyzed using what is called the ‘multiplicative model’ in Chen et al (1997). The ‘additive model’ is covered in the procedure that uses differences. Note that the multiplicative model is simply an additive model applied to the logarithmic transform of the original data.

Measurements are made on individuals that have been randomly assigned to one of several treatment sequences. This cross-over design may be analyzed by a TOST equivalence test to show that the two means do not differ by more than a small amount, called the margin of equivalence.
Cross-Over Designs

Senn (2002) defines a *cross-over* design as one in which each subject receives all treatments at least once and the object is to study differences among the treatments. The name *cross-over* comes from the most common case in which there are only two treatments. In this case, each subject *crosses over* from one treatment to another. It is assumed that there is a *washout* period between treatments during which the response returns to its baseline value. If this does not occur, there is said to be a *carryover* effect.

A 2x2 cross-over design refers to two treatments (periods) and two *sequences* (treatment orderings). One sequence of treatments is treatment A followed by treatment B. The other sequence is B and then A. The design includes a washout period between responses to make certain that the effects of the first drug do no carryover to the second. Thus, the groups of subjects in this design are defined by the sequence in which the two treatments are administered, not by the treatments they receive.

Higher-Order Cross-Over Designs

Chen et al. (1997) present the results for four cross-over designs that are more complicated than the 2x2 design. Assume that the two treatments are labeled A and B. The available designs are defined by the order and number of times the two treatments are administered.

**Balaam’s Design**

Balaam’s design has four sequences with two treatments each. It is popular because it allows the intrasubject variabilities to be estimated. The design is

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Period 1</th>
<th>Period 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>3</td>
<td>A</td>
<td>B</td>
</tr>
<tr>
<td>4</td>
<td>B</td>
<td>A</td>
</tr>
</tbody>
</table>

**Two-Sequence Dual Design**

This design has two sequences with three periods each. It is popular because it allows the intrasubject variabilities to be estimated. The design is

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>A</td>
<td>A</td>
</tr>
</tbody>
</table>

**Four-Period Design with Two Sequences**

This design has two sequences of four periods each. The design is

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
<th>Period 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
</tbody>
</table>
Four-Period Design with Four Sequences

This design has four sequences of four periods each. The design is:

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Period 1</th>
<th>Period 2</th>
<th>Period 3</th>
<th>Period 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>A</td>
<td>A</td>
<td>B</td>
<td>B</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>B</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>1</td>
<td>A</td>
<td>B</td>
<td>B</td>
<td>A</td>
</tr>
<tr>
<td>2</td>
<td>B</td>
<td>A</td>
<td>A</td>
<td>B</td>
</tr>
</tbody>
</table>

Advantages of Cross-Over Designs

A comparison of treatments on the same subject is expected to be more precise. The increased precision often translates into a smaller sample size. Also, patient enrollment may be easier to obtain because each patient will receive both treatments.

Disadvantages of Cross-Over Designs

The statistical analysis of a cross-over experiment is more complex than a parallel-group experiment and requires additional assumptions. In a cross-over experiment, it may be difficult to separate the treatment effect from the time effect and the carry-over effect of the previous treatment.

These cross-over designs cannot be used when the treatment (or the measurement of the response) alters the subject permanently. Hence, it cannot be used to compare treatments that are intended to provide a cure.

Because subjects must be measured at least twice, it may be more difficult to keep patients enrolled in the study. This is particularly true when the measurement process is painful, uncomfortable, embarrassing, or time consuming.

Outline of an Equivalence Test

PASS follows the two one-sided tests approach described by Schuirmann (1987) and Phillips (1990). It will be convenient to adopt the following specialized notation for the discussion of these tests.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>PASS Input/Output</th>
<th>Interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \mu_T )</td>
<td>Not used</td>
<td>Treatment mean. This is the treatment mean.</td>
</tr>
<tr>
<td>( \mu_R )</td>
<td>Not used</td>
<td>Reference mean. This is the mean of a reference population.</td>
</tr>
<tr>
<td>( \phi_L, \phi_U )</td>
<td>RL, RU</td>
<td>Margin of equivalence. These limits define an interval of the ratio of the means in which their difference is so small that it may be ignored.</td>
</tr>
<tr>
<td>( \phi )</td>
<td>R1</td>
<td>True ratio. This is the value of ( \phi = \frac{\mu_T}{\mu_R} ) at which the power is calculated.</td>
</tr>
</tbody>
</table>

Note that the actual values of \( \mu_T \) and \( \mu_R \) are not needed. Only their ratio is needed for power and sample size calculations.
The null hypothesis of non-equivalence is

\[ H_0: \phi \leq \phi_L \text{ or } \phi \geq \phi_U \] where \( \phi_L < 1, \phi_U > 1 \).

The alternative hypothesis of equivalence is

\[ H_1: \phi_L < \phi < \phi_U \]

**Log-Transformation**

In many cases, hypotheses stated in terms of ratios are more convenient than hypotheses stated in terms of differences. This is because ratios can be interpreted as percentages, but differences must be interpreted as actual amounts in their original scale. Also, the distributions of variables such as CMax or AUC are often skewed. Hence, a common practice is to take the following steps in hypothesis testing.

1. State the statistical hypothesis in terms of a ratio.
2. Transform this into a hypothesis about the difference by taking logarithms.
3. Analyze the logged data—that is, do the analysis in terms of the difference.
4. Draw the conclusion in the original ratio scale.

The details of step 2 for the alternative hypothesis are as follows.

\[ \phi_L < \phi < \phi_U \]

\[ \Rightarrow \phi_L < \left\{ \frac{\mu_T}{\mu_R} \right\} < \phi_U \]

\[ \Rightarrow \ln(\phi_L) < \left\{ \ln(\mu_T) - \ln(\mu_R) \right\} < \ln(\phi_U) \]

Thus, a hypothesis about the ratio of the means on the original scale can be translated into a hypothesis about the difference of two means on the logged scale.

When performing an equivalence test on the difference between means, the usual procedure is to set the equivalence limits symmetrically above and below zero. Thus, the equivalence limits will be plus or minus an appropriate amount. The common practice is to do the same when the data are being analyzed on the log scale. However, when symmetric limits are set on the log scale, they do not translate to symmetric limits on the original scale. Instead, they translate to limits that are the inverses of each other.

Perhaps these concepts can best be understood by considering an example. Suppose the researchers have determined that the lower equivalence limit should be 80% on the original scale. Since they are planning to use a log scale for their analysis, they transform this limit into the log scale by taking the logarithm of 0.80. The result is -0.223144. Wanting symmetric limits, they set the upper equivalence limit to 0.223144. Exponentiating this value, they find that \( \exp(0.223144) = 1.25 \). Note that \( 1/(0.80) = 1.25 \). Thus, the limits on the original scale are 80% and 125%, not 80% and 120%.

Using this procedure, appropriate equivalence limits for the ratio of two means can be easily determined. Here are a few sets of equivalence limits for ratios.
Note that negative percent-change values specify the lower limit first, while positive percent-change values specify the upper limit first. After the first limit is found, the other limit is calculated as its inverse.

### Coefficient of Variation

The coefficient of variation (COV or CV) as it is used here is the ratio of the standard deviation to the reference mean. This parameter can be used to represent the variation in the data because of a unique relationship that it has in the case of log-normal data.

Suppose the variable \( X \) is the logarithm of the original variable \( Y \). That is, \( X = \ln(Y) \) and \( Y = \exp(X) \). Label the mean and variance of \( X \) as \( \mu_X \) and \( \sigma_X^2 \), respectively. Similarly, label the mean and variance of \( Y \) as \( \mu_Y \) and \( \sigma_Y^2 \), respectively. If \( X \) is normally distributed, then \( Y \) is log-normally distributed. Julious (2004) presents the following well-known relationships between these two variables

\[
\begin{align*}
\mu_Y &= \left( e^{\mu_X + \frac{\sigma_X^2}{2}} \right) \\
\sigma_Y^2 &= \mu_Y^2 \left( e^{\sigma_X^2} - 1 \right)
\end{align*}
\]

From this relationship, the coefficient of variation of \( Y \) can be expressed as

\[
COV_Y = \frac{\sqrt{\mu_Y^2 \left( e^{\sigma_X^2} - 1 \right)}}{\mu_Y} = \sqrt{e^{\sigma_X^2} - 1}
\]

Solving this relationship for \( \sigma_X^2 \), the standard deviation of \( X \) can be stated in terms of the coefficient of variation of \( Y \). This equation is

\[
\sigma_X = \sqrt{\ln \left( COV_Y^2 + 1 \right)}
\]

Similarly, the mean of \( X \) is

\[
\mu_X = \frac{\mu_Y}{\ln \left( COV_Y^2 + 1 \right)}
\]
Thus, the hypotheses can be stated in the original (Y) scale and then the power can be analyzed in the transformed (X) scale.

**Test Statistics**

The analysis for assessing equivalence using higher-order cross-over designs is discussed in detail in Chapter 9 of Chow and Liu (2000). Unfortunately, their presentation is too lengthy to give here. Their method involves the computation of an analysis of variance to estimate the error variance. It also describes the construction of confidence limits for appropriate contrasts. These confidence limits can then be compared to the equivalence limits to test for equivalence. We refer you to their book for details.

**Power Calculation**

The power is given by

\[
Power(\phi) = T_{(1-\alpha)} \left( \frac{\ln(\phi_U) - \ln(\phi_L)}{CV_{m} \sqrt{b/n}} - t_{\alpha,1} \right) - T_{(1-\alpha)} \left( \frac{\ln(\phi_U) - \ln(\phi_L)}{CV_{m} \sqrt{b/n}} \right)
\]

where

\[
CV_{m} = \sqrt{\exp(MSE_{\log}) - 1},
\]

\(T\) represents the cumulative \(t\) distribution, \(V\) and \(b\) depend on the design, and \(n\) is the average number of subjects per sequence. Note that the constants \(V\) and \(b\) depend on the design as follows.

<table>
<thead>
<tr>
<th>Design Type</th>
<th>Parameters ((V,b))</th>
</tr>
</thead>
<tbody>
<tr>
<td>Balaam’s Design</td>
<td>(V = 4n - 3, b = 2).</td>
</tr>
<tr>
<td>Two-Sequence Dual Design</td>
<td>(V = 4n - 4, b = 3/4).</td>
</tr>
<tr>
<td>Four-Period Design with Two Sequences</td>
<td>(V = 6n - 5, b = 11/20).</td>
</tr>
<tr>
<td>Four-Period Design with Four Sequences</td>
<td>(V = 12n - 5, b = 1/4).</td>
</tr>
</tbody>
</table>
Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains the parameters associated with this test such as the means, sample sizes, alpha, and beta.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for from the other parameters. Under most situations, you will select either Power and Beta for a power analysis or $N$ for sample size determination.

Select $N$ when you want to calculate the sample size needed to achieve a given power and alpha level. Note that there are two choices for finding $N$. Select $N$ (Equal Per Sequence) when you want the design to have an equal number of subjects per sequence. Select $N$ (Exact) when you want to find the exact sample size even though the number of subjects cannot be dividing equally among the sequences.

Select Power and Beta when you want to calculate the power of an experiment.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.
Sample Size

N (Total Sample Size)
This option specifies one or more values of the sample size, the number of individuals in the study (total subjects in all sequences). This value must be an integer greater than one.

You may enter a list of values using the syntax 50,100,150,200,250 or 50 to 250 by 50.

Effect Size – Equivalence Limits

RU (Upper Equivalence Limit)
Enter the upper equivalence limit for the ratio of the two means. When the ratio of the means is between this value and RL, the two means are said to be equivalent. The value must be greater than one. A popular choice is 1.25. Note that this value is not a percentage.

If you enter 1/RL, then 1/RL will be calculated and used here. This choice is commonly used because RL and 1/RL give limits that are of equal magnitude on the log scale.

RL (Lower Equivalence Limit)
Enter the lower equivalence limit for the ratio of the two means. When the ratio of the means is between this value and RU, the two means are said to be equivalent. The value must be less than one. A popular choice is 0.80. Note that this value is not a percentage.

If you enter 1/RL, then 1/RL will be calculated and used here. This choice is commonly used because RU and 1/RU give limits that are of equal magnitude on the log scale.

Effect Size – True Ratio

R1 (True Ratio)
This is the value of the ratio of the two means at which the power is to be calculated. Usually, the ratio will be assumed to be one. However, some authors recommend calculating the power using a ratio of 1.05 since this will require a larger, more conservative, sample size.

Effect Size – Coefficient of Variation

COV (Coefficient of Variation)
The coefficient of variation is used to specify the variability (standard deviation). It is important to realize that this is the COV defined on the original (not logged) scale. This value must be determined from past experience or from a pilot study. It is most easily calculated from the within mean-square error of the analysis of variance of the logged data using the relationship

\[ CV_m = \sqrt{\exp(MSE_{\text{logged}})} - 1. \]

Test

Design Type
Specify the type of cross-over design that you are analyzing. Note that all of these designs assume that you are primarily interested in the overall difference between two means.
Example 1 – Finding Power

A company has opened a new manufacturing plant and wants to show that the drug produced in the new plant is equivalent to that produced in the older plant. A two-sequence, dual cross-over design will be used to test the equivalence of drugs produced at the two plants.

Researchers have decided to set the equivalence limits for the ratio at 0.80 and 1.25. Past experience leads the researchers to set the COV to 0.40. The significance level is 0.05. The power will be computed assuming that the true ratio is 0.96. Sample sizes between 10 and 80 will be included in the analysis.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Equivalence Tests for Two Means in a Higher-Order Cross-Over Design [Ratios] procedure window by expanding Means, then MxK Cross-Over Design, then clicking on Equivalence, and then clicking on Equivalence Tests for Two Means in a Higher-Order Cross-Over Design [Ratios]. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>10 20 30 40 60 80</td>
</tr>
<tr>
<td>RU (Upper Equivalence Limit)</td>
<td>1.25</td>
</tr>
<tr>
<td>RL (Lower Equivalence Limit)</td>
<td>1/RU</td>
</tr>
<tr>
<td>R1 (True Ratio)</td>
<td>0.96</td>
</tr>
<tr>
<td>COV (Coefficient of Variation)</td>
<td>0.40</td>
</tr>
<tr>
<td>Design Type</td>
<td>2x3 (Two-Sequence Dual)</td>
</tr>
</tbody>
</table>

Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>Total Sample Size (N)</th>
<th>Sequences and Periods (SxP)</th>
<th>Lower Equiv. Limit (RL)</th>
<th>Upper Equiv. Limit (RU)</th>
<th>Mean Ratio for Power (R1)</th>
<th>Coef. of Variation (COV)</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0000</td>
<td>10</td>
<td>2x3</td>
<td>0.80</td>
<td>1.25</td>
<td>0.96</td>
<td>0.40</td>
<td>0.050</td>
<td>1.0000</td>
</tr>
<tr>
<td>0.3051</td>
<td>20</td>
<td>2x3</td>
<td>0.80</td>
<td>1.25</td>
<td>0.96</td>
<td>0.40</td>
<td>0.050</td>
<td>0.6949</td>
</tr>
<tr>
<td>0.5868</td>
<td>30</td>
<td>2x3</td>
<td>0.80</td>
<td>1.25</td>
<td>0.96</td>
<td>0.40</td>
<td>0.050</td>
<td>0.4142</td>
</tr>
<tr>
<td>0.7483</td>
<td>40</td>
<td>2x3</td>
<td>0.80</td>
<td>1.25</td>
<td>0.96</td>
<td>0.40</td>
<td>0.050</td>
<td>0.2517</td>
</tr>
<tr>
<td>0.9035</td>
<td>60</td>
<td>2x3</td>
<td>0.80</td>
<td>1.25</td>
<td>0.96</td>
<td>0.40</td>
<td>0.050</td>
<td>0.0965</td>
</tr>
<tr>
<td>0.9627</td>
<td>80</td>
<td>2x3</td>
<td>0.80</td>
<td>1.25</td>
<td>0.96</td>
<td>0.40</td>
<td>0.050</td>
<td>0.0373</td>
</tr>
</tbody>
</table>
545-10 Equivalence of Two Means in a Higher-Order Cross-Over Design using Ratios

**Report Definitions**
- Power is the probability of rejecting non-equivalence when the means are equivalent.
- N is the total number of subjects. They are divided evenly among all sequences.
- RU & RL are the upper and lower equivalence limits. Ratios between these limits are equivalent.
- R1 is the ratio of the means at which the power is computed.
- COV is the coefficient of variation on the original scale.
- Alpha is the probability of rejecting non-equivalence when they are non-equivalent.
- Beta is the probability of accepting non-equivalence when they are equivalent.

Two-Sequence Dual Cross-Over Design with pattern: ABB; BAA

**Summary Statements**
In an equivalence test of means using two one-sided tests on data from a two-sequence dual cross-over design, a total sample size of 10 achieves 0% power at a 5% significance level when the true ratio of the means is 0.96, the coefficient of variation on the original (unlogged) scale is 0.40, and the equivalence limits are 0.80 and 1.25.

This report shows the power for the indicated scenarios. Note that 60 subjects will yield a power of just over 90%.

**Plots Section**

This plot shows the power versus the sample size.
Example 2 – Finding Sample Size

Continuing with Example 1, the researchers want to find the exact sample size needed to achieve both 80% power and 90% power.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Equivalence Tests for Two Means in a Higher-Order Cross-Over Design [Ratios] procedure window by expanding Means, then MxK Cross-Over Design, then clicking on Equivalence, and then clicking on Equivalence Tests for Two Means in a Higher-Order Cross-Over Design [Ratios]. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N (Equal per Sequence)</td>
</tr>
<tr>
<td>Power</td>
<td>0.80 0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>RU (Upper Equivalence Limit)</td>
<td>1.25</td>
</tr>
<tr>
<td>RL (Lower Equivalence Limit)</td>
<td>1/RU</td>
</tr>
<tr>
<td>R1 (True Ratio)</td>
<td>0.96</td>
</tr>
<tr>
<td>COV (Coefficient of Variation)</td>
<td>0.40</td>
</tr>
<tr>
<td>Design Type</td>
<td>2x3 (Two-Sequence Dual)</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>Total Sample Size (N)</th>
<th>Total Sequences and Periods (SxP)</th>
<th>Lower Equiv. Limit (RL)</th>
<th>Upper Equiv. Limit (RU)</th>
<th>Mean Ratio for Power (R1)</th>
<th>Coef. of Variation (COV)</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9035</td>
<td>60</td>
<td>2x3</td>
<td>0.80</td>
<td>1.25</td>
<td>0.96</td>
<td>0.40</td>
<td>0.0500</td>
<td>0.0965</td>
</tr>
<tr>
<td>0.8119</td>
<td>46</td>
<td>2x3</td>
<td>0.80</td>
<td>1.25</td>
<td>0.96</td>
<td>0.40</td>
<td>0.0500</td>
<td>0.1881</td>
</tr>
</tbody>
</table>

We note that 60 subjects are needed to achieve 90% power and 46 subjects are needed to achieve at least 80% power.
Example 3 – Validation using Chen

Chen et al. (1997) page 761 presents a table of sample sizes for various parameter values for Balaam’s design. We will reproduce entries from the first and seventeenth lines of the table in which the COV is 10%, the equivalence limits are 0.8 and 1.25, the actual ratio of between the means is 1, the power values are 80% and 90%, and the significance level is 0.05. Chen reports total sample sizes of 16 and 20. We will now setup this example in PASS.

The COV entered by Chen is the COV of the logged data. Since PASS requires the COV of the original data, we must use the relationship

\[
\text{COV}_Y = \sqrt{e^{\sigma^2_Y} - 1} = \sqrt{e^{0.1^2} - 1} = \sqrt{e^{0.01} - 1} = 0.10025
\]

to obtain the appropriate value of COV.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Equivalence Tests for Two Means in a Higher-Order Cross-Over Design [Ratios] procedure window by expanding Means, then MxK Cross-Over Design, then clicking on Equivalence, and then clicking on Equivalence Tests for Two Means in a Higher-Order Cross-Over Design [Ratios]. You may then make the appropriate entries as listed below, or open Example 3 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>N (Equal Per Sequence)</td>
</tr>
<tr>
<td>Power</td>
<td>0.80 0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>N (Total Sample Size)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>RU (Upper Equivalence Limit)</td>
<td>1.25</td>
</tr>
<tr>
<td>RL (Lower Equivalence Limit)</td>
<td>1/RU</td>
</tr>
<tr>
<td>R1 (True Ratio)</td>
<td>1</td>
</tr>
<tr>
<td>COV (Coefficient of Variation)</td>
<td>0.10025</td>
</tr>
<tr>
<td>Design Type</td>
<td>4x2 (Balaam)</td>
</tr>
</tbody>
</table>
## Output

Click the Run button to perform the calculations and generate the following output.

### Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>Sample Size (N)</th>
<th>Total Sequences and Periods (SxP)</th>
<th>Lower Equiv. Limit (RL)</th>
<th>Upper Equiv. Limit (RU)</th>
<th>Mean Ratio for Power (R)</th>
<th>Coef. of Variation (COV)</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.9085</td>
<td>20</td>
<td>4x2</td>
<td>0.80</td>
<td>1.25</td>
<td>1.00</td>
<td>0.10</td>
<td>0.0500</td>
<td>0.1894</td>
</tr>
<tr>
<td>0.8106</td>
<td>16</td>
<td>4x2</td>
<td>0.80</td>
<td>1.25</td>
<td>1.00</td>
<td>0.10</td>
<td>0.0500</td>
<td>0.0915</td>
</tr>
</tbody>
</table>

Note that *PASS* has obtained the same samples sizes as Chen et al. (1997).
545-14 Equivalence of Two Means in a Higher-Order Cross-Over Design using Ratios
Chapter 550

One-Way Analysis of Variance

Introduction

A common task in research is to compare the averages of two or more populations (groups). We might want to compare the income level of two regions, the nitrogen content of three lakes, or the effectiveness of four drugs. The one-way analysis of variance compares the means of two or more groups to determine if at least one mean is different from the others. The $F$ test is used to determine statistical significance. $F$ tests are non-directional in that the null hypothesis specifies that all means are equal and the alternative hypothesis simply states that at least one mean is different.

The methods described here are usually applied to the one-way experimental design. This design is an extension of the design used for the two-sample $t$ test. Instead of two groups, there are three or more groups. With careful modifications, this procedure may be used to test interaction terms as well.

Planned Comparisons

$PASS$ performs power and sample size calculations for user-specified contrasts.

The usual $F$ test tests the hypothesis that all means are equal versus the alternative that at least one mean is different from the rest. Often, a more specific alternative is desired. For example, you might want to test whether the treatment means are different from the control mean, the low dose is different from the high dose, a linear trend exists across dose levels, and so on. These questions are tested using planned comparisons.

We call the comparison planned because it was determined before the experiment was conducted. We planned to test the comparison.

A comparison is a weighted average of the means, in which the weights may be negative. When the weights sum to zero, the comparison is called a contrast. $PASS$ provides results for contrasts. To specify a contrast, we need only specify the weights. Statisticians call these weights the contrast coefficients.

For example, suppose an experiment conducted to study a drug will have three dose levels: none (control), 20 mg., and 40 mg. The first question is whether the drug made a difference. If it did, the average response for the two groups receiving the drug should be different from the control. If we label the group means $M_0$, $M_20$, and $M_40$, we are interested in comparing $M_0$ with $M_20$ and $M_40$. This can be done in two ways. One way is to construct two tests, one comparing $M_0$ and
M20 and the other comparing M0 and M40. Another method is to perform one test comparing M0 with the average of M20 and M40. These tests are conducted using planned comparisons. The coefficients are as follows:

**M0 vs. M20**

To compare M0 versus M20, use the coefficients -1,1,0. When applied to the group means, these coefficients result in the comparison $M0(-1)+M20(1)+M40(0)$ which reduces to $M20-M0$. That is, this contrast results in the difference between the two group means. We can test whether this difference is non-zero using the $t$ test (or $F$ test since the square of the $t$ test is an $F$ test).

**M0 vs. M40**

To compare M0 versus M40, use the coefficients -1,0,1. When applied to the group means, these coefficients result in the comparison $M0(-1)+M20(0)+M40(1)$ which reduces to $M40-M0$. That is, this contrast results in the difference between the two group means.

**M0 vs. Average of M20 and M40**

To compare M0 versus the average of M20 and M40, use the coefficients -2,1,1. When applied to the group means, these coefficients result in the comparison $M0(-2)+M20(1)+M40(1)$ which is equivalent to $M40+M20-2(M0)$.

To see how these coefficients were obtained, consider the following manipulations. Beginning with the difference between the average of M20 and M40 and M0, we obtain the coefficients above—aside from a scale factor of one-half.

$$\frac{M20 + M40}{2} - M0 = \frac{M20}{2} + \frac{M40}{2} - \frac{M0}{1}$$

$$= \frac{1}{2} M20 + \frac{1}{2} M40 - M0$$

$$= \frac{1}{2} (M20 + M40 - 2M0)$$

**Assumptions**

Using the $F$ test requires certain assumptions. One reason for the popularity of the $F$ test is its robustness in the face of assumption violation. However, if an assumption is not even approximately met, the significance levels and the power of the $F$ test are invalidated. Unfortunately, in practice it often happens that several assumptions are not met. This makes matters even worse. Hence, steps should be taken to check the assumptions before important decisions are made.

The assumptions of the one-way analysis of variance are:

1. The data are continuous (not discrete).
2. The data follow the normal probability distribution. Each group is normally distributed about the group mean.
3. The variances of the populations are equal.
4. The groups are independent. There is no relationship among the individuals in one group as compared to another.

5. Each group is a simple random sample from its population. Each individual in the population has an equal probability of being selected in the sample.

**Technical Details for the One-Way ANOVA**

Suppose $k$ groups each have a normal distribution and equal means ($\mu_1 = \mu_2 = \cdots = \mu_k$). Let $n_1 = n_2 = \cdots = n_k$ denote the number of subjects in each group and let $N$ denote the total sample size of all groups. Let $\overline{\mu}_w$ denote the weighted mean of all groups. That is

$$\overline{\mu}_w = \sum_{i=1}^{k} \left( \frac{n_i}{N} \right) \mu_i$$

Let $\sigma$ denote the common standard deviation of all groups.

Given the above terminology, the ratio of the mean square between groups to the mean square within groups follows a central $F$ distribution with two parameters matching the degrees of freedom of the numerator mean square and the denominator mean square. When the null hypothesis of mean equality is rejected, the above ratio has a noncentral $F$ distribution which also depends on the noncentrality parameter, $\lambda$. This parameter is calculated as

$$\lambda = N \frac{\sigma_m^2}{\sigma^2}$$

where

$$\sigma_m = \sqrt{\frac{\sum_{i=1}^{k} n_i (\mu_i - \overline{\mu}_w)^2}{N}}$$

Some authors use the symbol $\phi$ for the noncentrality parameter. The relationship between the two noncentrality parameters is

$$\phi = \frac{\lambda}{\sqrt{k}}.$$

The process of planning an experiment should include the following steps:

1. Determine an estimate of the within group standard deviation, $\sigma$. This may be done from prior studies, from experimentation with the Standard Deviation Estimation module, from pilot studies, or from crude estimates based on the range of the data. See the chapter on estimating the standard deviation for more details.

2. Determine a set of means that represent the group differences that you want to detect.

3. Determine the appropriate group sample sizes that will ensure desired levels of $\alpha$ and $\beta$. Although it is tempting to set all group sample sizes equal, it is easy to show that putting more subjects in some groups than in others may have better power than keeping group sizes equal (see Example 4).
Power Calculations for One-Way ANOVA

The calculation of the power of a particular test proceeds as follows:

1. Determine the critical value, $F_{k-1,N-k,\alpha}$ where $\alpha$ is the probability of a type-I error and $k$ and $N$ are defined above. Note that this is a two-tailed test as no direction is assigned in the alternative hypothesis.

2. From a hypothesized set of $\mu_i$'s, calculate the noncentrality parameter $\lambda$ based on the values of $N, k, \sigma_m,$ and $\sigma$.

3. Compute the power as the probability of being greater than $F_{k-1,N-k,\alpha}$ on a noncentral-$F$ distribution with noncentrality parameter $\lambda$.

Technical Details for a Planned Comparison

The terminology of planned comparisons is identical to that of the one-way AOV, so the notation used above will be repeated here.

Suppose you want to test whether the contrast $C$

$$C = \sum_{i=1}^{k} c_i \mu_i$$

is significantly different from zero. Here the $c_i$'s are the contrast coefficients.

Define

$$\sigma_{mc} = \sqrt{\frac{\sum_{i=1}^{k} c_i \mu_i}{\sqrt{N \sum_{i=1}^{k} c_i^2 / n_i}}}$$

Define the noncentrality parameter $\lambda_c$, as

$$\lambda_c = N \frac{\sigma_{mc}^2}{\sigma^2}$$

Power Calculations for Planned Comparisons

The calculation of the power of a particular test proceeds as follows:

1. Determine the critical value, $F_{1,N-k,\alpha}$ where $\alpha$ is the probability of a type-I error and $k$ and $N$ are defined above. Note that this is a two-tailed test as no direction is assigned in the alternative hypothesis.

2. From a hypothesized set of $\mu_i$'s, calculate the noncentrality parameter $\lambda_c$ based on the values of $N, k, \sigma_{mc},$ and $\sigma$. 
3. Compute the power as the probability of being greater than $F_{i,N-k,\alpha}$ on a noncentral-$F$ distribution with noncentrality parameter $\lambda_c$.

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains most of the parameters and options that you will be concerned with.

Solve For

Find (Solve For)

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are $SM$, $S$, $k$, $n$, $\text{Alpha}$, and $\text{Power and Beta}$. Under most situations, you will select either $\text{Power and Beta}$ for a power analysis or $n$ for sample size determination.

Error Rates

Power or Beta

This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of equal means when in fact the means are different.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

Alpha (Significance Level)

This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when you reject the null hypothesis of equal means when in fact the means are equal.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.
### Sample Size / Groups – Groups

#### k (Number of Groups)
This is the number of group means being compared. It must be greater than or equal to two.

You can enter a list of values, in which case, a separate analysis will be calculated for each value. Commas or blanks may separate the numbers. A TO-BY list may be used.

Note that the number of items used in the Hypothesized Means box and the Group Sample Size Pattern box is controlled by this number.

Examples:
- 2, 3, 4
- 2 3 4
- 2 to 10 by 2

#### Group Allocation Ratios
A set of positive, numeric values, one for each group, is entered here. The sample size of group $i$ is found by multiplying the $i^{th}$ number from this list times the value of $n$ and rounding up to the next whole number. The number of values must match the number of groups, $k$. When too few numbers are entered, 1’s are added. When too many numbers are entered, the extras are ignored.

- **Equal**
  If all sample sizes are to be equal, enter “Equal” here and the desired sample size in $n$. A set of $k$ 1’s will be used. This will result in $N1 = N2 = N3 = n$. That is, all sample sizes are equal to $n$.

**n (Sample Size per Group)**
This is the base, per group, sample size. One or more values, separated by blanks or commas, may be entered. A separate analysis is performed for each value listed here.

The group samples sizes are determined by multiplying this number by each of the Group Sample Size Pattern numbers. If the Group Sample Size Pattern numbers are represented by $m1, m2, m3, ..., mk$ and this value is represented by $n$, the group sample sizes $N1, N2, N3, ..., Nk$ are calculated as follows:

- $N1 = [n(m1)]$
- $N2 = [n(m2)]$
- $N3 = [n(m3)]$
- etc.

where the operator, $[X]$ means the next integer after $X$, e.g. $[3.1]=4$.

For example, suppose there are three groups and the Group Sample Size Pattern is set to 1, 2, 3. If $n$ is 5, the resulting sample sizes will be 5, 10, and 15. If $n$ is 50, the resulting group sample sizes will be 50, 100, and 150. If $n$ is set to 2, 4, 6, 8, 10, five sets of group sample sizes will be generated and an analysis run for each. These sets are:

<p>| | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>12</td>
</tr>
<tr>
<td>6</td>
<td>12</td>
<td>18</td>
</tr>
<tr>
<td>8</td>
<td>16</td>
<td>24</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>30</td>
</tr>
</tbody>
</table>
As a second example, suppose there are three groups and the Group Sample Size Pattern is 0.2, 0.3, 0.5. When the fractional Pattern values sum to one, \( n \) can be interpreted as the total sample size of all groups and the Pattern values as the proportion of the total in each group.

- If \( n \) is 10, the three group sample sizes would be 2, 3, and 5.
- If \( n \) is 20, the three group sample sizes would be 4, 6, and 10.
- If \( n \) is 12, the three group sample sizes would be
  - \((0.2)12 = 2.4\) which is rounded up to the next whole integer, 3.
  - \((0.3)12 = 3.6\) which is rounded up to the next whole integer, 4.
  - \((0.5)12 = 6\).

Note that in this case, 3+4+6 does not equal \( n \) (which is 12). This can happen because of rounding.

**Effect Size – Means**

**Hypothesized Means**
Enter a set of hypothesized means, one for each group. These means represent the group centers under the alternative hypothesis (the null hypothesis is that they are equal). The standard deviation of these means (\( \sigma_m \)) is used in the power calculations to represent the average size of the differences among the means. The standard deviation of the means is calculated using the formula:

\[
\sigma_m = \sqrt{\frac{\sum_{i=1}^{k} (\mu_i - \bar{\mu})^2}{k}}
\]

This quantity gives the magnitude of the differences among the group means. Note that when all means are equal, \( \sigma_m \) is zero.

You should enter a set of means that give the pattern of differences you expect or the pattern that you wish to detect. For example, in a particular study involving three groups, your research might be “meaningful” if either of two treatment means is 50\% larger than the control mean. If the control mean is 50, then you would enter 50, 75, 75 as the three means.

It is usually more intuitive to enter a set of mean values. However, it is possible to enter the standard deviation of the means directly by placing an \( S \) in front of the number (see below).

Some might wish to specify the alternative hypothesis as the effect size, \( f \), which is defined as

\[
f = \frac{\sigma_m}{\sigma}
\]

If so, set \( \sigma = 1 \) and \( \sigma_m = f \). Cohen (1988) has designated values of \( f \) less than 0.1 as small, values around 0.25 to be medium, and values over 0.4 to be large.


Entering a List of Means
If a set of numbers is entered without a leading $S$, they are assumed to be the hypothesized group means under the alternative hypothesis. Their standard deviation will be calculated and used in the calculations. Blanks or commas may separate the numbers. Note that it is not the values of the means themselves that is important, but only their differences. Thus, the mean values $0, 1, 2$ produce the same results as the values $100, 101, 102$.

If too few means are entered to match the number of groups, the last mean is repeated. For example, suppose that four means are needed and you enter $1, 2$ (only two means). $PASS$ will treat this as $1, 2, 2, 2$. If too many values are entered, $PASS$ will truncate the list to the number of means needed.

 Examples:
   
   5 20 60 
   2,5,7 
   -4,0,6,9 

S Option
If an $S$ is entered before the list of numbers, they are assumed to be values of $\sigma_m$, the standard deviations of the group means. A separate power calculation is made for each value. Note that this list can be a TO-BY phrase.

 Examples:
   
   S 4.7 
   S 4.3 5.7 4.2 
   S 10 to 20 by 2 

Effect Size – Standard Deviation

S (Standard Deviation of Subjects)
This is $\sigma$, the standard deviation within a group. It represents the variability from subject to subject that occurs when the subjects are treated identically. It is assumed to be the same for all groups. This value is approximated in an analysis of variance table by the square root of the mean square error.

Since they are positive square roots, the numbers must be strictly greater than zero. You can press the $SD$ button to obtain further help on estimating the standard deviation.

Note that if you are using this procedure to test a factor (such as an interaction) from a more complex design, the value of standard deviation is estimated by the square root of the mean square of the term that is used as the denominator in the $F$ test.
You can enter a list of values separated by blanks or commas, in which case, a separate analysis will be calculated for each value.

Examples of valid entries:
1, 4, 7, 10
1 4 7 10
1 to 10 by 3

---

**Planned Comparisons**

**Contrast Coefficients**

If you want to analyze a specific planned comparison, enter a set of contrast coefficients here. The calculations will then refer to the hypothesis that the corresponding contrast of the means is zero versus the alternative that it is non-zero (two-sided test). These are often called Planned Comparisons.

A contrast is a weighted average of the means in which the weights sum to zero. For example, suppose you are studying four groups and that the main hypothesis of interest is whether there is a linear trend across the groups. You would enter -3, -1, 1, 3 here. This would form the weighted average of the means:

\[-3(\text{Mean}_1)-\text{Mean}_2+(\text{Mean}_3)+3(\text{Mean}_3)\]

The point to realize is that these numbers (the coefficients) are used to calculate a specific weighted average of the means which is to be compared against zero using a standard \(F\) (or \(t\)) test.

- **NONE or blank**
  When the box is left blank or the word None is entered, this option is ignored.

- **Linear Trend**
  A set of coefficients is generated appropriate for testing the alternative hypothesis that there is a linear (straight-line) trend across the means. These coefficients assume that the means are equally spaced across the trend variable.

- **Quadratic**
  A set of coefficients is generated appropriate for testing the alternative hypothesis that the means follow a quadratic model. These coefficients assume that the means are equally spaced across the implicit \(X\) variable.

- **Cubic**
  A set of coefficients is generated appropriate for testing the alternative hypothesis that the means follow a cubic model. These coefficients assume that the means are equally spaced across the implicit \(X\) variable.

- **First Against Others**
  A set of coefficients is generated appropriate for testing the alternative hypothesis that the first mean is different from the average of the remaining means. For example, if there were four groups, the generated coefficients would be -3, 1, 1, 1.
List of Coefficients

A list of coefficients, separated by commas or blanks, may be entered. If the number of items in the list does not match the number of groups \(k\), zeros are added or extra coefficients are truncated.

Remember that these coefficients must sum to zero. Also, the scale of the coefficients does not matter. That is \(0.5,0.25,0.25; -2,1,1\); and \(-200,100,100\) will yield the same results.

To avoid rounding problems, it is better to use \(-3,1,1,1\) than the equivalent \(-1,0.333,0.333,0.333\). The second set does not sum to zero.

Example 1 – Finding the Statistical Power

An experiment is being designed to compare the means of four groups using an \(F\) test with a significance level of either 0.01 or 0.05. Previous studies have shown that the standard deviation within a group is 18. Treatment means of 40, 10, 10, and 10 represent clinically important treatment differences. To better understand the relationship between power and sample size, the researcher wants to compute the power for several group sample sizes between 2 and 14. The sample sizes will be equal across all groups.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the One-Way Analysis of Variance procedure window by expanding Means, then ANOVA, then clicking on Analysis of Variance, and then clicking on One-Way Analysis of Variance. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Power</td>
<td></td>
</tr>
<tr>
<td>Alpha</td>
<td>0.01 0.05</td>
</tr>
<tr>
<td>(k) (Number of Groups)</td>
<td>4</td>
</tr>
<tr>
<td>Group Allocation Ratios</td>
<td>Equal</td>
</tr>
<tr>
<td>(n) (Sample Size per Group)</td>
<td>2 to 14 by 2</td>
</tr>
<tr>
<td>Hypothesized Means</td>
<td>40 10 10 10</td>
</tr>
<tr>
<td>(S) (Standard Deviation of Subjects)</td>
<td>18</td>
</tr>
<tr>
<td>Contrast Coefficients</td>
<td>None</td>
</tr>
</tbody>
</table>
Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>Average n</th>
<th>k</th>
<th>Total N</th>
<th>Alpha</th>
<th>Beta</th>
<th>Std Dev of Means (Sm)</th>
<th>Standard Deviation ($)</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.04238</td>
<td>2.00</td>
<td>4</td>
<td>8</td>
<td>0.01000</td>
<td>0.95762</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
</tr>
<tr>
<td>0.17513</td>
<td>2.00</td>
<td>4</td>
<td>8</td>
<td>0.05000</td>
<td>0.82487</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
</tr>
<tr>
<td>0.23886</td>
<td>4.00</td>
<td>4</td>
<td>16</td>
<td>0.01000</td>
<td>0.76114</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
</tr>
<tr>
<td>0.52165</td>
<td>4.00</td>
<td>4</td>
<td>16</td>
<td>0.05000</td>
<td>0.47835</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
</tr>
<tr>
<td>0.50581</td>
<td>6.00</td>
<td>4</td>
<td>24</td>
<td>0.01000</td>
<td>0.49419</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
</tr>
<tr>
<td>0.77327</td>
<td>6.00</td>
<td>4</td>
<td>24</td>
<td>0.05000</td>
<td>0.22673</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
</tr>
<tr>
<td>0.72695</td>
<td>8.00</td>
<td>4</td>
<td>32</td>
<td>0.01000</td>
<td>0.27305</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
</tr>
<tr>
<td>0.90642</td>
<td>8.00</td>
<td>4</td>
<td>32</td>
<td>0.05000</td>
<td>0.09358</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
</tr>
<tr>
<td>0.86702</td>
<td>10.00</td>
<td>4</td>
<td>40</td>
<td>0.01000</td>
<td>0.13298</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
</tr>
<tr>
<td>0.96514</td>
<td>10.00</td>
<td>4</td>
<td>40</td>
<td>0.05000</td>
<td>0.03486</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
</tr>
<tr>
<td>0.94143</td>
<td>12.00</td>
<td>4</td>
<td>48</td>
<td>0.01000</td>
<td>0.05857</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
</tr>
<tr>
<td>0.98802</td>
<td>12.00</td>
<td>4</td>
<td>48</td>
<td>0.05000</td>
<td>0.01198</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
</tr>
<tr>
<td>0.97623</td>
<td>14.00</td>
<td>4</td>
<td>56</td>
<td>0.01000</td>
<td>0.02377</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
</tr>
<tr>
<td>0.99614</td>
<td>14.00</td>
<td>4</td>
<td>56</td>
<td>0.05000</td>
<td>0.00386</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
</tr>
</tbody>
</table>

Report Definitions

Power is the probability of rejecting a false null hypothesis. It should be close to one.

n is the average group sample size.
k is the number of groups.
Total N is the total sample size of all groups.
Alpha is the probability of rejecting a true null hypothesis. It should be small.
Beta is the probability of accepting a false null hypothesis. It should be small.
Sm is the standard deviation of the group means under the alternative hypothesis.
Standard deviation is the within group standard deviation.
The Effect Size is the ratio of Sm to standard deviation.

Summary Statements

In a one-way ANOVA study, sample sizes of 2, 2, 2, and 2 are obtained from the 4 groups whose
means are to be compared. The total sample of 8 subjects achieves 4% power to detect
differences among the means versus the alternative of equal means using an F test with a
0.01000 significance level. The size of the variation in the means is represented by their
standard deviation which is 12.99. The common standard deviation within a group is assumed to
be 18.00.

This report shows the numeric results of this power study. Following are the definitions of the
columns of the report.

Power
The probability of rejecting a false null hypothesis.

Average n
The average of the group sample sizes.

k
The number of groups.

Total N
The total sample size of the study.

Alpha
The probability of rejecting a true null hypothesis. This is often called the significance level.
Beta
The probability of accepting a false null hypothesis that $Sm$ is zero when $Sm$ is actually equal to the value shown in the next column.

**Std Dev of Means (Sm)**
This is the standard deviation of the hypothesized means. It was computed from the hypothesized means. It is roughly equal to the average difference between the group means and the overall mean.

Once you have computed this, you can enter a range of values to determine the effect of the hypothesized means on the power.

**Standard Deviation (S)**
This is the within-group standard deviation. It was set in the Data window.

**Effect Size**
The effect size is the ratio of $SM$ to $S$. It is an index of relative difference between the means that can be compared from study to study.

**Detailed Results Report**

<table>
<thead>
<tr>
<th>Group</th>
<th>Ni</th>
<th>Percent Ni of Total Ni</th>
<th>Mean</th>
<th>Deviation From Mean</th>
<th>Times Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>25.00</td>
<td>40.00</td>
<td>22.50</td>
<td>45.00</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>25.00</td>
<td>10.00</td>
<td>7.50</td>
<td>15.00</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>25.00</td>
<td>10.00</td>
<td>7.50</td>
<td>15.00</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>25.00</td>
<td>10.00</td>
<td>7.50</td>
<td>15.00</td>
</tr>
<tr>
<td>ALL</td>
<td>8</td>
<td>100.00</td>
<td>17.50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Details when Alpha = 0.050000, Power = 0.17513, SM = 12.99, S = 18.00**

<table>
<thead>
<tr>
<th>Group</th>
<th>Ni</th>
<th>Percent Ni of Total Ni</th>
<th>Mean</th>
<th>Deviation From Mean</th>
<th>Times Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>25.00</td>
<td>40.00</td>
<td>22.50</td>
<td>45.00</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>25.00</td>
<td>10.00</td>
<td>7.50</td>
<td>15.00</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>25.00</td>
<td>10.00</td>
<td>7.50</td>
<td>15.00</td>
</tr>
<tr>
<td>4</td>
<td>2</td>
<td>25.00</td>
<td>10.00</td>
<td>7.50</td>
<td>15.00</td>
</tr>
<tr>
<td>ALL</td>
<td>8</td>
<td>100.00</td>
<td>17.50</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This report shows the details of each row of the previous report.

**Group**
The number of the group shown on this line. The last line, labeled *ALL*, gives the average or the total as appropriate.

**Ni**
This is the sample size of each group. This column is especially useful when the sample sizes are unequal.

**Percent Ni of Total Ni**
This is the percentage of the total sample that is allocated to each group.

**Mean**
The is the value of the Hypothesized Mean. The final row gives the average for all groups.
Deviation From Mean
This is the absolute value of the mean minus the overall mean. Since $Sm$ is the sum of the squared deviations, these values show the relative contribution to $Sm$.

Ni Times Deviation
This is the group sample size times the absolute deviation. It shows the combined influence of the size of the deviation and the sample size on $Sm$.

Plots Section
This plot gives a visual presentation to the results in the Numeric Report. We can quickly see the impact on the power of increasing the sample size and increase the significance level. When you create one of these plots, it is important to use trial and error to find an appropriate range for the horizontal variable so that you have results with both low and high power.

Example 2 – Power after a Study
This example will cover the situation in which you are calculating the power of a one-way analysis of variance $F$ test on data that have already been collected and analyzed.

An experiment included a control group and two treatment groups. Each group had seven individuals. A single response was measured for each individual and recorded in the following table.

<table>
<thead>
<tr>
<th>Control</th>
<th>T1</th>
<th>T2</th>
</tr>
</thead>
<tbody>
<tr>
<td>452</td>
<td>646</td>
<td>685</td>
</tr>
<tr>
<td>674</td>
<td>547</td>
<td>658</td>
</tr>
<tr>
<td>554</td>
<td>774</td>
<td>786</td>
</tr>
<tr>
<td>447</td>
<td>465</td>
<td>536</td>
</tr>
<tr>
<td>356</td>
<td>759</td>
<td>653</td>
</tr>
<tr>
<td>654</td>
<td>665</td>
<td>669</td>
</tr>
<tr>
<td>558</td>
<td>767</td>
<td>557</td>
</tr>
</tbody>
</table>
When analyzed using the one-way analysis of variance procedure in *NCSS*, the following results were obtained.

**Analysis of Variance Table**

<table>
<thead>
<tr>
<th>Source Term</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-Ratio</th>
<th>Prob Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>A ( ... )</td>
<td>2</td>
<td>75629.8</td>
<td>37814.9</td>
<td>3.28</td>
<td>0.061167</td>
</tr>
<tr>
<td>S(A)</td>
<td>18</td>
<td>207743.4</td>
<td>11541.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (Adjusted)</td>
<td>20</td>
<td>283373.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Means Section**

<table>
<thead>
<tr>
<th>Group</th>
<th>Count</th>
<th>Mean</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>7</td>
<td>527.8571</td>
</tr>
<tr>
<td>T1</td>
<td>7</td>
<td>660.4286</td>
</tr>
<tr>
<td>T2</td>
<td>7</td>
<td>649.1429</td>
</tr>
</tbody>
</table>

The significance level (Prob Level) was only 0.061—not enough for statistical significance. The researcher had hoped to show that the treatment groups had higher response levels than the control group. He could see that the group means followed this pattern since the mean for T1 was about 25% higher than the control mean and the mean for T2 was about 23% higher than the control mean. He decided to calculate the power of the experiment using these values of the means. (We do not recommend this approach because the power should be calculated for the minimum difference among the means that is of interest, not at the values of the sample means.)

The data entry for this problem is simple. The only entry that is not straightforward is finding an appropriate value for the standard deviation. Since the standard deviation is estimated by the square root of the mean square error, it is calculated as $\sqrt{11541.3} = 107.4304$.

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **One-Way Analysis of Variance** procedure window by expanding **Means**, then **ANOVA**, then clicking on **Analysis of Variance**, and then clicking on **One-Way Analysis of Variance**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>k (Number of Groups)</td>
<td>3</td>
</tr>
<tr>
<td>Group Allocation Ratios</td>
<td>Equal</td>
</tr>
<tr>
<td>n (Sample Size per Group)</td>
<td>7</td>
</tr>
<tr>
<td>Hypothesized Means</td>
<td>527.8571 660.4286 649.1429</td>
</tr>
<tr>
<td>S (Standard Deviation of Subjects)</td>
<td>107.4304</td>
</tr>
<tr>
<td>Contrast Coefficients</td>
<td>None</td>
</tr>
</tbody>
</table>
Output
Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>Average n</th>
<th>k</th>
<th>Total N</th>
<th>Alpha</th>
<th>Beta</th>
<th>Std Dev of Means (Sm)</th>
<th>Standard Deviation (S)</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.54788</td>
<td>7.00</td>
<td>3</td>
<td>21</td>
<td>0.05000</td>
<td>0.45212</td>
<td>60.01</td>
<td>107.43</td>
<td>0.5586</td>
</tr>
</tbody>
</table>

The power is only 0.55. That is, there was only a 55% chance of rejecting the false null hypothesis. It is important to understand this power statement is conditional, so we will state it in detail. Given that the population means are equal to the sample means (that $Sm$ is 60.01) and the population standard deviation is equal to 107.43, the probability of rejecting the false null hypothesis is 0.55. If the population means are different from the sample means (which they must be), the power is different. However, the sample means provide a reasonable place to begin.

Example 3 – Finding the Sample Size Necessary to Reject

Continuing with the last example, we will determine how large the sample size would need to have been for alpha = 0.05 and beta = 0.20.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the One-Way Analysis of Variance procedure window by expanding Means, then ANOVA, then clicking on Analysis of Variance, and then clicking on One-Way Analysis of Variance. You may then make the appropriate entries as listed below, or open Example 3 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td>Data Tab</td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>n (Sample Size)</td>
</tr>
<tr>
<td>Power</td>
<td>0.80</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>k (Number of Groups)</td>
<td>3</td>
</tr>
<tr>
<td>Group Allocation Ratios</td>
<td>Equal</td>
</tr>
<tr>
<td>n (Sample Size per Group)</td>
<td>Ignore since this is the Find setting</td>
</tr>
<tr>
<td>Hypothesized Means</td>
<td>527.8571 660.4286 649.1429</td>
</tr>
<tr>
<td>S (Standard Deviation of Subjects)</td>
<td>107.4304</td>
</tr>
<tr>
<td>Contrast Coefficients</td>
<td>None</td>
</tr>
</tbody>
</table>
Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Numeric Results for One-Way Analysis of Variance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power</td>
</tr>
<tr>
<td>-------</td>
</tr>
<tr>
<td>0.82511</td>
</tr>
</tbody>
</table>

The required sample size is 12 per group or 36 subjects.

Example 4 – Using Unequal Sample Sizes

Continuing with the last example, consider the impact of allowing the group sample sizes to be unequal. Since the control group is being compared to two treatment groups, the mean of the control group is assumed to be different from those of the treatment groups. In this situation, experience has shown that adding extra subjects to the control group can increase the power. In a separate analysis, the power with 11 subjects per group was found to be 0.7851—not quite the required 0.80.

We will try moving two subjects from each treatment group into the control group. This will give an experimental design with 15 in the control group and 9 in each of the treatment groups.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the One-Way Analysis of Variance procedure window by expanding Means, then ANOVA, then clicking on Analysis of Variance, and then clicking on One-Way Analysis of Variance. You may then make the appropriate entries as listed below, or open Example 4 by going to the File menu and choosing Open Example Template.

Pay particular attention to how the sample size parameters were changed. The value of n is set to one so that it is essentially ignored. The Group Sample Size Pattern contains the three unequal sample sizes.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>k (Number of Groups)</td>
<td>3</td>
</tr>
<tr>
<td>Group Allocation Ratios</td>
<td>15 9 9</td>
</tr>
<tr>
<td>n (Sample Size per Group)</td>
<td>1</td>
</tr>
<tr>
<td>Hypothesized Means</td>
<td>527.8571 660.4286 649.1429</td>
</tr>
<tr>
<td>S (Standard Deviation of Subjects)</td>
<td>107.4304</td>
</tr>
<tr>
<td>Contrast Coefficients</td>
<td>None</td>
</tr>
</tbody>
</table>
Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>Average n</th>
<th>k</th>
<th>Total N</th>
<th>Alpha</th>
<th>Beta</th>
<th>Std Dev of Means (Sm)</th>
<th>Standard Deviation (S)</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.82967</td>
<td>11.00</td>
<td>3</td>
<td>33</td>
<td>0.05000</td>
<td>0.17033</td>
<td>63.34</td>
<td>107.43</td>
<td>0.5896</td>
</tr>
</tbody>
</table>

The power of 0.82967 achieved with the 33 subjects in this design is slightly higher than the power of 0.82511 that was achieved with the 36 subjects in the equal group size design. Apparently, unequal sample allocation can achieve better power!

We suggest that you try several different sample allocations. You will find that the optimum sample allocation depends on the values of the hypothesized means.

You should keep in mind that power may not be the only goal of the experiment. Other goals may include finding confidence intervals for each of the group means. And the narrowness of the width of the confidence interval is directly related to the sample size.

Example 5 – Minimum Detectable Difference

It may be useful to determine the minimum detectable difference among the means that can be found at the experimental conditions. This amounts to finding $\sigma_m$ (which we call $Sm$ on the windows and printouts).

Continuing with the previous example, find $Sm$ for a wide range of sample sizes when alpha is 0.05 and beta is 0.10 or 0.20.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the One-Way Analysis of Variance procedure window by expanding Means, then ANOVA, then clicking on Analysis of Variance, and then clicking on One-Way Analysis of Variance. You may then make the appropriate entries as listed below, or open Example 5 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Sm (Std Dev of Means)</td>
</tr>
<tr>
<td>Power</td>
<td>0.80 0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>k (Number of Groups)</td>
<td>3</td>
</tr>
<tr>
<td>Group Allocation Ratios</td>
<td>Equal</td>
</tr>
<tr>
<td>n (Sample Size per Group)</td>
<td>2 3 5 8 10 15 20 40 60 80 100</td>
</tr>
<tr>
<td>Hypothesized Means</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>S (Standard Deviation of Subjects)</td>
<td>107.4304</td>
</tr>
<tr>
<td>Contrast Coefficients</td>
<td>None</td>
</tr>
</tbody>
</table>
Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results and Plots

<table>
<thead>
<tr>
<th>Power</th>
<th>Average</th>
<th>n</th>
<th>k</th>
<th>Total N</th>
<th>Alpha</th>
<th>Beta</th>
<th>Std Dev of Means (Sm)</th>
<th>Standard Deviation (S)</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.90000 2.00</td>
<td>3</td>
<td>6</td>
<td>9</td>
<td>0.05000</td>
<td>0.10000</td>
<td>287.18</td>
<td>107.43</td>
<td>2.6732</td>
<td></td>
</tr>
<tr>
<td>0.80000 2.00</td>
<td>3</td>
<td>6</td>
<td>9</td>
<td>0.05000</td>
<td>0.20000</td>
<td>244.31</td>
<td>107.43</td>
<td>2.2741</td>
<td></td>
</tr>
<tr>
<td>0.90000 3.00</td>
<td>3</td>
<td>9</td>
<td>15</td>
<td>0.05000</td>
<td>0.20000</td>
<td>145.82</td>
<td>107.43</td>
<td>1.3573</td>
<td></td>
</tr>
<tr>
<td>0.80000 3.00</td>
<td>3</td>
<td>9</td>
<td>15</td>
<td>0.05000</td>
<td>0.10000</td>
<td>112.62</td>
<td>107.43</td>
<td>1.0483</td>
<td></td>
</tr>
<tr>
<td>0.90000 5.00</td>
<td>3</td>
<td>15</td>
<td>30</td>
<td>0.05000</td>
<td>0.20000</td>
<td>64.42</td>
<td>107.43</td>
<td>0.5997</td>
<td></td>
</tr>
<tr>
<td>0.80000 5.00</td>
<td>3</td>
<td>15</td>
<td>30</td>
<td>0.05000</td>
<td>0.10000</td>
<td>59.07</td>
<td>107.43</td>
<td>0.5499</td>
<td></td>
</tr>
<tr>
<td>0.90000 8.00</td>
<td>3</td>
<td>24</td>
<td>60</td>
<td>0.05000</td>
<td>0.20000</td>
<td>51.54</td>
<td>107.43</td>
<td>0.4797</td>
<td></td>
</tr>
<tr>
<td>0.80000 8.00</td>
<td>3</td>
<td>24</td>
<td>60</td>
<td>0.05000</td>
<td>0.10000</td>
<td>50.67</td>
<td>107.43</td>
<td>0.4716</td>
<td></td>
</tr>
<tr>
<td>0.90000 10.00</td>
<td>3</td>
<td>30</td>
<td>90</td>
<td>0.05000</td>
<td>0.20000</td>
<td>44.21</td>
<td>107.43</td>
<td>0.4115</td>
<td></td>
</tr>
<tr>
<td>0.80000 10.00</td>
<td>3</td>
<td>30</td>
<td>90</td>
<td>0.05000</td>
<td>0.10000</td>
<td>43.10</td>
<td>107.43</td>
<td>0.3889</td>
<td></td>
</tr>
<tr>
<td>0.90000 15.00</td>
<td>3</td>
<td>45</td>
<td>135</td>
<td>0.05000</td>
<td>0.20000</td>
<td>28.73</td>
<td>107.43</td>
<td>0.2674</td>
<td></td>
</tr>
<tr>
<td>0.80000 15.00</td>
<td>3</td>
<td>45</td>
<td>135</td>
<td>0.05000</td>
<td>0.10000</td>
<td>27.62</td>
<td>107.43</td>
<td>0.2433</td>
<td></td>
</tr>
<tr>
<td>0.90000 20.00</td>
<td>3</td>
<td>60</td>
<td>180</td>
<td>0.05000</td>
<td>0.20000</td>
<td>20.14</td>
<td>107.43</td>
<td>0.2111</td>
<td></td>
</tr>
<tr>
<td>0.80000 20.00</td>
<td>3</td>
<td>60</td>
<td>180</td>
<td>0.05000</td>
<td>0.10000</td>
<td>19.03</td>
<td>107.43</td>
<td>0.1901</td>
<td></td>
</tr>
<tr>
<td>0.90000 40.00</td>
<td>3</td>
<td>120</td>
<td>360</td>
<td>0.05000</td>
<td>0.20000</td>
<td>11.86</td>
<td>107.43</td>
<td>0.0904</td>
<td></td>
</tr>
<tr>
<td>0.80000 40.00</td>
<td>3</td>
<td>120</td>
<td>360</td>
<td>0.05000</td>
<td>0.10000</td>
<td>10.75</td>
<td>107.43</td>
<td>0.0794</td>
<td></td>
</tr>
<tr>
<td>0.90000 60.00</td>
<td>3</td>
<td>180</td>
<td>540</td>
<td>0.05000</td>
<td>0.20000</td>
<td>6.35</td>
<td>107.43</td>
<td>0.0490</td>
<td></td>
</tr>
<tr>
<td>0.80000 60.00</td>
<td>3</td>
<td>180</td>
<td>540</td>
<td>0.05000</td>
<td>0.10000</td>
<td>5.24</td>
<td>107.43</td>
<td>0.0380</td>
<td></td>
</tr>
<tr>
<td>0.90000 80.00</td>
<td>3</td>
<td>240</td>
<td>720</td>
<td>0.05000</td>
<td>0.20000</td>
<td>2.98</td>
<td>107.43</td>
<td>0.0280</td>
<td></td>
</tr>
<tr>
<td>0.80000 80.00</td>
<td>3</td>
<td>240</td>
<td>720</td>
<td>0.05000</td>
<td>0.10000</td>
<td>2.87</td>
<td>107.43</td>
<td>0.0208</td>
<td></td>
</tr>
<tr>
<td>0.90000 100.00</td>
<td>3</td>
<td>300</td>
<td>900</td>
<td>0.05000</td>
<td>0.20000</td>
<td>2.00</td>
<td>107.43</td>
<td>0.0168</td>
<td></td>
</tr>
<tr>
<td>0.80000 100.00</td>
<td>3</td>
<td>300</td>
<td>900</td>
<td>0.05000</td>
<td>0.10000</td>
<td>1.90</td>
<td>107.43</td>
<td>0.0128</td>
<td></td>
</tr>
</tbody>
</table>

This plot shows the relationships between power, sample size, and detectable difference. Several conclusions are possible, but the most impressive is the sharp elbow in the curve that occurs near \( n = 10 \) when \( Sm \) is about 64.

How do you interpret an \( Sm \) of 64? The easiest way is to generate a set of means that have a standard deviation of 64. To do this, press the SD button in the lower right corner of the One Way ANOVA panel to load the Standard Deviation Estimator module. Set \( N = 3 \), Mean = 0, and Standard Deviation = 64. Press the Two Unique Values button. This results in three means equal to -91, 45, and 45. The differences among these means are the minimum detectable differences that can be detecting with a sample size of 9 when the power is 80%.
Example 6 – Validation using Fleiss

Fleiss (1986) page 374 presents an example of determining a sample size in an experiment with 4 groups; means of 9.775, 12, 12, and 14.225; standard deviation of 3; alpha of 0.05, and beta of 0.20. He finds a sample size of 11 per group.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the One-Way Analysis of Variance procedure window by expanding Means, then ANOVA, then clicking on Analysis of Variance, and then clicking on One-Way Analysis of Variance. You may then make the appropriate entries as listed below, or open Example 6 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For) ........................</td>
<td>n (Sample Size)</td>
</tr>
<tr>
<td>Power</td>
<td>0.80</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>k (Number of Groups) ...............</td>
<td>4</td>
</tr>
<tr>
<td>Group Allocation Ratios .............</td>
<td>Equal</td>
</tr>
<tr>
<td>n (Sample Size per Group) ..........</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Hypothesized Means ..................</td>
<td>9.775 12 12 14.225</td>
</tr>
<tr>
<td>S (Standard Deviation of Subjects)</td>
<td>3</td>
</tr>
<tr>
<td>Contrast Coefficients ..............</td>
<td>None</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

### Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>Average</th>
<th>n</th>
<th>k</th>
<th>Total N</th>
<th>Alpha</th>
<th>Beta</th>
<th>Std Dev of Means (Sm)</th>
<th>Standard Deviation (S)</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.80273</td>
<td>11.00</td>
<td>4</td>
<td>44</td>
<td>0.05000</td>
<td>0.19727</td>
<td>1.57</td>
<td>3.00</td>
<td>0.5244</td>
<td></td>
</tr>
</tbody>
</table>

Details when Alpha = 0.05000, Power = 0.80273, SM = 1.57, S = 3.00

<table>
<thead>
<tr>
<th>Group</th>
<th>Ni</th>
<th>Percent</th>
<th>Ni of Total Ni</th>
<th>Mean</th>
<th>From Mean</th>
<th>Deviation</th>
<th>Ni Times Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>11</td>
<td>25.00</td>
<td>11</td>
<td>9.78</td>
<td>2.23</td>
<td>24.48</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>25.00</td>
<td>12.00</td>
<td>12.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>25.00</td>
<td>12.00</td>
<td>12.00</td>
<td>0.00</td>
<td>0.00</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>25.00</td>
<td>14.23</td>
<td>14.23</td>
<td>2.23</td>
<td>24.48</td>
<td></td>
</tr>
<tr>
<td>ALL</td>
<td>44</td>
<td>100.00</td>
<td>12.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

PASS also found n = 11. Note that Fleiss used calculations based on a normal approximation, but PASS uses exact calculations based on the non-central F distribution.
Example 7 – Validation using Desu

Desu (1990) page 48 presents an example of determining a sample size in an experiment with 3 groups; means of 0, -0.2553, and 0.2553; standard deviation of 1; alpha of 0.05, and beta of 0.10. He finds a sample size of 99 per group.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the One-Way Analysis of Variance procedure window by expanding Means, then ANOVA, then clicking on Analysis of Variance, and then clicking on One-Way Analysis of Variance. You may then make the appropriate entries as listed below, or open Example 7 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>n (Sample Size)</td>
</tr>
<tr>
<td>Power</td>
<td>0.90</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>k (Number of Groups)</td>
<td>3</td>
</tr>
<tr>
<td>Group Allocation Ratios</td>
<td>Equal</td>
</tr>
<tr>
<td>n (Sample Size per Group)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Hypothesized Means</td>
<td>0 -0.2553 0.2553</td>
</tr>
<tr>
<td>S (Standard Deviation of Subjects)</td>
<td>1</td>
</tr>
<tr>
<td>Contrast Coefficients</td>
<td>None</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>Average N</th>
<th>k</th>
<th>Total N</th>
<th>Alpha</th>
<th>Beta</th>
<th>Std Dev of Means (Sm)</th>
<th>Standard Deviation (S)</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.90285</td>
<td>99.00</td>
<td>3</td>
<td>297</td>
<td>0.05000</td>
<td>0.09715</td>
<td>0.21</td>
<td>1.00</td>
<td>0.2085</td>
</tr>
</tbody>
</table>

Details when Alpha = 0.05000, Power = 0.90285, SM = 0.21, S = 1.00

<table>
<thead>
<tr>
<th>Group</th>
<th>Ni</th>
<th>Percent Ni of Total Ni</th>
<th>Mean</th>
<th>Deviation From Mean</th>
<th>Ni Times Deviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>99</td>
<td>33.33</td>
<td>0.00</td>
<td>0.00</td>
<td>33.33</td>
</tr>
<tr>
<td>2</td>
<td>99</td>
<td>33.33</td>
<td>-0.26</td>
<td>0.26</td>
<td>25.27</td>
</tr>
<tr>
<td>3</td>
<td>99</td>
<td>33.33</td>
<td>0.26</td>
<td>0.26</td>
<td>25.27</td>
</tr>
<tr>
<td>ALL</td>
<td>297</td>
<td>100.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>

PASS also found $n = 99$. 
Example 8 – Validation using Kirk

Kirk (1982) pages 140-144 presents an example of determining a sample size in an experiment with 4 groups; means of 2.75, 3.50, 6.25, and 9.0; standard deviation of 1.20995; alpha of 0.05, and beta of 0.05. He finds a sample size of 3 per group.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the One-Way Analysis of Variance procedure window by expanding Means, then ANOVA, then clicking on Analysis of Variance, and then clicking on One-Way Analysis of Variance. You may then make the appropriate entries as listed below, or open Example 8 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>n (Sample Size)</td>
</tr>
<tr>
<td>Power</td>
<td>0.95</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>k (Number of Groups)</td>
<td>4</td>
</tr>
<tr>
<td>Group Allocation Ratios</td>
<td>Equal</td>
</tr>
<tr>
<td>n (Sample Size per Group)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Hypothesized Means</td>
<td>2.75 3.5 6.25 9</td>
</tr>
<tr>
<td>S (Standard Deviation of Subjects)</td>
<td>1.20995</td>
</tr>
<tr>
<td>Contrast Coefficients</td>
<td>None</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Numeric Results</th>
<th>Power</th>
<th>Average n</th>
<th>k</th>
<th>Total N</th>
<th>Alpha</th>
<th>Beta</th>
<th>Std Dev of Means (Sm)</th>
<th>Standard Deviation (S)</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.99767</td>
<td>3.00</td>
<td>4</td>
<td>12</td>
<td>0.05000</td>
<td>0.00233</td>
<td>2.47</td>
<td>1.21</td>
<td>2.0376</td>
<td></td>
</tr>
</tbody>
</table>

PASS also found \( n = 3 \).
Example 9 – Power of a Planned Comparison

An experiment is being designed to study the response to different doses of a drug. Three groups, receiving a dose of 0, 10, and 20 milligrams of the drug, are anticipated. An $F$ test will be used to test the hypothesis that the means exhibit a linear trend across the doses. The significance level is 0.05. Previous studies have shown the within group standard deviation to be 18. Treatment means of 5, 16, and 30 represent clinically important treatment differences. To better understand the relationship between power and sample size, the researcher wants to compute the power for several group sample sizes between 2 and 18. The sample sizes will be equal across all groups.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the One-Way Analysis of Variance procedure window by expanding Means, then ANOVA, then clicking on Analysis of Variance, and then clicking on One-Way Analysis of Variance. You may then make the appropriate entries as listed below, or open Example 9 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>k (Number of Groups)</td>
<td>3</td>
</tr>
<tr>
<td>Group Allocation Ratios</td>
<td>Equal</td>
</tr>
<tr>
<td>n (Sample Size per Group)</td>
<td>2 to 18 by 2</td>
</tr>
<tr>
<td>Hypothesized Means</td>
<td>5 16 30</td>
</tr>
<tr>
<td>S (Standard Deviation of Subjects)</td>
<td>18</td>
</tr>
<tr>
<td>Contrast Coefficients</td>
<td>Linear Trend</td>
</tr>
<tr>
<td>Axes/Legend/Grid Tab</td>
<td></td>
</tr>
<tr>
<td>Vertical Range</td>
<td>User (Given Below)</td>
</tr>
<tr>
<td>Minimum</td>
<td>0</td>
</tr>
<tr>
<td>Maximum</td>
<td>1</td>
</tr>
</tbody>
</table>
### Annotated Output

Click the Run button to perform the calculations and generate the following output.

#### Numeric Results Report

<table>
<thead>
<tr>
<th>Power</th>
<th>n</th>
<th>k</th>
<th>N</th>
<th>Alpha</th>
<th>Beta</th>
<th>Std Dev of Means (Sm)</th>
<th>Standard Deviation (S)</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.16781</td>
<td>2.00</td>
<td>3</td>
<td>6</td>
<td>0.05000</td>
<td>0.83219</td>
<td>10.21</td>
<td>18.00</td>
<td>0.5670</td>
</tr>
<tr>
<td>0.41889</td>
<td>4.00</td>
<td>3</td>
<td>12</td>
<td>0.05000</td>
<td>0.58111</td>
<td>10.21</td>
<td>18.00</td>
<td>0.5670</td>
</tr>
<tr>
<td>0.61410</td>
<td>6.00</td>
<td>3</td>
<td>18</td>
<td>0.05000</td>
<td>0.38590</td>
<td>10.21</td>
<td>18.00</td>
<td>0.5670</td>
</tr>
<tr>
<td>0.75458</td>
<td>8.00</td>
<td>3</td>
<td>24</td>
<td>0.05000</td>
<td>0.24542</td>
<td>10.21</td>
<td>18.00</td>
<td>0.5670</td>
</tr>
<tr>
<td>0.84932</td>
<td>10.00</td>
<td>3</td>
<td>30</td>
<td>0.05000</td>
<td>0.15068</td>
<td>10.21</td>
<td>18.00</td>
<td>0.5670</td>
</tr>
<tr>
<td>0.91013</td>
<td>12.00</td>
<td>3</td>
<td>36</td>
<td>0.05000</td>
<td>0.08987</td>
<td>10.21</td>
<td>18.00</td>
<td>0.5670</td>
</tr>
<tr>
<td>0.94768</td>
<td>14.00</td>
<td>3</td>
<td>42</td>
<td>0.05000</td>
<td>0.05232</td>
<td>10.21</td>
<td>18.00</td>
<td>0.5670</td>
</tr>
<tr>
<td>0.97017</td>
<td>16.00</td>
<td>3</td>
<td>48</td>
<td>0.05000</td>
<td>0.02983</td>
<td>10.21</td>
<td>18.00</td>
<td>0.5670</td>
</tr>
<tr>
<td>0.98329</td>
<td>18.00</td>
<td>3</td>
<td>54</td>
<td>0.05000</td>
<td>0.01671</td>
<td>10.21</td>
<td>18.00</td>
<td>0.5670</td>
</tr>
</tbody>
</table>

**Summary Statements**

In a one-way ANOVA study, sample sizes of 2, 2, and 2 are obtained from the 3 groups whose means are to be compared using a planned comparison (contrast). The total sample of 6 subjects achieves 17% power to detect a non-zero contrast of the means versus the alternative that the contrast is zero using an F test with a 0.05000 significance level. The value of the contrast of the means is 25.00. The common standard deviation within a group is assumed to be 18.00.

This report shows the numeric results of this power study. Most of the definitions are the same as with the one-way ANOVA test. Following are the definitions that are different.

### Std Dev of Means (Sm)

When displaying results for planned comparisons, this is not the standard deviation of the hypothesized means. Instead, it is a special function of the coefficients and the hypothesized means given by the equation

\[
\sigma_{mc} = \frac{\sum_{i=1}^{k} c_i \mu_i}{\sqrt{N \sum_{i=1}^{k} c_i^2 \over n_i}}
\]

**Effect Size**

The effect size is the ratio of SM to S. It is an index of relative difference between the means that can be compared from study to study.
Details Report

<table>
<thead>
<tr>
<th>Group</th>
<th>Ni</th>
<th>Ni of Total N</th>
<th>Mean</th>
<th>Contrast Coefficient</th>
<th>Mean Times Contrast</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>33.33</td>
<td>5.00</td>
<td>-1.000</td>
<td>-5.00</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>33.33</td>
<td>16.00</td>
<td>0.000</td>
<td>0.00</td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>33.33</td>
<td>30.00</td>
<td>1.000</td>
<td>30.00</td>
</tr>
<tr>
<td>ALL</td>
<td>6</td>
<td>100.00</td>
<td>17.00</td>
<td>0.000</td>
<td>25.00</td>
</tr>
</tbody>
</table>

This report shows the details of each row of the previous report. It is especially useful because it shows the values of the contrast coefficients and the contrast (which is the value in the lower right corner of the table).

Plots Section

This plot gives a visual presentation to the results in the Numeric Report. We can quickly see the impact on the power of increasing the sample size.

When you create one of these plots, it is important to use trial and error to find an appropriate range for the horizontal variable so that you have results with both low and high power.
Chapter 551

Analysis of Covariance

Introduction

A common task in research is to compare the averages of two or more populations (groups). We might want to compare the income level of two regions, the nitrogen content of three lakes, or the effectiveness of four drugs. The one-way analysis of variance compares the means of two or more groups to determine if at least one mean is different from the others. The $F$ test is used to determine statistical significance.

Analysis of Covariance (ANCOVA) is an extension of the one-way analysis of variance model that adds quantitative variables (covariates). When used, it is assumed that their inclusion will reduce the size of the error variance and thus increase the power of the design.

Covariates can only be used if the assumption of parallel slopes is viable.

Planned Comparisons

$PASS$ performs power and sample size calculations for user-specified contrasts.

The usual $F$ test tests the hypothesis that all means are equal versus the alternative that at least one mean is different from the rest. Often, a more specific alternative is desired. For example, you might want to test whether the treatment means are different from the control mean, the low dose is different from the high dose, a linear trend exists across dose levels, and so on. These questions are tested using planned comparisons.

We call the comparison planned because it was determined before the experiment was conducted. We planned to test the comparison.

A comparison is a weighted average of the means, in which the weights may be negative. When the weights sum to zero, the comparison is called a contrast. $PASS$ provides results for contrasts. To specify a contrast, we need only specify the weights. Statisticians call these weights the contrast coefficients.

For example, suppose an experiment conducted to study a drug will have three dose levels: none (control), 20 mg., and 40 mg. The first question is whether the drug made a difference. If it did, the average response for the two groups receiving the drug should be different from the control. If we label the group means $M_0$, $M_{20}$, and $M_{40}$, we are interested in comparing $M_0$ with $M_{20}$ and $M_{40}$. This can be done in two ways. One way is to construct two tests, one comparing $M_0$ and $M_{20}$ and the other comparing $M_0$ and $M_{40}$. Another method is to perform one test comparing
M0 with the average of M20 and M40. These tests are conducted using planned comparisons. The coefficients are as follows:

**M0 vs. M20**

To compare M0 versus M20, use the coefficients -1,1,0. When applied to the group means, these coefficients result in the comparison M0(-1)+M20(1)+M40(0) which reduces to M20-M0. That is, this contrast results in the difference between the two group means. We can test whether this difference is non-zero using the \( t \) test (or \( F \) test since the square of the \( t \) test is an \( F \) test).

**M0 vs. M40**

To compare M0 versus M40, use the coefficients -1,0,1. When applied to the group means, these coefficients result in the comparison M0(-1)+M20(0)+M40(1) which reduces to M40-M0. That is, this contrast results in the difference between the two group means.

**M0 vs. Average of M20 and M40**

To compare M0 versus the average of M20 and M40, use the coefficients -2,1,1. When applied to the group means, these coefficients result in the comparison M0(-2)+M20(1)+M40(1) which is equivalent to M40+M20-2(M0).

To see how these coefficients were obtained, consider the following manipulations. Beginning with the difference between the average of M20 and M40 and M0, we obtain the coefficients above—aside from a scale factor of one-half.

\[
\frac{M_{20} + M_{40}}{2} - M_0 = \frac{M_{20}}{2} + \frac{M_{40}}{2} - M_0
\]

\[
= \frac{1}{2} (M_{20} + M_{40} - 2M_0)
\]

**Assumptions**

Using the \( F \) test requires certain assumptions. One reason for the popularity of the \( F \) test is its robustness in the face of assumption violation. However, if an assumption is not even approximately met, the significance levels and the power of the \( F \) test are invalidated. Unfortunately, in practice it often happens that several assumptions are not met. This makes matters even worse. Hence, steps should be taken to check the assumptions before important decisions are made.

The following assumptions are needed for a one-way analysis of variance:

1. The data are continuous (not discrete).
2. The data follow the normal probability distribution. Each group is normally distributed about the group mean.
3. The variances of the populations are equal.
4. The groups are independent. There is no relationship among the individuals in one group as compared to another.

5. Each group is a simple random sample from its population. Each individual in the population has an equal probability of being selected in the sample.

Additional assumptions are needed for an analysis of covariance:

1. The covariates have a linear relationship the response variable.

2. The slopes of these linear relationships between the covariate and the response variable are approximately equal across all groups.

**Technical Details for ANCOVA**

We found two, slightly different, formulations for computing power for analysis of covariance. Keppel (1991) gives results that modify the standard deviation by an amount proportional to its reduction because of the covariate. Borm et al. (2007) give results that use a normal approximation to the noncentral F distribution. We use the Keppel approach in **PASS**.

Suppose that each observation consists of a response measurement, $Y_i$, and one or more covariate measurements: $X_{i1}, X_{i2}, \ldots, X_{ip}$. Further suppose that samples of $n_1, n_2, \ldots, n_k$ observations will be obtained from each of $k$ groups. The multiple regression equation relating $Y$ to the $X$'s within the $i^{th}$ group is

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \cdots + \beta_p X_{ip} + \varepsilon$$

The $\beta$'s are the regression coefficients or slopes. Analysis of covariance assumes that, except for the intercept $\beta_0$, the slopes are equal across all groups. Thus, the difference between the means of any two groups is equal to the difference between their intercepts.

Let $\sigma^2$ denote the common variance of all groups ignoring the covariates and $\sigma^2_\varepsilon$ the within-group variance after considering the covariates. These values are related according to the formula

$$\sigma^2_\varepsilon = (1 - \rho^2)\sigma^2$$

where $\rho^2$ is the coefficient of multiple determination (estimated by $R^2$).

Given the above terminology, the ratio of the mean square between groups to the mean square within groups follows a central $F$ distribution with two parameters matching the degrees of freedom of the numerator mean square and the denominator mean square. When the null hypothesis of mean equality is rejected, the above ratio has a noncentral $F$ distribution which also depends on the noncentrality parameter, $\lambda$. This parameter is calculated as

$$\lambda = \bar{\mu}k \frac{\sigma_m^2}{\sigma^2_\varepsilon}$$

where

$$\sigma_m^2 = \frac{1}{k} \sum_{i=1}^{k} n_i (\mu_i - \bar{\mu})^2$$
Analysis of Covariance

\[ \bar{\mu} = \sum_{i=1}^{k} \frac{n_i \mu_i}{N} , \]

\[ N = \sum_{i=1}^{k} n_i , \]

and

\[ \bar{n} = \frac{N}{k} . \]

Some authors use the symbol \( \phi \) for the noncentrality parameter. The relationship between the two noncentrality parameters is

\[ \phi = \sqrt{\frac{\lambda}{k}} . \]

The process of planning an experiment should include the following steps:

1. Determine an estimate of the within group standard deviation, \( \sigma \). This may be done from prior studies, from experimentation with the Standard Deviation Estimation module, from pilot studies, or from crude estimates based on the range of the data. See the chapter on estimating the standard deviation for more details.
2. Determine a set of means that represent the group differences that you want to detect.
3. Determine the R-squared value between the response and the covariates.
4. Determine the appropriate group sample sizes that will ensure desired levels of \( \alpha \) and \( \beta \).

Power Calculations for ANCOVA

The calculation of the power of a particular test proceeds as follows:

1. Determine the critical value, \( F_{k-1,N-k-p,\alpha} \), where \( \alpha \) is the probability of a type-I error and \( k, p, \) and \( N \) are defined above. Note that this is a two-tailed test as no direction is assigned in the alternative hypothesis.
2. From a hypothesized set of \( \mu_i \)'s, calculate the noncentrality parameter \( \lambda \) based on the values of \( N, k, \sigma_m, \rho^2 \), and \( \sigma \).
3. Compute the power as the probability of being greater than \( F_{k-1,N-k-p,\alpha} \) on a noncentral-\( F \) distribution with noncentrality parameter \( \lambda \).

Technical Details for a Planned Comparison

The terminology of planned comparisons is identical to that of the one-way AOV, so the notation used above will be repeated here.

Suppose you want to test whether the contrast \( C \)
Analysis of Covariance

\[ C = \sum_{i=1}^{k} c_i \mu_i \]

is significantly different from zero. Here the \( c_i \)'s are the contrast coefficients.

Define

\[ \sigma_{mc} = \sqrt{\frac{\sum_{i=1}^{k} c_i \mu_i}{\sum_{i=1}^{k} c_i^2}} \]

Define the noncentrality parameter \( \lambda_c \), as

\[ \lambda = \bar{n} k \frac{\sigma_{mc}^2}{\sigma^2} \]

**Power Calculations for Planned Comparisons**

The calculation of the power of a particular test proceeds as follows:

1. Determine the critical value, \( F_{(1,N-k-p,\alpha)} \) where \( \alpha \) is the probability of a type-I error and \( k \) and \( N \) are defined above. Note that this is a two-tailed test as no direction is assigned in the alternative hypothesis.

2. From a hypothesized set of \( \mu_i \)'s, calculate the noncentrality parameter \( \lambda_c \) based on the values of \( N, k, \sigma_{mc}, \rho^2 \), and \( \sigma \).

3. Compute the power as the probability of being greater than \( F_{(1,N-k-p,\alpha)} \) on a noncentral-\( F \) distribution with noncentrality parameter \( \lambda_c \).

**Procedure Options**

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

**Data Tab**

The Data tab contains most of the parameters and options that you will be concerned with.

**Solve For**

**Find (Solve For)**

This option specifies the parameter to be solved for from the other parameters. The parameters that may be selected are \( SM, S, k, n, Alpha \), and \( Power \ and \ Beta \). Under most situations, you will select either \( Power \ and \ Beta \) for a power analysis or \( n \) for sample size determination.
Error Rates

Power or Beta
This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of equal means when in fact the means are different.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

Alpha (Significance Level)
This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when you reject the null hypothesis of equal means when in fact the means are equal.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.

Sample Size / Groups

k (Number of Groups)
This is the number of group means being compared. It must be greater than or equal to two.

Note that the number of items used in the Hypothesized Means box and the Group Allocation Ratios box is controlled by this number.

Group Allocation Ratios
A set of positive, numeric values, one for each group, is entered here. The sample size of group $i$ is found by multiplying the $i^{th}$ number from this list times the value of $n$ and rounding up to the next whole number. The number of values must match the number of groups, $k$. When too few numbers are entered, 1’s are added. When too many numbers are entered, the extras are ignored.

- Equal
  If all sample sizes are to be equal, enter “Equal” here and the desired sample size in $n$. A set of $k$ 1’s will be used. This will result in $N_1 = N_2 = N_3 = n$. That is, all sample sizes are equal to $n$.

n (Sample Size per Group)
This is the base, per group, sample size. One or more values, separated by blanks or commas, may be entered. A separate analysis is performed for each value listed here.

The group samples sizes are determined by multiplying this number by each of the Group Allocation Ratios numbers. If the Group Allocation Ratios numbers are represented by $m_1, m_2,$
$m_3, \ldots, m_k$ and this value is represented by $n$, the group sample sizes $N_1, N_2, N_3, \ldots, N_k$ are calculated as follows:

\[
N_1 = \lfloor n(m_1) \rfloor \\
N_2 = \lfloor n(m_2) \rfloor \\
N_3 = \lfloor n(m_3) \rfloor \\
\text{etc.}
\]

where the operator, $\lfloor X \rfloor$ means the next integer after $X$, e.g. $\lfloor 3.1 \rfloor = 4$.

For example, suppose there are three groups and the Group Allocation Ratios is set to $1, 2, 3$. If $n$ is 5, the resulting sample sizes will be 5, 10, and 15. If $n$ is 50, the resulting group sample sizes will be 50, 100, and 150. If $n$ is set to $2, 4, 6, 8, 10$, five sets of group sample sizes will be generated and an analysis run for each. These sets are:

<table>
<thead>
<tr>
<th>n</th>
<th>N1</th>
<th>N2</th>
<th>N3</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>4</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>12</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>12</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>16</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>20</td>
<td>30</td>
<td></td>
</tr>
</tbody>
</table>

As a second example, suppose there are three groups and the Group Allocation Ratios is $0.2, 0.3, 0.5$. When fractional allocation ratios values sum to one, $n$ can be interpreted as the total sample size of all groups and the allocation ratios values as the proportion of the total in each group.

- If $n$ is 10, the three group sample sizes would be 2, 3, and 5.
- If $n$ is 20, the three group sample sizes would be 4, 6, and 10.
- If $n$ is 12, the three group sample sizes would be
  - $(0.2)12 = 2.4$ which is rounded up to the next whole integer, 3.
  - $(0.3)12 = 3.6$ which is rounded up to the next whole integer, 4.
  - $(0.5)12 = 6$.

Note that in this case, $3+4+6$ does not equal $n$ (which is 12). This can happen because of rounding.

**Effect Size – Means**

**Hypothesized Means**

Enter a set of hypothesized means, one for each group. These means represent the group centers under the alternative hypothesis (the null hypothesis is that they are equal). The standard deviation of these means ($\bar{\mu}$) is used in the power calculations to represent the average size of the differences among the means. The standard deviation of the means is calculated using the formula:

\[
\sigma_m = \sqrt{\frac{\sum_{i=1}^{k} (\mu_i - \bar{\mu})^2}{k}}
\]

This quantity gives the magnitude of the differences among the group means. Note that when all means are equal, $\sigma_m$ is zero.
**Analysis of Covariance**

You should enter a set of means that give the pattern of differences you expect or the pattern that you wish to detect. For example, in a particular study involving three groups, your research might be “meaningful” if either of two treatment means is 50% larger than the control mean. If the control mean is 50, then you would enter 50, 75, 75 as the three means.

It is usually more intuitive to enter a set of mean values. However, it is possible to enter the standard deviation of the means directly by placing an $S$ in front of the number (see below).

Some might wish to specify the alternative hypothesis as the effect size, $f$, which is defined as

$$f = \frac{\sigma_m}{\sigma}$$

If so, set $\sigma = 1$ and $\sigma_m = f$. Cohen (1988) has designated values of $f$ less than 0.1 as small, values around 0.25 to be medium, and values over 0.4 to be large.

**Entering a List of Means**

If a set of numbers is entered without a leading $S$, they are assumed to be the hypothesized group means under the alternative hypothesis. Their standard deviation will be calculated and used in the calculations. Blanks or commas may separate the numbers. Note that it is not the values of the means themselves that is important, but only their differences. Thus, the mean values 0, 1, 2 produce the same results as the values 100, 101, 102.

If too few means are entered to match the number of groups, the last mean is repeated. For example, suppose that four means are needed and you enter 1, 2 (only two means). **PASS** will treat this as 1, 2, 2, 2. If too many values are entered, **PASS** will truncate the list to the number of means needed.

**Examples:**
- 5 2 0 6 0
- 2, 5, 7
- -4, 0, 6, 9

**S Option**

If an $S$ is entered before the list of numbers, they are assumed to be values of $\sigma_m$, the standard deviations of the group means. A separate power calculation is made for each value. Note that this list can be a TO-BY phrase.

**Examples:**
- S 4.7
- S 4.3 5.7 4.2
- S 10 to 20 by 2

**Effect Size – Standard Deviation**

**S (Standard Deviation of Subjects)**

This is $\sigma$, the standard deviation within a group. It represents the variability from subject to subject that occurs when the subjects are treated identically. It is assumed to be the same for all groups. This value is approximated in an analysis of variance table by the square root of the mean square error.
Since they are positive square roots, the numbers must be strictly greater than zero. You can press the SD button to obtain further help on estimating the standard deviation.

Note that if you are using this procedure to test a factor (such as an interaction) from a more complex design, the value of standard deviation is estimated by the square root of the mean square of the term that is used as the denominator in the $F$ test.

You can enter a list of values separated by blanks or commas, in which case, a separate analysis will be calculated for each value.

Examples of valid entries:

- 1,4,7,10
- 1 4 7 10
- 1 to 10 by 3

**Effect Size – Covariates**

**Number of Covariates**
This is the number of covariates (X’s) in the study. Because of the stringent assumptions, this value is usually set to 1 or 2.

**R2 (R-Squared with Covariates)**
This is the average R-squared value that is achieved within a group by the covariates. It must be between 0 and 1.

**Planned Comparisons**

**Contrast Coefficients**
If you want to analyze a specific planned comparison, enter a set of contrast coefficients here. The calculations will then refer to the hypothesis that the corresponding contrast of the means is zero versus the alternative that it is non-zero (two-sided test). These are often called Planned Comparisons.

A contrast is a weighted average of the means in which the weights sum to zero. For example, suppose you are studying four groups and that the main hypothesis of interest is whether there is a linear trend across the groups. You would enter -3, -1, 1, 3 here. This would form the weighted average of the means:

$$-3(\text{Mean}_1)-(\text{Mean}_2)+(\text{Mean}_3)+3(\text{Mean}_3)$$

The point to realize is that these numbers (the coefficients) are used to calculate a specific weighted average of the means which is to be compared against zero using a standard $F$ (or $t$) test.

- **NONE or blank**
  When the box is left blank or the word *None* is entered, this option is ignored.

- **Linear Trend**
  A set of coefficients is generated appropriate for testing the alternative hypothesis that there is a linear (straight-line) trend across the means. These coefficients assume that the means are equally spaced across the trend variable.
• **Quadratic**
A set of coefficients is generated appropriate for testing the alternative hypothesis that the means follow a quadratic model. These coefficients assume that the means are equally spaced across the implicit X variable.

• **Cubic**
A set of coefficients is generated appropriate for testing the alternative hypothesis that the means follow a cubic model. These coefficients assume that the means are equally spaced across the implicit X variable.

• **First Against Others**
A set of coefficients is generated appropriate for testing the alternative hypothesis that the first mean is different from the average of the remaining means. For example, if there were four groups, the generated coefficients would be -3, 1, 1, 1.

• **List of Coefficients**
A list of coefficients, separated by commas or blanks, may be entered. If the number of items in the list does not match the number of groups (k), zeros are added or extra coefficients are truncated.

Remember that these coefficients must sum to zero. Also, the scale of the coefficients does not matter. That is 0.5,0.25,0.25; -2,1,1; and -200,100,100 will yield the same results.

To avoid rounding problems, it is better to use -3,1,1,1 than the equivalent -1,0.333,0.333,0.333. The second set does not sum to zero.

---

### Example 1 – Finding the Statistical Power

An experiment is being designed to compare the means of four groups using an $F$ test with a significance level of 0.05. A covariate is available that is estimated to have an R-squared of 0.4 with the response. Previous studies have shown that the standard deviation within a group is 18. Note that this value ignores the covariate.

Treatment means of 40, 10, 10, and 10 represent clinically important treatment differences. To better understand the relationship between power, sample size, and R-squared, the researcher wants to compute the power for R-squared's of 0.2, 0.3, 0.4, and 0.5, and for several group sample sizes between 2 and 10. The sample sizes will be equal across all groups.

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Analysis of Covariance** procedure window by expanding **Means**, then **ANOVA**, then clicking on **Analysis of Covariance**, and then clicking on **Analysis of Covariance**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.
Analysis of Covariance  551-11

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>k (Number of Groups)</td>
<td>4</td>
</tr>
<tr>
<td>Group Allocation Ratios</td>
<td>Equal</td>
</tr>
<tr>
<td>n (Sample Size per Group)</td>
<td>2 to 14 by 2</td>
</tr>
<tr>
<td>Hypothesized Means</td>
<td>40 10 10 10</td>
</tr>
<tr>
<td>S (Standard Deviation of Subjects)</td>
<td>18</td>
</tr>
<tr>
<td>Number of Covariates</td>
<td>1</td>
</tr>
<tr>
<td>R2 (R-Squared with Covariates)</td>
<td>0.2 to 0.5 by 0.1</td>
</tr>
<tr>
<td>Contrast Coefficients</td>
<td>None</td>
</tr>
</tbody>
</table>

Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>Average</th>
<th>Groups</th>
<th>Total</th>
<th>Alpha</th>
<th>Beta</th>
<th>Std Dev of Means</th>
<th>Standard Deviation</th>
<th>Effect Size</th>
<th>Cov's</th>
<th>R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.17245</td>
<td>2.0</td>
<td>4</td>
<td>8</td>
<td>0.05000</td>
<td>0.82755</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
<td>1</td>
<td>0.200</td>
</tr>
<tr>
<td>0.19041</td>
<td>2.0</td>
<td>4</td>
<td>8</td>
<td>0.05000</td>
<td>0.80959</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
<td>1</td>
<td>0.300</td>
</tr>
<tr>
<td>0.21428</td>
<td>2.0</td>
<td>4</td>
<td>8</td>
<td>0.05000</td>
<td>0.78572</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
<td>1</td>
<td>0.400</td>
</tr>
<tr>
<td>0.24742</td>
<td>2.0</td>
<td>4</td>
<td>8</td>
<td>0.05000</td>
<td>0.75258</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
<td>1</td>
<td>0.500</td>
</tr>
<tr>
<td>0.61111</td>
<td>4.0</td>
<td>4</td>
<td>16</td>
<td>0.05000</td>
<td>0.38889</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
<td>1</td>
<td>0.200</td>
</tr>
<tr>
<td>0.67475</td>
<td>4.0</td>
<td>4</td>
<td>16</td>
<td>0.05000</td>
<td>0.32525</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
<td>1</td>
<td>0.300</td>
</tr>
<tr>
<td>0.74725</td>
<td>4.0</td>
<td>4</td>
<td>16</td>
<td>0.05000</td>
<td>0.25275</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
<td>1</td>
<td>0.400</td>
</tr>
<tr>
<td>0.82656</td>
<td>4.0</td>
<td>4</td>
<td>16</td>
<td>0.05000</td>
<td>0.17344</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
<td>1</td>
<td>0.500</td>
</tr>
<tr>
<td>0.86165</td>
<td>6.0</td>
<td>4</td>
<td>24</td>
<td>0.05000</td>
<td>0.13835</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
<td>1</td>
<td>0.200</td>
</tr>
<tr>
<td>0.90662</td>
<td>6.0</td>
<td>4</td>
<td>24</td>
<td>0.05000</td>
<td>0.09338</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
<td>1</td>
<td>0.300</td>
</tr>
<tr>
<td>0.94625</td>
<td>6.0</td>
<td>4</td>
<td>24</td>
<td>0.05000</td>
<td>0.05375</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
<td>1</td>
<td>0.400</td>
</tr>
<tr>
<td>0.97632</td>
<td>6.0</td>
<td>4</td>
<td>24</td>
<td>0.05000</td>
<td>0.02368</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
<td>1</td>
<td>0.500</td>
</tr>
<tr>
<td>0.95878</td>
<td>8.0</td>
<td>4</td>
<td>32</td>
<td>0.05000</td>
<td>0.04122</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
<td>1</td>
<td>0.200</td>
</tr>
<tr>
<td>0.97817</td>
<td>8.0</td>
<td>4</td>
<td>32</td>
<td>0.05000</td>
<td>0.02183</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
<td>1</td>
<td>0.300</td>
</tr>
<tr>
<td>0.99102</td>
<td>8.0</td>
<td>4</td>
<td>32</td>
<td>0.05000</td>
<td>0.00898</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
<td>1</td>
<td>0.400</td>
</tr>
<tr>
<td>0.99758</td>
<td>8.0</td>
<td>4</td>
<td>32</td>
<td>0.05000</td>
<td>0.00242</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
<td>1</td>
<td>0.500</td>
</tr>
<tr>
<td>0.98919</td>
<td>10.0</td>
<td>4</td>
<td>40</td>
<td>0.05000</td>
<td>0.01081</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
<td>1</td>
<td>0.200</td>
</tr>
<tr>
<td>0.99559</td>
<td>10.0</td>
<td>4</td>
<td>40</td>
<td>0.05000</td>
<td>0.00441</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
<td>1</td>
<td>0.300</td>
</tr>
<tr>
<td>0.99873</td>
<td>10.0</td>
<td>4</td>
<td>40</td>
<td>0.05000</td>
<td>0.00127</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
<td>1</td>
<td>0.400</td>
</tr>
<tr>
<td>0.99980</td>
<td>10.0</td>
<td>4</td>
<td>40</td>
<td>0.05000</td>
<td>0.00020</td>
<td>12.99</td>
<td>18.00</td>
<td>0.7217</td>
<td>1</td>
<td>0.500</td>
</tr>
</tbody>
</table>

References

Borm, Fransen, and Lemmens. 2007. 'A simple sample size formula for analysis of covariance in randomized clinical trials.' J of Clinical Epidemiology, 60, 1234-1238.

Report Definitions

Power is the probability of rejecting a false null hypothesis. It should be close to one.
n is the average group sample size.
k is the number of groups.
Total N is the total sample size of all groups combined.
Alpha is the probability of rejecting a true null hypothesis. It should be small.
**551-12 Analysis of Covariance**

Beta is the probability of accepting a false null hypothesis. It should be small. Sm is the standard deviation of the group means under the alternative hypothesis. Standard deviation is the within group standard deviation. The Effect Size is the ratio of Sm to standard deviation. Cov's is the number of covariates. R2 (R-Squared) gives the strength of the relationship between the response and the covariates.

**Summary Statements**

In an analysis of covariance study, sample sizes of 2, 2, 2, and 2 are obtained from each of the 4 groups whose means are to be compared. The covariate has an R-squared of 0.200. The total sample of 8 subjects achieves 17% power to detect differences among the means versus the alternative of equal means using an F test with a 0.05000 significance level. The size of the variation in the means is represented by their standard deviation which is 12.99. The common standard deviation within a group is assumed to be 18.00.

This report shows the numeric results of this power study. Following are the definitions of the columns of the report.

**Power**
The probability of rejecting a false null hypothesis.

**Average n**
The average of the group sample sizes.

**k**
The number of groups.

**Total N**
The total sample size of the study.

**Alpha**
The probability of rejecting a true null hypothesis. This is often called the significance level.

**Beta**
The probability of accepting a false null hypothesis that Sm is zero when Sm is actually equal to the value shown in the next column.

**Std Dev of Means (Sm)**
This is the standard deviation of the hypothesized means. It was computed from the hypothesized means. It is roughly equal to the average difference between the group means and the overall mean.

Once you have computed this, you can enter a range of values to determine the effect of the hypothesized means on the power.

**Standard Deviation (S)**
This is the within-group standard deviation. It was set in the Data window.

**Effect Size**
The effect size is the ratio of SM to S. It is an index of relative difference between the means that can be compared from study to study.

**Cov’s**
This is the number of covariates.
R2
This is the value of R-squared.

Detailed Results Report

These reports show details of each row of the previous report.

Group
The number of the group shown on this line. The last line, labeled ALL, gives the average or the total as appropriate.

Ni
This is the sample size of each group. This column is especially useful when the sample sizes are unequal.

Percent Ni of Total Ni
This is the percentage of the total sample that is allocated to each group.

Mean
The is the value of the Hypothesized Mean. The final row gives the average for all groups.

Deviation From Mean
This is the absolute value of the mean minus the overall mean. Since Sm is the sum of the squared deviations, these values show the relative contribution to Sm.

Ni Times Deviation
This is the group sample size times the absolute deviation. It shows the combined influence of the size of the deviation and the sample size on Sm.
Plots Section

This plot gives a visual presentation to the results in the Numeric Report. We can quickly see the impact on the power of increasing the sample size and increasing the R-squared. Note that the value of R-squared has a large impact on power for small sample sizes, but now for larger ones.

Example 2 – Validation using Borm, et al.

Borm, Fransen, and Lemmens (2007) page 1237 presents an example of determining a sample size in an experiment with 2 groups, mean difference of 0.6, standard deviation of 1.2, alpha of 0.05, one covariate with an R-squared of 0.25, and power of 0.80. They find a total sample size of 95.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Analysis of Covariance procedure window by expanding Means, then ANOVA, then clicking on Analysis of Covariance, and then clicking on Analysis of Covariance. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.
<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>( n ) (Sample Size)</td>
</tr>
<tr>
<td>Power</td>
<td>0.80</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>( k ) (Number of Groups)</td>
<td>2</td>
</tr>
<tr>
<td>Group Allocation Ratios</td>
<td>Equal</td>
</tr>
<tr>
<td>Hypothesized Means</td>
<td>0.0 0.6</td>
</tr>
<tr>
<td>( S ) (Standard Deviation of Subjects)</td>
<td>1.2</td>
</tr>
<tr>
<td>( n ) (Sample Size per Group)</td>
<td>Ignored since this is to be solved for</td>
</tr>
<tr>
<td>Number of Covariates</td>
<td>1</td>
</tr>
<tr>
<td>R2 (R-Squared with Covariates)</td>
<td>0.25</td>
</tr>
<tr>
<td>Contrast Coefficients</td>
<td>None</td>
</tr>
</tbody>
</table>

### Output

Click the Run button to perform the calculations and generate the following output.

### Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>Average ( n )</th>
<th>Groups ( k )</th>
<th>Total N</th>
<th>Alpha</th>
<th>Beta</th>
<th>Std Dev Mean of Means ( (Sm) )</th>
<th>Standard Deviation ( (S) )</th>
<th>Effect Size</th>
<th>Cov's</th>
<th>R2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.80752</td>
<td>49.0</td>
<td>2</td>
<td>98</td>
<td>0.05000</td>
<td>0.19248</td>
<td>0.30</td>
<td>1.20</td>
<td>0.2500</td>
<td>1</td>
<td>0.250</td>
</tr>
</tbody>
</table>

**PASS** also found \( N = 98 \). Note that Borm (2007) used calculations based on a normal approximation and obtained \( N = 95 \), but **PASS** uses exact calculations based on the non-central \( F \) distribution.
551-16  Analysis of Covariance
Chapter 555

One-Way Analysis of Variance (Simulation)

Introduction

This procedure analyzes the power and significance level of the parametric F-Test and the nonparametric Kruskal-Wallis test which are used to test statistical hypotheses in a one-way experimental design. For each scenario that is set up, two simulations are run. One simulation estimates the significance level and the other estimates the power.

Technical Details

Computer simulation allows us to estimate the power and significance level that is actually achieved by a test procedure in situations that are not mathematically tractable. Computer simulation was once limited to mainframe computers. But, in recent years, as computer speeds have increased, simulation studies can be completed on desktop and laptop computers in a reasonable period of time.

The steps to a simulation study are

1. Specify how the test is carried out. This includes indicating how the test statistic is calculated and how the significance level is specified.

2. Generate random samples from the distributions specified by the alternative hypothesis. Calculate the test statistics from the simulated data and determine if the null hypothesis is accepted or rejected. Tabulate the number of rejections and use this to calculate the test’s power.

3. Generate random samples from the distributions specified by the null hypothesis. Calculate each test statistic from the simulated data and determine if the null hypothesis is accepted or rejected. Tabulate the number of rejections and use this to calculate the test’s significance level.

4. Repeat steps 2 and 3 several thousand times, tabulating the number of times the simulated data leads to a rejection of the null hypothesis. The power is the proportion of simulated samples in step 2 that lead to rejection. The significance level is the proportion of simulated samples in step 3 that lead to rejection.
Generating Random Distributions

Two methods are available in PASS to simulate random samples. The first method generates the random variates directly, one value at a time. The second method generates a large pool (over 10,000) of random values and then draw the random numbers from this pool. This second method can cut the running time of the simulation by 70%.

As mentioned above, the second method begins by generating a large pool of random numbers from the specified distributions. Each of these pools is evaluated to determine if its mean is within a small relative tolerance (0.0001) of the target mean. If the actual mean is not within the tolerance of the target mean, individual members of the population are replaced with new random numbers if the new random number moves the mean towards its target. Only a few hundred such swaps are required to bring the actual mean to within tolerance of the target mean. This population is then sampled with replacement using the uniform distribution. We have found that this method works well as long as the size of the pool is the maximum of twice the number of simulated samples desired and 10,000.

Test Statistics

Suppose \( g \) groups each have a normal distribution and means \( \mu_1, \mu_2, \ldots, \mu_g \) and common standard deviation \( \sigma \). Let \( n_1, n_2, \ldots, n_g \) denote the number of subjects in each group and let \( N \) denote the total sample size of all groups. The tests that follow assume that the data are obtained by taking simple random samples from the \( g \) populations.

F-Test

The formula for the calculation of the F-test is

\[
F_{g-1, N-g} = \frac{MSR}{MSE}
\]

where

\[
MSR = \frac{\sum_{k=1}^{g} n_k (\bar{X}_k - \bar{X})^2}{g - 1}
\]

and

\[
MSE = \frac{\sum_{k=1}^{g} \sum_{j=1}^{n_k} (X_{kj} - \bar{X}_k)^2}{N - g}
\]

\[
\bar{X}_k = \frac{\sum_{j=1}^{n_k} X_{kj}}{n_k}
\]
One-Way Analysis of Variance (Simulation)

\[
\bar{X} = \frac{\sum_{k=1}^{g} n_k \bar{X}_k}{N}
\]

\[
N = \sum_{k=1}^{g} n_k
\]

If the assumptions are met, the distribution of this test statistic follows the \( F \) distribution with degrees of freedom \( g-1 \) and \( N-g \).

**Kruskal-Wallis Test**

The Kruskal-Wallis test corrected for ties is calculated using the formula

\[
W = \frac{H}{T_c}
\]

where

\[
H = \frac{12}{N(N+1)} \sum_{k=1}^{g} \frac{R_k^2}{n_k} - 3(N+1)
\]

\[
T_c = 1 - \sum \frac{t(t^2-1)}{N(N^2-1)}
\]

\( R_k \) is the sum of the ranks of the \( k \)th group, and \( t \) is the count of a particular tie. The distribution of \( W \) is approximately Chi-square with \( g-1 \) degrees of freedom.

**Procedure Options**

This section describes the options that are specific to this procedure. These are located on the Data1, Data2, Reports, and Options tabs. For more information about the options of other tabs, go to the Procedure Window chapter.

**Data 1 Tab**

The Data 1 tab contains most of the parameters and options that you will be concerned with.

**Solve For**

**Find (Solve For)**

This option specifies the parameter to be calculated using the values of the other parameters.

Under most conditions, you would select either Power or \( n \).

Select Power when you want to estimate the power of a certain scenario.

Select \( n \) when you want to determine the sample size needed to achieve a given power and alpha error level. This option is very computationally intensive, so it may take a long time to complete.
Error Rates

Power or Beta
This option specifies one or more values for power or for beta (depending on the chosen setting). Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when a false null hypothesis is not rejected. In this procedure, a type-II error occurs when you fail to reject the null hypothesis of equal means when in fact the means are different.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

A single value may be entered here or a range of values such as 0.8 to 0.95 by 0.05 may be entered.

Alpha (Significance Level)
This option specifies one or more values for the probability of a type-I error. A type-I error occurs when a true null hypothesis is rejected. In this procedure, a type-I error occurs when you reject the null hypothesis of equal means when in fact the means are equal.

Values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This means that about one test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You may enter a range of values such as 0.01 0.05 0.10 or 0.01 to 0.10 by 0.01.

Sample Size

n (Sample Size Multiplier)
This is the base, per group, sample size. One or more values, separated by blanks or commas, may be entered. A separate analysis is performed for each value listed here.

The group samples sizes are determined by multiplying this number by each of the Group Sample Size Pattern numbers. If the Group Sample Size Pattern numbers are represented by m1, m2, m3, ..., mk and this value is represented by n, the group sample sizes N1, N2, N3, ..., Nk are calculated as follows:

\[ n1=[n(m1)] \]
\[ n2=[n(m2)] \]
\[ n3=[n(m3)] \]

etc.

where the operator, \([X]\) means the next integer after \(X\), e.g. \([3.1]=4\). This is required since sample sizes must be whole numbers.
For example, suppose there are three groups and the Group Sample Size Pattern is set to 1,2,3. If n is 5, the resulting sample sizes will be 5, 10, and 15. If n is 50, the resulting group sample sizes will be 50, 100, and 150. If n is set to 2,4,6,8,10; five sets of group sample sizes will be generated and an analysis run for each. These sets are:

2  4  6
4  8 12
6 12 18
8 16 24
10 20 30

As a second example, suppose there are three groups and the Group Sample Size Pattern is 0.2,0.3,0.5. When the fractional Pattern values sum to one, n can be interpreted as the total sample size N of all groups and the Pattern values as the proportion of the total in each group.

If n is 10, the three group sample sizes would be 2, 3, and 5.
If n is 20, the three group sample sizes would be 4, 6, and 10.
If n is 12, the three group sample sizes would be

$(0.2)12 = 2.4$ which is rounded up to the next whole integer, 3.
$(0.3)12 = 3.6$ which is rounded up to the next whole integer, 4.
$(0.5)12 = 6$.

Note that in this case, 3+4+6 does not equal n (which is 12). This can happen because of rounding.

**Group Sample Size Pattern**

A set of positive, numeric values, one for each row of distributions, is entered here. Each item specified in this list applies to the whole row of distributions. For example, suppose the entry is 1 2 1 and Grps 1 = 3, Grps 2 = 1, Grps 3 = 2. The sample size pattern used would be 1 1 2 1 1.

The sample size of group $i$ is found by multiplying the $i^{th}$ number from this list by the value of $n$ and rounding up to the next whole number. The number of values must match the number of groups, $g$. When too few numbers are entered, 1’s are added. When too many numbers are entered, the extras are ignored.

- **Equal**
  
  If all sample sizes are to be equal, enter Equal here and the desired sample size in $n$. A set of $g$ 1’s will be used. This will result in $n1 = n2 = \ldots = ng = n$. That is, all sample sizes are equal to $n$.

**Effect Size**

These options specify the distributions to be used in the two simulations, one set per row. The first option specifies the number of groups represented by the two distributions that follow. The second option specifies the distribution to be used in simulating the null hypothesis to determine the significance level (alpha). The third option specifies the distribution to be used in simulating the alternative hypothesis to determine the power.

**Grps [1 – 3] (Grps 4 – 9 are found on the Data 2 tab)**

This value specifies the number of groups specified by the H0 and H1 distribution statements to the right. Usually, you will enter ‘1’ to specify a single H0 and a single H1 distribution, or you
One-Way Analysis of Variance (Simulation)

will enter ‘0’ to indicate that the distributions specified on this line are to be ignored. This option lets you easily specify many identical distributions with a single phrase.

The total number of groups \( g \) is equal to the sum of the values for the three rows of distributions shown under the Data1 tab and the six rows of distributions shown under the Data2 tab.

Note that each item specified in the ‘Group Sample Size Pattern’ option applies to the whole row of entries here. For example, suppose the ‘Group Sample Size Pattern’ was ‘1 2 1’ and ‘Grps 1’ = 3, ‘Grps 2’ = 1, and ‘Grps 3’ = 2. The sample size pattern would be ‘1 1 1 2 1 1’.

**Group Distribution(s)**

This entry specifies the distribution of one or more groups under the null hypothesis, \( H_0 \). The magnitude of the differences of the means of these distributions, which is often summarized as the standard deviation of the means, represents the magnitude of the mean differences specified under \( H_0 \). Usually, the means are assumed to be equal under \( H_0 \), so their standard deviation should be zero except for rounding.

These distributions are used in the simulations that estimate the actual significance level. They also specify the value of the mean under the null hypothesis, \( H_0 \). Usually, these distributions will be identical. The parameters of each distribution are specified using numbers or letters. If letters are used, their values are specified in the boxes below. The value \( M_0 \) is reserved for the value of the mean under the null hypothesis.

Following is a list of the distributions that are available and the syntax used to specify them. Note that, except for the multinomial, the distributions are parameterized so that the mean is entered first.

- Beta\(=A(M_0,A,B,\text{Minimum})\)
- Binomial\(=B(M_0,N)\)
- Cauchy\(=C(M_0,\text{Scale})\)
- Constant\(=K(\text{Value})\)
- Exponential\(=E(M_0)\)
- F\(=F(M_0,\text{DF1})\)
- Gamma\(=G(M_0,A)\)
- Multinomial\(=M(P_1,P_2,\ldots,P_k)\)
- Normal\(=N(M_0,\text{SD})\)
- Poisson\(=P(M_0)\)
- Student's T\(=T(M_0,D)\)
- Tukey's Lambda\(=L(M_0,S,\text{Skewness},\text{Elongation})\)
- Uniform\(=U(M_0,\text{Minimum})\)
- Weibull\(=W(M_0,B)\)

Details of writing mixture distributions, combined distributions, and compound distributions are found in the chapter on Data Simulation and will not be repeated here.

**Finding the Value of the Mean of a Specified Distribution**

Except for the multinomial distribution, the distributions have been parameterized in terms of their means, since this is the parameter being tested. The mean of a distribution created as a linear combination of other distributions is found by applying the linear combination to the individual means. However, the mean of a distribution created by multiplying or dividing other distributions is not necessarily equal to applying the same function to the individual means. For example, the mean of \( 4N(4,5) + 2N(5,6) \) is \( 4*4 + 2*5 = 26 \), but the mean of \( 4N(4,5) * 2N(5,6) \) is not exactly \( 4*4*2*5 = 160 \) (although it is close).
Group Distribution(s)|H1

Specify the distribution of this group under the alternative hypothesis, H1. This distribution is used in the simulation that determines the power. A fundamental quantity in a power analysis is the amount of variation among the group means. In fact, classical power analysis formulas, this variation is summarized as the standard deviation of the means.

The important point to realize is that you must pay particular attention to the values you give to the means of these distributions because they are fundamental to the interpretation of the simulation.

For convenience in specifying a range of values, the parameters of the distribution can be specified using numbers or letters. If letters are used, their values are specified in the boxes below. The value $M1$ is reserved for the value of the mean under the alternative hypothesis.

Following is a list of the distributions that are available and the syntax used to specify them. Note that, except for the multinomial, the distributions are parameterized so that the mean, $M1$, is entered first.

```
Beta=A(M1,A,B,Minimum)
Binomial=B(M1,N)
Cauchy=C(M1,Scale)
Constant=K(Value)
Exponential=E(M1)
F=F(M1,DF1)
Gamma=G(M1,A)
Multinomial=M(P1,P2,..,Pk)
Normal=N(M1,SD)
Poisson=P(M1)
Student's T=T(M1,D)
Tukey's Lambda=L(M1,S,Skewness,Elongation)
Uniform=U(M1,Minimum)
Weibull=W(M1,B)
```

Details of writing mixture distributions, combined distributions, and compound distributions are found in the chapter on Data Simulation and will not be repeated here.

Effect Size – Distribution Parameters

M0 (Mean|H0)

These values are substituted for $M0$ in the distribution specifications given above. $M0$ is intended to be the value of the mean hypothesized by the null hypothesis, H0.

You can enter a list of values using the syntax 0 1 2 3 or 0 to 3 by 1.

M1 (Mean|H1)

These values are substituted for $M1$ in the distribution specifications given above. Although it can be used wherever you want, $M1$ is intended to be the value of the mean hypothesized by the alternative hypothesis, H1.

You can enter a list of values using the syntax 0 1 2 3 or 0 to 3 by 1.
Parameter Values (S, A, B, C)
Enter the numeric value(s) of the parameters listed above. These values are substituted for the corresponding letter in all four distribution specifications.

You can enter a list of values for each letter using the syntax 0 1 2 3 or 0 to 3 by 1.

You can also change the letter that is used as the name of this parameter using the pull-down menu to the side.

Test
Test Statistic
Specify which test is to be simulated. Although the F-test is the most commonly used test, it is based on assumptions that may not be viable in some situations. For your data, you may find that the Kruskal-Wallis test is more accurate (actual alpha = target alpha) and more precise (better power).

Simulations
Simulations
This option specifies the number of iterations, $M$, used in the simulation. The larger the number of iterations, the longer the running time, and, the more accurate the results.

The precision of the simulated power estimates are calculated from the binomial distribution. Thus, confidence intervals may be constructed for various power values. The following table gives an estimate of the precision that is achieved for various simulation sizes when the power is either 0.50 or 0.95. The table values are interpreted as follows: a 95% confidence interval of the true power is given by the power reported by the simulation plus and minus the ‘Precision’ amount given in the table.

<table>
<thead>
<tr>
<th>Simulation Size M</th>
<th>Precision when Power = 0.50</th>
<th>Precision when Power = 0.95</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.100</td>
<td>0.044</td>
</tr>
<tr>
<td>500</td>
<td>0.045</td>
<td>0.019</td>
</tr>
<tr>
<td>1000</td>
<td>0.032</td>
<td>0.014</td>
</tr>
<tr>
<td>2000</td>
<td>0.022</td>
<td>0.010</td>
</tr>
<tr>
<td>5000</td>
<td>0.014</td>
<td>0.006</td>
</tr>
<tr>
<td>10000</td>
<td>0.010</td>
<td>0.004</td>
</tr>
<tr>
<td>50000</td>
<td>0.004</td>
<td>0.002</td>
</tr>
<tr>
<td>100000</td>
<td>0.003</td>
<td>0.001</td>
</tr>
</tbody>
</table>

Notice that a simulation size of 1000 gives a precision of plus or minus 0.01 when the true power is 0.95. Also note that as the simulation size is increased beyond 5000, there is only a small amount of additional accuracy achieved.
Reports Tab
The Reports tab contains settings about the format of the output.

Select Output – Numeric Reports

Show Numeric Reports & Plots
These options let you specify whether you want to generate the standard reports and plots.

Show Inc's & 95% C.I.
Checking this option causes an additional line to be printed showing a 95% confidence interval for both the power and actual alpha and half the width of the confidence interval (the increment).

Select Output – Plots

Show Comparative Reports & Plots
These options let you specify whether you want to generate reports and plots that compare the test statistics that are available.

Iterations Tab
The Options tab contains limits on the number of iterations and various options about individual tests.

Maximum Iterations

Maximum Iterations Before Search Termination
Specify the maximum number of iterations before the search for the sample size is aborted. When the maximum number of iterations is reached without convergence, the sample size is left blank. We recommend a value of at least 500.

Random Numbers

Random Number Pool Size
This is the size of the pool of values from which the random samples will be drawn. Pools should be at least the maximum of 10,000 and twice the number of simulations. You can enter Automatic and an appropriate value will be calculated.

If you do not want to draw numbers from a pool, enter 0 here.
Example 1 – Power at Various Sample Sizes

For this first example we repeat Example 1 of the regular One-Way ANOVA procedure. This will allow you to compare the values obtained by simulation with the actual values obtained from the theoretical results.

An experiment is being designed to compare the means of four groups using an $F$ test with a significance level of 0.05. Previous studies have shown that the standard deviation within a group is 18. Treatment means of 40, 10, 10, and 10 represent clinically important treatment differences. To better understand the relationship between power and sample size, the researcher wants to compute the power for group sample sizes of 4, 8, and 12. The group sample sizes are equal.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the One-Way Analysis of Variance (Simulation) procedure window by expanding Means, then ANOVA, then clicking on Analysis of Variance, and then clicking on One-Way Analysis of Variance (Simulation). You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>n (Sample Size Multiplier)</td>
<td>4 8 12</td>
</tr>
<tr>
<td>Group Sample Size Pattern</td>
<td>Equal</td>
</tr>
<tr>
<td>Grps 1</td>
<td>1</td>
</tr>
<tr>
<td>Group 1 Distribution(s)</td>
<td>H0</td>
</tr>
<tr>
<td>Group 1 Distribution(s)</td>
<td>H1</td>
</tr>
<tr>
<td>Grps 2</td>
<td>3</td>
</tr>
<tr>
<td>Group 2 Distribution(s)</td>
<td>H0</td>
</tr>
<tr>
<td>Group 2 Distribution(s)</td>
<td>H1</td>
</tr>
<tr>
<td>Grps 3</td>
<td>0</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H1)</td>
</tr>
<tr>
<td>S</td>
<td>18</td>
</tr>
<tr>
<td>Test Type</td>
<td>F-Test</td>
</tr>
<tr>
<td>Simulations</td>
<td>2000</td>
</tr>
</tbody>
</table>

| Iterations Tab              |                            |
| Random Number Pool Size     | Automatic                  |
Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results Report

<table>
<thead>
<tr>
<th>Power</th>
<th>Average Group Size</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>S.D. of Means Sm</th>
<th>H1</th>
<th>S.D. of Data SD</th>
<th>H1</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.536</td>
<td>4.0</td>
<td>0.050</td>
<td>0.047</td>
<td>0.465</td>
<td>13.0</td>
<td>18.0</td>
<td>10.0</td>
<td>40.0</td>
<td>18.0</td>
<td></td>
</tr>
<tr>
<td>(0.022)</td>
<td>[0.514 0.557]</td>
<td>(0.009)</td>
<td>[0.038 0.056]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.896</td>
<td>8.0</td>
<td>0.050</td>
<td>0.053</td>
<td>0.105</td>
<td>13.0</td>
<td>18.1</td>
<td>10.0</td>
<td>40.0</td>
<td>18.0</td>
<td></td>
</tr>
<tr>
<td>(0.013)</td>
<td>[0.882 0.909]</td>
<td>(0.010)</td>
<td>[0.043 0.063]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>0.982</td>
<td>12.0</td>
<td>0.050</td>
<td>0.051</td>
<td>0.019</td>
<td>13.0</td>
<td>18.1</td>
<td>10.0</td>
<td>40.0</td>
<td>18.0</td>
<td></td>
</tr>
<tr>
<td>(0.006)</td>
<td>[0.976 0.987]</td>
<td>(0.010)</td>
<td>[0.041 0.061]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:
- **Pool Size**: 10000. **Simulations**: 2000. **Run Time**: 12.84 seconds.
- **H0 Distributions**: Normal(M0 S); Normal(M0 S); Normal(M0 S); and Normal(M0 S)
- **H1 Distributions**: Normal(M1 S); Normal(M0 S); Normal(M0 S); and Normal(M0 S)

Report Definitions
- H0 represents the null hypothesis.
- H1 represents the alternative hypothesis.
- Power is the probability of rejecting a false null hypothesis. It should be close to one.
- 'n' is the average of the group sample sizes.
- Total N is the total sample size of all groups combined.
- Target Alpha is the desired probability of rejecting a true null hypothesis.
- Actual Alpha is the alpha achieved by this simulation.
- Beta is the probability of accepting a false null hypothesis.
- Sm|H1 is the standard deviation of the group means under the alternative hypothesis.
- SD|H1 is the within group standard deviation under the alternative hypothesis.
- Second Row: (Power Prec.) [95% LCL and UCL Power] (Alpha Prec.) [95% LCL and UCL Alpha]

Summary Statements
A one-way design with 4 groups has sample sizes of 4, 4, 4, and 4. The null hypothesis is that the standard deviation of the group means is 0.1 and the alternative standard deviation of the group means is 13.0. The total sample of 16 subjects achieves a power of 0.536 using the F-Test with a target significance level of 0.050 and an actual significance level of 0.047. The average within group standard deviation assuming the alternative distribution is 18.0. These results are based on 2000 Monte Carlo samples from the null distributions: Normal(M0 S); Normal(M0 S); Normal(M0 S); and Normal(M0 S) and the alternative distributions: Normal(M1 S); Normal(M0 S); Normal(M0 S); and Normal(M0 S). Other parameters used in the simulation were: M0 = 10.0, M1 = 40.0, and S = 18.0.

This report shows the estimated power for each scenario. The first row shows the parameter settings and the estimated power and significance level (Actual Alpha).

The second row shows two 95% confidence intervals in brackets: the first for the power and the second for the significance level. Half the width of each confidence interval is given in parentheses as a fundamental measure of the accuracy of the simulation. As the number of simulations is increased, the width of the confidence intervals will decrease.

**Power**

This is the probability of rejecting a false null hypothesis. This value is estimated by the simulation using the H1 distributions.
Note that a precision value (half the width of its confidence interval) and a confidence interval are shown on the line below this row. These values are provided to help you understand the precision of the estimated power.

**Average Group Size n**
This is the average of the group sample sizes.

**Total Sample Size N**
This is the total sample size of the study.

**Target Alpha**
The target value of alpha: the probability of rejecting a true null hypothesis. This is often called the significance level.

**Actual Alpha**
This is the value of alpha estimated by the simulation using the H0 distributions. It should be compared with the Target Alpha to determine if the test statistic is accurate in this scenario.

Note that a precision value (half the width of its confidence interval) and a confidence interval are shown on the line below this row. These values are provided to help you understand the precision of the Actual Alpha.

**Beta**
Beta is the probability of accepting a false null hypothesis. This is the value of beta estimated by the simulation using the H1 distributions.

**S.D. of Means Sm|H1**
This is the standard deviation of the hypothesized means of the alternative distributions. Under the null hypothesis, this value is zero. So this value represents the magnitude of the difference among the means that is being tested. It is roughly equal to the average difference between the group means and the overall mean.

Note that the effect size is the ratio of Sm|H1 and SD|H1.

**S.D. of Data SD|H1**
This is the within-group standard deviation calculated from samples from the alternative distributions.

**M0**
This is the value entered for M0, the group means under H0.

**M1**
This is the value entered for M1, the group means under H1.

**S**
This is the value entered for S, the standard deviation.
### Detailed Results Report

#### Details when Target Alpha = 0.050, M0 = 10.0, M1 = 40.0, S = 18.0

**Test Statistic: F-Test**

<table>
<thead>
<tr>
<th>Group</th>
<th>Ni</th>
<th>Ni of N</th>
<th>H0 Mean</th>
<th>H1 Mean</th>
<th>H0 S.D.</th>
<th>H1 S.D.</th>
<th>H0 Sm</th>
<th>H1 Sm</th>
<th>Alpha</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>16</td>
<td>100.00</td>
<td>10.0</td>
<td>17.5</td>
<td>18.1</td>
<td>18.0</td>
<td>0.0</td>
<td>13.0</td>
<td>0.051</td>
<td>0.536</td>
</tr>
<tr>
<td>1</td>
<td>4</td>
<td>25.00</td>
<td>10.0</td>
<td>40.0</td>
<td>17.7</td>
<td>17.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>25.00</td>
<td>10.0</td>
<td>10.0</td>
<td>18.4</td>
<td>18.2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>25.00</td>
<td>10.0</td>
<td>10.0</td>
<td>18.2</td>
<td>17.8</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>4</td>
<td>25.00</td>
<td>10.0</td>
<td>10.0</td>
<td>18.0</td>
<td>18.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Details when Target Alpha = 0.050, M0 = 10.0, M1 = 40.0, S = 18.0

**Test Statistic: F-Test**

<table>
<thead>
<tr>
<th>Group</th>
<th>Ni</th>
<th>Ni of N</th>
<th>H0 Mean</th>
<th>H1 Mean</th>
<th>H0 S.D.</th>
<th>H1 S.D.</th>
<th>H0 Sm</th>
<th>H1 Sm</th>
<th>Alpha</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>32</td>
<td>100.00</td>
<td>10.0</td>
<td>17.5</td>
<td>18.1</td>
<td>18.0</td>
<td>0.0</td>
<td>13.0</td>
<td>0.051</td>
<td>0.896</td>
</tr>
<tr>
<td>1</td>
<td>8</td>
<td>25.00</td>
<td>10.0</td>
<td>40.0</td>
<td>17.9</td>
<td>18.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>8</td>
<td>25.00</td>
<td>10.0</td>
<td>10.0</td>
<td>18.1</td>
<td>18.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>8</td>
<td>25.00</td>
<td>10.0</td>
<td>10.0</td>
<td>18.1</td>
<td>18.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>8</td>
<td>25.00</td>
<td>10.0</td>
<td>10.0</td>
<td>18.0</td>
<td>18.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Details when Target Alpha = 0.050, M0 = 10.0, M1 = 40.0, S = 18.0

**Test Statistic: F-Test**

<table>
<thead>
<tr>
<th>Group</th>
<th>Ni</th>
<th>Ni of N</th>
<th>H0 Mean</th>
<th>H1 Mean</th>
<th>H0 S.D.</th>
<th>H1 S.D.</th>
<th>H0 Sm</th>
<th>H1 Sm</th>
<th>Alpha</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>48</td>
<td>100.00</td>
<td>10.0</td>
<td>17.5</td>
<td>18.1</td>
<td>18.0</td>
<td>0.0</td>
<td>13.0</td>
<td>0.051</td>
<td>0.982</td>
</tr>
<tr>
<td>1</td>
<td>12</td>
<td>25.00</td>
<td>10.0</td>
<td>40.0</td>
<td>18.3</td>
<td>18.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>12</td>
<td>25.00</td>
<td>10.0</td>
<td>10.0</td>
<td>18.1</td>
<td>18.4</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>12</td>
<td>25.00</td>
<td>10.0</td>
<td>10.0</td>
<td>18.2</td>
<td>18.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>12</td>
<td>25.00</td>
<td>10.0</td>
<td>10.0</td>
<td>17.9</td>
<td>17.9</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

This report shows the details of each row of the previous report.

**Group**

This is the number of the group shown on this line. The first line, labeled *All*, gives the average or the total as appropriate.

**Ni**

This is the sample size of each group. This column is especially useful when the sample sizes are unequal.

**Percent Ni of N**

This is the percentage of the total sample that is allocated to each group.

**H0 and H1 Means**

These are the means that were used in the simulations for H0 and H1, respectively.

**H0 and H1 S.D.’s**

These are the standard deviations that were obtained by the simulations for H0 and H1, respectively. Note that they often are not exactly equal to what was specified because of simulation error.

**H0 and H1 Sm’s**

These are the standard deviations of the means that were obtained by the simulations for H0 and H1, respectively. Under H0, the value of Sm should be near zero. It lets you determine if your simulation of H0 was correctly specified.
Plots Section

![Graph showing power vs n](image)

This plot gives a visual presentation to the results in the Numeric Report. We can quickly see the impact on the power of increasing the sample size.

Example 2 – Comparative Results

Continuing with Example 1, the researchers want to study the characteristics of alternative test statistics. They want to compare the results of the F-test and the Kruskal-Wallis test.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **One-Way Analysis of Variance (Simulation)** procedure window by expanding **Means**, then **ANOVA**, then clicking on **Analysis of Variance**, and then clicking on **One-Way Analysis of Variance (Simulation)**. You may then make the appropriate entries as listed below, or open **Example 2** by going to the **File** menu and choosing **Open Example Template**.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>n (Sample Size Multiplier)</td>
<td>4 8 12</td>
</tr>
<tr>
<td>Group Sample Size Pattern</td>
<td>Equal</td>
</tr>
<tr>
<td>Grps 1</td>
<td>1</td>
</tr>
<tr>
<td>Group 1 Distribution(s)</td>
<td>H0</td>
</tr>
<tr>
<td>Group 1 Distribution(s)</td>
<td>H1</td>
</tr>
</tbody>
</table>
Data Tab (continued)
Grps 2 ...................................................... 3  
Group 2 Distribution(s) | H0 ..................... N(M0 S)  
Group 2 Distribution(s) | H1 ..................... N(M0 S)  
Grps 3 ...................................................... 0  
M0 (Mean|H0) ............................................ 10  
M1 (Mean|H1) ............................................ 40  
S .............................................................. 18  
Test Type ................................................ F-Test  
Simulations .............................................. 2000  

Reports Tab  
Show Comparative Reports .................... Checked  
Show Comparative Plots ......................... Checked  

Output  
Click the Run button to perform the calculations and generate the following output.

Numeric Results and Plots

<table>
<thead>
<tr>
<th>Power</th>
<th>n</th>
<th>N</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>Beta</th>
<th>S.D. of Means</th>
<th>S.D. of SD</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.510</td>
<td>4.0</td>
<td>16</td>
<td>0.050</td>
<td>0.051</td>
<td>0.490</td>
<td>13.0</td>
<td>17.8</td>
<td>10.0</td>
<td>40.0</td>
<td>18.0</td>
</tr>
<tr>
<td>0.901</td>
<td>8.0</td>
<td>32</td>
<td>0.050</td>
<td>0.047</td>
<td>0.100</td>
<td>13.0</td>
<td>18.0</td>
<td>10.0</td>
<td>40.0</td>
<td>18.0</td>
</tr>
<tr>
<td>0.986</td>
<td>12.0</td>
<td>48</td>
<td>0.050</td>
<td>0.046</td>
<td>0.015</td>
<td>13.0</td>
<td>18.1</td>
<td>10.0</td>
<td>40.0</td>
<td>18.0</td>
</tr>
</tbody>
</table>

Notes:  
H0 Distributions: Normal(M0 S); Normal(M0 S); Normal(M0 S); and Normal(M0 S)  
H1 Distributions: Normal(M1 S); Normal(M0 S); Normal(M0 S); and Normal(M0 S)

Power Comparison for Testing the g = 4 Group Means

<table>
<thead>
<tr>
<th>Total Sample Size</th>
<th>S.D. of Means Sm</th>
<th>H1</th>
<th>S.D. of SD</th>
<th>H1</th>
<th>Target Alpha</th>
<th>F-Test Power</th>
<th>Kruskal Wallis Power</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>13.0</td>
<td>17.8</td>
<td>0.050</td>
<td>0.510</td>
<td>0.366</td>
<td>10.0</td>
<td>40.0</td>
<td>18.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>13.0</td>
<td>18.0</td>
<td>0.050</td>
<td>0.901</td>
<td>0.860</td>
<td>10.0</td>
<td>40.0</td>
<td>18.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>13.0</td>
<td>18.1</td>
<td>0.050</td>
<td>0.986</td>
<td>0.979</td>
<td>10.0</td>
<td>40.0</td>
<td>18.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Alpha Comparison for Testing the g = 4 Group Means

<table>
<thead>
<tr>
<th>Total Sample Size</th>
<th>S.D. of Means Sm</th>
<th>H1</th>
<th>S.D. of SD</th>
<th>H1</th>
<th>Target Alpha</th>
<th>F-Test Alpha</th>
<th>Kruskal Wallis Alpha</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>16</td>
<td>13.0</td>
<td>16.5</td>
<td>0.050</td>
<td>0.050</td>
<td>0.026</td>
<td>10.0</td>
<td>40.0</td>
<td>18.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>13.0</td>
<td>17.8</td>
<td>0.050</td>
<td>0.051</td>
<td>0.041</td>
<td>10.0</td>
<td>40.0</td>
<td>18.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>32</td>
<td>13.0</td>
<td>18.0</td>
<td>0.050</td>
<td>0.047</td>
<td>0.042</td>
<td>10.0</td>
<td>40.0</td>
<td>18.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>48</td>
<td>13.0</td>
<td>18.1</td>
<td>0.050</td>
<td>0.046</td>
<td>0.039</td>
<td>10.0</td>
<td>40.0</td>
<td>18.0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
We notice that the power of the F-test is much greater than the Kruskal-Wallis test for \( n = 4 \). However, when \( n = 12 \), the powers of the two tests are almost equal. Note that the alpha value of the Kruskal-Wallis test is almost half that of the F-test for \( n = 4 \). This is probably why the power is also low.

**Example 3 – Validation using Fleiss**

Fleiss (1986) page 374 presents an example of determining an appropriate sample size when using an F-test in an experiment with 4 groups; means of 9.775, 12, 12, and 14.225; standard deviation of 3; alpha of 0.05, and beta of 0.20. He finds a sample size of 11 per group.

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the One-Way Analysis of Variance (Simulation) procedure window by expanding Means, then ANOVA, then clicking on Analysis of Variance, and then clicking on One-Way Analysis of Variance (Simulation). You may then make the appropriate entries as listed below, or open Example 3 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Data Tab Data Tab</td>
<td></td>
</tr>
<tr>
<td>Data Tab Find (Solve For)</td>
<td>( n )</td>
</tr>
<tr>
<td>Data Tab Power</td>
<td>0.80</td>
</tr>
<tr>
<td>Data Tab Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>Data Tab n (Sample Size Multiplier)</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Data Tab Group Sample Size Pattern</td>
<td>Equal</td>
</tr>
<tr>
<td>Data Tab Grps 1</td>
<td>1</td>
</tr>
<tr>
<td>Data Tab Group 1 Distribution(s)</td>
<td>H0</td>
</tr>
<tr>
<td>Data Tab Group 1 Distribution(s)</td>
<td>H1</td>
</tr>
</tbody>
</table>
Data Tab (continued)
Grps 2...................................................... 2
Group 2 Distribution(s) | H0 .................... N(M0 S)
Group 2 Distribution(s) | H1 .................... N(M0 S)
Grps 3...................................................... 1
Group 3 Distribution(s) | H0 .................... N(M0 S)
Group 3 Distribution(s) | H1 .................... N(14.225 S)
M0 (Mean|H0) ........................................ 12
M1 (Mean|H1) ........................................ 0
S.............................................................. 3
Test Type ................................................ F-Test
Simulations.............................................. 2000

Output
Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Average Group Size</th>
<th>Total Sample Size</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>Beta</th>
<th>S.D. of Means Sm</th>
<th>H1</th>
<th>S.D. of Data SD</th>
<th>H1</th>
<th>M0</th>
<th>S</th>
</tr>
</thead>
<tbody>
<tr>
<td>Power</td>
<td>n</td>
<td>N</td>
<td>0.050</td>
<td>0.052</td>
<td>0.179</td>
<td>1.6</td>
<td>3.0</td>
<td>12.0</td>
<td>3.0</td>
<td></td>
</tr>
<tr>
<td>0.822 (0.017)</td>
<td>11.0</td>
<td>44</td>
<td>[0.805 0.838]</td>
<td>[0.042</td>
<td>[0.061]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:
H0 Distributions: Normal(M0 S); Normal(M0 S); Normal(M0 S); and Normal(M0 S)
H1 Distributions: Normal(9.775 S); Normal(M0 S); Normal(M0 S); and Normal(14.225 S)

Details when Target Alpha = 0.050, M0 = 12.0, S = 3.0

<table>
<thead>
<tr>
<th>Group</th>
<th>Ni</th>
<th>Percent of N</th>
<th>H0 Mean</th>
<th>H1 Mean</th>
<th>H0 S.D.</th>
<th>H1 S.D.</th>
<th>H0 Sm</th>
<th>H1 Sm</th>
<th>Actual Alpha</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>44</td>
<td>100.00</td>
<td>12.0</td>
<td>12.0</td>
<td>3.0</td>
<td>3.0</td>
<td>0.0</td>
<td>1.6</td>
<td>0.052</td>
<td>0.822</td>
</tr>
<tr>
<td>1</td>
<td>11</td>
<td>25.00</td>
<td>12.0</td>
<td>9.8</td>
<td>3.0</td>
<td>3.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>11</td>
<td>25.00</td>
<td>12.0</td>
<td>12.0</td>
<td>3.0</td>
<td>3.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>11</td>
<td>25.00</td>
<td>12.0</td>
<td>12.0</td>
<td>3.0</td>
<td>3.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>11</td>
<td>25.00</td>
<td>12.0</td>
<td>14.2</td>
<td>3.0</td>
<td>3.0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Note that PASS has also found the group sample size to be 11.
Example 4 – Selecting a Test Statistic when the Data Contain Outliers

The F-test is known to be robust to the violation of some assumptions, but it is susceptible to inaccuracy when the data contain outliers. This example will investigate the impact of outliers on the power and precision of the F-test and the Kruskal-Wallis test.

A mixture of two normal distributions will be used to randomly generate outliers. The mixture will draw 95% of the data from a normal distribution with mean zero and variance one. The other 5% of the data will come from a normal distribution with mean zero and variance that ranges from one to ten. In the alternative distributions, two will have a mean of zero and one will have a mean of one.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the One-Way Analysis of Variance (Simulation) procedure window by expanding Means, then ANOVA, then clicking on Analysis of Variance, and then clicking on One-Way Analysis of Variance (Simulation). You may then make the appropriate entries as listed below, or open Example 4 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>n (Sample Size Multiplier)</td>
<td>10 20</td>
</tr>
<tr>
<td>Group Sample Size Pattern</td>
<td>Equal</td>
</tr>
<tr>
<td>Grps 1</td>
<td>2</td>
</tr>
<tr>
<td>Group 1 Distribution(s)</td>
<td>H0</td>
</tr>
<tr>
<td>Group 1 Distribution(s)</td>
<td>H1</td>
</tr>
<tr>
<td>Grps 2</td>
<td>1</td>
</tr>
<tr>
<td>Group 2 Distribution(s)</td>
<td>H0</td>
</tr>
<tr>
<td>Group 2 Distribution(s)</td>
<td>H1</td>
</tr>
<tr>
<td>Grps 3</td>
<td>0</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H1)</td>
</tr>
<tr>
<td>S</td>
<td>1</td>
</tr>
<tr>
<td>A</td>
<td>1 5 10</td>
</tr>
<tr>
<td>Test Type</td>
<td>F-Test</td>
</tr>
<tr>
<td>Simulations</td>
<td>2000</td>
</tr>
</tbody>
</table>

Reports Tab

Show Comparative Reports        Checked
Show Comparative Plots          Checked
Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Power</th>
<th>n</th>
<th>N</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>Beta</th>
<th>S.D. of Means Sm</th>
<th>H1</th>
<th>S.D. of Means SD</th>
<th>H1</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.588</td>
<td>10.0</td>
<td>30</td>
<td>0.050</td>
<td>0.054</td>
<td>0.413</td>
<td>0.5</td>
<td>1.0</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>0.381</td>
<td>10.0</td>
<td>30</td>
<td>0.050</td>
<td>0.035</td>
<td>0.619</td>
<td>0.5</td>
<td>1.4</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>5.0</td>
<td>1.0</td>
</tr>
<tr>
<td>0.321</td>
<td>10.0</td>
<td>30</td>
<td>0.050</td>
<td>0.026</td>
<td>0.680</td>
<td>0.5</td>
<td>2.4</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>10.0</td>
<td>1.0</td>
</tr>
<tr>
<td>0.897</td>
<td>20.0</td>
<td>60</td>
<td>0.050</td>
<td>0.049</td>
<td>0.104</td>
<td>0.5</td>
<td>1.0</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>0.618</td>
<td>20.0</td>
<td>60</td>
<td>0.050</td>
<td>0.047</td>
<td>0.383</td>
<td>0.5</td>
<td>1.5</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>5.0</td>
<td>1.0</td>
</tr>
<tr>
<td>0.408</td>
<td>20.0</td>
<td>60</td>
<td>0.050</td>
<td>0.026</td>
<td>0.592</td>
<td>0.5</td>
<td>2.4</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>10.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Notes:
H0 Distributions: Normal(M0 S)[95]; Normal(M0 A)[5]; Normal(M0 S)[95]; Normal(M0 A)[5]; and Normal(M0 S)[95]; Normal(M0 A)[5]
H1 Distributions: Normal(M1 S)[95]; Normal(M1 A)[5]; Normal(M0 S)[95]; Normal(M0 A)[5]; and Normal(M0 S)[95]; Normal(M0 A)[5]

Power Comparison for Testing the g = 3 Group Means

<table>
<thead>
<tr>
<th>Total Sample Size</th>
<th>S.D. of Means Sm</th>
<th>H1</th>
<th>S.D. of Data SD</th>
<th>H1</th>
<th>Target Alpha</th>
<th>F-Test Power</th>
<th>Kruskal Wallis Power</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.5</td>
<td>1.0</td>
<td>0.050</td>
<td>0.588</td>
<td>0.544</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>30</td>
<td>0.5</td>
<td>1.4</td>
<td>0.050</td>
<td>0.381</td>
<td>0.458</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>5.0</td>
<td>1.0</td>
</tr>
<tr>
<td>60</td>
<td>0.5</td>
<td>2.4</td>
<td>0.050</td>
<td>0.321</td>
<td>0.473</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>10.0</td>
<td>1.0</td>
</tr>
<tr>
<td>60</td>
<td>0.5</td>
<td>1.5</td>
<td>0.050</td>
<td>0.897</td>
<td>0.880</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>5.0</td>
</tr>
<tr>
<td>60</td>
<td>0.5</td>
<td>2.4</td>
<td>0.050</td>
<td>0.618</td>
<td>0.813</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>10.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Alpha Comparison for Testing the g = 3 Group Means

<table>
<thead>
<tr>
<th>Total Sample Size</th>
<th>S.D. of Means Sm</th>
<th>H1</th>
<th>S.D. of Data SD</th>
<th>H1</th>
<th>Target Alpha</th>
<th>F-Test Power</th>
<th>Kruskal Wallis Power</th>
<th>M0</th>
<th>M1</th>
<th>S</th>
<th>A</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.5</td>
<td>1.0</td>
<td>0.050</td>
<td>0.054</td>
<td>0.055</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>30</td>
<td>0.5</td>
<td>1.4</td>
<td>0.050</td>
<td>0.035</td>
<td>0.041</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>5.0</td>
<td>1.0</td>
</tr>
<tr>
<td>60</td>
<td>0.5</td>
<td>2.4</td>
<td>0.050</td>
<td>0.026</td>
<td>0.041</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>10.0</td>
<td>1.0</td>
</tr>
<tr>
<td>60</td>
<td>0.5</td>
<td>1.5</td>
<td>0.050</td>
<td>0.049</td>
<td>0.050</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>5.0</td>
<td>1.0</td>
</tr>
<tr>
<td>60</td>
<td>0.5</td>
<td>2.4</td>
<td>0.050</td>
<td>0.047</td>
<td>0.054</td>
<td>0.0</td>
<td>1.0</td>
<td>1.0</td>
<td>1.0</td>
<td>10.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>
We note that when the variances are equal (A = 1), the F-Test is slightly better than the Kruskal-Wallis test. However, as the number of outliers is increased, the F-test does increasingly worse both in terms of power and significance, but the Kruskal-Wallis test is considerably less affected.

**Example 5 – Selecting a Test Statistic when the Data are Skewed**

The F-test is known to be robust to the violation of some assumptions, but it is susceptible to inaccuracy when the underlying distributions are skewed. This example will investigate the impact of skewness on the power and precision of the F-test and the Kruskal-Wallis test.

Tukey’s lambda distribution will be used because it allows the amount of skewness to be gradually increased.

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the One-Way Analysis of Variance (Simulation) procedure window by expanding Means, then ANOVA, then clicking on Analysis of Variance, and then clicking on One-Way Analysis of Variance (Simulation). You may then make the appropriate entries as listed below, or open Example 5 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Tab</strong></td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>n (Sample Size Multiplier)</td>
<td>10 20</td>
</tr>
<tr>
<td>Group Sample Size Pattern</td>
<td>Equal</td>
</tr>
<tr>
<td>Grps 1</td>
<td>2</td>
</tr>
<tr>
<td>Group 1 Distribution(s)</td>
<td>H0</td>
</tr>
<tr>
<td>Group 1 Distribution(s)</td>
<td>H1</td>
</tr>
<tr>
<td>Grps 2</td>
<td>1</td>
</tr>
<tr>
<td>Group 2 Distribution(s)</td>
<td>H0</td>
</tr>
<tr>
<td>Group 2 Distribution(s)</td>
<td>H1</td>
</tr>
<tr>
<td>Grps 3</td>
<td>0</td>
</tr>
<tr>
<td>M0 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>M1 (Mean</td>
<td>H0)</td>
</tr>
<tr>
<td>S</td>
<td>0.05 0.9</td>
</tr>
<tr>
<td>Test Type</td>
<td>F-Test</td>
</tr>
<tr>
<td>Simulations</td>
<td>2000</td>
</tr>
<tr>
<td><strong>Reports Tab</strong></td>
<td></td>
</tr>
<tr>
<td>Show Comparative Reports</td>
<td>Checked</td>
</tr>
<tr>
<td>Show Comparative Plots</td>
<td>Checked</td>
</tr>
</tbody>
</table>
Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results and Plots

Numeric Results for Testing the g = 3 Group Means
Test Statistic: F-Test

<table>
<thead>
<tr>
<th>Average Group Size</th>
<th>Total Sample Size</th>
<th>Target Alpha</th>
<th>Actual Alpha</th>
<th>S.D. of Means</th>
<th>S.D. of Data</th>
<th>Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>n</td>
<td>N</td>
<td></td>
<td></td>
<td>Sm[H1]</td>
<td>SD[H1]</td>
<td></td>
</tr>
<tr>
<td>0.596</td>
<td>10.0</td>
<td>30</td>
<td>0.050</td>
<td>0.405</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.615</td>
<td>10.0</td>
<td>30</td>
<td>0.050</td>
<td>0.386</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.719</td>
<td>10.0</td>
<td>30</td>
<td>0.050</td>
<td>0.281</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.899</td>
<td>20.0</td>
<td>60</td>
<td>0.050</td>
<td>0.010</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.902</td>
<td>20.0</td>
<td>60</td>
<td>0.050</td>
<td>0.008</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.895</td>
<td>20.0</td>
<td>60</td>
<td>0.050</td>
<td>0.105</td>
<td>1.0</td>
<td>0.0</td>
</tr>
</tbody>
</table>

Notes:
H0 Distributions: Tukey(M0 S G 0); Tukey(M0 S G 0); and Tukey(M0 S G 0)
H1 Distributions: Tukey(M0 S G 0); Tukey(M0 S G 0); and Tukey(M1 S G 0)

Power Comparison for Testing the g = 3 Group Means

<table>
<thead>
<tr>
<th>Total Sample Size</th>
<th>S.D. of Means Sm[H1]</th>
<th>S.D. of Data SD[H1]</th>
<th>Target Alpha</th>
<th>F-Test Power</th>
<th>Kruskal Wallis Power</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.5</td>
<td>1.0</td>
<td>0.050</td>
<td>0.596</td>
<td>0.554</td>
</tr>
<tr>
<td>30</td>
<td>0.5</td>
<td>1.0</td>
<td>0.050</td>
<td>0.615</td>
<td>0.735</td>
</tr>
<tr>
<td>30</td>
<td>0.5</td>
<td>1.0</td>
<td>0.050</td>
<td>0.719</td>
<td>0.932</td>
</tr>
<tr>
<td>60</td>
<td>0.5</td>
<td>1.0</td>
<td>0.050</td>
<td>0.899</td>
<td>0.871</td>
</tr>
<tr>
<td>60</td>
<td>0.5</td>
<td>1.0</td>
<td>0.050</td>
<td>0.902</td>
<td>0.978</td>
</tr>
<tr>
<td>60</td>
<td>0.5</td>
<td>1.0</td>
<td>0.050</td>
<td>0.895</td>
<td>1.000</td>
</tr>
</tbody>
</table>

Alpha Comparison for Testing the g = 3 Group Means

<table>
<thead>
<tr>
<th>Total Sample Size</th>
<th>S.D. of Means Sm[H1]</th>
<th>S.D. of Data SD[H1]</th>
<th>Target Alpha</th>
<th>F-Test Alpha</th>
<th>Kruskal Wallis Alpha</th>
</tr>
</thead>
<tbody>
<tr>
<td>30</td>
<td>0.5</td>
<td>1.0</td>
<td>0.050</td>
<td>0.050</td>
<td>0.045</td>
</tr>
<tr>
<td>30</td>
<td>0.5</td>
<td>1.0</td>
<td>0.050</td>
<td>0.034</td>
<td>0.039</td>
</tr>
<tr>
<td>30</td>
<td>0.5</td>
<td>1.0</td>
<td>0.050</td>
<td>0.036</td>
<td>0.047</td>
</tr>
<tr>
<td>60</td>
<td>0.5</td>
<td>1.0</td>
<td>0.050</td>
<td>0.046</td>
<td>0.041</td>
</tr>
<tr>
<td>60</td>
<td>0.5</td>
<td>1.0</td>
<td>0.050</td>
<td>0.050</td>
<td>0.046</td>
</tr>
<tr>
<td>60</td>
<td>0.5</td>
<td>1.0</td>
<td>0.050</td>
<td>0.034</td>
<td>0.055</td>
</tr>
</tbody>
</table>

Power vs G by Test
M0=0.0 M1=1.0 S=1.0 Alpha=0.0 n=20

Alpha vs G by Test
M0=0.0 M1=1.0 S=1.0 Alpha=0.0 n=20
We note that as the skewness increases, the power of the Kruskal-Wallis test increases substantially as compared to the F-test.
Chapter 560

Fixed Effects Analysis of Variance

Introduction

A common task in research is to compare the average response across levels of one or more factor variables. Examples of factor variables are income level of two regions, nitrogen content of three lakes, or drug dosage. The fixed-effects analysis of variance compares the means of two or more factors. $F$ tests are used to determine statistical significance of the factors and their interactions. The tests are nondirectional in that the null hypothesis specifies that all means are equal and the alternative hypothesis simply states that at least one mean is different. This PASS module performs power analysis and sample size estimation for an analysis of variance design with up to three factors.

In the following example, the responses of a weight loss experiment are arranged in a two-factor, fixed-effect, design. The first factor is diet (D1 and D2) and the second factor is dose level of a dietary drug (low, medium, and high). The twelve individuals available for this study were assigned at random to one of the six treatment groups (cells) so that there were two per group. The response variable was an individual’s weight loss after four months.

<table>
<thead>
<tr>
<th>Table of Individual Weight Losses</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dietary Drug Dose Level</td>
</tr>
<tr>
<td>Diet</td>
</tr>
<tr>
<td>D1</td>
</tr>
<tr>
<td>D2</td>
</tr>
</tbody>
</table>

Important features to note are that each table entry represents a different individual and that the response variable (weight loss) is continuous, while the factors (Diet and Dose) are discrete.
Means can be calculated for each cell of the table. These means are shown in the table below. Note that we have added an additional row and column for the row, column, and overall means. The six means in the interior of this table are called the cell means.

<table>
<thead>
<tr>
<th>Table of Means</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dietary Drug Dose Level</strong></td>
</tr>
<tr>
<td><strong>Diet</strong></td>
</tr>
<tr>
<td>D1</td>
</tr>
<tr>
<td>D2</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

The Linear Model

A mathematical model may be formulated that underlies this experimental design. This model expresses each cell mean, $\mu_{ij}$, as the sum of parameters called effects. A common linear model for a two-factor experiment is

$$
\mu_{ij} = m + a_i + b_j + (ab)_{ij}
$$

where $i = 1, 2, ..., I$ and $j = 1, 2, ..., J$. This model expresses the value of a cell mean as the sum of four components:

- $m$ the grand mean.
- $a_i$ the effect of the $i^{th}$ level of factor $A$. Note that $\sum a_i = 0$.
- $b_j$ the effect of the $j^{th}$ level of factor $B$. Note that $\sum b_j = 0$.
- $(ab)_{ij}$ the combined effect of the $i^{th}$ level of factor $A$ and the $j^{th}$ level of factor $B$. Note that $\sum (ab)_{ij} = 0$.

Another way of stating this model for the two factor case is

*Cell Mean = Overall Effect + Row Effect + Column Effect + Interaction Effect.*

Since this model is the sum of various constants, it is called a linear model.

Calculating the Effects

We will now calculate the effects for our example. We will let Drug Dose correspond to factor $A$ and Diet correspond to factor $B$. 
**Step 1 – Remove the Grand Mean**

Remove the grand mean from the table of means by subtracting 22.50 from each entry. The values in the margins are the effects of the corresponding factors.

<table>
<thead>
<tr>
<th>Dietary Drug Dose Level</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>-7.50</td>
<td>-6.00</td>
<td>3.00</td>
<td>-3.50</td>
</tr>
<tr>
<td>D2</td>
<td>-3.00</td>
<td>-2.50</td>
<td>16.00</td>
<td>3.50</td>
</tr>
<tr>
<td>Overall</td>
<td>-5.25</td>
<td>-4.25</td>
<td>9.50</td>
<td>22.50</td>
</tr>
</tbody>
</table>

**Step 2 – Remove the Effects of Factor B (Diet)**

Subtract the Diet effects (-3.50 and 3.50) from the entries in those rows.

<table>
<thead>
<tr>
<th>Dietary Drug Dose Level</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>-4.00</td>
<td>-2.50</td>
<td>6.50</td>
<td>-3.50</td>
</tr>
<tr>
<td>D2</td>
<td>-6.50</td>
<td>-6.00</td>
<td>12.50</td>
<td>3.50</td>
</tr>
<tr>
<td>Overall</td>
<td>-5.25</td>
<td>-4.25</td>
<td>9.50</td>
<td>22.50</td>
</tr>
</tbody>
</table>

**Step 3 – Remove the Effects of Factor A (Drug Dose)**

Subtract the Drug Dose effects (-5.25, -4.25, and 9.50) from the rest of the entries in those columns. This will result in a table of effects.

<table>
<thead>
<tr>
<th>Dietary Drug Dose Level</th>
<th>Low</th>
<th>Medium</th>
<th>High</th>
<th>Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1</td>
<td>1.25</td>
<td>1.75</td>
<td>-3.00</td>
<td>-3.50</td>
</tr>
<tr>
<td>D2</td>
<td>-1.25</td>
<td>-1.75</td>
<td>3.00</td>
<td>3.50</td>
</tr>
<tr>
<td>Overall</td>
<td>-5.25</td>
<td>-4.25</td>
<td>9.50</td>
<td>22.50</td>
</tr>
</tbody>
</table>
We have calculated a table of effects for the two-way linear model. Each cell mean can be calculated by summing the appropriate entries from this table.

The estimated linear effects are:

\[ m = 22.50 \]
\[ a_1 = -5.25 \quad a_2 = -4.25 \quad a_3 = 9.50 \]
\[ b_1 = -3.50 \quad b_2 = 3.50 \]
\[ ab_{11} = 1.25 \quad ab_{21} = 1.75 \quad ab_{31} = -3.00 \]
\[ ab_{12} = -1.25 \quad ab_{22} = -1.75 \quad ab_{32} = 3.00. \]

The six cell means are calculated from these effects as follows:

\[ 15.00 = 22.50 - 5.25 - 3.50 + 1.25 \]
\[ 19.50 = 22.50 - 5.25 + 3.50 - 1.25 \]
\[ 16.50 = 22.50 - 4.25 - 3.50 + 1.75 \]
\[ 20.00 = 22.50 - 4.25 + 3.50 - 1.75 \]
\[ 25.50 = 22.50 + 9.50 - 3.50 - 3.00 \]
\[ 38.50 = 22.50 + 9.50 + 3.50 + 3.00 \]

**Analysis of Variance Hypotheses**

The hypotheses that are tested in an analysis of variance table concern the effects, so in order to conduct a power analysis you must have a firm grasp of their meaning. For example, we would usually test the following hypotheses:

1. Are there differences in weight loss among the three drug doses? That is, are the drug dose effects all zero? This hypothesis is tested by the \( F \) test for factor \( A \), which tests whether the standard deviation of the \( a_i \) is zero.

2. Is there a difference in weight loss between the two diets? That is, are the diet effects all zero? This hypothesis is tested by the \( F \) test for factor \( B \), which tests whether the standard deviation of the \( b_j \) is zero.

3. Are there any diet-dose combinations that exhibit a weight loss that cannot be explained by diet and/or drug dose singly? This hypothesis is tested by the \( F \) test for the \( AB \) interaction, which tests whether the standard deviation of the \( (ab)_{ij} \) is zero.

Each of these hypotheses can be tested at a different alpha level and different precision. Hence each can have a different power. One of the tasks in planning such an experiment is to determine a sample size that yields necessary power values for each of these hypothesis tests. This is accomplished using this program module.
Definition of Terms

Factorial designs evaluate the effect of two or more categorical variables (called factors) on a response variable by testing hypotheses about various averages. These designs are popular because they allow experimentation across a wide variety of conditions and because they evaluate the interaction of two or more factors. Interaction is the effect that may be attributed to a combination of two or more factors, but not to one factor singly.

A factor is a variable that relates to the response. Either the factor is discrete by nature (as in location or gender) or has been made discrete by collapsing a continuous variable (as in income level or age group). The term factorial implies that all possible combinations of the factors being studied are included in the design.

A fixed factor is one in which all possible levels (categories) are considered. Examples of fixed factors are gender, dose level, and country of origin. They are different from random factors which represent a random selection of individuals from the population described by the factor. Examples of random factors are people living within a region, a sample of schools in a state, or a selection of labs. Again, a fixed factor includes the range of interest while a random factor includes only a sample of all possible levels.

A factorial design is analyzed using the analysis of variance. When only fixed factors are used in the design, the analysis is said to be a fixed-effects analysis of variance. Other types of designs will be discussed in later chapters.

Suppose a group of individuals have agreed to be in a study involving six treatments. In a completely randomized factorial design, each individual is assigned at random to one of the six groups and then the treatments are applied. In some situations, the randomization occurs by randomly selecting individuals from the populations defined by the treatment groups. The designs analyzed by this module are completely randomized factorial designs.

Power Calculations

The calculation of the power of a particular test proceeds as follows

1. Determine the critical value, $F_{df_1, df_2, \alpha}$ where $df_1$ is the numerator degrees of freedom, $df_2$ is the denominator degrees of freedom, and $\alpha$ is the probability of a type-I error (significance level). Note that the $F$ test is a two-tailed test as no logical direction is assigned in the alternative hypothesis.

2. Calculate the standard deviation of the hypothesized effects, using the formula:

$$\sigma_m = \sqrt{\frac{\sum_{i=1}^{k} (e_i - \bar{e})^2}{k}}$$

where the $e_i$ are effect values and $k$ is the number of effects. Note that the average effect will be zero by construction, so this formula reduces to

$$\sigma_m = \sqrt{\frac{\sum_{i=1}^{k} (e_i)^2}{k}}$$
3. Compute the noncentrality parameter $\lambda$ using the relationship:

$$\lambda = N \frac{\sigma^2_m}{\sigma^2}$$

where $N$ is the total number of subjects.

4. Compute the power as the probability of being greater than $F_{df_1,df_2,\alpha}$ on a noncentral-F distribution with noncentrality parameter $\lambda$.

**Example**

In the example discussed earlier, the standard deviation of the dose effects is

$$\sigma_m(A) = \sqrt{\frac{(-5.25)^2 + (-4.25)^2 + 9.50^2}{3}} = 6.729908$$

the standard deviation of the diet effects is

$$\sigma_m(B) = \sqrt{\frac{(-3.5)^2 + 3.5^2}{2}} = 3.5$$

and the standard deviation of the interaction effects is

$$\sigma_m(AB) = \sqrt{\frac{1.25^2 + (-1.25)^2 + 1.75^2 + (-1.75)^2 + (-3.00)^2 + 3.00^2}{6}} = 2.131119$$

**Change in Calculation from PASS 6.0**

In **PASS 6.0**, we used the approach of Cohen (1988) to calculate $\lambda$. However, we have found that Cohen’s method is less accurate in some situations. Here’s why. Cohen produced a set of tables for the one-way AOV which he extended to the two-way and three-way cases by adjusting the per group sample size (his $n'$) so that the denominator degrees of freedom were accurate. However, his adjustment also causes a change in $\lambda$ which can cause a substantial difference in the calculated power. By using the formula

$$\lambda = N \frac{\sigma^2_m}{\sigma^2}$$

we now calculate the correct power. This is why our calculations differ from that of Cohen (1988) for fixed factorial models.
Standard Deviation of Effects (of Means)

In the two-sample t-test case, the alternative hypothesis was represented as the difference between two group means. Unfortunately, for three or more groups, there is no simple extension of the two group difference. Instead, you must hypothesize a set of effects and calculate the value of $\sigma_m$.

Some might wish to specify the alternative hypothesis as the effect size, $f$, which is defined as

$$f = \frac{\sigma_m}{\sigma}$$

where $\sigma$ is the standard deviation of values within a cell (see Sigma below). If you want to use $f$, set $\sigma = 1$ and then $f$ is always equal to $\sigma_m$ so that the values you enter for $\sigma_m$ will be the values of $f$. Cohen (1988) has designated values of $f$ less than 0.1 as small, values around 0.25 to be medium, and values over 0.4 to be large. You should come up with your own cutoff values for low, medium, and high.

When you are analyzing the power of an existing analysis of variance table, you can compute the values of $\sigma_m$ for each term from its mean square or $F$ ratio using the following formulas:

$$\sigma_m = \sqrt{\frac{df_{\text{numerator}} MS_{\text{numerator}}}{N}}$$

or

$$\sigma_m = \sqrt{\frac{df_{\text{numerator}} (F)(MSE)}{N}}$$

where $N$ is the total number of observations, $MSE$ is the mean square error, $df$ is the numerator degrees of freedom, $MS$ is the mean square of the term, and $F$ is the $F$ ratio of the term. If you do this, you are setting the sample effects equal to the population effects for the purpose of computing the power.

Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains most of the parameters and options that you will be concerned with.

Sample Size

N per Cell

This is the sample size within a cell. Fractional values are allowed. When you have an unequal number of observations per cell, enter the average cell sample size.

If you enter more than one value, a separate analysis will be generated for each value.
Effect Size – Main Effects & Interactions

Factors (A, B, C) & Interactions (AB, ..., ABC)

These check boxes specify which terms are included in the analysis of variance model. Check a term to signify that it must be included in the analysis.

The three factors are assigned the labels A, B, and C. The interaction between factors A and B is labeled AB. The three-way interaction is labeled ABC.

You cannot include an interaction term without including all shorter terms that make up that interaction. For example, if you include the interaction AC, you must also include the terms A and C. Similarly, if you include the term ABC, you must also include the terms A, B, C, AB, AC, and BC.

Effect Size – Main Effects

Categories (A, B, and C)

These options specify the number of categories (levels) contained in each factor. Since the total sample size is equal to the product of the number of levels in each factor and the number of observations per cell (N Per Cell), increasing the number of levels of a factor increases the total sample size of the experiment.

Hypothesized Means (A, B, C)

Enter a set of hypothesized means (or effects), one for each factor level. The standard deviation of these means is used in the power calculations. The standard deviation is calculated using the formula:

$$\sigma_m = \sqrt{\frac{\sum_{i=1}^{k} (\mu_i - \overline{\mu})^2}{k}}$$

where $k$ is the number of effects. Note that the standard deviation will be the same whether you enter means or effects since the average of the effects is zero by definition.

Enter a set of means that give the pattern of differences you expect or the pattern that you wish to detect. For example, in a particular study involving a factor with three categories, your research might be meaningful if either of two treatment means is 50% larger than the control mean. If the control mean is 50, then you would enter 50, 75, 75 as the three means.

It is usually more intuitive to enter a set of mean values. However, it is possible to enter the standard deviation of the means directly by placing an S in front of the number.

Entering a List of Means

If numbers are entered without a leading S, they are assumed to be the hypothesized group means under the alternative hypothesis. Their standard deviation will be calculated and used in the calculations. Blanks or commas may separate the numbers. Note that it is not the values of the means themselves that is important, but only their differences. Thus, the mean values 0,1,2 produce the same results as the values 100,101,102.
If not enough means are entered to match the number of groups, the last mean is repeated. For example, suppose that four means are needed and you enter 1,2 (only two means). \textit{PAS} will treat this as 1,2,2,2. If too many values are entered, \textit{PASS} will truncate the list to the number of means needed.

\textbf{Examples:}

- 5 20 60
- 2,5,7
- -4,0,6,9

\textbf{S Option}

If an \textit{S} is entered before a number, the number is assumed to be the value of $\sigma_m$, the standard deviation of the means.

\textbf{Examples:}

- S 4.7
- S 5.7

\section*{Effect Size – Interactions}

\textbf{Hypothesized Effects}

Specify the standard deviation of the interaction effects using one of the following methods:

1. Enter a set of effects and let the program calculate their standard deviation (see below).
2. Enter the standard deviation directly.
3. Instruct the program to make the standard deviation proportional to one of the main effect terms.

The standard deviation of the effects is calculated using the formula:

$$
\sigma_m = \sqrt{\frac{\sum_{i=1}^{k} (e_i - \bar{e})^2}{k}}
$$

where $k$ is the number of effects and $e_1, e_2, \ldots, e_k$ are the effect values. The value of $\bar{e}$ may be ignored because it is zero by definition.

\textbf{Entering a List of Effects}

If numbers are entered without a leading letter, they are assumed to be the hypothesized effects under the alternative hypothesis (they are all assumed to be zero under the null hypothesis). Their standard deviation will be calculated and used in the calculations. Blanks or commas may separate the numbers.

If not enough effects are entered to match the number of levels in the term, the last effect is repeated. For example, suppose that four effects are needed and you enter 1,2 (only two effects). \textit{PAS} will treat this as 1,2,2,2. If too many values are entered, \textit{PASS} will truncate the list to the number of effects needed.
For interactions, the number of effects is equal to the product of the number of levels of each factor in the interaction. For example, suppose a two-factor design has one factor with three levels and another factor with five levels. The number of effects in the two-factor interaction is \((3)(5) = 15\).

Examples (note that they sum to zero):
-1 1 -3 3
2 2 0 -1 -1 -2
-4,0,1,3

**S Option**
If an \(S\) is followed by a number, the number is assumed to be the value of \(\sigma_m\), the standard deviation of the effects.

When a set of effects are equal to either \(e\) or \(-e\), the formula for the standard deviation may be simplified as follows:

\[
\sigma_m = \sqrt{\frac{\sum_{i=1}^{k} (e_i - 0)^2}{k}}
\]

\[
= \sqrt{\sum_{i=1}^{k} e_i^2}
\]

\[
= \frac{e}{k}
\]

Hence, another interpretation of \(\sigma_m\) is the absolute value of a set of effects that are equal, except for the sign.

**Example:**
S 4.7

**Enter a Term Followed by a Percentage**
You can enter the name of a previous term followed by a percentage. This instructs the program to set this standard deviation to \(x\)% of the term you specify, where \(x\) is a positive integer. This allows you to set the magnitude of the interaction standard deviation as a percentage of another term without specifying the interaction in detail.

Note that the term you are taking a percentage of must appear above the term you are specifying. That is, you cannot specify \(AB 50\) for factor \(C\) (since only \(A\) and \(B\) occur above \(C\) on the screen).

For example, if the standard deviation of factor \(A\) is 16, the command

A 75

will set the standard deviation of the current term to \((16)(75)/(100) = 12.0\).

Other examples of this syntax are:
A 50
B 25
AB 125
AC 150
Fixed Effects Analysis of Variance  560-11

Discussion

The general formula for the calculation of the standard deviation is

$$
\sigma_m = \sqrt{\frac{\sum_{i=1}^{k} (e_i)^2}{k}}
$$

where $k$ is the number of effects. In the case of a two-way interaction, the standard deviation is calculated using the formula:

$$
\sigma_m(AB) = \sqrt{\frac{\sum_{i=1}^{I} \sum_{j=1}^{J} (\mu_{ij} - \mu_{..} - \mu_{.j} + \mu) )^2}{IJ}}
$$

where $i$ is the factor $A$ index (from 1 to $I$), $j$ is the factor $B$ index (from 1 to $J$), $\mu_{ij}$ is the mean in the $ij^{th}$ cell, $\mu_{..}$ is the $i^{th}$ mean of factor $A$ across all levels of other factors, $\mu_{.j}$ is the $j^{th}$ mean when factor $B$ across all levels of other factors, and $\mu$ is the overall mean of the means.

To see how this works, consider the following table of means from an experiment with $I = 2$ and $J = 3$:

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.0</td>
<td>4.0</td>
<td>3.0</td>
</tr>
<tr>
<td>2</td>
<td>4.0</td>
<td>6.0</td>
<td>5.0</td>
</tr>
<tr>
<td>3</td>
<td>6.0</td>
<td>11.0</td>
<td>8.5</td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td>Total</td>
<td>4.0</td>
<td>7.0</td>
<td>5.5</td>
</tr>
</tbody>
</table>

Now, if we subtract the factor $A$ means, subtract the factor $B$ means, and add the overall mean, we get the interaction effects:

$$
0.5 \quad 0.5 \quad 0.5 \quad -0.5 \quad 1.0
$$

Next, we sum the squares of these six values:

$$
(0.5)^2 + (-0.5)^2 + (0.5)^2 + (-0.5)^2 + (-1.0)^2 + (1.0)^2 = 3
$$

Next we divide this value by $(2)(3) = 6$:

$$
3 / 6 = 0.5
$$
Finally, we take the square root of this value:

$$\sqrt{0.5} = 0.7071$$

Hence, for this configuration of means,

$$\sigma_m(AB) = 0.7071.$$  

Notice that the average of the absolute values of the interaction effects is:

$$\frac{0.5 + 0.5 + 0.5 + 0.5 + 1.0 + 1.0}{6} = 0.6667$$

We see that SD(interaction) is close to the average absolute interaction effect. That is, 0.7071 is close to 0.6667. This will usually be the case. Hence, one way to interpret the interaction standard deviation is as a number a little larger than the average absolute interaction effect.

**Alpha**

These options specify the significance levels (the probability of a type-I error) of each term. A type-I error occurs when you reject the null hypothesis of that all effects are zero when in fact they are.

Since they are probabilities, alpha values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This value may be interpreted as meaning that about one $F$ test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You can select different alpha values for different terms. For example, although you have three factors in an experiment, you might be mainly interested in only one of them. Hence, you could increase the alpha level of the tests from, for example, 0.05 to 0.10 and thereby increase their power. Also, you may want to increase the alpha level of the interaction terms, since these will often have poor power otherwise.

**Effect Size – Standard Deviation**

**S (Standard Deviation of Subjects)**

This option specifies the value of the standard deviation ($\sigma$) within a cell (the analysis of variance assumes that $\sigma$ is constant across all cells). Since they are positive square roots, the numbers must be strictly greater than zero. You can press the SD button to obtain further help on estimating the standard deviation.

This value may be estimated from a previous analysis of variance table by the square root of the mean square error.

If you want to use the effect size, $f$, as the measure of the variability of the effects, you can use 1.0 for $\sigma$. 


Example 1 – Power after a Study

This example will explain how to calculate the power of \( F \) tests from data that have already been collected and analyzed.

Analyze the power of the experiment that was given at the beginning of this chapter. These data were analyzed using the analysis of variance procedure in \( \text{NCSS} \) and the following results were obtained.

<table>
<thead>
<tr>
<th>Analysis of Variance Table</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Source</strong></td>
</tr>
<tr>
<td>------</td>
</tr>
<tr>
<td>A (Dose)</td>
</tr>
<tr>
<td>B (Diet)</td>
</tr>
<tr>
<td>AB</td>
</tr>
<tr>
<td>S</td>
</tr>
<tr>
<td>Total (Adjusted)</td>
</tr>
<tr>
<td>Total</td>
</tr>
</tbody>
</table>

* Term significant at alpha = 0.05

<table>
<thead>
<tr>
<th>Means and Effects Section</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Term</strong></td>
</tr>
<tr>
<td>All</td>
</tr>
<tr>
<td>A: Dose</td>
</tr>
<tr>
<td>High</td>
</tr>
<tr>
<td>Medium</td>
</tr>
<tr>
<td>Low</td>
</tr>
<tr>
<td>B: Diet</td>
</tr>
<tr>
<td>D1</td>
</tr>
<tr>
<td>D2</td>
</tr>
<tr>
<td>AB: Dose,Diet</td>
</tr>
<tr>
<td>High,D1</td>
</tr>
<tr>
<td>High,D2</td>
</tr>
<tr>
<td>Low,D1</td>
</tr>
<tr>
<td>Low,D2</td>
</tr>
<tr>
<td>Medium,D1</td>
</tr>
<tr>
<td>Medium,D2</td>
</tr>
</tbody>
</table>

Setup

To analyze these data, we can enter the means for factors \( A \) and \( B \) as well as the \( AB \) interaction effects.

Alternatively, we could have calculated the standard deviation of the interaction. This can be done in either of two ways.

Using mean square for \( AB \) (27.25), the degrees of freedom for \( AB \) (2), and the total sample size (12), the standard deviation of the \( AB \)-interaction effects is calculated as follows

\[
\sigma_{\text{AB}} = \sqrt{\frac{2(27.25)}{12}} = 2.1311
\]

Using the formula based on the effects, the standard deviation of the \( AB \)-interaction effects is calculated as follows

\[
\sigma_{\text{AB}} = \sqrt{\frac{(25.50 - 22.50)^2 + (25.50 - 22.50)^2 + (25.50 - 22.50)^2}{2}} = 2.1311
\]
\[
\sigma_m(AB) = \sqrt{\frac{3^2 + 3^2 + 1.25^2 + 1.25^2 + 1.75^2 + 1.75^2}{6}} = 2.1311
\]

The value of \(\sigma\) is estimated from the square root of the mean square error:

\[
\sigma = \sqrt{5.333333} = 2.3094
\]

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Fixed Effects Analysis of Variance procedure window by expanding Means, then ANOVA, then clicking on Analysis of Variance, and then clicking on Fixed Effects Analysis of Variance. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>N Per Cell</td>
<td>2</td>
</tr>
<tr>
<td>Factors (A, B, AB)</td>
<td>Checked</td>
</tr>
<tr>
<td>Factors (C, AC, BC, ABC)</td>
<td>Not checked</td>
</tr>
<tr>
<td>Categories (A)</td>
<td>3</td>
</tr>
<tr>
<td>Categories (B)</td>
<td>2</td>
</tr>
<tr>
<td>Hypothesized Means (A)</td>
<td>17.25 18.25 32</td>
</tr>
<tr>
<td>Hypothesized Means (B)</td>
<td>19 26</td>
</tr>
<tr>
<td>Hypothesized Effects (AB)</td>
<td>-3 3 1.25 -1.25 1.75 -1.75</td>
</tr>
<tr>
<td>Alpha</td>
<td>All are set to 0.05</td>
</tr>
<tr>
<td>S (Std Dev of Subjects)</td>
<td>2.3094</td>
</tr>
<tr>
<td>Report Tab</td>
<td></td>
</tr>
<tr>
<td>Numeric Report Prob Decimals</td>
<td>6</td>
</tr>
<tr>
<td>Std Dev Decimals</td>
<td>4</td>
</tr>
</tbody>
</table>

Annotated Output

Click the Run button to perform the calculations and generate the following output.

**Numeric Results**

<table>
<thead>
<tr>
<th>Term</th>
<th>Power</th>
<th>n</th>
<th>N</th>
<th>df1</th>
<th>df2</th>
<th>Std Dev of Means (Sm)</th>
<th>Effect Size</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1.000000</td>
<td>2.00</td>
<td>12</td>
<td>2</td>
<td>6</td>
<td>6.7299</td>
<td>2.914</td>
<td>0.000000</td>
<td>0.000000</td>
</tr>
<tr>
<td>B</td>
<td>0.990499</td>
<td>2.00</td>
<td>12</td>
<td>1</td>
<td>6</td>
<td>3.5000</td>
<td>1.516</td>
<td>0.050000</td>
<td>0.009501</td>
</tr>
<tr>
<td>AB</td>
<td>0.588914</td>
<td>2.00</td>
<td>12</td>
<td>2</td>
<td>6</td>
<td>2.1311</td>
<td>0.923</td>
<td>0.050000</td>
<td>0.411086</td>
</tr>
</tbody>
</table>

Standard Deviation Within Subjects = 2.3094

**Summary Statements**

A factorial design with two factors at 3 and 2 levels has 6.0 cells (treatment combinations). A total of 12.0 subjects are required to provide 2.0 subjects per cell. The within-cell standard deviation is 2.3094. This design achieves 100% power when an F test is used to test factor A at a 5% significance level and the actual standard deviation among the appropriate means is 6.7299 (an effect size of 2.914), achieves 99% power when an F test is used to test factor B at a 5% significance level and the actual standard deviation among the appropriate means is 3.5000 (an effect size of 1.516), and achieves 99% power when an F test is used to test the AB interaction at a 5% significance level and the actual standard deviation among the appropriate means is 2.1312 (an effect size of 0.923).
This report shows the power for each of the three factors. Note that these power values match those given by the *NCSS* program in the analysis of variance report.

It is important to emphasize that these power values are for the case when the effects associated with the alternative hypotheses are equal to those given by the data. It will often be informative to calculate the power for other values as well.

**Term**

This is the term (main effect or interaction) from the analysis of variance model being displayed on this line.

**Power**

This is the power of the $F$ test for this term. Note that since adding and removing terms changes the denominator degrees of freedom ($df_2$), the power depends on which other terms are included in the model.

**n**

This is the sample size per cell (treatment combination). Fractional values indicate an unequal allocation among the cells.

**Total N**

This is the total sample size for the complete design.

**df1**

This is the numerator degrees of freedom of the $F$ test.

**df2**

This is the denominator degrees of freedom of the $F$ test. This value depends on which terms are included in the AOV model.

**Std Dev of Means (Sm)**

This is the standard deviation of the means (or effects). It represents the size of the differences among the effects that is to be detected by the analysis. If you have entered hypothesized means, only their standard deviation is displayed here.

**Effect Size**

This is the standard deviation of the means divided by the standard deviation of subjects. It provides an index of the magnitude of the difference among the means that can be detected by this design.

**Alpha**

This is the significance level of the $F$ test. This is the probability of a type-I error given the null hypothesis of equal means and zero effects.

**Beta**

This is the probability of the type-II error for this test given the sample size, significance level, and effect size.
Example 2 – Finding the Sample Size

In this example, we will investigate the impact of increasing the sample size on the power of each of the seven tests in the analysis of variance table of a three factor experiment. The first factor (A) has two levels, the second factor (B) has three levels, and the third factor (C) has four levels. This creates a design with \(2 \times 3 \times 4 = 24\) treatment combinations.

All values of \(\sigma_m\) will be set equal to 0.2, \(\sigma\) is set equal to 1.0, and alpha is set to 0.05.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Fixed Effects Analysis of Variance procedure window by expanding Means, then ANOVA, then clicking on Analysis of Variance, and then clicking on Fixed Effects Analysis of Variance. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>N Per Cell</td>
<td>2 8 16 22</td>
</tr>
<tr>
<td>Factors and Interactions</td>
<td>All Checked</td>
</tr>
<tr>
<td>Categories (A)</td>
<td>2</td>
</tr>
<tr>
<td>Categories (B)</td>
<td>3</td>
</tr>
<tr>
<td>Categories (C)</td>
<td>4</td>
</tr>
<tr>
<td>Hypothesized Means (A, B, &amp; C)</td>
<td>S 0.2</td>
</tr>
<tr>
<td>Hypothesized Effects (AB to ABC)</td>
<td>A 100 (so they will equal that of factor A)</td>
</tr>
<tr>
<td>Alpha</td>
<td>All are set to 0.05</td>
</tr>
<tr>
<td>S (Std Dev of Subjects)</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Term</th>
<th>Power</th>
<th>n</th>
<th>Total N</th>
<th>df1</th>
<th>df2</th>
<th>Std Dev of Means (Sm)</th>
<th>Effect Size</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.26502</td>
<td>2.00</td>
<td>48</td>
<td>1</td>
<td>24</td>
<td>0.2</td>
<td>0.200</td>
<td>0.05000</td>
<td>0.73498</td>
</tr>
<tr>
<td>B</td>
<td>0.19674</td>
<td>2.00</td>
<td>48</td>
<td>2</td>
<td>24</td>
<td>0.2</td>
<td>0.200</td>
<td>0.05000</td>
<td>0.80326</td>
</tr>
<tr>
<td>C</td>
<td>0.16369</td>
<td>2.00</td>
<td>48</td>
<td>3</td>
<td>24</td>
<td>0.2</td>
<td>0.200</td>
<td>0.05000</td>
<td>0.83631</td>
</tr>
<tr>
<td>AB</td>
<td>0.19674</td>
<td>2.00</td>
<td>48</td>
<td>2</td>
<td>24</td>
<td>0.2</td>
<td>0.200</td>
<td>0.05000</td>
<td>0.80326</td>
</tr>
<tr>
<td>AC</td>
<td>0.16369</td>
<td>2.00</td>
<td>48</td>
<td>3</td>
<td>24</td>
<td>0.2</td>
<td>0.200</td>
<td>0.05000</td>
<td>0.83631</td>
</tr>
<tr>
<td>BC</td>
<td>0.11945</td>
<td>2.00</td>
<td>48</td>
<td>6</td>
<td>24</td>
<td>0.2</td>
<td>0.200</td>
<td>0.05000</td>
<td>0.88055</td>
</tr>
<tr>
<td>ABC</td>
<td>0.11945</td>
<td>2.00</td>
<td>48</td>
<td>6</td>
<td>24</td>
<td>0.2</td>
<td>0.200</td>
<td>0.05000</td>
<td>0.88055</td>
</tr>
<tr>
<td>A</td>
<td>0.78682</td>
<td>8.00</td>
<td>192</td>
<td>1</td>
<td>168</td>
<td>0.2</td>
<td>0.200</td>
<td>0.05000</td>
<td>0.21318</td>
</tr>
<tr>
<td>B</td>
<td>0.69038</td>
<td>8.00</td>
<td>192</td>
<td>2</td>
<td>168</td>
<td>0.2</td>
<td>0.200</td>
<td>0.05000</td>
<td>0.30962</td>
</tr>
<tr>
<td>C</td>
<td>0.62299</td>
<td>8.00</td>
<td>192</td>
<td>3</td>
<td>168</td>
<td>0.2</td>
<td>0.200</td>
<td>0.05000</td>
<td>0.37701</td>
</tr>
<tr>
<td>AB</td>
<td>0.69038</td>
<td>8.00</td>
<td>192</td>
<td>2</td>
<td>168</td>
<td>0.2</td>
<td>0.200</td>
<td>0.05000</td>
<td>0.30962</td>
</tr>
<tr>
<td>AC</td>
<td>0.62299</td>
<td>8.00</td>
<td>192</td>
<td>3</td>
<td>168</td>
<td>0.2</td>
<td>0.200</td>
<td>0.05000</td>
<td>0.37701</td>
</tr>
<tr>
<td>BC</td>
<td>0.49353</td>
<td>8.00</td>
<td>192</td>
<td>6</td>
<td>168</td>
<td>0.2</td>
<td>0.200</td>
<td>0.05000</td>
<td>0.50647</td>
</tr>
<tr>
<td>ABC</td>
<td>0.49353</td>
<td>8.00</td>
<td>192</td>
<td>6</td>
<td>168</td>
<td>0.2</td>
<td>0.200</td>
<td>0.05000</td>
<td>0.50647</td>
</tr>
</tbody>
</table>
A few interesting features of this report stand out. First note the range of power values across the range of sample size values tested. Reasonable power is not reached until \( n \) is 16. Also note that as the number of numerator degrees of freedom (\( df_1 \)) increases, the power decreases, other things being equal. We must use this knowledge when planning for appropriate power in tests of important interaction terms.

There are a lot of additional runs that you might try. For example, you might look at the impact of setting the alpha level of interaction terms 0.008. You might look at varying \( \sigma_m \) across the different terms. You might try varying the number of levels of a factor. All of these will impact the power of the \( F \) tests and will thus be important to consider during the planning stage of an experiment.

**Example 3 – Latin Square Design**

This example shows how to study the power of a complicated experimental design like a Latin square. Suppose you want to run a Five-Level Latin square design. Recall that a Five-Level Latin square design consists of three factors each at five levels. One factor is associated with the columns of the square, a second factor is associated with the rows of the square, and a third factor is associated with the letters of the square. In all there are only \( 5 \times 5 = 25 \) observations used instead of the \( 5 \times 5 \times 5 = 125 \) that would normally be required. The Latin square design has reduced the number of observations by 80%.

The 80% decrease in observations comes at a price—the interaction terms must be ignored. If you can legitimately assume that the interactions are zero, the Latin square (or some other design which reduces the number of observations) is an efficient design to use. We will now show you how to analyze the power of the \( F \) tests from such a design.

The key is to enter 0.2 (which is 25/125) for \( n \) and set all the interaction indicators off. Since all three factors have five levels, the power of the three \( F \) tests will be the same if \( \sigma_m \) is the same. Hence, we can try three different sets of hypothesized means. The first set will be five means 0.1 units apart. The second set will be five means 0.5 units apart. The third set will be five means 1.0 unit apart. The standard deviation will be set to 1.0. All alpha levels will be set at 0.05.

The sample size per cell is set at 0.2 and 0.4. This will result in total sample sizes of 25 (one replication) and 50 (two replications).
Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Fixed Effects Analysis of Variance procedure window by expanding Means, then ANOVA, then clicking on Analysis of Variance, and then clicking on Fixed Effects Analysis of Variance. You may then make the appropriate entries as listed below, or open Example 3 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>N Per Cell</td>
<td>0.2 0.4</td>
</tr>
<tr>
<td>Factors (A, B, C)</td>
<td>Checked</td>
</tr>
<tr>
<td>Interactions (AB, AC, BC, ABC)</td>
<td>Not checked</td>
</tr>
<tr>
<td>Categories (A, B, C)</td>
<td>5</td>
</tr>
<tr>
<td>Hypothesized Means (A)</td>
<td>1.0 1.1 1.2 1.3 1.4</td>
</tr>
<tr>
<td>Hypothesized Means (B)</td>
<td>1.0 1.5 2.0 2.5 3.0</td>
</tr>
<tr>
<td>Hypothesized Means (C)</td>
<td>1.0 2.0 3.0 4.0 5.0</td>
</tr>
<tr>
<td>Alpha</td>
<td>All are set to 0.05</td>
</tr>
<tr>
<td>S (Std Dev of Subjects)</td>
<td>1.0</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Term</th>
<th>Power</th>
<th>n</th>
<th>N</th>
<th>df1</th>
<th>df2</th>
<th>Std Dev of Means (Sm)</th>
<th>Effect Size</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.06807</td>
<td>0.20</td>
<td>25</td>
<td>4</td>
<td>12</td>
<td>0.141</td>
<td>0.14</td>
<td>0.05000</td>
<td>0.93193</td>
</tr>
<tr>
<td>B</td>
<td>0.63675</td>
<td>0.20</td>
<td>25</td>
<td>4</td>
<td>12</td>
<td>0.707</td>
<td>0.707</td>
<td>0.05000</td>
<td>0.36325</td>
</tr>
<tr>
<td>C</td>
<td>0.99867</td>
<td>0.20</td>
<td>25</td>
<td>4</td>
<td>12</td>
<td>1.414</td>
<td>1.414</td>
<td>0.05000</td>
<td>0.00133</td>
</tr>
<tr>
<td>A</td>
<td>0.09842</td>
<td>0.40</td>
<td>50</td>
<td>4</td>
<td>37</td>
<td>0.141</td>
<td>0.141</td>
<td>0.05000</td>
<td>0.90158</td>
</tr>
<tr>
<td>B</td>
<td>0.97743</td>
<td>0.40</td>
<td>50</td>
<td>4</td>
<td>37</td>
<td>0.707</td>
<td>0.707</td>
<td>0.05000</td>
<td>0.02257</td>
</tr>
<tr>
<td>C</td>
<td>1.00000</td>
<td>0.40</td>
<td>50</td>
<td>4</td>
<td>37</td>
<td>1.414</td>
<td>1.414</td>
<td>0.05000</td>
<td>0.00000</td>
</tr>
</tbody>
</table>

Standard Deviation of Subjects = 1.000

In the first design in which $N = 25$, only the power of the test for $C$ is greater than 0.8. Of course, this power value also depends on the value of the standard deviation of subjects within a cell.

It is interesting to note that doubling the sample size did not double the power!
Example 4 – Validation using Winer

Winer (1991) pages 428-429 presents the power calculations for a two-way design in which factor A has two levels and factor B has three levels. Winer provides estimates of the sum of squared A effects (1.0189), sum of squared B effects (5.06), and sum of squared interaction effects (42.11). The mean square error is 8.83 and the per cell sample size is 3. All alpha levels are set to 0.05.

Winer’s results are approximate because he has to interpolate in the tables that he is using. He finds the power of the $F$ test for factor A to be between 0.10 and 0.26. He estimates it as 0.17. The exact power of the $F$ test for factor B is not given. Instead, the range is found to be between 0.26 and 0.36. The power of the $F$ test for the AB interaction is “approximately” 0.86.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Fixed Effects Analysis of Variance procedure window by expanding Means, then ANOVA, then clicking on Analysis of Variance, and then clicking on Fixed Effects Analysis of Variance. You may then make the appropriate entries as listed below, or open Example 4 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>N Per Cell</td>
<td>3</td>
</tr>
<tr>
<td>Factors (A and B)</td>
<td>Checked</td>
</tr>
<tr>
<td>Factor (C)</td>
<td>Not checked</td>
</tr>
<tr>
<td>Interaction (AB)</td>
<td>Checked</td>
</tr>
<tr>
<td>Interactions (AC, BC, ABC)</td>
<td>Not checked</td>
</tr>
<tr>
<td>Categories (A)</td>
<td>2</td>
</tr>
<tr>
<td>Categories (B)</td>
<td>3</td>
</tr>
<tr>
<td>Hypothesized Means (A)</td>
<td>S 0.714</td>
</tr>
<tr>
<td>Hypothesized Means (B)</td>
<td>S 1.3</td>
</tr>
<tr>
<td>Hypothesized Effects (AB)</td>
<td>S 2.65</td>
</tr>
<tr>
<td>Alpha</td>
<td>All are set to 0.05</td>
</tr>
<tr>
<td>S (Std Dev of Subjects)</td>
<td>2.97</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

### Numeric Results

<table>
<thead>
<tr>
<th>Term</th>
<th>Power</th>
<th>n</th>
<th>N</th>
<th>df1</th>
<th>df2</th>
<th>Std Dev of Means (Sm)</th>
<th>Effect Size</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.15576</td>
<td>3.00</td>
<td>18</td>
<td>1</td>
<td>12</td>
<td>0.714</td>
<td>0.240</td>
<td>0.05000</td>
<td>0.84424</td>
</tr>
<tr>
<td>B</td>
<td>0.29178</td>
<td>3.00</td>
<td>18</td>
<td>2</td>
<td>12</td>
<td>1.300</td>
<td>0.438</td>
<td>0.05000</td>
<td>0.70822</td>
</tr>
<tr>
<td>AB</td>
<td>0.85338</td>
<td>3.00</td>
<td>18</td>
<td>2</td>
<td>12</td>
<td>2.650</td>
<td>0.892</td>
<td>0.05000</td>
<td>0.14662</td>
</tr>
</tbody>
</table>

Standard Deviation of Subjects = 2.970
The power of the test for factor A is 0.16 which is between 0.10 and 0.26. It is close to the interpolated 0.17 that Winer obtained from his tables.

The power of the test for factor B is 0.29 which is between 0.26 and 0.36.

The power of the test for the AB interaction is 0.85 which is close to the interpolated 0.86 that Winer obtained from his tables.

**Example 5 – Validation using Prihoda**

Prihoda (1983) pages 7-8 presents the power calculations for a two-way design with the following pattern of means:

<table>
<thead>
<tr>
<th>Factor B</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor A</td>
<td>1</td>
<td>41</td>
<td>34</td>
<td>30</td>
<td>27</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>33</td>
<td>24</td>
<td>22</td>
<td>29</td>
</tr>
<tr>
<td>All</td>
<td>37</td>
<td>29</td>
<td>26</td>
<td>28</td>
<td>30</td>
</tr>
</tbody>
</table>

The means may be manipulated to show the overall mean, the main effects, and the interaction effects:

<table>
<thead>
<tr>
<th>Factor B</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>All</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor A</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>-4</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>-1</td>
<td>-2</td>
<td>-1</td>
<td>4</td>
</tr>
<tr>
<td>All</td>
<td>7</td>
<td>-1</td>
<td>-4</td>
<td>-2</td>
<td>30</td>
</tr>
</tbody>
</table>

Based on the above effects, Prihoda calculates the power of the interaction test when the sample size per cell is 6, 8, 10, 12, and 14 to be 0.34, 0.45, 0.56, 0.65, and 0.73. The mean square error is 64 and the alpha level is 0.05.

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Fixed Effects Analysis of Variance procedure window by expanding Means, then ANOVA, then clicking on Analysis of Variance, and then clicking on Fixed Effects Analysis of Variance. You may then make the appropriate entries as listed below, or open Example 5 by going to the File menu and choosing Open Example Template.
### Output

Click the Run button to perform the calculations and generate the following output.

### Numeric Results

<table>
<thead>
<tr>
<th>Term</th>
<th>Power</th>
<th>n</th>
<th>Total N</th>
<th>df1</th>
<th>df2</th>
<th>Std Dev of Means (Sm)</th>
<th>Effect Size</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.71746</td>
<td>6.00</td>
<td>48</td>
<td>1</td>
<td>40</td>
<td>3.000</td>
<td>0.375</td>
<td>0.05000</td>
<td>0.28254</td>
</tr>
<tr>
<td>B</td>
<td>0.83676</td>
<td>6.00</td>
<td>48</td>
<td>3</td>
<td>40</td>
<td>4.183</td>
<td>0.523</td>
<td>0.05000</td>
<td>0.16324</td>
</tr>
<tr>
<td>AB</td>
<td>0.33722</td>
<td>6.00</td>
<td>48</td>
<td>3</td>
<td>40</td>
<td>2.345</td>
<td>0.293</td>
<td>0.05000</td>
<td>0.66278</td>
</tr>
<tr>
<td>A</td>
<td>0.83848</td>
<td>8.00</td>
<td>64</td>
<td>1</td>
<td>56</td>
<td>3.000</td>
<td>0.375</td>
<td>0.05000</td>
<td>0.16152</td>
</tr>
<tr>
<td>B</td>
<td>0.93871</td>
<td>8.00</td>
<td>64</td>
<td>3</td>
<td>56</td>
<td>4.183</td>
<td>0.523</td>
<td>0.05000</td>
<td>0.06129</td>
</tr>
<tr>
<td>AB</td>
<td>0.45099</td>
<td>8.00</td>
<td>64</td>
<td>3</td>
<td>56</td>
<td>2.345</td>
<td>0.293</td>
<td>0.05000</td>
<td>0.54901</td>
</tr>
<tr>
<td>A</td>
<td>0.91134</td>
<td>10.00</td>
<td>80</td>
<td>1</td>
<td>72</td>
<td>3.000</td>
<td>0.375</td>
<td>0.05000</td>
<td>0.08866</td>
</tr>
<tr>
<td>B</td>
<td>0.97917</td>
<td>10.00</td>
<td>80</td>
<td>3</td>
<td>72</td>
<td>4.183</td>
<td>0.523</td>
<td>0.05000</td>
<td>0.02083</td>
</tr>
<tr>
<td>AB</td>
<td>0.55558</td>
<td>10.00</td>
<td>80</td>
<td>3</td>
<td>72</td>
<td>2.345</td>
<td>0.293</td>
<td>0.05000</td>
<td>0.44442</td>
</tr>
<tr>
<td>A</td>
<td>0.95292</td>
<td>12.00</td>
<td>96</td>
<td>1</td>
<td>88</td>
<td>3.000</td>
<td>0.375</td>
<td>0.05000</td>
<td>0.04708</td>
</tr>
<tr>
<td>B</td>
<td>0.99346</td>
<td>12.00</td>
<td>96</td>
<td>3</td>
<td>88</td>
<td>4.183</td>
<td>0.523</td>
<td>0.05000</td>
<td>0.00654</td>
</tr>
<tr>
<td>AB</td>
<td>0.64749</td>
<td>12.00</td>
<td>96</td>
<td>3</td>
<td>88</td>
<td>2.345</td>
<td>0.293</td>
<td>0.05000</td>
<td>0.35251</td>
</tr>
<tr>
<td>A</td>
<td>0.97568</td>
<td>14.00</td>
<td>112</td>
<td>1</td>
<td>104</td>
<td>3.000</td>
<td>0.375</td>
<td>0.05000</td>
<td>0.02432</td>
</tr>
<tr>
<td>B</td>
<td>0.99807</td>
<td>14.00</td>
<td>112</td>
<td>3</td>
<td>104</td>
<td>4.183</td>
<td>0.523</td>
<td>0.05000</td>
<td>0.00193</td>
</tr>
<tr>
<td>AB</td>
<td>0.72541</td>
<td>14.00</td>
<td>112</td>
<td>3</td>
<td>104</td>
<td>2.345</td>
<td>0.293</td>
<td>0.05000</td>
<td>0.27459</td>
</tr>
</tbody>
</table>

Standard Deviation of Subjects = 8.000

Prihoda only presents the power for the interaction test at each sample size. You can check to see that the results match Prihoda’s exactly.
Example 6 – Validation using Neter, Kutner, Nachtsheim, and Wasserman

Neter, Kutner, Nachtsheim, and Wasserman (1996) page 1057 presents a power analysis of a two-factor experiment in which factor A has three levels and factor B has two levels. The significance level is 0.05, the standard deviation is 3.0, and N is 2. They calculate a power of about 0.89 for the test of factor A when the three means are 50, 55, and 45.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Fixed Effects Analysis of Variance procedure window by expanding Means, then ANOVA, then clicking on Analysis of Variance, and then clicking on Fixed Effects Analysis of Variance. You may then make the appropriate entries as listed below, or open Example 6 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>N Per Cell</td>
<td>2</td>
</tr>
<tr>
<td>Factors (A and B)</td>
<td>Checked</td>
</tr>
<tr>
<td>Factor (C)</td>
<td>Not checked</td>
</tr>
<tr>
<td>Interaction (AB)</td>
<td>Checked</td>
</tr>
<tr>
<td>Interactions (AC, BC, ABC)</td>
<td>Not checked</td>
</tr>
<tr>
<td>Categories (A)</td>
<td>3</td>
</tr>
<tr>
<td>Categories (B)</td>
<td>2</td>
</tr>
<tr>
<td>Hypothesized Means (A)</td>
<td>50 55 45</td>
</tr>
<tr>
<td>Hypothesized Means (B)</td>
<td>S 1</td>
</tr>
<tr>
<td>Hypothesized Effects (AB)</td>
<td>S 1</td>
</tr>
<tr>
<td>Alpha</td>
<td>All are set to 0.05</td>
</tr>
<tr>
<td>S (Std Dev of Subjects)</td>
<td>3.0</td>
</tr>
</tbody>
</table>

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Numeric Results</th>
<th>Power 0.090162</th>
<th>Total N 12 2 6</th>
<th>Std Dev of Means (Sm) 4.082</th>
<th>Effect Size 1.361</th>
<th>Alpha 0.05600</th>
<th>Beta 0.09838</th>
</tr>
</thead>
<tbody>
<tr>
<td>Term</td>
<td>Effect</td>
<td>df1</td>
<td>df2</td>
<td>Alpha</td>
<td>Beta</td>
<td></td>
</tr>
<tr>
<td>A</td>
<td>0.90162</td>
<td>2.00</td>
<td>12</td>
<td>2</td>
<td>6</td>
<td>0.05000</td>
</tr>
<tr>
<td>B</td>
<td>0.16479</td>
<td>2.00</td>
<td>12</td>
<td>1</td>
<td>6</td>
<td>0.05000</td>
</tr>
<tr>
<td>AB</td>
<td>0.11783</td>
<td>2.00</td>
<td>12</td>
<td>2</td>
<td>6</td>
<td>0.05000</td>
</tr>
</tbody>
</table>

Note that the power of 0.90 that PASS has calculated is within rounding of the 0.89 that Neter et al. calculated.
Chapter 565

Randomized Block Analysis of Variance

Introduction

This module analyzes a randomized block analysis of variance with up to two treatment factors and their interaction. It provides tables of power values for various configurations of the randomized block design.

The Randomized Block Design

The randomized block design (RBD) may be used when a researcher wants to reduce the experimental error among observations of the same treatment by accounting for the differences among blocks. If three treatments are arranged in two blocks, the RBD might appear as follows:

<table>
<thead>
<tr>
<th>Block A</th>
<th>Block B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment 1</td>
<td>Treatment 2</td>
</tr>
<tr>
<td>Treatment 3</td>
<td>Treatment 1</td>
</tr>
<tr>
<td>Treatment 2</td>
<td>Treatment 3</td>
</tr>
</tbody>
</table>

This diagram shows the main features of a RBD:

1. Each block is divided into k sub-blocks, where k is the number of treatments.
2. Each block receives all the treatments.
3. The treatments are assigned to the sub-blocks in random order.
4. There is some reason to believe that the blocks are the same internally, but different from each other.
RBD Reduces Random Error

The random error component of a completely randomized design (such as a one-way or a fixed-effects factorial design) represents the influence of all possible variables in the universe on the response except for the controlled (treatment) variables. This random error component is called the standard deviation or $\sigma$ (sigma).

As we have discussed, the sample size required to meet alpha and beta error requirements depends directly on the standard deviation. As the standard deviation increases, the sample size increases. Hence, researchers are always looking for ways to reduce the standard deviation. Since the random error component contains the variation due to all possible variables other than treatment variables, one of the most obvious ways to reduce the standard deviation is to remove one or more of these nuisance variables from the random error component. One of the simplest ways of doing this is by blocking on them.

For example, an agricultural experiment is often blocked on fields so that differences among fields are explicitly accounted for and removed from the error component. Since these field differences are caused by variations in variables such as soil type, sunlight, temperature, and water, blocking on fields removes the influence of several variables.

Blocks are constructed so that the response is as alike (homogeneous) as possible within a block, but as different as possible between blocks. In many situations, there are obvious natural blocking factors such as schools, seasons, individual farms, families, times of day, etc. In other situations, the blocks may be somewhat artificially constructed.

Once the blocks are defined, they are divided into $k$ smaller sections called subblocks, where $k$ is the number of treatment levels. The $k$ treatments are randomly assigned to the subblocks, one block at a time. Hence the order of treatment application will be different from block to block.

Measurement of Random Error

The measurement of the random error component ($\sigma$) is based on the assumption that there is no fundamental relationship between the treatment variable and the blocking variable. When this is true, the interaction component between blocks and treatment is zero. If the interaction component is zero, then the amount measured by the interaction is actually random error and can be used as an estimate of $\sigma$.

Hence, the randomized block design makes the assumption that there is no interaction between treatments and blocks. The block by treatment mean square is still calculated, but it is used as the estimated standard deviation. This means that the degrees of freedom associated with the block-treatment interaction are the degrees of freedom of the error estimate. If the experimental design has $k$ treatments and $b$ blocks, the interaction degrees of freedom are equal to $(k-1)(b-1)$. Hence the sample size of this type of experiment is measured in terms of the number of blocks.

Treatment Effects

Either one or two treatment variables may be specified. If two are used, their interaction may also be measured. The null hypothesis in the $F$ test states that the effects of the treatment variable are zero. The magnitude of the alternative hypothesis is represented as the size of the standard deviation ($\sigma_m$) of these effects. The larger the size of the effects, the larger their standard deviation.
When there are two factors, the block-treatment interaction may be partitioned just as the treatment may be partitioned. For example, if we let $C$ and $D$ represent two treatments, an analysis of variance will include the terms $C$, $D$, and $CD$. If we represent the blocking factor as $B$, there will be three interactions with blocks: $BC$, $BD$, and $BCD$. Since all three of these terms are assumed to measure the random error, the overall estimate of random error is found by averaging (or pooling) these three interactions. The pooling of these interactions increases the power of the experiment by effectively increasing the sample size on which the estimate of $\sigma$ is based. However, it is based on the assumption that $\sigma = \sigma_{BC} = \sigma_{BD} = \sigma_{BCD}$, which may or may not be true.

### An Example

Following is an example of data from a randomized block design. The block factor has four blocks ($B1$, $B2$, $B3$, $B4$) while the treatment factor has three levels (low, medium, and high). The response is shown within the table.

<table>
<thead>
<tr>
<th>Blocks</th>
<th>Treatments</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Low</td>
</tr>
<tr>
<td>B1</td>
<td>16</td>
</tr>
<tr>
<td>B2</td>
<td>18</td>
</tr>
<tr>
<td>B3</td>
<td>15</td>
</tr>
<tr>
<td>B4</td>
<td>14</td>
</tr>
</tbody>
</table>

### Analysis of Variance Hypotheses

The $F$ test for treatments in a randomized block design tests the hypothesis that the treatment effects are zero. (See the beginning of the Fixed-Effects Analysis of Variance chapter for a discussion of the meaning of effects.)

### Single-Factor Repeated Measures Designs

The randomized block design is often confused with a single-factor repeated measures design because the analysis of each is similar. However, the randomization pattern is different. In a randomized block design, the treatments are applied in random order within each block. In a repeated measures design, however, the treatments are usually applied in the same order through time. You should not mix the two. If you are analyzing a repeated measures design, we suggest that you use that module of PASS to do the sample size and power calculations.
Procedure Options

This section describes the options that are specific to this procedure. These are located on the Data tab. For more information about the options of other tabs, go to the Procedure Window chapter.

Data Tab

The Data tab contains most of the parameters and options that you will be concerned with.

Sample Size

Number of Blocks
This specifies one or more values for the number of blocks. If a list of values is entered, a separate calculation will be made for each value.

Effect Size – Main Effects & Interactions

Factors (A, B, and AB)
These check boxes specify which terms are included in the analysis of variance model. Check a term to indicate that it is included.

The two factors are assigned the labels A and B. The interaction between factors A and B is labeled AB. You cannot include the interaction term without including both A and B.

Effect Size – Main Effects

Categories (A and B)
This option specifies the number of categories (levels) contained in each factor. Since the effective sample size is equal to the product of the number of levels in each factor and the number of blocks, increasing the number of levels of a factor increases the sample size of the experiment.

Hypothesized Means (A and B)
Enter a set of hypothesized means (or effects), one for each factor level. The standard deviation of these means is used in the power calculations. The standard deviation is calculated using the formula:

\[ \sigma_m = \sqrt{\frac{\sum_{i=1}^{k} (e_i - \bar{e})^2}{k}} \]

where \( k \) is the number of levels. Note that the standard deviation will be the same whether you enter means or effects since the average of the effects is zero by definition.

Enter a set of means that give the pattern of differences you expect or the pattern that you wish to detect. For example, in a particular study involving a factor with three categories, your research might be meaningful if either of two treatment means is 50% larger than the control mean. If the control mean is 50, then you would enter 50, 75, 75 as the three means.
It is usually more intuitive to enter a set of mean values. However, it is possible to enter the standard deviation of the means directly by placing an $S$ in front of the number.

**Entering a List of Means**

If numbers are entered without a leading $S$, they are assumed to be the hypothesized group means under the alternative hypothesis. Their standard deviation will be calculated and used in the calculations. Blanks or commas may separate the numbers. Note that it is not the values of the means themselves that is important, but only their differences. Thus, the mean values $0, 1, 2$ produce the same results as the values $100, 101, 102$.

If not enough means are entered to match the number of groups, the last mean is repeated. For example, suppose that four means are needed and you enter $1, 2$ (only two means). $PASS$ will treat this as $1, 2, 2, 2$. If too many values are entered, $PASS$ will truncate the list to the number of means needed.

- Examples:
  - 5 20 60
  - 2, 5, 7
  - -4, 0, 6, 9

**S Option**

If an $S$ is entered before a number, the number is assumed to be the value of $\sigma_m$, the standard deviation of the means.

- Examples:
  - S 4.6
  - S 5.8

---

**Effect Size – Interactions**

**Hypothesized Effects**

Specify the standard deviation of the interaction effects using one of the following methods:

1. Enter a set of effects and let the program calculate their standard deviation.
2. Enter the standard deviation directly.
3. Instruct the program to make the standard deviation proportional to one of the main effect terms.

The standard deviation of the effects is calculated using the formula:

$$\sigma_m = \sqrt{\frac{\sum_{i=1}^{k} (e_i - \bar{e})^2}{k}}$$

where $k$ is the number of effects and $e_1, e_2, \ldots, e_k$ are the effect values. The value of $\bar{e}$ may be ignored because it is zero by definition.
**Entering a List of Effects**

If numbers are entered without a leading letter, they are assumed to be the hypothesized effects under the alternative hypothesis (they are all assumed to be zero under the null hypothesis). Their standard deviation will be calculated and used in the calculations. Blanks or commas may separate the numbers.

If not enough effects are entered to match the number of levels in the term, the last effect is repeated. For example, suppose that four effects are needed and you enter 1,2 (only two effects). **PASS** will treat this as 1,2,2,2. If too many values are entered, **PASS** will truncate the list to the number of effects needed.

For interactions, the number of effects is equal to the product of the number of levels of each factor in the interaction. For example, suppose a two-factor design has one factor with three levels and another factor with five levels. The number of effects in the two-factor interaction is (3)(5) = 15.

**Examples (note that they sum to zero):**
- -1 1 -3 3
- 2 2 0 -1 -1 -2
- -4,0,1,3

**S Option**

If an S is followed by a number, the number is assumed to be the value of \( \sigma_m \), the standard deviation of the effects.

When a set of effects are equal to either e or -e, the formula for the standard deviation may be simplified as follows:

\[
\sigma_m = \sqrt{\frac{\sum_{i=1}^{k} (e_i - \bar{e})^2}{k}}
\]

Hence, another interpretation of \( \sigma_m \) is the absolute value of a set of effects that are equal, except for the sign.

**Example:**

S 4.7

**Enter a Term Followed by a Percentage**

You can enter the name of a previous term followed by a percentage. This instructs the program to set this standard deviation to \( x \% \) of the term you specify, where \( x \) is a positive integer. This allows you to set the magnitude of the interaction standard deviation as a percentage of the standard deviation of one of the factors without specifying the interaction in detail.

For example, if the standard deviation of factor A is 16, the command

A 75

will set the standard deviation of the current term to \((16)(75)/(100) = 12.0\).

**Other examples of this syntax are:**

A 50
B 25
Discussion

The general formula for the calculation of the standard deviation is

$$\sigma_m = \sqrt{\frac{\sum_{i=1}^{k} (e_i)^2}{k}}$$

where $k$ is the number of effects. In the case of a two-way interaction, the standard deviation is calculated using the formula:

$$\sigma_{m}(AB) = \sqrt{\frac{\sum_{i=1}^{I} \sum_{j=1}^{J} (\mu_{ij} - \mu_i - \mu_j + \overline{\mu})^2}{IJ}}$$

where $i$ is the factor A index (from 1 to $I$), $j$ is the factor B index (from 1 to $J$), $\mu_{ij}$ is the mean in the $ij^{th}$ cell, $\mu_i$ is the mean of factor $A$ across all levels of other factors, $\mu_j$ is the $j^{th}$ mean when factor $B$ across all levels of other factors, and $\overline{\mu}$ is the overall mean of the means.

To see how this works, consider the following table of means from an experiment with $I = 2$ and $J = 3$:

<table>
<thead>
<tr>
<th>$i$</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2.0</td>
<td>4.0</td>
</tr>
<tr>
<td>2</td>
<td>4.0</td>
<td>6.0</td>
</tr>
<tr>
<td>3</td>
<td>6.0</td>
<td>11.0</td>
</tr>
</tbody>
</table>

--- --- --- --- ---

Total 4.0 7.0 5.5

Now, if we subtract the factor $A$ means, subtract the factor $B$ means, and add the overall mean, we get the interaction effects:

0.5 -0.5 
0.5 -0.5
-1.0 1.0

Next, we sum the squares of these six values:

$$(0.5)^2 + (-0.5)^2 + (0.5)^2 + (-0.5)^2 + (-1.0)^2 + (1.0)^2 = 3$$

Next we divide this value by $(2)(3) = 6$: 

$$\frac{3}{6} = 0.5$$
Finally, we take the square root of this value:
\[ \sqrt{0.5} = 0.7071 \]

Hence, for this configuration of means,
\[ \sigma_m(AB) = 0.7071. \]

Notice that the average of the absolute values of the interaction effects is:
\[ \frac{0.5 + 0.5 + 0.5 + 0.5 + 1.0 + 1.0}{6} = 0.6667. \]

We see that SD(interaction) is close to the average absolute interaction effect. That is, 0.7071 is close to 0.6667. This will usually be the case. Hence, one way to interpret the interaction standard deviation is as a number a little larger than the average absolute interaction effect.

**Alpha**
This option specifies the probability of a type-I error (alpha) for each term. A type-I error occurs when you reject the null hypothesis that the effects are zero when in fact they are.

Since they are probabilities, alpha values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This value may be interpreted as meaning that about one \( F \) test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You can select different alpha values for different terms. For example, although you have three factors in an experiment, you might be mainly interested in only one of them. Hence, you could increase the alpha level of the tests from, for example, 0.05 to 0.10 and thereby increase their power. Also, you may want to increase the alpha level of the interaction terms, since these will often have poor power otherwise.

**S (Standard Deviation)**
This option specifies the value of the standard deviation. In a randomized block design, this value is estimated by the square root of the mean square error (which may be listed as the mean square of the block-by-treatment interaction). This value will usually have to be determined from a previous study.

Assuming that each block is divided into several subblocks, this is an estimate of the standard deviation that would result when the subblocks within the same block received the same treatment.

If you want to use the effect size, \( f \), as the measure of the variability of the effects, you can use 1.0 for \( \sigma \).

Estimation of the standard deviation is discussed in detail in the Standard Deviation Estimator chapter.
Example 1 – Power after a Study

This example will explain how to calculate the power of $F$ tests from data that have already been collected and analyzed.

We will analyze the power of the experiment that was given at the beginning of this chapter.

These data were analyzed using the analysis of variance procedure in NCSS and the following results were obtained.

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-Ratio</th>
<th>Prob Level</th>
<th>Power (Alpha=0.05)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Term</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>A (Blocks)</td>
<td>3</td>
<td>13.66667</td>
<td>4.555555</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B (Treatment)</td>
<td>2</td>
<td>45.16667</td>
<td>22.58333</td>
<td>19.83</td>
<td>0.002269*</td>
<td>0.991442</td>
</tr>
<tr>
<td>AB</td>
<td>6</td>
<td>6.833333</td>
<td>1.138889</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>S</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (Adjusted)</td>
<td>11</td>
<td>65.66666</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* Term significant at alpha = 0.05

Means and Effects Section

<table>
<thead>
<tr>
<th>Term</th>
<th>Count</th>
<th>Mean</th>
<th>Standard Error</th>
<th>Effect</th>
</tr>
</thead>
<tbody>
<tr>
<td>B: Treatment</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High</td>
<td>4</td>
<td>20.5</td>
<td>0.5335937</td>
<td>2.333333</td>
</tr>
<tr>
<td>Low</td>
<td>4</td>
<td>15.75</td>
<td>0.5335937</td>
<td>-2.416667</td>
</tr>
<tr>
<td>Medium</td>
<td>4</td>
<td>18.25</td>
<td>0.5335937</td>
<td>8.333334E-02</td>
</tr>
</tbody>
</table>

We will now calculate the power of the $F$ test. Note that factor $B$ in this printout becomes factor $A$ on the PASS template.

Setup

To analyze these data, we enter the means for factor $A$. The value of $\sigma$ is estimated as the square root of the mean square error:

$$\sigma = \sqrt{1.138889} = 1.0672$$

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Randomized Block Analysis of Variance procedure window by expanding Means, then ANOVA, then clicking on Analysis of Variance, and then clicking on Randomized Block Analysis of Variance. You may then make the appropriate entries as listed below, or open Example 1 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Number of Blocks</td>
<td>2 3 4 5</td>
</tr>
<tr>
<td>Factor (A)</td>
<td>Checked</td>
</tr>
<tr>
<td>Factors (B, AB)</td>
<td>Not checked</td>
</tr>
<tr>
<td>Categories (A)</td>
<td>3</td>
</tr>
<tr>
<td>Hypothesized Means</td>
<td>15.75 18.25 20.50</td>
</tr>
<tr>
<td>Alpha</td>
<td>All are set to 0.05</td>
</tr>
<tr>
<td>S (Standard Deviation)</td>
<td>1.0672</td>
</tr>
</tbody>
</table>
Annotated Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Term</th>
<th>Power</th>
<th>Blocks</th>
<th>Units</th>
<th>df1</th>
<th>df2</th>
<th>Std Dev of Means (Sm)</th>
<th>Effect Size</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>0.42132</td>
<td>2</td>
<td>6</td>
<td>2</td>
<td>2</td>
<td>1.940</td>
<td>1.8179</td>
<td>0.0500</td>
<td>0.57868</td>
</tr>
<tr>
<td>A</td>
<td>0.89376</td>
<td>3</td>
<td>9</td>
<td>2</td>
<td>4</td>
<td>1.940</td>
<td>1.8179</td>
<td>0.0500</td>
<td>0.10624</td>
</tr>
<tr>
<td>A</td>
<td>0.99144</td>
<td>4</td>
<td>12</td>
<td>2</td>
<td>6</td>
<td>1.940</td>
<td>1.8179</td>
<td>0.05000</td>
<td>0.00856</td>
</tr>
<tr>
<td>A</td>
<td>0.99956</td>
<td>5</td>
<td>15</td>
<td>2</td>
<td>8</td>
<td>1.940</td>
<td>1.8179</td>
<td>0.05000</td>
<td>0.00044</td>
</tr>
</tbody>
</table>

Standard Deviation Within Blocks (block-treatment interaction) = 1.067

Summary Statements

A randomized-block design with one treatment factor at 3 levels has 2.0 blocks each with 3.0 treatment combinations. The square root of the block-treatment interaction is 1.067. This design achieves 42% power when an F test is used to test factor A at a 5% significance level and the actual standard deviation among the appropriate means is 1.940 (an effect size of 1.8179).

This report shows the power for each of the five block counts. We see that adequate power of about 0.9 would have been achieved by three blocks.

It is important to emphasize that these power values are for the case when the effects associated with the alternative hypotheses are equal to those given by the data. It will often be informative to calculate the power for other values as well.

Term

This is the term (main effect or interaction) from the analysis of variance model being displayed on this line.

Power

This is the power of the F test for this term. Note that since adding and removing terms changes the denominator degrees of freedom (df2), the power depends on which other terms are included in the model.

Blocks

This is the number of blocks in the design.

Units

This is the number of subblocks (plots) in the design. It is the product of the number of treatment levels and the number of blocks.

df1

This is the numerator degrees of freedom of the F test.

df2

This is the denominator degrees of freedom of the F test. This value depends on which terms are included in the AOV model.
**Std Dev of Means (Sm)**
This is the standard deviation of the means (or effects). It represents the size of the differences among the effects that is to be detected by the analysis. If you have entered hypothesized means, only their standard deviation is displayed here.

**Effect Size**
This is the standard deviation of the means divided by the standard deviation of subjects. It provides an index of the magnitude of the difference among the means that can be detected by this design.

**Alpha**
This is the significance level of the \( F \) test. This is the probability of a type-I error given the null hypothesis of equal means and zero effects.

**Beta**
This is the probability of the type-II error for this test given the sample size, significance level, and effect size.

---

**Example 2 – Validation using Prihoda**

Prihoda (1983) presents details of an example that is given in Odeh and Fox (1991). In this example, \( \text{Alpha} \) is 0.025, \( Sm \) of \( A \) is 0.577, the number of treatments in factor \( A \) is 6, the number of treatments in factor \( B \) is 3, \( S \) is 1.0, and the Number of Blocks is 2, 3, 4, 5, 6, 7, and 8. Prihoda gives the power values for the \( F \) test on factor \( A \) as 0.477, 0.797, 0.935, 0.982, 0.995, 0.999, and 1.000.

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Randomized Block Analysis of Variance procedure window by expanding Means, then ANOVA, then clicking on Analysis of Variance, and then clicking on Randomized Block Analysis of Variance. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Tab</strong></td>
<td></td>
</tr>
<tr>
<td>Number of Blocks</td>
<td>2 3 4 5 6 7 8</td>
</tr>
<tr>
<td>Terms (A, B, AB)</td>
<td>Checked</td>
</tr>
<tr>
<td>Categories (A)</td>
<td>6</td>
</tr>
<tr>
<td>Categories (B)</td>
<td>3</td>
</tr>
<tr>
<td>Hypothesized Means (A)</td>
<td>S 0.577</td>
</tr>
<tr>
<td>Hypothesized Means (B)</td>
<td>S 1</td>
</tr>
<tr>
<td>Hypothesized Effects (AB)</td>
<td>S 1</td>
</tr>
<tr>
<td>Alpha</td>
<td>All are set to 0.025</td>
</tr>
<tr>
<td>S (Standard Deviation)</td>
<td>1.0</td>
</tr>
</tbody>
</table>
Output

Click the Run button to perform the calculations and generate the following output.

**Numeric Results**

<table>
<thead>
<tr>
<th>Term</th>
<th>Power</th>
<th>Blocks</th>
<th>Units</th>
<th>df1</th>
<th>df2</th>
<th>Std Dev of Means (Sm)</th>
<th>Effect Size</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>A 0.47622</td>
<td>2</td>
<td>36</td>
<td>5</td>
<td>17</td>
<td>17</td>
<td>0.577</td>
<td>0.577</td>
<td>0.02500</td>
<td>0.52378</td>
</tr>
<tr>
<td>B 0.99997</td>
<td>2</td>
<td>36</td>
<td>2</td>
<td>17</td>
<td>17</td>
<td>1.000</td>
<td>1.000</td>
<td>0.02500</td>
<td>0.00303</td>
</tr>
<tr>
<td>AB 0.85337</td>
<td>2</td>
<td>36</td>
<td>10</td>
<td>17</td>
<td>17</td>
<td>1.000</td>
<td>1.000</td>
<td>0.02500</td>
<td>0.14663</td>
</tr>
<tr>
<td>A 0.79521</td>
<td>3</td>
<td>54</td>
<td>5</td>
<td>34</td>
<td>17</td>
<td>0.577</td>
<td>0.577</td>
<td>0.02500</td>
<td>0.20479</td>
</tr>
<tr>
<td>B 0.99999</td>
<td>3</td>
<td>54</td>
<td>2</td>
<td>34</td>
<td>17</td>
<td>1.000</td>
<td>1.000</td>
<td>0.02500</td>
<td>0.00001</td>
</tr>
<tr>
<td>AB 0.99615</td>
<td>3</td>
<td>54</td>
<td>10</td>
<td>34</td>
<td>17</td>
<td>1.000</td>
<td>1.000</td>
<td>0.02500</td>
<td>0.00385</td>
</tr>
<tr>
<td>A 0.93479</td>
<td>4</td>
<td>72</td>
<td>5</td>
<td>51</td>
<td>17</td>
<td>0.577</td>
<td>0.577</td>
<td>0.02500</td>
<td>0.06521</td>
</tr>
<tr>
<td>B 1.00000</td>
<td>4</td>
<td>72</td>
<td>2</td>
<td>51</td>
<td>17</td>
<td>1.000</td>
<td>1.000</td>
<td>0.02500</td>
<td>0.00000</td>
</tr>
<tr>
<td>AB 0.99995</td>
<td>4</td>
<td>72</td>
<td>10</td>
<td>51</td>
<td>17</td>
<td>1.000</td>
<td>1.000</td>
<td>0.02500</td>
<td>0.00005</td>
</tr>
<tr>
<td>A 0.98226</td>
<td>5</td>
<td>90</td>
<td>5</td>
<td>68</td>
<td>17</td>
<td>0.577</td>
<td>0.577</td>
<td>0.02500</td>
<td>0.01774</td>
</tr>
<tr>
<td>B 1.00000</td>
<td>5</td>
<td>90</td>
<td>2</td>
<td>68</td>
<td>17</td>
<td>1.000</td>
<td>1.000</td>
<td>0.02500</td>
<td>0.00000</td>
</tr>
<tr>
<td>AB 1.00000</td>
<td>5</td>
<td>90</td>
<td>10</td>
<td>68</td>
<td>17</td>
<td>1.000</td>
<td>1.000</td>
<td>0.02500</td>
<td>0.00000</td>
</tr>
<tr>
<td>A 0.99573</td>
<td>6</td>
<td>108</td>
<td>5</td>
<td>85</td>
<td>17</td>
<td>0.577</td>
<td>0.577</td>
<td>0.02500</td>
<td>0.00427</td>
</tr>
<tr>
<td>B 1.00000</td>
<td>6</td>
<td>108</td>
<td>2</td>
<td>85</td>
<td>17</td>
<td>1.000</td>
<td>1.000</td>
<td>0.02500</td>
<td>0.00000</td>
</tr>
<tr>
<td>AB 1.00000</td>
<td>6</td>
<td>108</td>
<td>10</td>
<td>85</td>
<td>17</td>
<td>1.000</td>
<td>1.000</td>
<td>0.02500</td>
<td>0.00000</td>
</tr>
<tr>
<td>A 0.99907</td>
<td>7</td>
<td>126</td>
<td>5</td>
<td>102</td>
<td>17</td>
<td>0.577</td>
<td>0.577</td>
<td>0.02500</td>
<td>0.00093</td>
</tr>
<tr>
<td>B 1.00000</td>
<td>7</td>
<td>126</td>
<td>2</td>
<td>102</td>
<td>17</td>
<td>1.000</td>
<td>1.000</td>
<td>0.02500</td>
<td>0.00000</td>
</tr>
<tr>
<td>AB 1.00000</td>
<td>7</td>
<td>126</td>
<td>10</td>
<td>102</td>
<td>17</td>
<td>1.000</td>
<td>1.000</td>
<td>0.02500</td>
<td>0.00000</td>
</tr>
<tr>
<td>A 0.99981</td>
<td>8</td>
<td>144</td>
<td>5</td>
<td>119</td>
<td>17</td>
<td>0.577</td>
<td>0.577</td>
<td>0.02500</td>
<td>0.00019</td>
</tr>
<tr>
<td>B 1.00000</td>
<td>8</td>
<td>144</td>
<td>2</td>
<td>119</td>
<td>17</td>
<td>1.000</td>
<td>1.000</td>
<td>0.02500</td>
<td>0.00000</td>
</tr>
<tr>
<td>AB 1.00000</td>
<td>8</td>
<td>144</td>
<td>10</td>
<td>119</td>
<td>17</td>
<td>1.000</td>
<td>1.000</td>
<td>0.02500</td>
<td>0.00000</td>
</tr>
</tbody>
</table>

Standard Deviation Within Blocks (block-treatment interaction) = 1.000

We have bolded the power values on this report that should match Prihoda’s results. You see that they do match.
Chapter 570

Repeated Measures Analysis of Variance

Introduction

This module calculates the power for repeated measures designs having up to three within factors and up to three between factors. It computes power for various test statistics including the $F$ test with the Geisser-Greenhouse correction, Wilks’ lambda, Pillai-Bartlett trace, and Hotelling-Lawley trace. It can be used to calculate the power of crossover designs.

Repeated measures designs are popular because they allow a subject to serve as their own control. This usually improves the precision of the experiment. However, when the analysis of the data uses the traditional $F$ tests, additional assumptions concerning the structure of the error variance must be made. When these assumptions do not hold, the Geisser-Greenhouse correction provides reasonable adjustments so that significance levels are accurate.

An alternative to using the $F$ test with repeated measures designs is to use one of the multivariate tests: Wilks’ lambda, Pillai-Bartlett trace, or Hotelling-Lawley trace. These alternatives are appealing because they do not make the strict, often unrealistic, assumptions about the structure of the error variance. Unfortunately, they may have less power than the $F$ test and they cannot be used in all situations.

An example of a two-factor repeated measures design that can be analyzed by this procedure is shown by the following diagram.

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Subject 1</th>
<th>Subject 2</th>
<th>Month</th>
<th>Group 2</th>
<th>Subject 3</th>
<th>Subject 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Treatment L</td>
<td>Treatment L</td>
<td>1</td>
<td>Treatment L</td>
<td>Treatment L</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Treatment M</td>
<td>Treatment M</td>
<td>2</td>
<td>Treatment M</td>
<td>Treatment M</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Treatment H</td>
<td>Treatment H</td>
<td>3</td>
<td>Treatment H</td>
<td>Treatment H</td>
<td></td>
</tr>
</tbody>
</table>

Groups 1 and 2 form the between factor. The within factor has three levels: $L$, $M$, and $H$ (low, medium, and high). There are four subjects in this experiment. The three treatments are applied to
each subject, one treatment per month. Note that the three treatments are applied to each subject in the same order. Although the order of treatment application should be randomized, it is often the same for all subjects.

This diagram shows the main features of a repeated measures design, which are

1. Each subject receives all treatments.
2. The treatments are applied through time. When the treatments are applied in the same order across all subjects, it is impossible to separate the treatment effects from the sequence effects. Some processes that can cause sequence effects are learning, practice, or fatigue—any pattern in the responses across time that occurs without the treatment. If you think the possibility for sequence effects exists, you must make sure that the effects of prior treatments have been washed out before applying the next treatment.
3. Unlike other designs, the repeated measures design has two experimental units: between and within. In this example, the first (between) experimental unit is a subject. Subject-to-subject variability is used to test the between factor (groups). The second (within) experimental unit is the time period. In the above example, the month-to-month variability within a subject is used to test the treatment. The important point to realize is that the repeated measures design has two error components, the between and the within.

Assumptions

The following assumptions are made when using the $F$ test to analyze a factorial experimental design.

1. The response variable is continuous.
2. The residuals follow the normal probability distribution with mean equal to zero and constant variance.
3. The subjects are independent.

Since in a within-subject design responses coming from the same subject are not independent, assumption 3 must be modified for responses within a subject. Independence between subjects is still assumed.

4. The within-subject covariance matrices are equal for all between-subject groups. In this type of experiment, the repeated measurements on a subject may be thought of as a multivariate response vector having a certain covariance structure. This assumption states that these covariance matrices are constant from group to group.

5. When using an $F$ test, the within-subject covariance matrices are assumed to be circular. One way of defining circularity is that the variances of differences between any two measurements within a subject are constant for all measurements. Since responses that are close together in time often have a higher correlation than those that are far apart, it is common for this assumption to be violated. This assumption is not necessary for the validity of the three multivariate tests: Wilks’ lambda, Pillai-Bartlett trace, or Hotelling-Lawley trace.
Advantages of Within-Subjects Designs

Because the response to stimuli usually varies less within an individual than between individuals, the within-subject variability is usually less than (or at most equal to) the between-subject variability. By reducing the underlying variability, the same power can be achieved with a smaller number of subjects.

Disadvantages of Within-Subjects Designs

1. **Practice effect.** In some experiments, subjects systematically improve as they practice the task being studied. In other cases, subjects may systematically get worse as they get fatigued or bored with the experimental task. Note that only the treatment administered first is immune to practice effects. Hence, experimenters should make an effort to balance the number of subjects receiving each treatment first.

2. **Carryover effect.** In many drug studies, it is important to wash out the influence of one drug completely before the next drug is administered. Otherwise, the influence of the first drug carries over into the response to the second drug.

3. **Statistical analysis.** The statistical model is more restrictive than in a regular factorial design since the individual responses must have certain mathematical properties.

Even in the face of all these disadvantages, repeated measures (within-subject) designs are popular in many areas of research. It is important that you recognize these problems going in so you can make sure that the design is appropriate, rather than learning of them later after the research has been conducted.

Technical Details

**General Linear Multivariate Model**

This section provides the technical details of the repeated measures designs that can be analyzed by *PASS*. The approximate power calculations outlined in Muller, LaVange, Ramey, and Ramey (1992) are used. Using their notation, for *N* subjects, the usual general linear multivariate model is

$$ Y = XM + R $$

where each row of the residual matrix *R* is distributed as a multivariate normal

$$ \text{row}_i(R) \sim N_p(0, \Sigma) $$

Note that *p* is the product of the number of levels of each of the within-subject factors, *q* is the number of design variables, *Y* is the matrix of responses, *X* is the design matrix, *M* is the matrix of regression parameters (means), and *R* is the matrix of residuals.

Hypotheses about various sets of regression parameters are tested using

$$ H_0: \Theta_{a \times b} = \Theta_0 $$

$$ CMD_{a \times q \times p \times b} = \Theta $$
where $C$ and $D$ are orthonormal contrast matrices and $\Theta_0$ is a matrix of hypothesized values, usually zeros. Note that $C$ defines contrasts among the between-subject factor levels and $D$ defines contrast among the within-subject factor levels.

Tests of the various main effects and interactions may be constructed with suitable choices for $C$ and $D$. These tests are based on

$$
\hat{M} = (X'X)^{-1}X'Y
$$

$$
\hat{\Theta} = C\hat{M}D
$$

$$
H = \left(\hat{\Theta} - \Theta_0\right)' \left[C(X'X)^{-1}C\right]^{-1} \left(\hat{\Theta} - \Theta_0\right)
$$

$$
E = D\hat{\Sigma}D \cdot (N-r)
$$

$$
T = H + E
$$

where $r$ is the rank of $X$.

**Geisser-Greenhouse F Test**

Upon the assumption that $\Sigma$ has compound symmetry, a size $\alpha$ test of $H_0: \Theta = \Theta_0$ is given by the $F$ ratio

$$
F = \frac{\text{tr}(H)/ab}{\text{tr}(E)/(b(N-r))}
$$

with degrees of freedom given by

$$
df_1 = ab \\
df_2 = b(N-r)
$$

and noncentrality parameter

$$
\lambda = df_1(F)
$$

The assumption that $\Sigma$ has compound symmetry is usually not viable. Box (1954a,b) suggested that adjusting the degrees of freedom of the above $F$-ratio could compensate for the lack of compound symmetry in $\Sigma$. His adjustment has become known as the Geisser-Greenhouse adjustment. Under this adjustment, the modified degrees of freedom and noncentrality parameter are given by

$$
df_1 = ab\varepsilon \\
df_2 = b(N-r)\varepsilon \\
\lambda = df_1(F)\varepsilon
$$

where

$$
\varepsilon = \frac{\text{tr}(D'\hat{\Sigma}D)^2}{b \text{tr}(D'\hat{\Sigma}DD'\hat{\Sigma}D)}
$$
The range of $\varepsilon$ is $\frac{1}{b-1}$ to 1. When $\varepsilon = 1$, the matrix is spherical. When $\varepsilon = \frac{1}{b-1}$, the matrix differs maximally from sphericity.

Note that the Geisser-Greenhouse adjustment is only needed for testing main effects and interactions involving within-subject factors. Main effects and interactions that involve only between-subject factors need no such adjustment.

The critical value $F_{\text{Crit}}$ is computed using the expected value of $\varepsilon$ to adjust the degrees of freedom. That is, the degrees of freedom of $F_{\text{Crit}}$ are given by

$$
\text{df}1 = abE(\varepsilon) \\
\text{df}2 = b(N - r)E(\varepsilon)
$$

where

$$
E(\hat{\varepsilon}) = \begin{cases} 
\varepsilon + \frac{g_1}{N - r} & \text{if } \varepsilon > \frac{g_1}{N - r} \\
\varepsilon / 2 & \text{otherwise}
\end{cases}
$$

$$
g_i = \sum_{i=1}^{T} f_i \xi_i^2 + \sum_{i \neq j} \frac{f_i \xi_i \xi_j}{(\xi_i - \xi_j)^2}
$$

$$
f_i = \frac{\partial \varepsilon}{\partial \xi_i}
$$

$$
= \frac{2\sum \xi_j}{df_i \sum \xi_j^2} - \frac{2\lambda_i (\sum \xi_j)^2}{df_i (\sum \xi_j^2)^2}
$$

$$
f_{ii} = \frac{\partial^2 \varepsilon}{\partial \xi_i^2}
$$

$$
= 2h_i - 8h_2 + 8h_3 - 2h_4
$$

$$
h_i = \frac{2}{df_i \sum \xi_j^2}
$$

$$
h_2 = \frac{\xi_i (\sum \xi_j)}{df_i (\sum \xi_j^2)^2}
$$

$$
h_3 = \frac{\xi_i^2 (\sum \xi_j)^2}{df_i (\sum \xi_j^3)^2}
$$

$$
h_4 = \frac{(\sum \xi_j)^2}{df_i (\sum \xi_j^3)^2}
$$
where the $\xi_j$’s are the ordered eigenvalues of $D'\Sigma D$.

### Wilks’ Lambda Approximate F Test

The hypothesis $H_0: \Theta = \Theta_0$ may be tested using Wilks’ likelihood ratio statistic $W$. This statistic is computed using

$$W = |ET^{-1}|$$

An $F$ approximation to the distribution of $W$ is given by

$$F_{df_1, df_2} = \frac{\eta / df_1}{(1 - \eta) / df_2}$$

where

$$\lambda = df_1F_{df_1, df_2}$$
$$\eta = 1 - W^{1/g}$$
$$df_1 = ab$$
$$df_2 = g[(N - r) - (b - a + 1) / 2 - (ab - 2) / 2$$
$$g = \left(\frac{a^2b^2 - 4}{a^2 + b^2 - 5}\right)^{1/2}$$

### Pillai-Bartlett Trace Approximate F Test

The hypothesis $H_0: \Theta = \Theta_0$ may be tested using the Pillai-Bartlett Trace. This statistic is computed using

$$T_{pb} = tr\left(HT^{-1}\right)$$

A non-central $F$ approximation to the distribution of $T_{pb}$ is given by

$$F_{df_1, df_2} = \frac{\eta / df_1}{(1 - \eta) / df_2}$$

where

$$\lambda = df_1F_{df_1, df_2}$$
$$\eta = \frac{T_{pb}}{s}$$
$$s = \min(a, b)$$
$$df_1 = ab$$
$$df_2 = s\left[(N - r) - b + s\right]$$
Hotelling-Lawley Trace Approximate F Test

The hypothesis $H_0: \Theta = \Theta_0$ may be tested using the Hotelling-Lawley Trace. This statistic is computed using

$$T_{HL} = \text{tr} \left( HE^{-1} \right)$$

An $F$ approximation to the distribution of $T_{HL}$ is given by

$$F_{df_1, df_2} = \frac{\eta / df_1}{(1 - \eta) / df_2}$$

where

$$\lambda = df_1 F_{df_1, df_2}$$

$$\eta = \frac{T_{HL}}{s + \frac{T_{HL}}{s}}$$

$$s = \min(a, b)$$

$$df'1 = ab$$

$$df'2 = s \left( (N - r) - b + s \right)$$

The M (Mean) Matrix

In the general linear multivariate model presented above, $M$ represents a matrix of regression coefficients. Since you must provide the elements of $M$, we will discuss its meaning in more detail. Although other structures and interpretations of $M$ are possible, in this module we assume that the elements of $M$ are the cell means. The rows of $M$ represent the between-subject categories and the columns of $M$ represent the within-group categories.

The $q$ rows of $M$ represent the $q$ groups into which the subjects can be classified. For example, if a design includes three between-subject factors with 2, 3, and 4 categories, the matrix $M$ would have $2 \times 3 \times 4 = 24$ rows. That is, $q = 24$. Similarly, if a design has three within-subject factors with 3, 3, and 3 categories, the matrix $M$ would have $3 \times 3 \times 3 = 27$ columns. That is, $p = 27$.

Consider now an example in which $q = 3$ and $p = 4$. That is, there are three groups into which subjects can be placed. Each subject is measured four times. The matrix $M$ would appear as follows.

$$M = \begin{bmatrix}
\mu_{11} & \mu_{12} & \mu_{13} & \mu_{14} \\
\mu_{21} & \mu_{22} & \mu_{23} & \mu_{24} \\
\mu_{31} & \mu_{32} & \mu_{33} & \mu_{34}
\end{bmatrix}$$

For example, the element $\mu_{12}$ is the mean of the second measurement of subjects in the first group. To calculate the power of this design, you would need to specify appropriate values of all twelve means.
As a second example, consider a design with three between-subject factors and three within-subject factors, all of which have two categories. The $M$ matrix for this design would be as follows.

$$
M = \begin{bmatrix}
W1 & W2 & W3 \\
B1 & B2 & B3 \\
1 & 1 & 1 & \mu_{11111} & \mu_{11112} & \mu_{11113} & \mu_{11121} & \mu_{11122} & \mu_{11123} & \mu_{11131} & \mu_{11132} & \mu_{11133} \\
1 & 1 & 2 & \mu_{12111} & \mu_{12112} & \mu_{12113} & \mu_{12121} & \mu_{12122} & \mu_{12123} & \mu_{12131} & \mu_{12132} & \mu_{12133} \\
1 & 2 & 1 & \mu_{21111} & \mu_{21112} & \mu_{21113} & \mu_{21121} & \mu_{21122} & \mu_{21123} & \mu_{21131} & \mu_{21132} & \mu_{21133} \\
1 & 2 & 2 & \mu_{22111} & \mu_{22112} & \mu_{22113} & \mu_{22121} & \mu_{22122} & \mu_{22123} & \mu_{22131} & \mu_{22132} & \mu_{22133} \\
2 & 1 & 1 & \mu_{21211} & \mu_{21212} & \mu_{21213} & \mu_{21221} & \mu_{21222} & \mu_{21223} & \mu_{21231} & \mu_{21232} & \mu_{21233} \\
2 & 1 & 2 & \mu_{22111} & \mu_{22112} & \mu_{22113} & \mu_{22121} & \mu_{22122} & \mu_{22123} & \mu_{22131} & \mu_{22132} & \mu_{22133} \\
2 & 2 & 1 & \mu_{22211} & \mu_{22212} & \mu_{22213} & \mu_{22221} & \mu_{22222} & \mu_{22223} & \mu_{22231} & \mu_{22232} & \mu_{22233} \\
2 & 2 & 2 & \mu_{22211} & \mu_{22212} & \mu_{22213} & \mu_{22221} & \mu_{22222} & \mu_{22223} & \mu_{22231} & \mu_{22232} & \mu_{22233}
\end{bmatrix}
$$

The subscripts for each mean follow the pattern $\mu_{B1B2B3W1W2W3}$. The first three subscripts indicate the between-subject categories and the second three subscripts indicate the within-subject categories. Notice that the first three subscripts are constant in each row and the second three subscripts are constant in each column.

### Specifying the M Matrix

When computing the power in a repeated measures analysis of variance, the specification of the $M$ matrix is one of your main tasks. The program cannot do this for you. The calculated power is directly related to your choice. So your choice for the elements of $M$ must be selected carefully and thoughtfully. When authorization and approval from a government organization is sought, you should be prepared to defend your choice of $M$. In this section, we will explain how you can specify $M$.

Before we begin, it is important that you have in mind exactly what $M$ is. $M$ is a table of means that represent the size of the differences among the means that you want the study or experiment to detect. That is, $M$ gives the means under the alternative hypothesis. Under the null hypothesis, these means are assumed to be equal. Because of the complexity of the repeated measures design, it is often difficult to choose reasonable values, so PASS will help you. But it is important to remember that you are responsible for these values and that the sample sizes calculated are based on them.

One way to specify the $M$ matrix is to do so directly into the spreadsheet. You might do this if you are calculating the ‘retrospective’ power of a study that has already been completed, or if it is simply easier to write the matrix directly. Usually, however, you will specify the $M$ matrix in portions.

We will begin our discussion of specifying the $M$ matrix with an example. Consider a study of two groups of subjects. Each subject was tested, then a treatment was administered, then the subject was tested again at the ten minute mark, and then tested a third time after sixty minutes. The researchers wanted the sample size to be large enough to detect the following pattern in the means.
To understand how they derived this table, we will perform some basic arithmetic on it.

### Step 1 – Remove the Overall Mean
Subtract 125, the overall mean, from each of the individual means.

#### Table of Hypothesized Means Adjusted for Overall Mean

<table>
<thead>
<tr>
<th>Group</th>
<th>T0</th>
<th>T10</th>
<th>T60</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-25</td>
<td>5</td>
<td>-25</td>
<td>-15</td>
</tr>
<tr>
<td>B</td>
<td>-5</td>
<td>55</td>
<td>-5</td>
<td>15</td>
</tr>
<tr>
<td>Average</td>
<td>-15</td>
<td>30</td>
<td>-15</td>
<td>125</td>
</tr>
</tbody>
</table>

### Step 2 – Remove the Group Effect
Subtract -15 from the first row and 15 from the second row.

#### Table of Hypothesized Means Adjusted for Group

<table>
<thead>
<tr>
<th>Group</th>
<th>T0</th>
<th>T10</th>
<th>T60</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-10</td>
<td>20</td>
<td>-10</td>
<td>-15</td>
</tr>
<tr>
<td>B</td>
<td>-20</td>
<td>40</td>
<td>-20</td>
<td>15</td>
</tr>
<tr>
<td>Total</td>
<td>-15</td>
<td>30</td>
<td>-15</td>
<td>125</td>
</tr>
</tbody>
</table>
Step 3 – Remove the Time Effect
Subtract -15 from the first column, 30 from the second column, and -15 from the third column.

<table>
<thead>
<tr>
<th>Group</th>
<th>Time Period</th>
<th>Effect</th>
<th>Effect + Overall</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>T0</td>
<td>T10</td>
<td>T60</td>
</tr>
<tr>
<td>A</td>
<td>5</td>
<td>-10</td>
<td>5</td>
</tr>
<tr>
<td>B</td>
<td>-5</td>
<td>10</td>
<td>-5</td>
</tr>
<tr>
<td>Effect</td>
<td>-15</td>
<td>30</td>
<td>-15</td>
</tr>
<tr>
<td>Effect + Overall</td>
<td>110</td>
<td>155</td>
<td>110</td>
</tr>
</tbody>
</table>

This table, called an effects table, lets us see the individual effect of each component of the model. For example, we can see that the hypothesized pattern across time is that T10 is 45 units higher than either endpoint. Similarly, we note that the hypothesized pattern for the two groups is that Group B is 30 units larger than Group A.

Understanding the interaction is more difficult. One interpretation focuses on T10. We note that in Group A the response for T10 is 10 less than expected while in Group B the response for T10 is 10 more than expected.

**Entering This Information into PASS**
Rather than enter the individual values into PASS, you can enter the group, time, and interaction effects directly. For this example, you could enter ‘110 140’ or ‘-15 15’ for the hypothesized means of the between factor and ‘110 155 110’ or ‘-15 30 -15’ for the hypothesized means of the within factor. For the interaction, you would enter the six interaction values ‘5 -10 5 -5 10 -5’.

Another way to enter the interaction information would be to indicate that the size of the interaction to be detected is about half that of the group factor or about a third of the time factor. For a complete discussion of the interpretation of various interactions, we suggest that you look at Kirk (1982).

**The C Matrix for Between-Subject Contrasts**
The C matrix is comprised of contrasts that are applied to the rows of $M$. That is, these are between-group contrasts. You do not have to specify these contrasts. They are generated for you. You should understand that a different $C$ matrix is generated for each between-subject term in the model. For example, in the six factor example above, the $C$ matrix that will be generated for testing the between-subject factor B1 is

$$C_{B1} = \begin{bmatrix} -\frac{1}{\sqrt{8}} & -\frac{1}{\sqrt{8}} & -\frac{1}{\sqrt{8}} & -\frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} & \frac{1}{\sqrt{8}} \end{bmatrix}$$
Note that the divisor $\sqrt{8}$ is used so that the total of the squared elements is one. This is required so that the contrast matrix is *orthonormal*.

When creating a test for B1, the matrix $D$ is created to average across all within-subject categories.

$$D_{B1} = \begin{bmatrix} 1 \\ \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} \\ \frac{1}{\sqrt{8}} \end{bmatrix}$$

**Generating the C Matrix when There are Multiple Between Factors**

Generating the $C$ matrix when there is more than one between factor is more difficult. We like the method of O’Brien and Kaiser (1985) which we briefly summarize here.

**Step 1.** Write a complete set of contrasts suitable for testing each factor separately. For example, if you have three factors with 2, 3, and 4 categories, you might use

$$\tilde{C}_{B1} = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \tilde{C}_{B2} = \begin{bmatrix} -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ 0 \end{bmatrix}, \text{ and } \tilde{C}_{B3} = \begin{bmatrix} -\frac{3}{\sqrt{12}} \\ \frac{1}{\sqrt{12}} \\ \frac{1}{\sqrt{12}} \end{bmatrix}$$

**Step 2.** Define appropriate $J_k$ matrices corresponding to each factor. These matrices comprised of one row and $k$ columns whose equal element is chosen so that the sum of its elements squared is one. In this example, we use

$$J_2 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}, J_3 = \begin{bmatrix} \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{bmatrix}, J_4 = \begin{bmatrix} \frac{1}{\sqrt{4}} \\ \frac{1}{\sqrt{4}} \\ \frac{1}{\sqrt{4}} \end{bmatrix}$$

**Step 3.** Create the appropriate contrast matrix using a direct (Kronecker) product of either the $\tilde{C}_{B_i}$ matrix if the factor is included in the term or the $J_i$ matrix when the factor is not in the term. Remember that the direct product is formed by multiplying each element of the second matrix by all members of the first matrix. Here is an example
As an example, we will compute the $C$ matrix suitable for testing factor $B2$

$$C_{B2} = J_2 \otimes C_{B2} \otimes J_4$$

Expanding the direct product results in

$$C_{xy} = J_4 \otimes C_{B2} \otimes J_4$$

Similarly, the $C$ matrix suitable for testing interaction $B2B3$ is

$$C_{B2B3} = J_2 \otimes C_{B2} \otimes C_{B3}$$

We leave the expansion of this matrix $PASS$, but we think you have the idea.

The D Matrix for Within-Subject Contrasts

The $D$ matrix is comprised of contrasts that are applied to the columns of $M$. That is, these are within-group contrasts. You do not have to specify these contrasts either. They will be generated for you. Specification of the $D$ matrix is similar to the specification of the $C$ matrix, except that now the matrices are all transposed.

Interactions of Between-Subject and Within-Subject Factors

Interactions that include both between-subject factors and within-subject factors require that between-subject portion be specified by the $C$ matrix and the within-subject portion be specified with the $D$ matrix.
Power Calculations

To calculate statistical power, we must determine distribution of the test statistic under the alternative hypothesis which specifies a different value for the regression parameter matrix $B$. The distribution theory in this case has not been worked out, so approximations must be used. We use the approximations given by Mueller and Barton (1989) and Muller, LaVange, Ramey, and Ramey (1992). These approximations state that under the alternative hypothesis, $F_U$ is distributed as a noncentral $F$ random variable with degrees of freedom and noncentrality shown above. The calculation of the power of a particular test may be summarized as follows

1. Specify values of $X, M, \Sigma, C, D,$ and $\Theta_0$.
2. Determine the critical value using $F_{crit} = FINV(1 - \alpha, df_1, df_2)$, where $FINV(\cdot)$ is the inverse of the central $F$ distribution and $\alpha$ is the significance level.
3. Compute the noncentrality parameter $\lambda$.
4. Compute the power as

$$\text{Power} = 1 - \text{NCFPROB}(F_{crit}, df_1, df_2, \lambda)$$

where NCFPROB( ) is the noncentral $F$ distribution.

Covariance Matrix Assumptions

The following assumptions are made when using the $F$ test. These assumptions are not needed when using one of the three multivariate tests.

In order to use the $F$ ratio to test hypotheses, certain assumptions are made about the distribution of the residuals $e_{ijk}$. Specifically, it is assumed that the residuals for each subject, $e_{i1}, e_{i2}, \ldots, e_{iT}$, are distributed as a multivariate normal with means equal to zero and covariance matrix $\Sigma_i$. Two additional assumptions are made about these covariance matrices. First, they are assumed to be equal for all subjects. That is, it is assumed that $\Sigma_1 = \Sigma_2 = \cdots = \Sigma_{Gn} = \Sigma$.

Second, the covariance matrix is assumed to have a particular form called circularity. A covariance matrix is circular if there exists a matrix $A$ such that

$$\Sigma = A + A' + \lambda I_T$$

where $I_T$ is the identity matrix of order $T$ and $\lambda$ is a constant.

This property may also be defined as

$$\sigma_{ii} + \sigma_{jj} - 2\sigma_{ij} = 2\lambda$$

One type of matrix that is circular is one that has compound symmetry. A matrix with this property has all elements on the main diagonal equal and all elements off the main diagonal equal. An example of a covariance matrix with compound symmetry is
Repeated Measures Analysis of Variance

\[
\Sigma = \begin{bmatrix}
\sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\
\rho\sigma^2 & \sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\
\rho\sigma^2 & \rho\sigma^2 & \sigma^2 & \cdots & \rho\sigma^2 \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\rho\sigma^2 & \rho\sigma^2 & \rho\sigma^2 & \cdots & \sigma^2 
\end{bmatrix}
\]

or, with actual numbers,

\[
\begin{bmatrix}
9 & 2 & 2 & 2 \\
2 & 9 & 2 & 2 \\
2 & 2 & 9 & 2 \\
2 & 2 & 2 & 9
\end{bmatrix}
\]

An example of a matrix which does not have compound symmetry but is still circular is

\[
\begin{bmatrix}
1 & 1 & 1 & 1 \\
2 & 2 & 2 & 2 \\
3 & 3 & 3 & 3 \\
4 & 4 & 4 & 4
\end{bmatrix} + \begin{bmatrix}
1 & 2 & 3 & 4 \\
1 & 2 & 3 & 4 \\
1 & 2 & 3 & 4 \\
1 & 2 & 3 & 4
\end{bmatrix} + \begin{bmatrix}
2 & 0 & 0 & 0 \\
0 & 2 & 0 & 0 \\
0 & 0 & 2 & 0 \\
0 & 0 & 0 & 2
\end{bmatrix} = \begin{bmatrix}
4 & 3 & 4 & 5 \\
3 & 6 & 5 & 6 \\
4 & 5 & 8 & 7 \\
5 & 6 & 7 & 10
\end{bmatrix}
\]

Needless to say, the need to have the covariance matrix circular is a very restrictive assumption.

**Between-Subject Standard Deviation**

The subject-to-subject variability is represented by \( \sigma^2_{\text{Between}} \). In the repeated measures AOV table, this quantity is estimated by the between subjects mean square (MSB). This quantity is calculated from \( \Sigma \) using the formula

\[
\sigma^2_{\text{Between}} = \frac{\sum_{i=1}^{T} \sum_{j=1}^{T} \sigma_{ij}}{T} = \frac{\sum_{i=1}^{T} \sum_{j=1}^{T} \sigma_{ii} \sigma_{jj} \rho_{ij}}{T}
\]

When \( \Sigma \) has compound symmetry, which requires all \( \sigma_{ii} = \sigma \) and all \( \rho_{ij} = \rho \), the above formula reduces to

\[
\sigma^2_{\text{Between}} = \sigma^2 \left( 1 + (T - 1)\rho \right)
\]

Note that \( F \) tests of between factors and their interactions do not require the circularity assumption so the Geisser-Greenhouse correction is not applied to these tests.
Within-Subject Standard Deviation

The within-subject variability is represented by \( \sigma^2_{\text{Within}} \). In the repeated measures AOV table, this quantity is estimated by the within-subjects mean square (MSW). This quantity is calculated from \( \Sigma \) using the formula

\[
\sigma^2_{\text{Within}} = \frac{\sum_{i=1}^{T} \sigma_{ii}}{T} - \frac{2 \sum_{i=1}^{T} \sum_{j=i+1}^{T} \sigma_{ij}}{T(T-1)}
\]

When \( \Sigma \) has compound symmetry, which requires all \( \sigma_{ii} = \sigma \) and all \( \rho_{ij} = \rho \), the above formula reduces to

\[
\sigma^2_{\text{Within}} = \sigma^2 (1 - \rho)
\]

Estimating Sigma and Rho from Existing Data

Using the above results for existing data, approximate values for \( \sigma \) and \( \rho \) may be estimated from a previous analysis of variance table that provides estimates of MSB and MSW. Solving the above equations for \( \sigma \) and \( \rho \) yields

\[
\rho = \frac{\sigma^2_{\text{Between}} - \sigma^2_{\text{Within}}}{\sigma^2_{\text{Between}} + (T-1)\sigma^2_{\text{Within}}}
\]

\[
\sigma^2 = \frac{\sigma^2_{\text{Within}}}{1 - \rho}
\]

Substituting MSB for \( \sigma^2_{\text{Between}} \) and MSW for \( \sigma^2_{\text{Within}} \) yields the estimates

\[
\hat{\rho} = \frac{\text{MSB} - \text{MSW}}{\text{MSB} + (T-1)\text{MSW}}
\]

\[
\hat{\sigma}^2 = \frac{\text{MSW}}{1 - \hat{\rho}}
\]

Note that these estimators assume that the design meets the circularity assumption, which is usually not the case. However, they provide crude estimates that can be used in planning.
**Procedure Options**

This section describes the options that are unique to this procedure. To find out more about using the other tabs, go to the Procedure Window chapter.

**Data Tab**

The Data tab contains most of the parameters and options that you will be concerned with.

**Solve For**

**Find (Solve For)**

This option specifies the parameter to be solved for. If you choose to solve for \( n \) (sample size), you must also specify which test statistic you want to use.

When you choose to solve for \( n \), the program searches for the lowest sample size that meets the alpha and power criterion you have specified for each of the terms. If you do not want a term to be used in the search, set its levels to empty or 0.

Also, when the ‘= n’s’ box is not checked, the search is made using unequal group sample sizes. The relative proportion of the sample in each group is set by the values of \( n \) given in the Subjects Per Group box. For example, if your design has three groups and you enter ‘1 1 2’ in the Subjects Per Group box, the search will only consider designs in which the size of the last group is twice the rest. That is, it will consider ‘2 2 4’, ‘3 3 6’, ‘4 4 8’, etc.

Note: no plots are generated when you solve for \( n \).

**Sample Size**

**\( n \) (Subjects Per Group)**

Specify one or more values for the number of subjects per group. The total sample size is the sum of the individual group sizes across all groups.

You can specify a list like ‘2 4 6’. The items in the list may be separated with commas or blanks. The interpretation of the list depends on the =n's check box. When the =n's box is checked, a separate analysis is calculated for each value of \( n \). When the =n's box is not checked, **PASS** uses the \( n 's \) as the actual group sizes. In this case, the number of items entered must match the number of groups in the design.

When you choose to solve for \( n \) and the ‘= n’s’ box is not checked, the search is made using unequal group sample sizes. The relative proportion of the sample in each group is set by the values of \( n \) given in this box. For example, if your design has three groups and you enter ‘1 1 2’ here, the search will only consider designs in which the size of the last group is twice the rest. That is, it will consider ‘2 2 4’, ‘3 3 6’, ‘4 4 8’, etc.

**= n’s**

This option controls whether the number of subjects per group is to be equal for all groups or not. When checked, the number of subjects per group is equal for all groups. A list of values such as ‘5 10 15’ represents three designs: one with five per group, one with ten per group, and one with fifteen per group.
Repeated Measures Analysis of Variance  570-17

When this option is not checked, the n’s are assumed to be unequal. A list of values represents the size of the individual groups. For example, ‘5 10 15’ represents a single, three-group design with five in the first group, ten in the second group, and fifteen in the third group.

**Effect Size – Means**

**Means Matrix**

Use this option to specify spreadsheet columns containing a hypothesized means matrix that will be used to compute the Sm values. All individual Sm values are ignored. You can obtain the spreadsheet by selecting ‘Window’, then ‘Data’, from the menus.

The between factors are represented across the columns of the spreadsheet and the within factors are represented down the rows. The number of columns specified must equal the number of groups. The number of rows with data in these columns must equal the number of times. For example, suppose you are designing an experiment that is to have two between factors (A & B) and two within factors (D & E). Suppose each of the four factors has two levels. The columns of the spreadsheet would be


The rows of the spreadsheet would represent

D1E1
D1E2
D2E1
D2E2

**Example**

To see how this option works, consider the following table of hypothesized means for an experiment with one between factor (A) having two groups and one within factor (B) having three time periods. The values in columns C1 and C2 of the spreadsheet are

<table>
<thead>
<tr>
<th></th>
<th>C1</th>
<th>C2</th>
<th>Means</th>
<th>Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row1</td>
<td>0.5</td>
<td>-0.5</td>
<td>3.0</td>
<td>-2.5</td>
</tr>
<tr>
<td>Row2</td>
<td>0.5</td>
<td>-0.5</td>
<td>5.0</td>
<td>-0.5</td>
</tr>
<tr>
<td>Row3</td>
<td>-1.0</td>
<td>1.0</td>
<td>8.5</td>
<td>3.0</td>
</tr>
<tr>
<td>Means</td>
<td>4.0</td>
<td>7.0</td>
<td>5.5</td>
<td></td>
</tr>
<tr>
<td>Effects</td>
<td>-1.5</td>
<td>1.5</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
The standard deviation of the A effects is calculated as
\[ \sigma_A = \sqrt{\frac{(-1.5)^2 + (1.5)^2}{2}} \]
\[ = \sqrt{2.25} \]
\[ = 1.5 \]

The standard deviation of the B effects is calculated as
\[ \sigma_B = \sqrt{\frac{(-2.5)^2 + (-0.5)^2 + (3.0)^2}{3}} \]
\[ = \sqrt{\frac{15.5}{3}} \]
\[ = 2.27 \]

The standard deviation of the interaction effects is found to be
\[ \sigma_{AB} = \sqrt{\frac{(0.5)^2 + (0.5)^2 + (-1.0)^2 + (-0.5)^2 + (-0.5)^2 + (1.0)^2}{6}} \]
\[ = \sqrt{\frac{3.0}{6}} \]
\[ = 0.71 \]

These three standard deviations are used to represent the effect sizes of the corresponding terms.

**Discussion**

When using this option, it is less confusing to concentrate on a single term at a time. For example, consider a 2-by-4 design in which your primary interest is in testing the AB interaction. Instead of trying to determine a means matrix the will represent factor A, factor B, and the AB interaction, ignore factor A and factor B and just consider the interaction. You might want to consider the following pattern

<table>
<thead>
<tr>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.0</td>
<td>1.0</td>
</tr>
<tr>
<td>0.0</td>
<td>2.0</td>
</tr>
<tr>
<td>0.0</td>
<td>3.0</td>
</tr>
</tbody>
</table>

That is, the first group remains constant while the second group increases for 0.0 to 3.0.

By specifying various values for K (the means multiplier), you can study to impact of increasing the values. For example, when K is set to 2, the above means matrix becomes

<table>
<thead>
<tr>
<th>C1</th>
<th>C2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.0</td>
<td>0.0</td>
</tr>
<tr>
<td>0.0</td>
<td>2.0</td>
</tr>
<tr>
<td>0.0</td>
<td>4.0</td>
</tr>
<tr>
<td>0.0</td>
<td>6.0</td>
</tr>
</tbody>
</table>

Thus, by simply changing K, several scenarios may be studied. (We wish to thank Keith Muller for suggesting this method of specifying the Sm values.)
**K (Means Multipliers)**

These values are multiplied times the means matrix to give you various effect sizes. A separate power calculation is generated for each value of K. These values become the horizontal axis in the second power chart. For example, if an Sm value is 80, setting this option to ‘50 100 150’ would result in three Sm values: 40, 80, and 120. If you want to ignore this setting, enter ‘1’.

Note that when you enter Sm values directly, PASS generates an appropriate means matrix and then multiplies this matrix by each of these K values.

---

**Effect Size – Between- and Within-Subject Factors**

**Label**

Specify a label for this factor. Although we suggest that only a single letter be used, the label can consist of several letters. When several letters are used, the labels for the interactions may be extra long and confusing. Of course, you must be careful not to use the same label for two factors.

One of the easiest sets of labels is to use A, B, and C for the between factors and D, E, and F for the within factors. A useful alternative is to use B1, B2, and B3 for the between factors and W1, W2, and W3 for the within factors.

**Levels**

Specify the number of levels (categories) in this factor. Typical values are from 2 to 8. Set this to a blank (or 0) to ignore the factor in the design.

**Alpha**

These options specify the probability of a type-I error (alpha) for each factor and interaction. A type-I error occurs when you reject the null hypothesis of zero effects when in fact they are zero. Since they are probabilities, alpha values must be between zero and one. Historically, the value of 0.05 has been used for alpha. This value may be interpreted as meaning that about one F-test in twenty will falsely reject the null hypothesis. You should pick a value for alpha that represents the risk of a type-I error you are willing to take in your experimental situation.

You can specify different alpha values for different terms. For example, although you have three terms in an experiment, you might be mainly interested in only one of them. Hence, you could increase the alpha level of the tests of the other terms and thereby increase their power. Also, you may want to increase the alpha level of the interaction terms, as these will often have poor power otherwise.

**Power or Beta**

These options specify the power or beta (depending on the chosen setting) for each factor and interaction. Power is the probability of rejecting a false null hypothesis, and is equal to one minus Beta. Beta is the probability of a type-II error, which occurs when you fail to reject the null hypothesis of equal effects when in fact they are different.

Values must be between zero and one. Historically, the value of 0.80 (Beta = 0.20) was used for power. Now, 0.90 (Beta = 0.10) is also commonly used.

**Sm (Standard Deviation of Effects)**

Enter the standard deviation of the effects ($\sigma_{\text{effects}}$) for this factor or interaction. This value represents the magnitude of the differences among the means (effects) that is to be detected.
The value of Sm may be entered in several ways: directly, as a list of numbers, or as a percentage of another term.

- **Directly**
  You can enter the value of Sm directly by specifying a single number. If only a single number is entered, it becomes the value of Sm.

  You can use the Standard Deviation Estimator window to calculate the value of Sm for various sets of means. This window is obtained by selecting PASS, then Other, and then Standard Deviation Estimator from the menus.

- **List of Numbers**
  When a list of numbers is entered, the standard deviation of those numbers is computed and used as the value of Sm. The numbers in the list may represent means or effects. The list may be a simple list, the STEP command, or the RANGE command.

  **Simple List**
  A Simple List is a set of numbers separated by blanks or commas. Examples a simple list are
  
  5 20 60
  2,5,7
  -4,0,6,9

  **STEP Command**
  The syntax of the STEP command is \texttt{STEP Start Inc}. The list begins at \texttt{Start} and increases by \texttt{Inc}. The number of values generated is determined by the number of levels in the term.

  Examples of the STEP command for a term with four levels are
  
  ‘STEP 0 2’ results in ‘0 2 4 6’.
  ‘STEP 1 -1’ results in ‘1 0 -1 -2’.
  ‘STEP 1 0.5’ results in ‘1 1.5 2 2.5’.

  **RANGE Command**
  The syntax of the RANGE command is \texttt{RANGE Minimum Maximum}. The list of numbers generated increase steadily from \texttt{Minimum} to \texttt{Maximum}. The RANGE command is handy when you want to vary the number of levels while keeping the values in a known range.

  Examples of the RANGE command for a term with four levels are
  
  ‘RANGE 10 70’ results in ‘10 30 50 70’.
  ‘RANGE 0 1’ results in ‘0 0.33 0.67 1’.
  ‘RANGE 1 4’ results in ‘1 2 3 4’.
**Percentage of another Term**

You can specify the value as a percentage of another term. The syntax of this command is `TERM PCT` where `TERM` is any other main effect or interaction in the model and `PCT` is a percentage. This method is often used to specify interaction Sm’s. Examples of the PERCENTAGE command are

- ‘A 100’ results in an Sm equal to the Sm of factor A.
- ‘B 50’ results in an Sm equal to one-half of the Sm of factor B.

---

**Interactions Tab**

The values of Alpha, Power, and Sm are entered for various groups of 2-way and higher-order interactions.

---

**Covariance Tab**

This tab specifies the covariance matrix.

---

**Covariance Matrix Specification**

**Specify Which Covariance Matrix Input Method to Use**

This option specifies which method will be used to define the covariance matrix.

- **Standard Deviations and Autocorrelations**
  
  This option generates a covariance matrix based on the settings for the standard deviations (SD’s) and the pattern of autocorrelations as specified in the options on this screen down to and including ‘R2’. More about this option is given below.

- **Covariance Matrix Variables**
  
  When this option is selected, the covariance matrix is read in from the columns of the spreadsheet. This is the most flexible method, but specifying a covariance matrix is tedious. You will usually only use this method when a specific covariance is given to you. More about this option is given below.

  Note that the spreadsheet is shown by selecting the menus: ‘Window’ and then ‘Data’.

---

**Time Metric**

This option is used when the ‘Specify How the Standard Deviations Change Across Time’ option is set to ‘Range from SD1 to SD2 using the Time Metric’ to help define the covariance matrix. It specifies a sequential list of time points at which measurements of the subjects are made. Often, measurements are made at equally-spaced points through time. This is not always the case. It is important to define a time metric that corresponds to the study. For example, measurements might be planned at the beginning, after one day, after one week, and after one month.

The number of time points is the product of the number of levels of all within factors.

The time metric influences the values of the SD’s as well as the correlations between two measurements on the same individual.
Entering a List of Times
A list of times can be entered in which the time values are separated by blanks or commas. The time metric can be defined in any time scale desired. For example, you could enter 0, 0.143, 1, 2, 3 if times were 0 weeks, 1/7 week (day 1), 1 week, 2 weeks, 3 weeks. The same values in days would be 0, 1, 7, 14, 21.

Using the RANGE Command
The RANGE command can be used to specify a list of times. The syntax of the RANGE command is

   RANGE Minimum Maximum

A set of equal-spaced time points is generated between Minimum and Maximum. The number of time points depends on the number of within-factor levels.

This setting is very useful when you want to study the impact of increasing/decreasing the number of measurements per subject during the same period of time. That is, if the study will last five weeks, will the power of the statistical tests increase if you take ten measurements rather than five?

For example, suppose there are six times. Entering

   RANGE 0 10

will generate the time metric: 0, 2, 4, 6, 8, 10. If the number of times is changed to eleven, the time metric will become: 0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10.

Using the STEP Command
The STEP command can be used to specify a list of times. The syntax of the STEP command is

   STEP Start, Inc

This command generates time points beginning at Start and incrementing by Inc. For example,

   STEP 0 2

would generate the values 0, 2, 4, 6, 8, …

Covariance Matrix Specification – Within-Subject Standard Deviation Pattern
The parameters in this section provide a flexible way to specify \( \Sigma \), the covariance matrix. Because the covariance matrix is symmetric, it can be represented as

\[
\Sigma = \begin{bmatrix}
\sigma_{11} & \sigma_{12} & \cdots & \sigma_{1T} \\
\sigma_{12} & \sigma_{22} & \cdots & \sigma_{2T} \\
\vdots & \vdots & \ddots & \vdots \\
\sigma_{1T} & \sigma_{2T} & \cdots & \sigma_{TT}
\end{bmatrix}
\]
Repeated Measures Analysis of Variance

\[ \begin{bmatrix} \sigma_1^2 & \sigma_1\sigma_2\rho_{12} & \cdots & \sigma_1\sigma_T\rho_{1T} \\ \sigma_1\sigma_2\rho_{12} & \sigma_2^2 & \cdots & \sigma_2\sigma_T\rho_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_1\sigma_T\rho_{1T} & \sigma_2\sigma_T\rho_{2T} & \cdots & \sigma_T^2 \end{bmatrix} \]

\[ \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & 1 & \rho_{12} & \cdots & \rho_{1T} \\ 0 & \sigma_2 & \cdots & 0 & \rho_{12} & 1 & \cdots & \rho_{2T} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_T & \rho_{1T} & \rho_{2T} & \cdots & 1 \end{bmatrix} \]

where \( T \) is the product of the number of levels of all of the within factors.

Thus, the covariance matrix can be represented with complete generality by specifying the standard deviations \( \sigma_1, \sigma_2, \ldots, \sigma_T \) and the correlation matrix

\[ R = \begin{bmatrix} 1 & \rho_{12} & \cdots & \rho_{1T} \\ \rho_{12} & 1 & \cdots & \rho_{2T} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1T} & \rho_{2T} & \cdots & 1 \end{bmatrix} \]

Specify How the Standard Deviations Change Across Time

This option specifies the method used to specify the standard deviations \( \sigma_1, \sigma_2, \ldots, \sigma_T \). Based on the method selected, the actual values are specified using SD1 and, in some cases, SD2.

Each \( \sigma \) is an estimate of the standard deviation that occurs when the same individual is measured at the same point in time under identical treatment conditions. It is a measure of the within-subject variability.

The available options are

- **Constant (Use SD1. Ignore SD2)**
  When this option is selected, the standard deviations are assumed to be equal. That is, it is assumed that \( \sigma_1 = \cdots = \sigma_T \). The value of \( \sigma_i \) is specified in the SD1 field. The value in the SD2 field is ignored.

- **List of Standard Deviations (Use list in SD1. Ignore SD2.)**
  When this option is selected, a list of standard deviations can be entered in SD1. The items in the list can be separated by commas or blanks. The first value in the list becomes \( \sigma_1 \), the second value becomes \( \sigma_2 \), and so on. If the number of values in the list is less than the number of standard deviations required, the last value in the list is repeated. Note that all standard deviations in the list must be positive numbers.
• **Range from SD1 to SD2 using the Time Metric**

  When this option is selected, the standard deviations are spread between \( \sigma_1 \) and \( \sigma_T \) according to the spread in the Time Metric. The value in SD1 becomes \( \sigma_1 \) and the value in SD2 becomes \( \sigma_T \).

  For example, suppose SD1 = 100, SD2 = 200, and the Time Metric values are 0, 2, 4, 10. The standard deviations would be

  \[
  \sigma_1 = 100.0 \quad \sigma_2 = 120.0 \quad \sigma_3 = 140.0 \quad \sigma_4 = 200.0
  \]

**SD1 (Standard Deviation 1)**

This option is used to generate the covariance matrix. Its interpretation depends on the ‘How the SD’s Change Across Time’ option’s setting. Fundamentally, this is the standard deviation that occurs when the same individual is measured at the same point in time under identical treatment conditions. It is a measure of the within-subject variability.

You may want to use the special window that has been prepared to estimate SD1 from the mean square between (MSB) and the mean square within (MSW) of an existing table. To display this special window, from the menus select ‘PASS’, then ‘Other’, and then ‘Standard Deviation Estimator’. Click on the ‘Covariance Matrix’ tab. Enter the values from the ANOVA table. The resulting value of ‘Sigma’ should be placed here.

**Specify How the Standard Deviations Change Across Time = Constant**

The value entered here is used as the standard deviation for all time points.

**Specify How the Standard Deviations Change Across Time = List of Standard Deviations**

The values in the list entered here become the values of the standard deviations. If the number in the list is less than the number required, the last value in the list is repeated.

**Specify How the Standard Deviations Change Across Time = Range from SD1 to SD2 using the Time Metric**

The value entered here is used as a beginning standard deviation, the value in SD2 is used as an ending standard deviation, and the intermediate standard deviations are spaced between SD1 and SD2 proportional to the values of the Time Metric.

For example, suppose SD1 = 100, SD2 = 200, and the Time Metric values are 0,2,4,10. The standard deviation values would be:

\[
S(1)=100 \\
S(2)=120 \\
S(3)=140 \\
S(4)=200
\]

**SD2 (Standard Deviation 2)**

This parameter is used when ‘How the SD’s Change Across Time’ option is set to ‘Range…’. In that case, this option specifies the value of \( \sigma_T \).
Covariance Matrix Specification – Autocorrelation Pattern

Specify How the Autocorrelations Change Across Time

This option specifies the pattern of the autocorrelations in the variance-covariance matrix. Three options are possible:

- **Constant**

  The value of \( R_1 \) is used as the constant autocorrelation until the maximum time difference is reached, then the value of \( R_2 \) is used. For example, if the maximum time difference is 3, \( R_1 = 0.6 \), \( R_2 = 0.1 \), and \( T = 6 \), the correlation matrix would appear as

  \[
  R = \begin{bmatrix}
  1 & 0.600 & 0.600 & 0.600 & 0.100 & 0.100 \\
  0.600 & 1 & 0.600 & 0.600 & 0.600 & 0.100 \\
  0.600 & 0.600 & 1 & 0.600 & 0.600 & 0.600 \\
  0.600 & 0.600 & 0.600 & 1 & 0.600 & 0.600 \\
  0.100 & 0.600 & 0.600 & 0.600 & 1 & 0.600 \\
  0.100 & 0.100 & 0.600 & 0.600 & 0.600 & 1 \\
  \end{bmatrix}
  \]

  Note that when all correlations are equal, this is the correlation pattern that is assumed by the repeated measure ANOVA F-test. It may be a good first approximation, but many researchers believe the next option (first-order autocorrelation) is more realistic.

- **1st Order Autocorrelation**

  The value of \( R_1 \) is used as the base autocorrelation in a first-order, serial correlation pattern. For example, if the maximum time difference is 3, \( R_1 = 0.6 \), \( R_2 = 0.1 \), and \( T = 6 \), the correlation matrix would appear as

  \[
  R = \begin{bmatrix}
  1 & 0.600 & 0.360 & 0.216 & 0.100 & 0.100 \\
  0.600 & 1 & 0.600 & 0.360 & 0.216 & 0.100 \\
  0.360 & 0.600 & 1 & 0.600 & 0.360 & 0.216 \\
  0.216 & 0.360 & 0.600 & 1 & 0.600 & 0.360 \\
  0.100 & 0.216 & 0.360 & 0.600 & 1 & 0.600 \\
  0.100 & 0.100 & 0.216 & 0.360 & 0.600 & 1 \\
  \end{bmatrix}
  \]

  This pattern is often chosen as the most realistic when little is known about the correlation pattern.

- **Custom**

  The values of \( R_1, R_2, A, V, \) and Max Time Diff are used to generate a custom autocorrelation pattern. This relationship is modeled using the equation

  \[
  \text{Corr}(Y_{ai}, Y_{sj}) = \rho_{ij} \\
  = d \left( R_1^{1 - At_j} \right) + (1 - d) R_2
  \]
where $R_1$ is the base correlation, $t_i$ and $t_j$ are two time points, and $A$ and $V$ are specified constants. The variable $d$ is one if $|t_i - t_j|$ is less than Max Time Diff and zero otherwise.

Machin, Campbell, Fayers, and Pinol (1997) state that values of $R_1$ between 0.60 and 0.75 are common.

We will present some examples to show you how this formula may be interpreted. For the moment, assume that the time metric is four, equally space time points of 1, 2, 3, and 4. Also, assume that Max Time Diff is set to 20.

**Example 1**
Let $A = 0$, $V = 1$, and $\rho_1 = \rho$. The correlation matrix becomes

$$R = \begin{bmatrix} 1 & \rho & \rho & \rho \\ \rho & 1 & \rho & \rho \\ \rho & \rho & 1 & \rho \\ \rho & \rho & \rho & 1 \end{bmatrix}$$

**Example 2**
Let $A = 1$, $V = 1$, and $\rho_1 = \rho$. The correlation matrix becomes

$$R = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

which is the first-order autoregression model, a popular model in time series analysis.

**Example 3**
Let $A = 1$, $V = 2$, and $\rho_1 = \rho$. The correlation matrix becomes

$$R = \begin{bmatrix} 1 & \rho^2 & \rho^4 & \rho^6 \\ \rho^2 & 1 & \rho^2 & \rho^4 \\ \rho^4 & \rho^2 & 1 & \rho^2 \\ \rho^6 & \rho^4 & \rho^2 & 1 \end{bmatrix}$$

which is similar to Example 2 except that the correlations die out much more quickly.
Example 4
Let $A = 0.5$, $V = 1$, and $\rho_1 = \rho$. The correlation matrix becomes

\[
R = \begin{bmatrix}
1 & \rho^{0.5} & \rho^{1.0} & \rho^{1.5} \\
\rho^{0.5} & 1 & \rho^{1.0} & \rho^{1.5} \\
\rho^{1.0} & \rho^{1.5} & 1 & \rho^{1.0} \\
\rho^{1.5} & \rho^{1.0} & \rho^{1.5} & 1
\end{bmatrix}
\]

which is similar to Example 2 except that the correlations die out much more slowly.

Example 5
Let $A = 1$ and $V = 1$. For this example, set the Max Time Diff option to 2. The correlation matrix becomes

\[
R = \begin{bmatrix}
1 & \rho_1 & \rho_2 & \rho_2 \\
\rho_1 & 1 & \rho_1 & \rho_2 \\
\rho_2 & \rho_1 & 1 & \rho_1 \\
\rho_2 & \rho_2 & \rho_1 & 1
\end{bmatrix}
\]

Notice that this scenario lets you create a banded correlation matrix with two unique correlations.

Example 6
This example shows how this formula works when the Max Time Diff is set to 7 and the time metric is 1, 2, 7, 15. Let $A = 1$ and $V = 1$. The correlation matrix becomes

\[
R = \begin{bmatrix}
1 & \rho_1 & \rho_1^6 & \rho_2 \\
\rho_1 & 1 & \rho_1^6 & \rho_2 \\
\rho_1^6 & \rho_1^6 & 1 & \rho_2 \\
\rho_2 & \rho_2 & \rho_2 & 1
\end{bmatrix}
\]

R1 (Autocorrelation)
This is the autocorrelation, $r_1$, between two measurements made on a subject at two time points that differ by one time unit. This value is combined with the other parameters in this section to form the covariance matrix.

Since this is a type of correlation, possible values range from -1 to 1. However, in this situation, a positive value is usually assumed, so the range is 0 to 1. A value near 0 indicates low autocorrelation. A value near 1 indicates high autocorrelation.

The value of this parameter depends on the Time Metric that is defined. Normally, you would expect a larger autocorrelation if the time metric units were in hours rather than days. In their book on sample size, Machin and Campbell comment the values between 0.60 and 0.75 are typical.
It is reasonable to assume that there is a correlation between two measurements made on the same subject at two points in time. It is often reasonable to assume that the size of this correlation diminishes as the two time points are further and further apart. That is, you would expect a much larger autocorrelation between two measurements taken one minute apart than between two measurements taken one week apart.

You may want to use the special window that has been prepared to estimate R1 from the mean square between (MSB) and the mean square within (MSW) of an existing table. To display this special window, from the menus select ‘PASS’, then ‘Other’, and then ‘Standard Deviation Estimator’. Click on the ‘Covariance Matrix’ tab. Enter the values from the ANOVA table. The resulting value of ‘Rho should be placed here.

**R2 (Second AC)**

This is the value of the secondary autocorrelation, R2. This value is used when the difference between two time points (see Time Metric) is greater than the value of Max Time Diff. Hence, if you set Max Time Diff to zero, this value will be used to calculate all correlations in the covariance matrix. When used, think of R2 as the correlation between measurements made on the same subject, regardless of how far apart in time they are. Since we are assuming a positive autocorrelation, this value ranges between 0 and 1.

**Max Time Diff**

This is the maximum time difference (MTD) between two measurement points before the autocorrelation is set to the constant correlation value, R2.

In the autocorrelation formula:

\[ \rho_{i,j} = dR1^{(t_i-t_j)^V} + (1-d)R2 \]

The parameter \(d\) is equal to 1 when \(|t_i-t_j| < \text{MTD}\) and 0 otherwise. Hence, this value controls when the autocorrelation is set to R2.

If you think of R2 as the correlation between measurements made on the same subject, regardless of how far apart in time they are, then this value should be set to that time difference at which the measurements times are no longer a factor.

For example, you might postulate that the autocorrelations are 0.9, 0.5, 0.2, 0.2, 0.2, and so on. That is, the autocorrelation is 0.9 for measurements taken one day apart, 0.5 for measurements taken two days apart, and 0.2 for all others. In this case, you would set Max Time Diff to 3.

**A, V**

These parameters are used when a 'Custom' autocorrelation pattern across time is specified. The formula used to calculate the autocorrelation is:

\[ \rho_{i,j} = dR1^{(t_i-t_j)^V} + (1-d)R2 \]

where \(d\) is 1 if the time difference is \(<=\) Max Time Diff and 0 otherwise.

A = 0, V = 1 gives constant autocorrelation.

A = 1, V = 1 gives first-order autocorrelation.

Usually, V is set to 1. V should only be set to 1 or 2. The value of 1 is recommended. Set V at 2 when you want the autocorrelations to taper off rapidly as time occurs between measurements.
Covariance Matrix Specification – Covariance Matrix Variables
This option instructs the program to read the covariance matrix from the spreadsheet.

Spreadsheet Columns Containing the Covariance Matrix
This option designates the columns on the current spreadsheet holding the covariance matrix. It is used when the ‘Specify Which Covariance Matrix Input Method to Use’ option is set to Covariance Matrix Variables.

The number of columns and number of rows must match the number of time periods at which the subjects are measured.

Reports Tab
This tab specifies which reports and graphs are displayed as well as their format.

Select Output – Numeric Reports
Test in Summary Statement(s)
Indicate the test that is to be reported on in the Summary Statements.

Select Output – Report Options
Maximum Term-Order Reported
Indicate the maximum order of terms to be reported on. Occasionally, higher-order interactions are of little interest and so they may be omitted. For example, enter a ‘2’ to limit output to individual factors and two-way interactions.

Skip Line After
The names of the terms can be too long to fit in the space provided. If the name contains more characters than this, the rest of the output is placed on a separate line. Enter ‘1’ when you want every term’s results printed on two lines. Enter ‘100’ when you want every variable’s results printed on one line.

Example 1 – Determining Power
Researchers are planning a study of the impact of a drug on heart rate. They want to evaluate the differences in heart rate among three age groups: 20-40, 41-60, and over 60. Their experimental protocol calls for a baseline heart rate measurement, followed by administration of a certain level of the drug, followed by three additional measurements 30 minutes apart. They want to be able to detect a 10% difference in heart rate among the age groups. They want to detect 5% difference in heart rate within an individual across time. They decide the experiment should detect interaction effects of the same magnitude as the time factor. They plan to analyze the data using a Geisser-Greenhouse corrected F test.

Similar studies have found an average heart rate of 93, a standard deviation of 4, and an autocorrelation between adjacent measurements on the same individual of 0.7. The researchers assume that first-order autocorrelation adequately represents the autocorrelation pattern.
From a heart rate of 93, a 10% reduction gives 84. They decide on the age-group means of 93, 87, and 84. Similarly, a 5% reduction within a subject would result in a heart rate of 88. They decide on time means of 93, 89, 88, and 91.

How many subjects per age group are needed to achieve 95% power and a 0.05 significance level?

**Setup**

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the **Repeated Measures Analysis of Variance** procedure window by expanding **Means**, then **ANOVA**, then clicking on **Analysis of Variance**, and then clicking on **Repeated Measures Analysis of Variance**. You may then make the appropriate entries as listed below, or open **Example 1** by going to the **File** menu and choosing **Open Example Template**.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Tab</strong></td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>n (Subjects Per Group)</td>
<td>2 to 8 by 1</td>
</tr>
<tr>
<td>=n’s</td>
<td>checked</td>
</tr>
<tr>
<td>Means Matrix</td>
<td>blank</td>
</tr>
<tr>
<td>K (Means Multipliers)</td>
<td>1.0</td>
</tr>
<tr>
<td><strong>For First Between-Subject Factor</strong></td>
<td></td>
</tr>
<tr>
<td>Label</td>
<td>B</td>
</tr>
<tr>
<td>Levels</td>
<td>3</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Sm (Standard Deviation of Effects)</td>
<td>93 87 84</td>
</tr>
<tr>
<td><strong>For First Within-Subject Factor</strong></td>
<td></td>
</tr>
<tr>
<td>Label</td>
<td>W</td>
</tr>
<tr>
<td>Levels</td>
<td>4</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Sm (Standard Deviation of Effects)</td>
<td>93 89 88 91</td>
</tr>
<tr>
<td><strong>Interactions Tab</strong></td>
<td></td>
</tr>
<tr>
<td>2-Way(Mixed) Interaction</td>
<td></td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Sm (Standard Deviation of Effects)</td>
<td>W 100</td>
</tr>
<tr>
<td><strong>Covariance Tab</strong></td>
<td></td>
</tr>
<tr>
<td>Specify Covariance Method</td>
<td>1) Standard Deviations and Autocorrelations</td>
</tr>
<tr>
<td>Time Metric</td>
<td>Ignored</td>
</tr>
<tr>
<td>How SD’s Change Across Time</td>
<td>Constant</td>
</tr>
<tr>
<td>SD1 (Standard Deviation 1)</td>
<td>4</td>
</tr>
<tr>
<td>Specify How Autocorr’s Change</td>
<td>1st Order Autocorr</td>
</tr>
<tr>
<td>R1 (Autocorrelation)</td>
<td>0.7</td>
</tr>
<tr>
<td>Max Time Diff.</td>
<td>100 (This large value will cause R2 to be ignored.)</td>
</tr>
</tbody>
</table>
Repeated Measures Analysis of Variance  570-31

Reports Tab
Numeric Results by Term...........................Checked
Numeric Results by Design.........................Checked
Regular F Test ........................................Not checked
GG F Test ...............................................Checked
Wilks’ Lambda ..........................................Not checked
Pillai-Bartlett .........................................Not checked
Hotelling-Lawley........................................Not checked
GG Detail Report ......................................Checked
Means Matrix...........................................Checked
Covariance Matrix .....................................Checked
Test in Summary Statement .......................GG F Test
Show Plot 1 .............................................Checked
Show Plot 2 .............................................Not checked
Test That is Plotted ..................................GG F Test
Max Term-Order Plotted .........................2
Max Term-Order Reported ....................2

Annotated Output

Click the Run button to perform the calculations and generate the following output.

Design Report

<table>
<thead>
<tr>
<th>Term</th>
<th>Test</th>
<th>Power</th>
<th>n</th>
<th>N</th>
<th>Multiply Means By</th>
<th>SD of Effects (Sm)</th>
<th>Standard Deviation (Sigma)</th>
<th>Effect Size</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>B(3)</td>
<td>GG F</td>
<td>0.3096</td>
<td>2</td>
<td>6</td>
<td>1.00</td>
<td>3.74</td>
<td>3.29</td>
<td>1.14</td>
<td>0.0500</td>
<td>0.6904</td>
</tr>
<tr>
<td>W(4)</td>
<td>GG F</td>
<td>0.3266</td>
<td>2</td>
<td>6</td>
<td>1.00</td>
<td>1.92</td>
<td>1.31</td>
<td>1.47</td>
<td>0.0500</td>
<td>0.6734</td>
</tr>
<tr>
<td>BW</td>
<td>GG F</td>
<td>0.1588</td>
<td>2</td>
<td>6</td>
<td>1.00</td>
<td>1.92</td>
<td>1.31</td>
<td>1.47</td>
<td>0.0500</td>
<td>0.8412</td>
</tr>
<tr>
<td>B(3)</td>
<td>GG F</td>
<td>0.6459</td>
<td>3</td>
<td>9</td>
<td>1.00</td>
<td>3.74</td>
<td>3.29</td>
<td>1.14</td>
<td>0.0500</td>
<td>0.3541</td>
</tr>
<tr>
<td>W(4)</td>
<td>GG F</td>
<td>0.8099</td>
<td>3</td>
<td>9</td>
<td>1.00</td>
<td>1.92</td>
<td>1.31</td>
<td>1.47</td>
<td>0.0500</td>
<td>0.1901</td>
</tr>
<tr>
<td>BW</td>
<td>GG F</td>
<td>0.6267</td>
<td>3</td>
<td>9</td>
<td>1.00</td>
<td>1.92</td>
<td>1.31</td>
<td>1.47</td>
<td>0.0500</td>
<td>0.3733</td>
</tr>
<tr>
<td>B(3)</td>
<td>GG F</td>
<td>0.8478</td>
<td>4</td>
<td>12</td>
<td>1.00</td>
<td>3.74</td>
<td>3.29</td>
<td>1.14</td>
<td>0.0500</td>
<td>0.1522</td>
</tr>
<tr>
<td>W(4)</td>
<td>GG F</td>
<td>0.9486</td>
<td>4</td>
<td>12</td>
<td>1.00</td>
<td>1.92</td>
<td>1.31</td>
<td>1.47</td>
<td>0.0500</td>
<td>0.0514</td>
</tr>
<tr>
<td>BW</td>
<td>GG F</td>
<td>0.8598</td>
<td>4</td>
<td>12</td>
<td>1.00</td>
<td>1.92</td>
<td>1.31</td>
<td>1.47</td>
<td>0.0500</td>
<td>0.1402</td>
</tr>
<tr>
<td>B(3)</td>
<td>GG F</td>
<td>0.9415</td>
<td>5</td>
<td>15</td>
<td>1.00</td>
<td>3.74</td>
<td>3.29</td>
<td>1.14</td>
<td>0.0500</td>
<td>0.0585</td>
</tr>
<tr>
<td>W(4)</td>
<td>GG F</td>
<td>0.9871</td>
<td>5</td>
<td>15</td>
<td>1.00</td>
<td>1.92</td>
<td>1.31</td>
<td>1.47</td>
<td>0.0500</td>
<td>0.0129</td>
</tr>
<tr>
<td>BW</td>
<td>GG F</td>
<td>0.9535</td>
<td>5</td>
<td>15</td>
<td>1.00</td>
<td>1.92</td>
<td>1.31</td>
<td>1.47</td>
<td>0.0500</td>
<td>0.0465</td>
</tr>
<tr>
<td>B(3)</td>
<td>GG F</td>
<td>0.9793</td>
<td>6</td>
<td>18</td>
<td>1.00</td>
<td>3.74</td>
<td>3.29</td>
<td>1.14</td>
<td>0.0500</td>
<td>0.0207</td>
</tr>
<tr>
<td>W(4)</td>
<td>GG F</td>
<td>0.9970</td>
<td>6</td>
<td>18</td>
<td>1.00</td>
<td>1.92</td>
<td>1.31</td>
<td>1.47</td>
<td>0.0500</td>
<td>0.0030</td>
</tr>
<tr>
<td>BW</td>
<td>GG F</td>
<td>0.9960</td>
<td>6</td>
<td>18</td>
<td>1.00</td>
<td>1.92</td>
<td>1.31</td>
<td>1.47</td>
<td>0.0500</td>
<td>0.0140</td>
</tr>
<tr>
<td>B(3)</td>
<td>GG F</td>
<td>0.9931</td>
<td>7</td>
<td>21</td>
<td>1.00</td>
<td>3.74</td>
<td>3.29</td>
<td>1.14</td>
<td>0.0500</td>
<td>0.0069</td>
</tr>
<tr>
<td>W(4)</td>
<td>GG F</td>
<td>0.9993</td>
<td>7</td>
<td>21</td>
<td>1.00</td>
<td>1.92</td>
<td>1.31</td>
<td>1.47</td>
<td>0.0500</td>
<td>0.0007</td>
</tr>
<tr>
<td>BW</td>
<td>GG F</td>
<td>0.9961</td>
<td>7</td>
<td>21</td>
<td>1.00</td>
<td>1.92</td>
<td>1.31</td>
<td>1.47</td>
<td>0.0500</td>
<td>0.0039</td>
</tr>
<tr>
<td>B(3)</td>
<td>GG F</td>
<td>0.9978</td>
<td>8</td>
<td>24</td>
<td>1.00</td>
<td>3.74</td>
<td>3.29</td>
<td>1.14</td>
<td>0.0500</td>
<td>0.0022</td>
</tr>
<tr>
<td>W(4)</td>
<td>GG F</td>
<td>0.9999</td>
<td>8</td>
<td>24</td>
<td>1.00</td>
<td>1.92</td>
<td>1.31</td>
<td>1.47</td>
<td>0.0500</td>
<td>0.0001</td>
</tr>
<tr>
<td>BW</td>
<td>GG F</td>
<td>0.9990</td>
<td>8</td>
<td>24</td>
<td>1.00</td>
<td>1.92</td>
<td>1.31</td>
<td>1.47</td>
<td>0.0500</td>
<td>0.0010</td>
</tr>
</tbody>
</table>

The Design Report gives the power for each term in the design for each value of n. It is useful when you want to compare the powers of the terms in the design at a specific sample size.
In this example, the design goals of 0.95 power on all terms are achieved for \( n = 6 \).

The definitions of each of the columns of the report are as follows.

**Term**
This column contains the identifying label of the term. The number of levels for a factor is given in parentheses.

**Test**
This column identifies the test statistic. Since the power depends on the test statistic, you should make sure that this is the test statistic that you will use.

**Power**
This is the computed power for the term.

\( n \)
The value of \( n \) is the number of subjects per group.

\( N \)
The value of \( N \) is the total number of subjects in the study.

**Multiply Means By**
This is the value of the means multiplier, \( K \).

**SD of Effects (Sm)**
This is the standard deviation of the effects \( \sigma_m \) for this term.

**Standard Deviation**
This is the value of \( \sigma \), the random variation that \( \sigma_m \) is compared against by the \( F \) test. See the Technical Details for details on how these values are calculated.

**Effect Size**
The Effect Size is calculated by the expression \( \sigma_m / \sigma \). It is an index of the size of the effect values relative to the standard deviation. Its value may be compared from experiment to experiment, regardless of the scale of the response variable.

**Alpha**
Alpha is the significance level of the test

**Beta**
Beta is the probability of failing to reject the null hypothesis when the alternative hypothesis is true.
## Term Reports

### Results for Factor B (Levels = 3)

<table>
<thead>
<tr>
<th>Test</th>
<th>Power</th>
<th>n</th>
<th>N</th>
<th>Multiply SD of Means Effects Deviation (Sm)</th>
<th>Standard Deviation (Sigma)</th>
<th>Effect Size</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>GG F</td>
<td>0.3096</td>
<td>2</td>
<td>6</td>
<td>1.00 3.74 3.29 3.29 1.14 0.14 0.0500 0.6904</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG F</td>
<td>0.6459</td>
<td>3</td>
<td>9</td>
<td>1.00 3.74 3.29 3.29 1.14 0.14 0.0500 0.3541</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG F</td>
<td>0.8478</td>
<td>4</td>
<td>12</td>
<td>1.00 3.74 3.29 3.29 1.14 0.14 0.0500 0.1522</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG F</td>
<td>0.9415</td>
<td>5</td>
<td>15</td>
<td>1.00 3.74 3.29 3.29 1.14 0.14 0.0500 0.0585</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG F</td>
<td>0.9793</td>
<td>6</td>
<td>18</td>
<td>1.00 3.74 3.29 3.29 1.14 0.14 0.0500 0.0207</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG F</td>
<td>0.9931</td>
<td>7</td>
<td>21</td>
<td>1.00 3.74 3.29 3.29 1.14 0.14 0.0500 0.0069</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG F</td>
<td>0.9978</td>
<td>8</td>
<td>24</td>
<td>1.00 3.74 3.29 3.29 1.14 0.14 0.0500 0.0022</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Results for Factor W (Levels = 4)

<table>
<thead>
<tr>
<th>Test</th>
<th>Power</th>
<th>n</th>
<th>N</th>
<th>Multiply SD of Means Effects Deviation (Sm)</th>
<th>Standard Deviation (Sigma)</th>
<th>Effect Size</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>GG F</td>
<td>0.3266</td>
<td>2</td>
<td>6</td>
<td>1.00 1.92 1.31 1.47 0.0500 0.6734</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG F</td>
<td>0.8099</td>
<td>3</td>
<td>9</td>
<td>1.00 1.92 1.31 1.47 0.0500 0.1901</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG F</td>
<td>0.9486</td>
<td>4</td>
<td>12</td>
<td>1.00 1.92 1.31 1.47 0.0500 0.0514</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG F</td>
<td>0.9871</td>
<td>5</td>
<td>15</td>
<td>1.00 1.92 1.31 1.47 0.0500 0.0129</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG F</td>
<td>0.9970</td>
<td>6</td>
<td>18</td>
<td>1.00 1.92 1.31 1.47 0.0500 0.0030</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG F</td>
<td>0.9993</td>
<td>7</td>
<td>21</td>
<td>1.00 1.92 1.31 1.47 0.0500 0.0007</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG F</td>
<td>0.9999</td>
<td>8</td>
<td>24</td>
<td>1.00 1.92 1.31 1.47 0.0500 0.0001</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Results for Term BW

<table>
<thead>
<tr>
<th>Test</th>
<th>Power</th>
<th>n</th>
<th>N</th>
<th>Multiply SD of Means Effects Deviation (Sm)</th>
<th>Standard Deviation (Sigma)</th>
<th>Effect Size</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>GG F</td>
<td>0.1588</td>
<td>2</td>
<td>6</td>
<td>1.00 1.92 1.31 1.47 0.0500 0.8412</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG F</td>
<td>0.6267</td>
<td>3</td>
<td>9</td>
<td>1.00 1.92 1.31 1.47 0.0500 0.3733</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG F</td>
<td>0.8598</td>
<td>4</td>
<td>12</td>
<td>1.00 1.92 1.31 1.47 0.0500 0.1402</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG F</td>
<td>0.9535</td>
<td>5</td>
<td>15</td>
<td>1.00 1.92 1.31 1.47 0.0500 0.0465</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG F</td>
<td>0.9860</td>
<td>6</td>
<td>18</td>
<td>1.00 1.92 1.31 1.47 0.0500 0.0140</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG F</td>
<td>0.9961</td>
<td>7</td>
<td>21</td>
<td>1.00 1.92 1.31 1.47 0.0500 0.0039</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG F</td>
<td>0.9990</td>
<td>8</td>
<td>24</td>
<td>1.00 1.92 1.31 1.47 0.0500 0.0010</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The *Term Reports* provide a complete report for each term at all sample sizes. They are especially useful when you are only interested in the power of one or two terms.

The definitions of each of the columns of the report are identical to the corresponding columns in the *Design Report*, so they are not repeated here.

## Geisser-Greenhouse Correction Detail Report

<table>
<thead>
<tr>
<th>Term (Levels)</th>
<th>Power</th>
<th>Critical F</th>
<th>Lambda</th>
<th>df1</th>
<th>df2</th>
<th>Epsilon</th>
<th>E(Epsilon)</th>
<th>G1</th>
</tr>
</thead>
<tbody>
<tr>
<td>n = 2 N = 6 Means x 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B (3)</td>
<td>0.3096</td>
<td>0.0500</td>
<td>9.55</td>
<td>7.74</td>
<td>2</td>
<td>3</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>W (4)</td>
<td>0.3266</td>
<td>0.0500</td>
<td>7.60</td>
<td>12.88</td>
<td>3</td>
<td>9</td>
<td>0.77</td>
<td>0.60</td>
</tr>
<tr>
<td>BW</td>
<td>0.1588</td>
<td>0.0500</td>
<td>6.92</td>
<td>12.88</td>
<td>6</td>
<td>9</td>
<td>0.77</td>
<td>0.60</td>
</tr>
<tr>
<td>n = 3 N = 9 Means x 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B (3)</td>
<td>0.6459</td>
<td>0.0500</td>
<td>5.14</td>
<td>11.62</td>
<td>2</td>
<td>6</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>W (4)</td>
<td>0.8099</td>
<td>0.0500</td>
<td>4.12</td>
<td>19.32</td>
<td>3</td>
<td>18</td>
<td>0.77</td>
<td>0.60</td>
</tr>
<tr>
<td>BW</td>
<td>0.6267</td>
<td>0.0500</td>
<td>4.12</td>
<td>19.32</td>
<td>6</td>
<td>18</td>
<td>0.77</td>
<td>0.60</td>
</tr>
<tr>
<td>n = 4 N = 12 Means x 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B (3)</td>
<td>0.8478</td>
<td>0.0500</td>
<td>4.26</td>
<td>15.49</td>
<td>2</td>
<td>9</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>W (4)</td>
<td>0.9486</td>
<td>0.0500</td>
<td>3.58</td>
<td>25.76</td>
<td>3</td>
<td>27</td>
<td>0.77</td>
<td>0.66</td>
</tr>
<tr>
<td>BW</td>
<td>0.8598</td>
<td>0.0500</td>
<td>2.95</td>
<td>25.76</td>
<td>6</td>
<td>27</td>
<td>0.77</td>
<td>0.66</td>
</tr>
<tr>
<td>n = 5 N = 15 Means x 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>B (3)</td>
<td>0.9415</td>
<td>0.0500</td>
<td>3.89</td>
<td>19.36</td>
<td>2</td>
<td>12</td>
<td>1.00</td>
<td>1.00</td>
</tr>
<tr>
<td>W (4)</td>
<td>0.9871</td>
<td>0.0500</td>
<td>3.36</td>
<td>32.20</td>
<td>3</td>
<td>36</td>
<td>0.77</td>
<td>0.68</td>
</tr>
<tr>
<td>BW</td>
<td>0.9535</td>
<td>0.0500</td>
<td>2.75</td>
<td>32.20</td>
<td>6</td>
<td>36</td>
<td>0.77</td>
<td>0.68</td>
</tr>
</tbody>
</table>
This report gives the details of the components of the Geisser-Greenhouse correction for each term and sample size. It is useful when you want to compare various aspects of this test.

The definitions of each of the columns of the report are as follows.

**Term**
This column contains the identifying label of the term. For factors, the number of levels is also given in parentheses.

**Power**
This is the computed power for the term.

**Alpha**
Alpha is the significance level of the test.

**Critical F**
This is the critical value of the F statistic. An $F$ value computed from the data that is larger than this value is statistically significant at the alpha level given.

**Lambda**
This is the value of the noncentrality parameter $\lambda$ of the approximate noncentral $F$ distribution.

**df1|df2**
These are the values of the numerator and denominator degrees of freedom of the approximate $F$ test that is used. These values are useful when comparing various designs. Other things being equal, you would like to have $df_2$ large and $df_1$ small.

**Epsilon**
The Geisser-Greenhouse epsilon is a measure of how far the covariance matrix departs from the assumption of circularity.

**E(Epsilon)**
This is the expected value of epsilon. It is a measure of how far the covariance matrix departs from the assumption of circularity.

**G1**
$G1$ is part of a correction factor used to convert $\varepsilon$ to $E(\hat{\varepsilon})$. It is reported for your convenience.
**Summary Statements**

A repeated measures design with 1 between factor and 1 within factor has 3 groups with 2 subjects each for a total of 6 subjects. Each subject is measured 4 times. This design achieves 31% power to test factor B if a Geisser-Greenhouse Corrected F Test is used with a 5% significance level and the actual effect standard deviation is 3.74 (an effect size of 1.14), achieves 33% power to test factor W if a Geisser-Greenhouse Corrected F Test is used with a 5% significance level and the actual effect standard deviation is 1.92 (an effect size of 1.47), and achieves 16% power to test the BW interaction if a Geisser-Greenhouse Corrected F Test is used with a 5% significance level and the actual effect standard deviation is 1.92 (an effect size of 1.47).

A summary statement can be generated for each sample size that is used. This statement gives the results in sentence form. The number of designs reported on textually is controlled by the Summary Statement option on the Reports Tab.

**Means Matrix**

<table>
<thead>
<tr>
<th>Name</th>
<th>B1</th>
<th>B2</th>
<th>B3</th>
</tr>
</thead>
<tbody>
<tr>
<td>W1</td>
<td>-10.62</td>
<td>5.19</td>
<td>-2.72</td>
</tr>
<tr>
<td>W2</td>
<td>1.46</td>
<td>3.97</td>
<td>2.72</td>
</tr>
<tr>
<td>W3</td>
<td>-4.58</td>
<td>4.58</td>
<td>0.00</td>
</tr>
<tr>
<td>W4</td>
<td>-4.58</td>
<td>4.58</td>
<td>0.00</td>
</tr>
</tbody>
</table>

This report shows the means matrix that was read in from the spreadsheet or generated by the Sm values that were given. It may be used to get an impression of the magnitude of the difference among the means that is being studied. When a Means Multiplier, $K$, is used, each value of $K$ is multiplied times each value of this matrix.

**Variance-Covariance Matrix Section**

<table>
<thead>
<tr>
<th>Time</th>
<th>W1</th>
<th>W2</th>
<th>W3</th>
<th>W4</th>
</tr>
</thead>
<tbody>
<tr>
<td>W1</td>
<td>4.00</td>
<td>0.70</td>
<td>0.49</td>
<td>0.34</td>
</tr>
<tr>
<td>W2</td>
<td>0.70</td>
<td>4.00</td>
<td>0.70</td>
<td>0.49</td>
</tr>
<tr>
<td>W3</td>
<td>0.49</td>
<td>0.70</td>
<td>4.00</td>
<td>0.70</td>
</tr>
<tr>
<td>W4</td>
<td>0.34</td>
<td>0.49</td>
<td>0.70</td>
<td>4.00</td>
</tr>
</tbody>
</table>

This report shows the variance-covariance matrix that was read in from the spreadsheet or generated by the settings of on the Covariance tab. The standard deviations are given on the diagonal and the autocorrelations are given off the diagonal.
Plots Section

The chart shows the relationship between power and \( n \) for the terms in the design. Note that high-order interactions may be omitted from the plot by reducing the Max Term-Order Plotted option on the Plot Setup tab.

Example 2 – Varying the Difference Between the Means

Continuing with Example 1, the researchers want to evaluate the impact on power of varying the size of the difference among the means for a range of sample sizes from 2 to 8 per groups. The researchers could try calculating various multiples of the means, inputting them, and recording the results. This can be accomplished very easily by using the \( K \) option.

Keeping all other settings as in Example 1, the value of \( K \) is varied from 0.2 to 3.0 in steps of 0.2. We determined these values by experimentation so that a full range of power values are shown on the plots.

In the output to follow, we only display the plots. You may want to display the numeric reports as well, but we do not here in order to save space.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Repeated Measures Analysis of Variance procedure window by expanding Means, then ANOVA, then clicking on Analysis of Variance, and then clicking on Repeated Measures Analysis of Variance. You may then make the appropriate entries as listed below, or open Example 2 by going to the File menu and choosing Open Example Template.
### Repeated Measures Analysis of Variance

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Tab</strong></td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>n (Subjects Per Group)</td>
<td>2 3 4 8</td>
</tr>
<tr>
<td>=n's</td>
<td>checked</td>
</tr>
<tr>
<td>Means Matrix</td>
<td>blank</td>
</tr>
<tr>
<td>K (Means Multipliers)</td>
<td>0.2 to 3.0 by 0.2</td>
</tr>
</tbody>
</table>

*For First Between-Subject Factor*

| Label                        | B     |
| Levels                       | 3     |
| Alpha                        | 0.05  |
| Power                        | Ignored since this is the Find setting |
| Sm (Standard Deviation of Effects) | 93 87 84 |

*For First Within-Subject Factor*

| Label                        | W     |
| Levels                       | 4     |
| Alpha                        | 0.05  |
| Power                        | Ignored since this is the Find setting |
| Sm (Standard Deviation of Effects) | 93 89 88 91 |

*Interactions Tab*

*2-Way(Mixed) Interaction*

| Alpha                        | 0.05  |
| Power                        | Ignored since this is the Find setting |
| Sm (Standard Deviation of Effects) | W 100 |

*Covariance Tab*

| Specify Covariance Method    | 1) Standard Deviations and Autocorrelations |
| How SD's Change Across Time  | Constant |
| SD1 (Standard Deviation 1)   | 4      |
| Time Metric                  | Ignored |
| Specify How Autocorr's Change| 1st Order Autocorr |
| R1 (Autocorrelation)         | 0.7    |
| Max Time Diff.               | 100 (This large value will cause R2 to be ignored.) |

*Reports Tab*

| Numeric Results by Term      | Not checked |
| Numeric Results by Design    | Not checked |
| Regular F Test               | Not checked |
| GG F Test                    | Not checked |
| Wilks' Lambda                | Not checked |
| Pillai-Bartlett              | Not checked |
| Hotelling-Lawley             | Not checked |
| GG Detail Report             | Not checked |
| Means Matrix                 | Not checked |
| Covariance Matrix            | Not checked |
Reports Tab (continued)
Show Plot 1 ............................................ Not checked
Show Plot 2 ............................................ Checked
Test That is Plotted ................................ GG F Test
Max Term-Order Plotted ......................... 2

Output
Click the Run button to perform the calculations and generate the following output.

Plots Section

These charts show how the power depends on the relative size of $Sm$ (i.e. $K$) as well as the group sample size $n$. 
Example 3 – Impact of the Number of Repeated Measurements

Continuing with Example 1, the researchers want to study the impact on power of changing the number of measurements made on each individual. Their experimental protocol calls for four measurements that are 30 minutes apart. They want to see the impact of taking twice that many measurements. To keep the output simple and two the point, they decide to look at the case when \( n = 4 \) and \( K = 1 \).

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Repeated Measures Analysis of Variance procedure window by expanding Means, then ANOVA, then clicking on Analysis of Variance, and then clicking on Repeated Measures Analysis of Variance. You may then make the appropriate entries as listed below, or open Example 3 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Tab</strong></td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>n (Subjects Per Group)</td>
<td>4</td>
</tr>
<tr>
<td>( n )'s</td>
<td>checked</td>
</tr>
<tr>
<td>Means Matrix</td>
<td>blank</td>
</tr>
<tr>
<td>K (Means Multipliers)</td>
<td>1.0</td>
</tr>
<tr>
<td><strong>For First Between-Subject Factor</strong></td>
<td></td>
</tr>
<tr>
<td>Label</td>
<td>B</td>
</tr>
<tr>
<td>Levels</td>
<td>3</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Sm (Standard Deviation of Effects)</td>
<td>93 \ 87 \ 84</td>
</tr>
<tr>
<td><strong>For First Within-Subject Factor</strong></td>
<td></td>
</tr>
<tr>
<td>Label</td>
<td>W</td>
</tr>
<tr>
<td>Levels</td>
<td>4</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Sm (Standard Deviation of Effects)</td>
<td>RANGE 88 \ 93</td>
</tr>
<tr>
<td><strong>Interactions Tab</strong></td>
<td></td>
</tr>
<tr>
<td>2-Way(Mixed) Interaction</td>
<td></td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting</td>
</tr>
<tr>
<td>Sm (Standard Deviation of Effects)</td>
<td>W 100</td>
</tr>
<tr>
<td><strong>Covariance Tab</strong></td>
<td></td>
</tr>
<tr>
<td>Specify Covariance Method</td>
<td>1) Standard Deviations and Autocorrelations</td>
</tr>
<tr>
<td>Time Metric</td>
<td>Range 0 \ 3</td>
</tr>
<tr>
<td>How SD’s Change Across Time</td>
<td>Constant</td>
</tr>
<tr>
<td>SD1 (Standard Deviation 1)</td>
<td>4</td>
</tr>
</tbody>
</table>
Covariance Tab (continued)
Specify How Autocorr’s Change ............ 1st Order Autocorr
R1 (Autocorrelation) ......................... 0.7
Max Time Diff. .................................. 100 (This large value will cause R2 to be ignored.)

Reports Tab
Numeric Results by Term .................... Not checked
Numeric Results by Design .................. Checked
Regular F Test .................................. Not checked
GG F Test ........................................ Checked
Wilks’ Lambda .................................. Not checked
Pillai-Bartlett ................................. Not checked
Hotelling-Lawley .............................. Not checked
GG Detail Report ............................... Not checked
Means Matrix .................................. Not checked
Covariance Matrix ............................. Not checked
Show Plot 1 ..................................... Not checked
Show Plot 2 ..................................... Not checked
Max Term-Order Reported .................... 2

Output
Click the Run button to perform the calculations and generate the following output.

Design Report

<table>
<thead>
<tr>
<th>Term (Results with 4 measurements)</th>
<th>Test</th>
<th>Power</th>
<th>n</th>
<th>Multiply Means By</th>
<th>SD of Effects (Sm)</th>
<th>Standard Deviation (Sigma)</th>
<th>Effect Size</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>B(3)</td>
<td>GG F</td>
<td>0.8478</td>
<td>4</td>
<td>12.00</td>
<td>3.74</td>
<td>3.29</td>
<td>1.14</td>
<td>0.0500</td>
<td>0.1522</td>
</tr>
<tr>
<td>W(4)</td>
<td>GG F</td>
<td>0.9350</td>
<td>4</td>
<td>12.00</td>
<td>1.86</td>
<td>1.31</td>
<td>1.42</td>
<td>0.0500</td>
<td>0.0650</td>
</tr>
<tr>
<td>BW</td>
<td>GG F</td>
<td>0.8348</td>
<td>4</td>
<td>12.00</td>
<td>1.86</td>
<td>1.31</td>
<td>1.42</td>
<td>0.0500</td>
<td>0.1652</td>
</tr>
</tbody>
</table>

(Results with 8 measurements)
| B(3)                              | GG F | 0.8375| 4 | 12.00            | 3.74              | 3.34                        | 1.12        | 0.0500| 0.1625|
| W(8)                              | GG F | 0.9354| 4 | 12.00            | 1.64              | 0.83                        | 1.97        | 0.0500| 0.0646|
| BW                                | GG F | 0.8321| 4 | 12.00            | 1.64              | 0.83                        | 1.97        | 0.0500| 0.1679|

Notice that the power of the between subjects factor decreased slightly, the power of the within-subjects factor increased slightly, and the power of the interaction test decreased slightly. This pattern of increase or decrease depends on all the settings.

We tried varying the value of the autocorrelation from 0.7 to 0.1 and found this to impact the direction of the change in the number of measurements. Hence, our conclusion is that there is no single answer. Changing the number of measurements may increase or decrease the power of a specific test depending on the values of the other parameters.
### Example 4 – Power after a Study

This example will show how to calculate the power of $F$ tests from data that have already been collected and analyzed using the analysis of variance. The following results were obtained using the analysis of variance procedure in NCSS. In this example, Gender is the between factor with two levels and Treatment is the within factor with three levels. The experiment was conducted with two subjects per group, but there is interest in the power for 2, 3, and 4 subjects per group. All significance levels are set to 0.05.

#### Analysis of Variance Table

<table>
<thead>
<tr>
<th>Source Term</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-Ratio</th>
<th>Prob Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>A (Gender)</td>
<td>1</td>
<td>21.33333</td>
<td>21.33333</td>
<td>32.00</td>
<td>0.029857</td>
</tr>
<tr>
<td>B(A)</td>
<td>2</td>
<td>1.333333</td>
<td>0.6666667</td>
<td>6.20</td>
<td>0.059488</td>
</tr>
<tr>
<td>C (Treatment)</td>
<td>2</td>
<td>5.166667</td>
<td>2.583333</td>
<td>6.20</td>
<td>0.059488</td>
</tr>
<tr>
<td>AC</td>
<td>2</td>
<td>5.166667</td>
<td>2.583333</td>
<td>6.20</td>
<td>0.059488</td>
</tr>
<tr>
<td>BC(A)</td>
<td>4</td>
<td>1.666667</td>
<td>0.4166667</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (Adjusted)</td>
<td>11</td>
<td>34.66667</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>12</td>
<td>34.66667</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

#### Means and Effects Section

<table>
<thead>
<tr>
<th>Term</th>
<th>Count</th>
<th>Mean</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>12</td>
<td>17.33333</td>
<td></td>
</tr>
<tr>
<td>A: Gender</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Females</td>
<td>6</td>
<td>16</td>
<td>0.3333333</td>
</tr>
<tr>
<td>Males</td>
<td>6</td>
<td>18.66667</td>
<td>0.3333333</td>
</tr>
<tr>
<td>C: Treatment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L</td>
<td>4</td>
<td>16.75</td>
<td>0.3227486</td>
</tr>
<tr>
<td>M</td>
<td>4</td>
<td>17</td>
<td>0.3227486</td>
</tr>
<tr>
<td>H</td>
<td>4</td>
<td>18.25</td>
<td>0.3227486</td>
</tr>
<tr>
<td>AC: Gender,Treatment</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Females,L</td>
<td>2</td>
<td>14.5</td>
<td>0.4564355</td>
</tr>
<tr>
<td>Females,M</td>
<td>2</td>
<td>16</td>
<td>0.4564355</td>
</tr>
<tr>
<td>Females,H</td>
<td>2</td>
<td>17.5</td>
<td>0.4564355</td>
</tr>
<tr>
<td>Males,L</td>
<td>2</td>
<td>19</td>
<td>0.4564355</td>
</tr>
<tr>
<td>Males,M</td>
<td>2</td>
<td>18</td>
<td>0.4564355</td>
</tr>
<tr>
<td>Males,H</td>
<td>2</td>
<td>19</td>
<td>0.4564355</td>
</tr>
</tbody>
</table>

Note that the treatment means (L, M, and H) show an increasing pattern from 16.75 to 18.25, but the hypothesis test of this factor is not statistically significant at the 0.05 level. We will now calculate the power of the three $F$ tests using PASS. We will use the regular F test since that is what was used in the above table.

From the above printout, we note that $MSB = 0.6666667$ and $MSW = 0.4166667$. Plugging these values into the estimating equations

\[
\hat{\rho} = \frac{MSB - MSW}{MSB + (T - 1)MSW}
\]

\[
\hat{\sigma}^2 = \frac{MSW}{1 - \hat{\rho}}
\]

yields

\[
\hat{\rho} = \frac{0.6666667 - 0.4166667}{0.6666667 + (3 - 1)0.4166667} = 0.1666667
\]
Repeated Measures Analysis of Variance

\[
\sigma^2 = \frac{0.4166667}{1-0.1666667} = 0.5
\]

so that

\[
\hat{\sigma} = \sqrt{0.5} = 0.70710681
\]

With these values calculated, we can setup PASS to calculate the power of the three \( F \) tests as follows.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Repeated Measures Analysis of Variance procedure window by expanding Means, then ANOVA, then clicking on Analysis of Variance, and then clicking on Repeated Measures Analysis of Variance. You may then make the appropriate entries as listed below, or open Example 4 by going to the File menu and choosing Open Example Template. You can see that the values have been loaded into the spreadsheet by clicking on the spreadsheet button.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>n (Subjects Per Group)</td>
<td>2 3 4</td>
</tr>
<tr>
<td>=n's</td>
<td>checked</td>
</tr>
<tr>
<td>Means Matrix</td>
<td>FEMALES-MALES</td>
</tr>
<tr>
<td>K (Means Multipliers)</td>
<td>1.0</td>
</tr>
<tr>
<td>For First Between-Subject Factor</td>
<td></td>
</tr>
<tr>
<td>Label</td>
<td>B</td>
</tr>
<tr>
<td>Levels</td>
<td>2</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting.</td>
</tr>
<tr>
<td>Sm (Standard Deviation of Effects)</td>
<td>Ignored since the Means Matrix is loaded.</td>
</tr>
<tr>
<td>For First Within-Subject Factor</td>
<td></td>
</tr>
<tr>
<td>Label</td>
<td>W</td>
</tr>
<tr>
<td>Levels</td>
<td>3</td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting.</td>
</tr>
<tr>
<td>Sm (Standard Deviation of Effects)</td>
<td>Ignored since the Means Matrix is loaded.</td>
</tr>
<tr>
<td>Interactions Tab</td>
<td></td>
</tr>
<tr>
<td>2-Way(Mixed) Interaction</td>
<td></td>
</tr>
<tr>
<td>Alpha</td>
<td>0.05</td>
</tr>
<tr>
<td>Power</td>
<td>Ignored since this is the Find setting.</td>
</tr>
<tr>
<td>Sm (Standard Deviation of Effects)</td>
<td>Ignored since the Means Matrix is loaded.</td>
</tr>
</tbody>
</table>
Covariance Tab
Specify Covariance Method ...................1) Standard Deviations and Autocorrelations
Time Metric ............................................Ignored
How SD’s Change Across Time ..............Constant
SD1 (Standard Deviation 1) ......................70710681
Specify How Autocorr’s Change ..........Constant
R1 (Autocorrelation) .................................0.16666667
Max Time Diff. ........................................100 (This large value will cause R2 to be ignored.)

Reports Tab
Numeric Results by Term..................Not Checked
Numeric Results by Design ...............Checked
Regular F Test .................................Checked
GG F Test .........................................Not checked
Wilks’ Lambda ..................................Not checked
Pillai-Bartlett ....................................Not checked
Hotelling-Lawley ................................Not checked
GG Detail Report ...............................Not checked
Means Matrix ......................................Not checked
Covariance Matrix .............................Not checked
Test in Summary Statement ...............Regular F Test
Show Plot 1 .......................................Not checked
Show Plot 2 .......................................Not checked
Max Term-Order Reported ...............2

Output
Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Term</th>
<th>Test</th>
<th>Power</th>
<th>n</th>
<th>Multiply Means By (Sm)</th>
<th>SD of Effects (Sigma)</th>
<th>Effect Size</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>B(2)</td>
<td>F</td>
<td>0.8004</td>
<td>2</td>
<td>4 1.00 1.33</td>
<td>0.47 2.83</td>
<td>0.0500</td>
<td>0.1996</td>
<td></td>
</tr>
<tr>
<td>W(3)</td>
<td>F</td>
<td>0.5536</td>
<td>2</td>
<td>4 1.00 0.66</td>
<td>0.37 1.76</td>
<td>0.0500</td>
<td>0.4464</td>
<td></td>
</tr>
<tr>
<td>BW</td>
<td>F</td>
<td>0.5536</td>
<td>2</td>
<td>4 1.00 0.66</td>
<td>0.37 1.76</td>
<td>0.0500</td>
<td>0.4464</td>
<td></td>
</tr>
<tr>
<td>B(2)</td>
<td>F</td>
<td>0.9985</td>
<td>3</td>
<td>6 1.00 1.33</td>
<td>0.47 2.83</td>
<td>0.0500</td>
<td>0.0015</td>
<td></td>
</tr>
<tr>
<td>W(3)</td>
<td>F</td>
<td>0.8933</td>
<td>3</td>
<td>6 1.00 0.66</td>
<td>0.37 1.76</td>
<td>0.0500</td>
<td>0.1067</td>
<td></td>
</tr>
<tr>
<td>BW</td>
<td>F</td>
<td>0.8933</td>
<td>3</td>
<td>6 1.00 0.66</td>
<td>0.37 1.76</td>
<td>0.0500</td>
<td>0.1067</td>
<td></td>
</tr>
<tr>
<td>B(2)</td>
<td>F</td>
<td>1.0000</td>
<td>4</td>
<td>8 1.00 1.33</td>
<td>0.47 2.83</td>
<td>0.0500</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>W(3)</td>
<td>F</td>
<td>0.9801</td>
<td>4</td>
<td>8 1.00 0.66</td>
<td>0.37 1.76</td>
<td>0.0500</td>
<td>0.0199</td>
<td></td>
</tr>
<tr>
<td>BW</td>
<td>F</td>
<td>0.9801</td>
<td>4</td>
<td>8 1.00 0.66</td>
<td>0.37 1.76</td>
<td>0.0500</td>
<td>0.0199</td>
<td></td>
</tr>
</tbody>
</table>

You can see that the power of the tests on W and BW was only 0.55 for an n of 2. However, if n would have been 3, a much more reasonable power of 0.89 would have been achieved.
Example 5 – Cross-Over Design

A crossover design is a special type of repeated measures design in which the treatments are applied to the subjects in different orders. The between-subjects (grouping) factor is defined by the specific sequence in which the treatments are applied. For example, suppose the treatments are represented by B1 and B2. Further suppose that half the subjects receive treatment B1 followed by treatment B2 (sequence B1B2), while the other half receive treatment B2 followed by treatment B1 (sequence B2B1). This is a two-group crossover design.

Crossover designs assume that a long enough period elapses between measurements so that the effects of one treatment are washed out before the next treatment is applied. This is known as the assumption of no carryover effects.

When a crossover design is analyzed using repeated measures, the interaction is the only term of interest. The $F$ test on the between factor tests whether averages across each sequence are equal—a test of little interest. The $F$ test on the within factor tests whether the response is different across the time periods—also of little interest. The $F$ test for interaction tests whether the change in response across time is the same for both sequences. The interaction can only be significant if the treatments effect the outcome differently. Hence, to specify a crossover design requires the careful specification of the interaction effects.

With this background, we present an example. Suppose researchers want to investigate the reduction in heart-beat rate caused by the administration of a certain drug using a simple two-period crossover design. The researchers want a sample size large enough to detect a drop in heart-beat rate from 95 to 90 with a power of 90% at the 0.05 significance level. Previous studies have shown a within-patient autocorrelation of 0.50 and a standard deviation of 3.98. They decide to consider sample sizes between 2 and 8.

The hypothesized interaction is specified by entering the mean heart-beat rates of the four treatment groups as 95, 90, 90, and 95. Since the standard deviation of these values is all that is used, the order of these values does not matter. In this case the sequences means are both 92.5 and the average time-period means are both 92.5. Hence, the interaction effects are 2.5, -2.5, -2.5, and 2.5. You can check that the set of numbers ‘95, 90, 90, 95’ has the same standard deviation as the set ‘2.5, -2.5, -2.5, 2.5’ or even ‘5, 0, 0, 5’. All of these will work.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Repeated Measures Analysis of Variance procedure window by expanding Means, then ANOVA, then clicking on Analysis of Variance, and then clicking on Repeated Measures Analysis of Variance. You may then make the appropriate entries as listed below, or open Example 5 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>n (Subjects Per Group)</td>
<td>2 to 8 by 1</td>
</tr>
<tr>
<td>=n's</td>
<td>checked</td>
</tr>
<tr>
<td>Means Matrix</td>
<td>blank</td>
</tr>
<tr>
<td>K (Means Multipliers)</td>
<td>1.0</td>
</tr>
</tbody>
</table>
Data Tab (continued)

For First Between-Subject Factor
Label..........................................................Seq
Levels..........................................................2
Alpha .........................................................0.05
Power ..........................................................Ignored since this is the Find setting
Sm (Standard Deviation of Effects)........0.1 (This value is arbitrary.)

For First Within-Subject Factor
Label..........................................................Time
Levels..........................................................2
Alpha .........................................................0.05
Power ..........................................................Ignored since this is the Find setting
Sm (Standard Deviation of Effects)........0.1 (This value is arbitrary.)

Interactions Tab
2-Way(Mixed) Interaction
Alpha ..........................................................0.05
Power ..........................................................Ignored since this is the Find setting
Sm (Standard Deviation of Effects)........95 90 90 95

Covariance Tab
Specify Covariance Method .................1) Standard Deviations and Autocorrelations
Time Metric .................................................Range 0 1
How SD’s Change Across Time..............Constant
SD1 (Standard Deviation 1).................3.98
Specify How Autocorr’s Change...........1st Order Autocorr
R1 (Autocorrelation).....................0.5
Max Time Diff. ...........................................100 (This large value will cause R2 to be ignored.)

Reports Tab
Numeric Results by Term.....................Checked
Numeric Results by Design...................Not checked
Regular F Test ........................................Not checked
GG F Test ..................................................Checked
Wilks’ Lambda ..........................................Not checked
Pillai-Bartlett ........................................Not checked
Hotelling-Lawley .....................................Not checked
GG Detail Report ...............................Not checked
Means Matrix .........................................Not checked
Covariance Matrix ..................................Not checked
Show Plot 1 .............................................Not checked
Show Plot 2 .............................................Not checked
Max Term-Order Reported...................2
Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Results for Term SeqTime</th>
<th>Multiply SD of Standard Means Effects</th>
<th>Standard</th>
<th>Effect</th>
<th>Power n N By (Sm) (Sigma) Size Alpha Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>GG F 0.3017 2 4 1.00 2.50 1.99 1.26 0.0500 0.6983</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG F 0.6401 3 6 1.00 2.50 1.99 1.26 0.0500 0.3599</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG F 0.8395 4 8 1.00 2.50 1.99 1.26 0.0500 0.1605</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG F 0.9338 5 10 1.00 2.50 1.99 1.26 0.0500 0.0662</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG F 0.9742 6 12 1.00 2.50 1.99 1.26 0.0500 0.0258</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG F 0.9903 7 14 1.00 2.50 1.99 1.26 0.0500 0.0097</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>GG F 0.9965 8 16 1.00 2.50 1.99 1.26 0.0500 0.0035</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

We only display the interaction term since that is the only term of interest. A quick glance at the plot shows that 90% power is achieved when \( n \) is five. This corresponds to a total sample size of ten subjects.

Example 6 – Power of a Completed Cross-Over Design

The following analysis of variance table was generated by NCSS for a set of crossover data. Find the power of the interaction \( F \) test assuming a significance level of 0.05.

Analysis of Variance Table

<table>
<thead>
<tr>
<th>Source</th>
<th>DF</th>
<th>Sum of Squares</th>
<th>Mean Square</th>
<th>F-Ratio</th>
<th>Prob Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: Sequence</td>
<td>1</td>
<td>89397.6</td>
<td>89397.6</td>
<td>1.19</td>
<td>0.285442</td>
</tr>
<tr>
<td>B(A): Subject</td>
<td>28</td>
<td>2110739</td>
<td>75383.54</td>
<td></td>
<td></td>
</tr>
<tr>
<td>C: Period</td>
<td>1</td>
<td>117395.3</td>
<td>117395.3</td>
<td>1.40</td>
<td>0.246854</td>
</tr>
<tr>
<td>AC</td>
<td>1</td>
<td>122401.7</td>
<td>122401.7</td>
<td>1.46</td>
<td>0.237263</td>
</tr>
<tr>
<td>BC(A)</td>
<td>28</td>
<td>2349752</td>
<td>83919.72</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total (Adjusted)</td>
<td>59</td>
<td>4789686</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>60</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Means Section

<table>
<thead>
<tr>
<th>Term</th>
<th>Count</th>
<th>Mean</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>All</td>
<td>60</td>
<td>492.2000</td>
<td>50.12768</td>
</tr>
<tr>
<td>A: Sequence</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>30</td>
<td>453.6000</td>
<td>50.12768</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>530.8000</td>
<td>50.12768</td>
</tr>
<tr>
<td>C: Period</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>30</td>
<td>447.9667</td>
<td>52.88973</td>
</tr>
<tr>
<td>2</td>
<td>30</td>
<td>536.4333</td>
<td>52.88973</td>
</tr>
<tr>
<td>AC: Sequence,Period</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1,1</td>
<td>15</td>
<td>364.2000</td>
<td>74.79738</td>
</tr>
<tr>
<td>1,2</td>
<td>15</td>
<td>543.0000</td>
<td>74.79738</td>
</tr>
<tr>
<td>2,1</td>
<td>15</td>
<td>531.7333</td>
<td>74.79738</td>
</tr>
<tr>
<td>2,2</td>
<td>15</td>
<td>529.8666</td>
<td>74.79738</td>
</tr>
</tbody>
</table>

One difficulty in analyzing an existing crossover design is determining an appropriate value for the hypothesized interaction effects. One method is to find the standard deviation of the interaction effects by taking the square root of the Sum of Squares for the interaction divided by the total number of observations. In this case,
Another method is to find the individual interaction effects by subtraction. This method proceeds as follows.

First, subtract the Period means from the Sequence by Period means.

\[
\begin{bmatrix}
364.2000 & 531.7333 \\
543.0000 & 529.8666
\end{bmatrix}
-\begin{bmatrix}
447.9667 \\
536.4333
\end{bmatrix}
= \begin{bmatrix}
-83.7667 & 83.7667 \\
6.5667 & -6.5667
\end{bmatrix}
\]

Next, compute the column means and subtract them from the current values. This results in the effects.

\[
\begin{bmatrix}
-83.7667 & 83.7667 \\
6.5667 & -6.5667
\end{bmatrix}
-\begin{bmatrix}
-38.6000 & 38.6000 \\
-38.6000 & 38.6000
\end{bmatrix}
= \begin{bmatrix}
-45.1667 & 45.1667 \\
45.1667 & -45.1667
\end{bmatrix}
\]

Finally, compute the standard deviation of the effects. Since the mean of the effects is zero, the standard deviation is

\[
\sigma_{\text{Interaction}} = \sqrt{\frac{(-45.1667)^2 + (45.1667)^2 + (45.1667)^2 + (-45.1667)^2}{4}}
= 45.1667
\]

Another difficulty that must be solved is to estimate the autocorrelation and within-subject standard deviation. From the above printout, we note that \(MSB = 75383.54\) and \(MSW = 83919.72\). Plugging these values into the estimating equations

\[
\hat{\rho} = \frac{MSB - MSW}{MSB + (T - 1)MSW}
\]

\[
\hat{\sigma}^2 = \frac{MSW}{1 - \hat{\rho}}
\]

yields

\[
\hat{\rho} = \frac{75383.54 - 83919.72}{75383.54 + (2 - 1)83919.72} = -0.05358447
\]

\[
\hat{\sigma}^2 = \frac{83919.72}{1 + 0.05358447} = 79651.63
\]

so that

\[
\hat{\sigma} = \sqrt{79651.63} = 282.2262
\]
Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Repeated Measures Analysis of Variance procedure window by expanding Means, then ANOVA, then clicking on Analysis of Variance, and then clicking on Repeated Measures Analysis of Variance. You may then make the appropriate entries as listed below, or open Example 6 by going to the File menu and choosing Open Example Template.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td>Find (Solve For) ..................................</td>
</tr>
<tr>
<td></td>
<td>n (Subjects Per Group) ................................</td>
</tr>
<tr>
<td></td>
<td>=n’s ..................................................</td>
</tr>
<tr>
<td></td>
<td>Means Matrix .........................................</td>
</tr>
<tr>
<td></td>
<td>K (Means Multipliers) ................................</td>
</tr>
<tr>
<td></td>
<td>For First Between-Subject Factor</td>
</tr>
<tr>
<td></td>
<td>Label ..................................................</td>
</tr>
<tr>
<td></td>
<td>Levels ..................................................</td>
</tr>
<tr>
<td></td>
<td>Alpha ..................................................</td>
</tr>
<tr>
<td></td>
<td>Power ..................................................</td>
</tr>
<tr>
<td></td>
<td>Sm (Standard Deviation of Effects) ..............</td>
</tr>
<tr>
<td></td>
<td>For First Within-Subject Factor</td>
</tr>
<tr>
<td></td>
<td>Label ..................................................</td>
</tr>
<tr>
<td></td>
<td>Levels ..................................................</td>
</tr>
<tr>
<td></td>
<td>Alpha ..................................................</td>
</tr>
<tr>
<td></td>
<td>Power ..................................................</td>
</tr>
<tr>
<td></td>
<td>Sm (Standard Deviation of Effects) ..............</td>
</tr>
<tr>
<td>Interactions Tab</td>
<td>2-Way(Mixed) Interaction</td>
</tr>
<tr>
<td></td>
<td>Alpha ..................................................</td>
</tr>
<tr>
<td></td>
<td>Power ..................................................</td>
</tr>
<tr>
<td></td>
<td>Sm (Standard Deviation of Effects) ..............</td>
</tr>
<tr>
<td>Covariance Tab</td>
<td>Specify Covariance Method ..........................</td>
</tr>
<tr>
<td></td>
<td>How SD’s Change Across Time .......................</td>
</tr>
<tr>
<td></td>
<td>SD1 (Standard Deviation 1) .........................</td>
</tr>
<tr>
<td></td>
<td>Time Metric ...........................................</td>
</tr>
<tr>
<td></td>
<td>Specify How Autocorr’s Change ....................</td>
</tr>
<tr>
<td></td>
<td>R1 (Autocorrelation) ................................</td>
</tr>
<tr>
<td></td>
<td>Max Time Diff. ......................................</td>
</tr>
<tr>
<td></td>
<td>(This large value will cause R2 to be ignored.)</td>
</tr>
<tr>
<td>Reports Tab</td>
<td>Numeric Results by Term ...........................</td>
</tr>
<tr>
<td></td>
<td>Numeric Results by Design .........................</td>
</tr>
<tr>
<td></td>
<td>Regular F Test ......................................</td>
</tr>
<tr>
<td></td>
<td>GG F Test ............................................</td>
</tr>
<tr>
<td></td>
<td>Wilks’ Lambda ......................................</td>
</tr>
</tbody>
</table>
Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Term</th>
<th>Test</th>
<th>Power</th>
<th>n</th>
<th>N</th>
<th>Multiply Means By</th>
<th>SD of Effects (Sm)</th>
<th>Standard Deviation (Sigma)</th>
<th>Effect Size</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>S(2)</td>
<td>GG F</td>
<td>0.1832</td>
<td>15</td>
<td>30</td>
<td>1.00</td>
<td>38.60</td>
<td>194.14</td>
<td>0.20</td>
<td>0.0500</td>
<td>0.8168</td>
</tr>
<tr>
<td>P(2)</td>
<td>GG F</td>
<td>0.2078</td>
<td>15</td>
<td>30</td>
<td>1.00</td>
<td>44.23</td>
<td>204.84</td>
<td>0.22</td>
<td>0.0500</td>
<td>0.7922</td>
</tr>
<tr>
<td>SP</td>
<td>GG F</td>
<td>0.2147</td>
<td>15</td>
<td>30</td>
<td>1.00</td>
<td>45.17</td>
<td>204.84</td>
<td>0.22</td>
<td>0.0500</td>
<td>0.7853</td>
</tr>
</tbody>
</table>

Notice that these power values are low. Fifteen was not a large enough sample size to detect Sm values near 40.

Example 7 – Validation using O’Brien and Muller

O’Brien and Muller’s article in the book edited by Edwards (1993) analyze the power of a two-group repeated-measures experiment in which three measurements are made on each subject.

The hypothesized means are

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Group 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time 1</td>
<td>3</td>
</tr>
<tr>
<td>Time 2</td>
<td>12</td>
</tr>
<tr>
<td>Time 3</td>
<td>8</td>
</tr>
</tbody>
</table>

The covariance matrix is

<table>
<thead>
<tr>
<th></th>
<th>Time 1</th>
<th>Time 2</th>
<th>Time 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Time 1</td>
<td>25</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>Time 2</td>
<td>16</td>
<td>64</td>
<td>30</td>
</tr>
<tr>
<td>Time 3</td>
<td>12</td>
<td>30</td>
<td>36</td>
</tr>
</tbody>
</table>
With \( n \)'s of 12, 18, and 24 and an alpha of 0.05, they obtained power values using the Wilks’ Lambda test. Their reported power values are

<table>
<thead>
<tr>
<th>( n )</th>
<th>Group</th>
<th>Time</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>0.326</td>
<td>0.983</td>
<td>0.461</td>
</tr>
<tr>
<td>18</td>
<td>0.467</td>
<td>0.999</td>
<td>0.671</td>
</tr>
<tr>
<td>24</td>
<td>0.589</td>
<td>0.999</td>
<td>0.814</td>
</tr>
</tbody>
</table>

O’Brien, in a private communication, re-ran these data using the Geisser-Greenhouse correction. His results were as follows:

<table>
<thead>
<tr>
<th>( n )</th>
<th>Group</th>
<th>Time</th>
<th>Interaction</th>
</tr>
</thead>
<tbody>
<tr>
<td>12</td>
<td>0.326</td>
<td>0.993</td>
<td>0.486</td>
</tr>
<tr>
<td>18</td>
<td>0.467</td>
<td>0.999</td>
<td>0.685</td>
</tr>
<tr>
<td>24</td>
<td>0.589</td>
<td>0.999</td>
<td>0.819</td>
</tr>
</tbody>
</table>

In order to run this example in \textit{PASS}, the values of the means and the covariance matrix (given above) must be entered on a spreadsheet. We have loaded these values into the database called OBRIEN. Either enter the values yourself, or load the OBRIEN database which should be in the Data directory. The instructions below assume that the means are in columns one and two, while the covariance matrix is in columns four through six of the current database.

### Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the \textit{Repeated Measures Analysis of Variance} procedure window by expanding \textit{Means}, then \textit{ANOVA}, then clicking on \textit{Analysis of Variance}, and then clicking on \textit{Repeated Measures Analysis of Variance}. You may then make the appropriate entries as listed below, or open Example 7 by going to the File menu and choosing \textit{Open Example Template}. You can see that the values have been loaded into the spreadsheet by clicking on the spreadsheet button.

<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data Tab</td>
<td></td>
</tr>
<tr>
<td>\textit{Find (Solve For)}</td>
<td>\textit{Power and Beta}</td>
</tr>
<tr>
<td>( n ) (Subjects Per Group)</td>
<td>12 18 24</td>
</tr>
<tr>
<td>( =n's )</td>
<td>checked</td>
</tr>
<tr>
<td>Means Matrix</td>
<td>M1-M2</td>
</tr>
<tr>
<td>( K ) (Means Multipliers)</td>
<td>1.0</td>
</tr>
</tbody>
</table>

\textit{For First Between-Subject Factor}

| Label                        | G                      |
| Levels                       | 2                      |
| Alpha                        | 0.05                   |
| Power                        | Ignored since this is the Find setting |
| \( Sm \) (Standard Deviation of Effects) | Ignored |
Data Tab (continued)

For First Within-Subject Factor
Label........................................T
Levels........................................3
Alpha ............................................0.05
Power ........................................Ignored since this is the Find setting
Sm (Standard Deviation of Effects)......Ignored

Interactions Tab
2-Way(Mixed) Interaction
Alpha ............................................0.05
Power ........................................Ignored since this is the Find setting
Sm (Standard Deviation of Effects)......Ignored

Covariance Tab
Specify Covariance Method ..............2) Covariance Matrix Variables
Spreadsheet Columns......................S1-S3

Reports Tab
Numeric Results by Term..................Checked
Numeric Results by Design................Not checked
Regular F Test .................................Not checked
GG F Test .........................................Checked
Wilks' Lambda ..................................Checked
Pillai-Bartlett ..................................Not checked
Hotelling-Lawley ..............................Not checked
GG Detail Report ..............................Not checked
Means Matrix ....................................Not checked
Covariance Matrix ..............................Not checked
Show Summary Statements .................Not checked
Show Plot 1 .....................................Not checked
Show Plot 2 .....................................Not checked
Max Term-Order Reported...............2

Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Test</th>
<th>Power</th>
<th>n</th>
<th>N</th>
<th>Multiply Means By</th>
<th>SD of Effects (Sm)</th>
<th>Standard Deviation (Sigma)</th>
<th>Effect Size</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>GG F</td>
<td>0.3263</td>
<td>12</td>
<td>24</td>
<td>1.00</td>
<td>1.67</td>
<td>5.17</td>
<td>0.32</td>
<td>0.0500</td>
<td>0.6737</td>
</tr>
<tr>
<td>Wilks</td>
<td>0.3263</td>
<td>12</td>
<td>24</td>
<td>1.00</td>
<td>1.67</td>
<td>5.17</td>
<td>0.32</td>
<td>0.0500</td>
<td>0.6737</td>
</tr>
<tr>
<td>GG F</td>
<td>0.4673</td>
<td>18</td>
<td>36</td>
<td>1.00</td>
<td>1.67</td>
<td>5.17</td>
<td>0.32</td>
<td>0.0500</td>
<td>0.5327</td>
</tr>
<tr>
<td>Wilks</td>
<td>0.4673</td>
<td>18</td>
<td>36</td>
<td>1.00</td>
<td>1.67</td>
<td>5.17</td>
<td>0.32</td>
<td>0.0500</td>
<td>0.5327</td>
</tr>
<tr>
<td>GG F</td>
<td>0.5889</td>
<td>24</td>
<td>48</td>
<td>1.00</td>
<td>1.67</td>
<td>5.17</td>
<td>0.32</td>
<td>0.0500</td>
<td>0.4111</td>
</tr>
<tr>
<td>Wilks</td>
<td>0.5889</td>
<td>24</td>
<td>48</td>
<td>1.00</td>
<td>1.67</td>
<td>5.17</td>
<td>0.32</td>
<td>0.0500</td>
<td>0.4111</td>
</tr>
</tbody>
</table>
Results for Factor T (Levels =3)

<table>
<thead>
<tr>
<th>Test</th>
<th>Power</th>
<th>n</th>
<th>N</th>
<th>Multiply Means By</th>
<th>SD of Effects (Sm)</th>
<th>Standard Deviation (Sigma)</th>
<th>Effect Size</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>GG F</td>
<td>0.9933</td>
<td>12</td>
<td>24</td>
<td>1.00</td>
<td>2.86</td>
<td>2.73</td>
<td>1.05</td>
<td>0.0500</td>
<td>0.0067</td>
</tr>
<tr>
<td>Wilks</td>
<td>0.9825</td>
<td>12</td>
<td>24</td>
<td>1.00</td>
<td>2.86</td>
<td>2.73</td>
<td>1.05</td>
<td>0.0500</td>
<td>0.0175</td>
</tr>
<tr>
<td>GG F</td>
<td>0.9999</td>
<td>18</td>
<td>36</td>
<td>1.00</td>
<td>2.86</td>
<td>2.73</td>
<td>1.05</td>
<td>0.0500</td>
<td>0.0001</td>
</tr>
<tr>
<td>Wilks</td>
<td>0.9995</td>
<td>18</td>
<td>36</td>
<td>1.00</td>
<td>2.86</td>
<td>2.73</td>
<td>1.05</td>
<td>0.0500</td>
<td>0.0005</td>
</tr>
<tr>
<td>GG F</td>
<td>1.0000</td>
<td>24</td>
<td>48</td>
<td>1.00</td>
<td>2.86</td>
<td>2.73</td>
<td>1.05</td>
<td>0.0500</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Results for Term GT

<table>
<thead>
<tr>
<th>Test</th>
<th>Power</th>
<th>n</th>
<th>N</th>
<th>Multiply Means By</th>
<th>SD of Effects (Sm)</th>
<th>Standard Deviation (Sigma)</th>
<th>Effect Size</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>GG F</td>
<td>0.4861</td>
<td>12</td>
<td>24</td>
<td>1.00</td>
<td>1.31</td>
<td>2.73</td>
<td>0.48</td>
<td>0.0500</td>
<td>0.5139</td>
</tr>
<tr>
<td>Wilks</td>
<td>0.4605</td>
<td>12</td>
<td>24</td>
<td>1.00</td>
<td>1.31</td>
<td>2.73</td>
<td>0.48</td>
<td>0.0500</td>
<td>0.5395</td>
</tr>
<tr>
<td>GG F</td>
<td>0.6850</td>
<td>18</td>
<td>36</td>
<td>1.00</td>
<td>1.31</td>
<td>2.73</td>
<td>0.48</td>
<td>0.0500</td>
<td>0.3150</td>
</tr>
<tr>
<td>Wilks</td>
<td>0.6706</td>
<td>18</td>
<td>36</td>
<td>1.00</td>
<td>1.31</td>
<td>2.73</td>
<td>0.48</td>
<td>0.0500</td>
<td>0.3294</td>
</tr>
<tr>
<td>GG F</td>
<td>0.8193</td>
<td>24</td>
<td>48</td>
<td>1.00</td>
<td>1.31</td>
<td>2.73</td>
<td>0.48</td>
<td>0.0500</td>
<td>0.1807</td>
</tr>
<tr>
<td>Wilks</td>
<td>0.8136</td>
<td>24</td>
<td>48</td>
<td>1.00</td>
<td>1.31</td>
<td>2.73</td>
<td>0.48</td>
<td>0.0500</td>
<td>0.1864</td>
</tr>
</tbody>
</table>

PASS agrees exactly with O’Brien’s calculations.

Example 8 – Unequal Group Sizes

Usually, in the planning stages, the group sample sizes are equal. Occasionally, however, you may want to plan for a situation in which one group will have a much larger sample size than the others. Also, when doing a power analysis on a study that has already been conducted, the group sample sizes are often unequal.

In this example, we will re-analyze the Example 4. However, we will now assume that there were four subjects in group 1 and eight subjects in group 2. The setup and output for this example are as follows.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Repeated Measures Analysis of Variance procedure window by expanding Means, then ANOVA, then clicking on Analysis of Variance, and then clicking on Repeated Measures Analysis of Variance. You may then make the appropriate entries as listed below, or open Example 8 by going to the File menu and choosing Open Example Template. You can see that the values have been loaded into the spreadsheet by clicking on the spreadsheet button.
**Option**          | **Value**
---|---
**Data Tab**
Find (Solve For) | Power and Beta
n (Subjects Per Group) | 4 8
= n's | Not checked
Means Matrix | FEMALES-MALES
K (Means Multipliers) | 1.0

*For First Between-Subject Factor*
Label | B
Levels | 2
Alpha | 0.05

*For First Within-Subject Factor*
Label | W
Levels | 3
Alpha | 0.05

*Interactions Tab*
2-Way (Mixed) Interaction
Alpha | 0.05

*Covariance Tab*
Specify Covariance Method | 1) Standard Deviations and Autocorrelations
Time Metric | Ignored
How SD’s Change Across Time | Constant
SD1 (Standard Deviation 1) | 70710681
Specify How Autocorr’s Change | Constant
R1 (Autocorrelation) | 0.16666667
Max Time Diff | 100 (This large value will cause R2 to be ignored.)

*Reports Tab*
Numeric Results by Term | Not Checked
Numeric Results by Design | Checked
Regular F Test | Not checked
GG F Test | Checked
Wilks’ Lambda | Not checked
Pillai-Bartlett | Not checked
Hotelling-Lawley | Not checked
GG Detail Report | Not checked
Means Matrix | Not checked
Covariance Matrix | Not checked
Show Plot 1 | Not checked
Show Plot 2 | Not checked
Max Term-Order Reported | 2
Output

Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Term</th>
<th>Test</th>
<th>Power</th>
<th>n</th>
<th>N</th>
<th>By</th>
<th>SD of Effects</th>
<th>Standard Deviation</th>
<th>Effect Size</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>B(2)</td>
<td>GG F</td>
<td>1.0000</td>
<td>6.0</td>
<td>12</td>
<td>1.0</td>
<td>1.26</td>
<td>0.47</td>
<td>2.67</td>
<td>0.0500</td>
<td>0.0000</td>
</tr>
<tr>
<td>W(3)</td>
<td>GG F</td>
<td>0.9983</td>
<td>6.0</td>
<td>12</td>
<td>1.0</td>
<td>0.62</td>
<td>0.37</td>
<td>1.66</td>
<td>0.0500</td>
<td>0.0017</td>
</tr>
<tr>
<td>BW</td>
<td>GG F</td>
<td>0.9983</td>
<td>6.0</td>
<td>12</td>
<td>1.0</td>
<td>0.62</td>
<td>0.37</td>
<td>1.66</td>
<td>0.0500</td>
<td>0.0017</td>
</tr>
</tbody>
</table>

n’s: 4 8

Notice that the values of $n$ are now shown to one decimal place. That is because the value reported is the average value of $n$. The actual $n$’s are shown following the report.

Example 9 – Designs with More Than Two Factors

Occasionally, you will have a design that has more than two factors. We will now show you how to compute the necessary sample size for such a design.

Suppose your design calls for two between-subject factors, Age (A) and Gender (G), and two within-subject factors, Dose-Level (D) and Application-Method (M). Suppose the number of levels of these four factors are, respectively, 3, 2, 4, and 2.

Our first task is to determine appropriate values of $S_m$ for each of the terms. We decide to ignore the interactions during the planning and only consider the factors themselves. The desired difference to be detected among the three age groups can be represented by the means 80, 88, and 96. The desired difference to be detected among the two genders can be represented by the means 80 and 96. The desired difference to be detected among the four dose levels is represented by the means 80, 82, 84, and 86. The desired difference to be detected among the two application methods is represented by the means 80 and 86.

Our next task is to specify the covariance matrix. From previous experience, we have found that a constant value of 20.0 is appropriate for SD1. An autocorrelation of 0.5 with a first-order autocorrelation pattern is also appropriate.

Finally, we decide to calculate the power using the GG F test at the following sample sizes: 2, 4, 6, 8, 10, 20, 30, and 40.

Setup

This section presents the values of each of the parameters needed to run this example. First, from the PASS Home window, load the Repeated Measures Analysis of Variance procedure window by expanding Means, then ANOVA, then clicking on Analysis of Variance, and then clicking on Repeated Measures Analysis of Variance. You may then make the appropriate entries as listed below, or open Example 9 by going to the File menu and choosing Open Example Template.
<table>
<thead>
<tr>
<th>Option</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Data Tab</strong></td>
<td></td>
</tr>
<tr>
<td>Find (Solve For)</td>
<td>Power and Beta</td>
</tr>
<tr>
<td>n (Subjects Per Group)</td>
<td>2 4 6 8 10 20 30 40</td>
</tr>
<tr>
<td>=n’s.</td>
<td>Checked</td>
</tr>
<tr>
<td>Means Matrix</td>
<td>blank</td>
</tr>
<tr>
<td>K (Means Multipliers)</td>
<td>1.0</td>
</tr>
</tbody>
</table>

*For First Between-Subject Factor*
- Label: A
- Levels: 3
- Alpha: 0.05
- Sm (Standard Deviation of Effects): 80 88 96

*For Second Between-Subject Factor*
- Label: G
- Levels: 2
- Alpha: 0.05
- Sm (Standard Deviation of Effects): 80 96

*For First Within-Subject Factor*
- Label: D
- Levels: 4
- Alpha: 0.05
- Sm (Standard Deviation of Effects): 80 82 84 86

*For Second Within-Subject Factor*
- Label: M
- Levels: 2
- Alpha: 0.05
- Sm (Standard Deviation of Effects): 80 86

**Interactions Tab**
- All Interactions
  - Alpha: 0.05
  - Sm (Standard Deviation of Effects): D 100

**Covariance Tab**
- Specify Covariance Method: 1) Standard Deviations and Autocorrelations
- Time Metric: Step 0 1
- How SD’s Change Across Time: Constant
- SD1 (Standard Deviation 1): 20
- Specify How Autocorr’s Change: 1st Order Autocorrelation
- R1 (Autocorrelation): 0.5
- Max Time Diff.: 100 (This large value will cause R2 to be ignored.)
Reports Tab
Numeric Results by Term..........................Not checked
Numeric Results by Design.......................Checked
Regular F Test......................................Not checked
GG F Test............................................Checked
Wilks’ Lambda........................................Not checked
Pillai-Bartlett......................................Not checked
Hotelling-Lawley....................................Not checked
GG Detail Report....................................Not checked
Means Matrix........................................Not checked
Covariance Matrix...................................Not checked
Test in Summary Statement......................Regular F Test
Show Plot 1.........................................Checked
Show Plot 2.........................................Not checked
Test That is Plotted................................GG F Test
Max Term-Order Plotted...........................1
Max Term-Order Reported.........................1

Output
Click the Run button to perform the calculations and generate the following output.

Numeric Results

<table>
<thead>
<tr>
<th>Term</th>
<th>Test</th>
<th>Power</th>
<th>n</th>
<th>N</th>
<th>By</th>
<th>SD of Effects (Sm)</th>
<th>Standard Deviation (Sigma)</th>
<th>Effect Size</th>
<th>Alpha</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>a(3)</td>
<td>GG F</td>
<td>0.2735</td>
<td>2</td>
<td>12</td>
<td>1.00</td>
<td>6.53</td>
<td>11.18</td>
<td>0.58</td>
<td>0.0500</td>
<td>0.7265</td>
</tr>
<tr>
<td>g(2)</td>
<td>GG F</td>
<td>0.5465</td>
<td>2</td>
<td>12</td>
<td>1.00</td>
<td>8.00</td>
<td>11.18</td>
<td>0.72</td>
<td>0.0500</td>
<td>0.4535</td>
</tr>
<tr>
<td>d(4)</td>
<td>GG F</td>
<td>0.0506</td>
<td>2</td>
<td>12</td>
<td>1.00</td>
<td>2.24</td>
<td>7.64</td>
<td>0.29</td>
<td>0.0500</td>
<td>0.9494</td>
</tr>
<tr>
<td>m(2)</td>
<td>GG F</td>
<td>0.5072</td>
<td>2</td>
<td>12</td>
<td>1.00</td>
<td>3.00</td>
<td>4.41</td>
<td>0.68</td>
<td>0.0500</td>
<td>0.4928</td>
</tr>
<tr>
<td>a(3)</td>
<td>GG F</td>
<td>0.6493</td>
<td>4</td>
<td>24</td>
<td>1.00</td>
<td>6.53</td>
<td>11.18</td>
<td>0.58</td>
<td>0.0500</td>
<td>0.3507</td>
</tr>
<tr>
<td>g(2)</td>
<td>GG F</td>
<td>0.9118</td>
<td>4</td>
<td>24</td>
<td>1.00</td>
<td>8.00</td>
<td>11.18</td>
<td>0.72</td>
<td>0.0500</td>
<td>0.0082</td>
</tr>
<tr>
<td>d(4)</td>
<td>GG F</td>
<td>0.1563</td>
<td>4</td>
<td>24</td>
<td>1.00</td>
<td>2.24</td>
<td>7.64</td>
<td>0.29</td>
<td>0.0500</td>
<td>0.8437</td>
</tr>
<tr>
<td>m(2)</td>
<td>GG F</td>
<td>0.8833</td>
<td>4</td>
<td>24</td>
<td>1.00</td>
<td>3.00</td>
<td>4.41</td>
<td>0.68</td>
<td>0.0500</td>
<td>0.1167</td>
</tr>
<tr>
<td>a(3)</td>
<td>GG F</td>
<td>0.8556</td>
<td>6</td>
<td>36</td>
<td>1.00</td>
<td>6.53</td>
<td>11.18</td>
<td>0.58</td>
<td>0.0500</td>
<td>0.1444</td>
</tr>
<tr>
<td>g(2)</td>
<td>GG F</td>
<td>0.9858</td>
<td>6</td>
<td>36</td>
<td>1.00</td>
<td>8.00</td>
<td>11.18</td>
<td>0.72</td>
<td>0.0500</td>
<td>0.0142</td>
</tr>
<tr>
<td>d(4)</td>
<td>GG F</td>
<td>0.2412</td>
<td>6</td>
<td>36</td>
<td>1.00</td>
<td>2.24</td>
<td>7.64</td>
<td>0.29</td>
<td>0.0500</td>
<td>0.7588</td>
</tr>
<tr>
<td>m(2)</td>
<td>GG F</td>
<td>0.9767</td>
<td>6</td>
<td>36</td>
<td>1.00</td>
<td>3.00</td>
<td>4.41</td>
<td>0.68</td>
<td>0.0500</td>
<td>0.0233</td>
</tr>
<tr>
<td>a(3)</td>
<td>GG F</td>
<td>0.9471</td>
<td>8</td>
<td>48</td>
<td>1.00</td>
<td>6.53</td>
<td>11.18</td>
<td>0.58</td>
<td>0.0500</td>
<td>0.0529</td>
</tr>
<tr>
<td>g(2)</td>
<td>GG F</td>
<td>0.9980</td>
<td>8</td>
<td>48</td>
<td>1.00</td>
<td>8.00</td>
<td>11.18</td>
<td>0.72</td>
<td>0.0500</td>
<td>0.0020</td>
</tr>
<tr>
<td>d(4)</td>
<td>GG F</td>
<td>0.3245</td>
<td>8</td>
<td>48</td>
<td>1.00</td>
<td>2.24</td>
<td>7.64</td>
<td>0.29</td>
<td>0.0500</td>
<td>0.6755</td>
</tr>
<tr>
<td>m(2)</td>
<td>GG F</td>
<td>0.9959</td>
<td>8</td>
<td>48</td>
<td>1.00</td>
<td>3.00</td>
<td>4.41</td>
<td>0.68</td>
<td>0.0500</td>
<td>0.0041</td>
</tr>
<tr>
<td>a(3)</td>
<td>GG F</td>
<td>0.9823</td>
<td>10</td>
<td>60</td>
<td>1.00</td>
<td>6.53</td>
<td>11.18</td>
<td>0.58</td>
<td>0.0500</td>
<td>0.0177</td>
</tr>
<tr>
<td>g(2)</td>
<td>GG F</td>
<td>0.9997</td>
<td>10</td>
<td>60</td>
<td>1.00</td>
<td>8.00</td>
<td>11.18</td>
<td>0.72</td>
<td>0.0500</td>
<td>0.0003</td>
</tr>
<tr>
<td>d(4)</td>
<td>GG F</td>
<td>0.4054</td>
<td>10</td>
<td>60</td>
<td>1.00</td>
<td>2.24</td>
<td>7.64</td>
<td>0.29</td>
<td>0.0500</td>
<td>0.5946</td>
</tr>
<tr>
<td>m(2)</td>
<td>GG F</td>
<td>0.9993</td>
<td>10</td>
<td>60</td>
<td>1.00</td>
<td>3.00</td>
<td>4.41</td>
<td>0.68</td>
<td>0.0500</td>
<td>0.0007</td>
</tr>
<tr>
<td>a(3)</td>
<td>GG F</td>
<td>1.0000</td>
<td>20</td>
<td>120</td>
<td>1.00</td>
<td>6.53</td>
<td>11.18</td>
<td>0.58</td>
<td>0.0500</td>
<td>0.0000</td>
</tr>
<tr>
<td>g(2)</td>
<td>GG F</td>
<td>1.0000</td>
<td>20</td>
<td>120</td>
<td>1.00</td>
<td>8.00</td>
<td>11.18</td>
<td>0.72</td>
<td>0.0500</td>
<td>0.0000</td>
</tr>
<tr>
<td>d(4)</td>
<td>GG F</td>
<td>0.7286</td>
<td>20</td>
<td>120</td>
<td>1.00</td>
<td>2.24</td>
<td>7.64</td>
<td>0.29</td>
<td>0.0500</td>
<td>0.2714</td>
</tr>
<tr>
<td>m(2)</td>
<td>GG F</td>
<td>1.0000</td>
<td>20</td>
<td>120</td>
<td>1.00</td>
<td>3.00</td>
<td>4.41</td>
<td>0.68</td>
<td>0.0500</td>
<td>0.0000</td>
</tr>
</tbody>
</table>
This report gives the power values for the various terms and sample sizes that were entered. It is much easier to consider the following plot to interpret the results.

### Plots Section

![Power vs n by Terms](image)

From this chart, we can see that the first within-subject factor, dose level, has a power much lower than the other factors. Looking at the Sm values in the numeric table, we find that the Sm value for factor D is much less than for the other values. This explains why its power is so poor. Our options are to either increase the sample size or increase the value of Sm for factor D.
Repeated Measures Analysis of Variance
References

A


B


experiments. You definitely should acquire and study this book if you plan anything but a casual acquaintance with experimental design. The book is loaded with examples and explanations.


---

**C**


This is one of the classic books on experimental design, first published in 1957.


Cohen, Jacob. 1988. *Statistical Power Analysis for the Behavioral Sciences*, Lawrence Erlbaum Associates, Hillsdale, New Jersey. This is a very nice, clearly written book. There are MANY examples. It is the largest of the sample size books. It does not deal with clinical trials.


Conlon, M. and Thomas, R. 1993. “The Power Function for Fisher’s Exact Test.” *Applied Statistics*, Volume 42, No. 1, pages 258-260. This article was used to validate the power calculations of Fisher’s Exact Test in PASS. Unfortunately, we could not use the algorithm to improve the speed because the algorithm requires equal sample sizes.


Davies, Owen L. 1971. *The Design and Analysis of Industrial Experiments*. Hafner Publishing Company, New York. This was one of the first books on experimental design and analysis. It has many examples and is highly recommended.


Donner, Allan. 1984. “Approaches to Sample Size Estimation in the Design of Clinical Trials--A Review.” *Statistics in Medicine*, Volume 3, pages 199-214. This is a well done review of the clinical trial literature. Although it is becoming out of date, it is still a good place to start.


**References-4**


---

**E**


---

**F**


**G**


**H**


Hsieh, F.Y. 1989. “Sample Size Tables for Logistic Regression,” Statistics in Medicine, Volume 8, pages 795-802. This is the article that was the basis for the sample size calculations in logistic regression in PASS 6.0. It has been superceded by the 1998 article.


Hsieh, F.Y. and Lavori, P.W. 2000. “Sample-Size Calculations for the Cox Proportional Hazards Regression Model with Nonbinary Covariates,” Controlled Clinical Trials, Volume 21, pages 552-560. The sample size calculation for Cox regression in PASS are based on this article.


J


Lachin, John M. 2000. *Biostatistical Methods*. John Wiley & Sons. New York. This is a graduate-level methods book that deals with statistical methods that are of interest to biostatisticians such as odds ratios, relative risks, regression analysis, case-control studies, and so on.


Lawson, John. 1987. *Basic Industrial Experimental Design Strategies*. Center for Statistical Research at Brigham Young University. Provo, Utah. 84602. This is a manuscript used by Dr. Lawson in courses and workshops that he provides to industrial engineers. It is the basis for many of our experimental design procedures.


Myers, R.H. 1990. Classical and Modern Regression with Applications. PWS-Kent Publishing Company. Boston, Massachusetts. This is one of the bibles on the topic of regression analysis.


Prihoda, Tom. 1983. “Convenient Power Analysis For Complex Analysis of Variance Models.” *Poster Session of the American Statistical Association Joint Statistical Meetings*, August 15-18, 1983, Toronto, Canada. Tom is currently at the University of Texas Health Science Center. This article includes FORTRAN code for performing power analysis.

Ramsey, Philip H. 1978 “Power Differences Between Pairwise Multiple Comparisons,” *JASA*, vol. 73, no. 363, pages 479-485.


Rencher, Alvin C. 1998. *Multivariate Statistical Inference and Applications*. John Wiley. New York, New York. This book provides a comprehensive mixture of theoretical and applied results in multivariate analysis. My evaluation may be biased since Al Rencher took me fishing when I was his student.


Shuster, Jonathan J. 1990. CRC Handbook of Sample Size Guidelines for Clinical Trials. CRC Press, Boca Raton, Florida. This is an expensive book ($300) of tables for running log-rank tests. It is well documented, but at this price it better be.


Therneau, T.M. and Grambsch, P.M. 2000. Modeling Survival Data. Springer: New York, New York. A the time of the writing of the Cox regression procedure, this book provides a thorough, up-to-date discussion of this procedure as well as many extensions to it. Recommended, especially to those with at least a masters in statistics.


Welch, B.L. 1938. "The significance of the difference between two means when the population variances are unequal.” *Biometrika*, 29, 350-362.
Wilson, E.B.1927. “Probable Inference, the Law of Succession, and Statistical Inference,” *Journal of the American Statistical Association*, Volume 22, pages 209-212. This article discusses the ‘score’ method that has become popular when dealing with proportions.


Z


## Index

Index entries are of the form “chapter-page”. A list of chapters is given in the Table of Contents.

### A

<table>
<thead>
<tr>
<th>Abbreviations</th>
<th>4-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute deviation</td>
<td>standard deviation estimator, 905-3</td>
</tr>
<tr>
<td>Accrual time</td>
<td>logrank, 705-6</td>
</tr>
<tr>
<td>A-efficiency</td>
<td>D-optimal designs, 888-13</td>
</tr>
<tr>
<td>Agreement between two raters - kappa</td>
<td>811-1</td>
</tr>
<tr>
<td>Alias</td>
<td>two-level designs, 881-2</td>
</tr>
<tr>
<td>Alignment</td>
<td>histogram, 945-4</td>
</tr>
<tr>
<td></td>
<td>scatter plot, 940-2</td>
</tr>
<tr>
<td>All-contrasts power</td>
<td>simulation, 590-4</td>
</tr>
<tr>
<td>All-pair power</td>
<td>multiple comparisons - simulation, 580-4</td>
</tr>
<tr>
<td></td>
<td>simulation, 585-4</td>
</tr>
<tr>
<td>Alpha</td>
<td>6-3</td>
</tr>
<tr>
<td></td>
<td>adjusting, 8-5</td>
</tr>
<tr>
<td>Alpha - simulation</td>
<td>multiple comparisons, 580-1, 585-2</td>
</tr>
<tr>
<td></td>
<td>multiple contrasts, 590-1</td>
</tr>
<tr>
<td>Alpha spending function</td>
<td>survival - group sequential, 710-1, 711-1</td>
</tr>
<tr>
<td></td>
<td>two means - group sequential, 475-1</td>
</tr>
<tr>
<td></td>
<td>two proportions - group sequential, 220-1</td>
</tr>
<tr>
<td>Alternative hypothesis</td>
<td>6-1, 8-1</td>
</tr>
<tr>
<td>Analysis of covariance</td>
<td>551-1</td>
</tr>
<tr>
<td>Analysis of variance</td>
<td>fixed effects ANOVA, 560-1</td>
</tr>
<tr>
<td></td>
<td>one-way, 550-1</td>
</tr>
<tr>
<td></td>
<td>one-way - simulation, 555-1</td>
</tr>
<tr>
<td></td>
<td>randomized block, 565-1</td>
</tr>
<tr>
<td></td>
<td>repeated measures, 570-1</td>
</tr>
<tr>
<td></td>
<td>three-way, 560-1</td>
</tr>
<tr>
<td></td>
<td>two-way, 560-1</td>
</tr>
<tr>
<td>ANOVA - fixed effects</td>
<td>examples, 560-13</td>
</tr>
<tr>
<td></td>
<td>validation, 560-19</td>
</tr>
<tr>
<td>ANOVA - one-way</td>
<td>assumptions, 550-2</td>
</tr>
<tr>
<td></td>
<td>examples, 550-10</td>
</tr>
<tr>
<td></td>
<td>validation, 550-19</td>
</tr>
<tr>
<td>ANOVA - one-way - simulation</td>
<td>examples, 555-10</td>
</tr>
<tr>
<td></td>
<td>validation, 555-16</td>
</tr>
<tr>
<td>ANOVA - random block</td>
<td>examples, 565-9</td>
</tr>
<tr>
<td></td>
<td>validation, 565-11</td>
</tr>
<tr>
<td>ANOVA - repeated measures</td>
<td>examples, 570-29</td>
</tr>
<tr>
<td></td>
<td>validation, 570-49</td>
</tr>
<tr>
<td>Any-contrast power</td>
<td>simulation, 590-4</td>
</tr>
<tr>
<td>Any-pair power</td>
<td>multiple comparisons - simulation, 580-4</td>
</tr>
<tr>
<td></td>
<td>simulation, 585-4</td>
</tr>
<tr>
<td>Assigning subjects to groups</td>
<td>880-1</td>
</tr>
<tr>
<td>AUC</td>
<td>ROC curve (one), 260-3</td>
</tr>
<tr>
<td></td>
<td>ROC curves (two), 265-1, 265-6</td>
</tr>
<tr>
<td>Autocorrelation</td>
<td>repeated measures ANOVA, 570-25</td>
</tr>
<tr>
<td>Average absolute deviation</td>
<td>standard deviation estimator, 905-3</td>
</tr>
<tr>
<td>Axis</td>
<td>histogram, 945-3</td>
</tr>
<tr>
<td></td>
<td>parameters, 4-10</td>
</tr>
<tr>
<td></td>
<td>scatter plot, 940-1</td>
</tr>
<tr>
<td>Axis line</td>
<td>histogram, 945-5</td>
</tr>
<tr>
<td></td>
<td>scatter plot, 940-3</td>
</tr>
</tbody>
</table>
B

Background color
  histogram, 945-11
  scatter plot, 940-11
Balaam's design, 528-1, 530-1, 533-1, 535-1, 540-1, 545-1
Balance, 880-1
Balanced incomplete block designs, 883-1
  examples, 883-3
Bars
  histogram, 945-1
  scatter plot, 940-11
Beta
  calculating, 6-4
  probability calculator, 915-1
  simulation, 920-3
Between standard deviation
  repeated measures ANOVA, 570-14
BIB designs, 883-1
Bimodal data
  simulation, 920-22
Bin width
  histogram, 945-2
Binomial distribution
  probability calculator, 915-2
  simulation, 920-5
Binomial model
  one proportion, 100-2, 270-2
Binomial probabilities, 6-2
Binomial test
  one proportion, 100-4
Binormal model
  ROC curve (one), 260-2
  ROC curves (two), 265-2
Bins
  histogram, 945-1
Bioequivalence
  two proportions, 215-1
Bivariate normal distribution
  probability calculator, 915-2
Blackwelder
  risk ratio, 205-26
Blocking
  two-level designs, 881-2
Bonferroni adjustment
  t-test - one group, 610-5
  t-test - two groups, 615-4
Bonferroni test
  multiple contrasts - simulation, 590-3
Bootstrap test
  simulation, 410-3
Bootstrap test - paired means
  simulation, 490-4
Bootstrap test - paired means - equivalence
  simulation, 495-5
Border plots
  histogram, 945-10
Bottom bar
  output window, 5-10
Boundaries
  histogram, 945-3
  scatter plot, 940-1
Box plot
  histogram, 945-11
Box-Behnken designs, 885-1
Brownian motion
  survival - group sequential, 710-5
  two means - group sequential, 475-4
  two proportions - group sequential, 220-4
Buttons
  output window, 5-8
  PASS home window, 3-3
  procedure window, 4-5

C

Candidate points
  D-optimal designs, 888-14
Carryover effects
  repeated measures ANOVA, 570-3
Case-control
  matched, 155-1
Cauchy distribution
  simulation, 920-5
Central-composite designs, 885-1
Chen
  three-stage phase II trials, 130-1
Chi-square
  estimator, 900-1
  one variance, 650-1
Chi-square distribution
  probability calculator, 915-2
Chi-Square estimator
  examples, 900-2
Chi-square test, 250-1
  examples, 250-6
  two proportions, 200-5, 223-6
  validation, 250-9
Circularity
  repeated measures ANOVA, 570-13
Clinical trial
  three-stage, 130-1
Clinical trials
  one proportion, 120-1
Clopper-Pearson confidence interval, 115-2
Cluster randomized - equivalence
  two proportions, 240-1
Cluster randomized - non-inferiority
  two proportions, 235-1
Cluster randomized - non-zero null
  two proportions, 233-1
Cluster randomized design
  two means, 480-1
  two proportions, 230-1
Cochran-Armitage test, 255-1
examples, 255-12
validation, 255-19
Cochran-Mantel-Haenszel, 225-1
Coefficient alpha
one, 815-1
two, 820-1
Coefficient alpha (one)
examples, 815-4
validation, 815-7
Coefficient alphas (two)
examples, 820-6
validation, 820-10
Coefficient of variation
cross-over, 505-2
cross-over - equivalence, 525-3
cross-over - higher-order - equivalence, 545-5
cross-over - higher-order - non-zero null, 533-4
cross-over - higher-order - non-inferiority, 535-4
cross-over - non-inferiority, 515-3
cross-over - non-unity null, 513-5
mean ratio, 445-2
mean ratio - equivalence, 470-3
mean ratio - non-zero null, 453-2
mean ratio - non-inferiority, 455-2
standard deviation estimator, 905-9

Cohort study
post-marketing surveillance, 135-2

Color
histogram, 945-11
scatter plot, 940-11

Comments
histogram, 945-10
scatter plot, 940-10

Comparisons
ANCOVA, 551-1
multiple comparisons, 575-1
one-way ANOVA, 550-1

Comparisonwise error rate
simulation, 580-1, 585-2, 590-1

Complete randomization, 880-3

Compound symmetry
repeated measures ANOVA, 570-13

Confidence interval
one proportion, 115-1
paired means, 496-1

Confidence intervals
one correlation, 801-1
one mean, 420-1
one mean - tolerance probability, 421-1
one standard deviation, 640-1
one standard deviation - relative error, 642-1
one standard deviation - tolerance probability, 641-1
one variance, 651-1
one variance - relative error, 653-1
one variance - tolerance probability, 652-1
paired means - tolerance probability, 497-1
slope - simple linear regression, 856-1
two means, 471-1
two means - tolerance probability, 472-1
two proportions - difference, 216-1
two proportions - odds ratio, 216-1
two proportions - ratio, 216-1
variance ratio, 656-1
variance ratio - relative error, 657-1

Confounding
two-level designs, 881-2

Connecting points
scatter plot, 940-10

Constant distribution
simulation, 920-6

Consumer’s risk
exponential mean (one), 405-2

Contaminated normal simulation, 920-19

Contingency table, 250-1

Chi-square estimator, 900-1

Continuity correction
one proportion, 100-5
two proportions - group sequential, 220-8
two proportions - offset, 205-7

Contrast matrix
repeated measures ANOVA, 570-10

Contrasts
ANCOVA, 551-1
MANOVA, 605-4
mixed models, 201-6, 431-7
one-way ANOVA, 550-1

Contrasts - multiple
simulation, 590-1

Control chart
Cusum, 290-1
EWMA, 290-1
FIR, 290-1
one mean - simulation, 290-1
R Chart, 295-1
S Chart, 295-1
S Chart with Probability Limits, 295-1
Shewhart, 290-1
variability - simulation, 295-1

Control charts
formulas, 290-2, 295-2

Control charts - means
examples, 290-10
validation, 290-14

Control charts - variability
examples, 295-8
validation, 295-11

Controlled variables
multiple regression, 865-4

Corners
histogram, 945-4
scatter plot, 940-1

Correlated proportions
matched case-control, 155-1
McNemar test, 150-1

Correlation
intraclass, 810-1

Correlation (one)
confidence interval, 801-1
examples, 800-4
validation, 800-7
Correlation (one) - confidence interval
examples, 801-4
validation, 801-7
Correlation coefficient, 800-1
Correlation coefficient distribution
probability calculator, 915-3
Correlation test
one, 800-1
two, 805-1
Correlations (two)
examples, 805-5
validation, 805-8
Covariance matrix
repeated measures ANOVA, 570-13
Covariance patterns
AR(1), 201-11, 431-4
banded, 201-11, 431-4
compound symmetry, 201-11, 431-4
simple, 201-12
simple, 431-5
Cross-over
higher-order - equivalence, 540-1, 545-1
higher-order - non-inferiority, 530-1
higher-order - non-zero null, 528-1
ratio - equivalence, 525-1
ratio - higher-order - non-inferiority, 535-1
ratio - higher-order - non-zero null, 533-1
ratio - non-inferiority, 515-1
ratio - non-unity null, 513-1
repeated measures ANOVA, 570-44
two means, 500-1
two means - equivalence, 520-1
two means - non-inferiority, 510-1
two means - non-zero null, 508-1
two means - ratio, 505-1
Cross-over - higher-order - equivalence
examples, 540-9
validation, 540-12
Cross-over - higher-order - non-zero null
examples, 528-8
validation, 528-11
Cross-over - higher-order - non-inferiority
examples, 530-8
validation, 530-11
Cross-over - ratio - equivalence
examples, 525-7
validation, 525-9
Cross-over - ratio - higher-order - equivalence
examples, 545-9
validation, 545-12
Cross-over - ratio - higher-order - non-zero null
examples, 533-9
validation, 533-12
Cross-over - ratio - higher-order - non-inferiority
examples, 535-9
validation, 535-12
Cross-over - ratio - non-inferiority
examples, 515-6
validation, 515-8
Cross-over - ratio - non-unity null
examples, 513-6
Cross-over (two means)
examples, 500-7
validation, 500-10
Cross-over (two means) - equivalence
examples, 520-7
validation, 520-10
Cross-over (two means) - non-unity null
validation, 513-8
Cross-over (two means) - non-zero null
examples, 508-7
validation, 508-10
Cross-over (two means) - non-inferiority
examples, 510-7
validation, 510-10
Cross-over (two means) - ratio
examples, 505-6
validation, 505-8
Crossover analysis
mixed models, 571-1
Cumulative scale
histogram, 945-4
CUSUM chart, 290-4
D’Agostino kurtosis, 670-2
D’Agostino omnibus test, 670-3
D’Agostino skewness test, 670-4
Data
histogram, 945-11
scatter plot, 940-11
simulation of, 920-1
Data bars
scatter plot, 940-11
Data entry, 925-1
Data simulation
examples, 920-18
Data simulator, 920-1
Data tab, 4-6
Data ticks
scatter plot, 940-10
Decimals
histogram, 945-6
scatter plot, 940-4
Default template, 4-2
D-efficiency
D-optimal designs, 888-12
Density trace
  histogram, 945-3
Design generator, 889-1
  examples, 889-3
Design of experiments
  randomization lists, 880-1
Designs
  Box-Behnken, 885-1
  central-composite, 885-1
  design generator, 889-1
  fractional factorial, 882-1
  Plackett-Burman, 886-1
  response surface, 885-1
  screening, 886-1
  Taguchi, 887-1
  two-level factorial, 881-1, 889-1
Determinant
  D-optimal designs, 888-13
Determinant analysis
  D-optimal designs, 888-11
Diagnostic testing
  ROC curve (one), 260-1
Difference
  proportions, 7-2
Difference data
  simulation, 920-25
Discordant pairs
  McNemar test, 150-2
Distributions
  combining, 920-13
  mixing, 920-13
  simulation, 920-1
Documentation
  printing, 1-7
Donner and Klar
  cluster randomized design, 230-2
D-optimal designs, 888-1
  examples, 888-8
Dot plot
  histogram, 945-11
Downloading updates, 1-2
Drift
  survival - group sequential, 710-5
  two means - group sequential, 475-4
  two proportions - group sequential, 220-4
Dunn’s test
  multiple contrasts - simulation, 590-3
Dunnett’s test
  multiple comparisons, 575-3
Dunnett's test
  simulation, 585-1, 585-3
Dunn's test
  simulation, 590-1
Duplicates
  D-optimal designs, 888-4
E
  Edit menu
    output window, 5-5
    spreadsheet, 925-2
Effect size, 6-5
  ANOVA, 560-2, 560-7
  chi-square estimator, 900-1
  chi-square test, 250-2
  multiple regression, 865-2
  one-way ANOVA, 550-3
  randomized block ANOVA, 565-2
Efficacy boundaries, 221-1, 222-1, 223-1, 476-1, 477-1, 478-1
Efron’s biased coin randomization, 880-3
Entering procedure options, 4-1
Entering your serial number, 1-2
Equivalence
  correlated proportions, 165-1
  means (two), 460-1
  means (two) - ratio, 470-1
  means (two) - simulation, 465-1
  one proportion, 110-1
  two proportions, 215-1
Equivalence - two correlated proportions
  examples, 165-11
  validation, 165-14
Equivalence hypothesis, 6-9
Equivalence limits - paired means
  simulation, 495-9
Equivalence test data
  simulating, 495-3
Error rates - simulation
  multiple comparison, 580-1, 585-2
  multiple contrasts, 590-1
Errors, 6-2
EWMA chart limits, 290-5
Exact binomial test
  one proportion, 100-4
Example data
  histogram, 945-11
  scatter plot, 940-11
Examples
  ANCOVA, 551-10
  balanced incomplete block designs, 883-3
  Chi-square estimator, 900-2
  Chi-square test, 250-6
  Cochran-Armitage test, 255-12
  control charts - means, 290-10
  control charts - variability, 295-8
  correlation (one) - confidence interval, 801-4
  Cox regression, 850-4
  cross-over - higher-order - equivalence, 540-9
  cross-over - higher-order - non-zero null, 528-8
  cross-over - higher-order - non-inferiority, 530-8
  cross-over - ratio - equivalence, 525-7
  cross-over - ratio - higher-order - equivalence, 545-9
  cross-over - ratio - higher-order - non-zero null, 533-9
cross-over - ratio - higher-order - non-inferiority, 535-9
cross-over - ratio - non-inferiority, 515-6
cross-over - ratio - non-unity null, 513-6
cross-over (two means), 500-7
cross-over (two means) - equivalence, 520-7
cross-over (two means) - non-zero null, 508-7
cross-over (two means) - non-inferiority, 510-7
cross-over (two means) - ratio, 505-6
data simulation, 920-18
design generator, 889-3
D-optimal designs, 888-8
equivalence - two correlated proportions, 165-11
exponential mean (one), 405-8
exponential means (two), 435-5
fixed effects ANOVA, 560-13
fractional factorial designs, 882-3
intraclass correlation, 810-4
kappa, 811-6
Latin square designs, 884-4
linear regression, 855-5
logistic regression, 860-9
logrank tests - non-inferiority, 706-7
logrank tests (Lakatos), 715-15
Mann-Whitney test, 430-16
MANOVA, 605-12
many proportions - trend, 255-12
matched case-control - proportions, 155-6
McNemar test - two correlated proportions, 150-7
mean ratio - non-zero null, 453-6
mean ratio - non-inferiority, 455-6
microarray one-sample or paired t-test, 610-12
microarray two-sample t-test, 615-12
minimum effective dos, 595-4
mixed models, 571-36
multiple comparisons, 575-13
multiple comparisons - simulation, 580-13
multiple comparisons - vs control - simulation, 585-12
multiple contrasts - simulation, 590-14
multiple regression, 865-6
non-inferiority - two correlated proportions, 160-10
normality - simulation, 670-8
odds ratio estimator, 910-2
one coefficient alpha, 815-4
one correlation, 800-4
one mean - confidence interval, 420-5
one mean - tolerance - confidence interval, 421-6
one proportion, 100-13
one proportion - confidence interval, 115-6
one proportion - equivalence, 110-11
one proportion - non-inferiority, 105-10
one proportion - non-zero null, 103-10
one sensitivity, 270-6
one specificity, 270-6
one standard deviation - confidence interval, 640-5
one standard deviation - relative error - confidence interval, 642-4
one standard deviation - tolerance - confidence interval, 641-6
one variance - confidence interval, 651-4
one variance - relative error - confidence interval, 653-4
one variance - tolerance - confidence interval, 652-6
one-way ANOVA, 550-10
one-way ANOVA - simulation, 555-10
paired means - confidence interval, 496-5
paired means - tolerance - confidence interval, 497-6
paired sensitivities, 276-7
Poisson mean (one), 412-4
Poisson means (two), 437-7
Poisson regression, 870-7
post-marketing surveillance, 135-6
proportions estimator, 910-2
randomization lists, 880-9
randomized block ANOVA, 565-9
ratio of two means, 445-6
ratio of two means - equivalence, 470-7
regression - confidence interval, 856-7
repeated measures - two means, 431-13
repeated measures - two proportions, 201-16
repeated measures ANOVA, 570-29
response surface designs, 885-3
ROC curve (one), 260-8
ROC curves (two), 265-8
screening designs, 886-3
single-stage phase II trials, 120-3
standard deviation estimator, 905-6
survival - group sequential, 710-11, 711-27
survival - logrank, 700-6
survival - logrank - advanced, 705-8
Taguchi designs, 887-4
three-stage phase II trials, 130-8
t-test (one mean), 400-8
t-test (one mean) - non-zero null, 413-8
t-test (one mean) - non-inferiority, 415-8
t-test (one mean) - simulation, 410-10
t-test (paired means) - equivalence - simulation, 495-11
t-test (paired means) - simulation, 490-11
t-test (paired), 400-13, 485-8
t-test (two means), 430-10
t-test (two means) - simulation, 440-12
two coefficient alphas, 820-6
two correlated proportions - equivalence, 165-11
two correlated proportions - non-inferiority, 160-10
two correlations, 805-5
two means - cluster randomized, 480-4
two means - confidence interval, 471-5
two means - equivalence, 460-5
two means - equivalence - simulation, 465-12
two means - group sequential, 475-10, 476-17, 477-15
two means - non-zero null, 448-10
two means - non-inferiority, 450-10
two means – non-inferiority – group sequential, 478-17
two means - ratio, 445-6
two means - ratio - equivalence, 470-7
two means - tolerance - confidence interval, 472-6
two ordered categorical variables, 253-5
two proportions - cluster - equivalence, 240-8
two proportions - cluster - non-inferiority, 235-7
two proportions - cluster - non-zero null, 233-7
two proportions - cluster randomized, 230-9
two proportions - confidence interval, 216-26
two proportions - equivalence, 215-17
two proportions - group sequential, 220-9, 221-18
two proportions - inequality, 200-14
two proportions - non-inferiority, 210-17
two proportions – non-inferiority – group sequential, 223-22
two proportions – non-zero null – group sequential, 222-22
two proportions - offset, 205-18
two proportions - stratified design, 225-8
two sensitivities, 275-9
two-level designs, 881-6
two-stage phase II trials, 125-8
variance (one), 650-4
variance ratio - confidence interval, 656-5
variance ratio - relative error - confidence interval, 657-4
variances (two), 655-4
Williams' test, 595-4
Exiting PASS, 4-3
Experiment (run)
two-level designs, 881-2
Experimental design, 881-1
two-level designs, 881-2
Experimental error
two-level designs, 881-2
Experimentwise error rate
simulation, 580-1, 585-2, 590-1
Exponential
log-rank, 705-3
log-rank, 715-2
Exponential distribution
simulation, 920-6
Exponential mean (one), 405-1
distances, 405-8
validation, 405-10
Exponential means (two), 435-1
distances, 435-5
validation, 435-7
Exponential test
simulation, 410-4
Exposure
Poisson regression, 870-1
Exposure probability
matched case-control, 155-4
Factor scaling
D-optimal designs, 888-2
Factorial ANOVA, 560-1
Factorial designs
two-level designs, 881-1, 881-2
False discovery rate adjustment
t-test

two groups, 615-4
t-test - one group, 610-6
Familywise error rate
simulation, 580-1, 585-2, 590-1
Farrington-Manning confidence interval
two proportions, 216-5
Farrington-Manning test
two proportions - equivalence, 215-9
two proportions - non-inferiority, 210-10
two proportions - offset, 205-9
Fast initial response, 290-4
File menu
output window, 5-2
procedure window, 4-3
spreadsheet, 925-2
Finite population correction
test, 400-8, 485-8
Finite population size
one mean - confidence interval, 420-2
FIR, 290-4
Fisher's exact test
two proportions, 200-4
Fisher-z transformation
two correlations, 805-1
Fixed effects ANOVA
two proportions, 216-12, 216-16
Fixed effects models
mixed models, 571-1
Fixed factor
ANOVA, 560-5
Fleiss Confidence intervals
two proportions, 216-12, 216-16
Fleming algorithm
single-stage phase II trials, 120-1
Fleming-Harrington tests
Group Sequential, 711-5
Follow-up
log-rank, 705-4
Font
histogram, 945-4
scatter plot, 940-2
Forced points
D-optimal designs, 888-5
FPR
ROC curve (one), 260-7
ROC curves (two), 265-1
Fractional factorial designs, 882-1
distances, 882-3
Freedman
log-rank, 700-1
Frequency polygon
histogram, 945-3
F-test
  Geisser-Greenhouse, 570-4
  simulation, 555-1
  two variances, 655-1
Futility boundaries, 221-1, 222-1, 223-1, 476-1, 477-1, 478-1

G

Gallery, 5-2
Games-Howell - simulation multiple comparison, 580-1
Games-Howell test simulation, 580-3
Gamma data simulation, 920-23
Gamma distribution probability calculator, 915-4
simulation, 920-7
Gart-Nam confidence interval two proportions, 216-7
Gart-Nam test two proportions - equivalence, 215-10
two proportions - non-inferiority, 210-11
two proportions - offset, 205-10
Gehan test Group Sequential, 711-5
Geisser-Greenhouse repeated measures ANOVA, 570-1
Geisser-Greenhouse F-test repeated measures ANOVA, 570-4
General linear multivariate model MANOVA, 605-2
repeated measures ANOVA, 570-3
Generating data, 920-1
Goodness of fit chi-square estimator, 900-1
Chi-square test, 250-1
Graeco-Latin square designs, 884-1
Grid lines histogram, 945-7
scatter plot, 940-5
Group sequential test log-rank, 710-1, 711-1
survival, 710-1, 711-1
two means, 475-1, 476-1, 477-1
two means – non-inferiority, 478-1
two means – non-inferiority (simulation), 478-1
two proportions, 220-1, 221-1
two proportions – non-inferiority, 223-1
two proportions – non-inferiority (simulation), 223-1
two proportions – non-zero null, 222-1
two proportions – non-zero null (simulation), 222-1
two proportions (simulation), 221-1

H

Hazard rate
  Cox regression, 850-1
Hazard rate conversion, 903-1
Hazard rate parameterization logrank, 711-2, 715-2
Hazard rates logrank, 700-2
logrank - advanced, 705-2
time dependent, 715-22
Hazard ratio
  group sequential, 710-2
Help menu
  output window, 5-7
  procedure window, 4-5
  spreadsheet, 925-3
Help system
  contents window, 1-5
  index window, 1-4
  navigating, 1-4
  printing documentation, 1-7
  search window, 1-6
Heterogeneous variances
  mixed models, 571-47, 571-49
Histogram
  alignment, 945-4
  axis, 945-3
  axis line, 945-5
  background color, 945-11
  bars, 945-1
  bin width, 945-2
  bins, 945-1
  border plots, 945-10
  boundaries, 945-3
  box plot, 945-11
  color, 945-11
  comments, 945-10
  corners, 945-4
  cumulative scale, 945-4
  data, 945-11
  decimals, 945-6
  density trace, 945-3
  dot plot, 945-11
  example data, 945-11
  font, 945-4
  frequency polygon, 945-3
  grid lines, 945-7
  interior lines, 945-1
  labels, 945-4
  layout, 945-4
  legend, 945-9
  lines, 945-10
  major ticks, 945-5
  maximum, 945-3
  minimum, 945-3
  minor ticks, 945-6
  normal density, 945-3
  notes, 945-10
  number bins, 945-2
painting order, 945-11
scale, 945-4
scaling, 945-2
settings, 945-1
size, 945-11
text, 945-10
ticks, 945-5
titles, 945-8
Histograms, 945-1
Hotelling’s T2, 600-1
validation, 600-11
Hotelling’s T2 distribution
probability calculator, 915-4
Hotelling-Lawley trace, 605-1
MANOVA, 605-3
repeated measures ANOVA, 570-1, 570-7
Hypergeometric distribution
probability calculator, 915-4
Hypergeometric model
one proportion, 100-2
Hypothesis
equivalence, 6-9
inequality, 6-7
introduction, 6-1
means, 8-1
non-inferiority, 6-7
superiority, 6-8
types, 6-6
Hypothesis testing
introduction, 6-1

I

Icons
output window, 5-8
procedure window, 4-5
Incidence rate
post-marketing surveillance, 135-2
Inclusion points
D-optimal designs, 888-6
Independence test, 250-1
Inequality
one proportion, 100-1
two means ratio, 445-1
two proportions, 200-1
two proportions - offset, 205-1
Inequality hypothesis, 6-7
Installation, 1-1
Interaction
two-level designs, 881-3
Intercept
linear regression, 855-1
Interim analysis
survival, 710-1, 711-1
three-stage trial, 130-1
two means, 475-1, 476-1, 477-1
two means – non-inferiority, 478-1
two proportions, 220-1, 221-1
two proportions – non-inferiority, 223-1
two proportions – non-zero null, 222-1
Interior lines
histogram, 945-1
Intraclass correlation, 810-1
eamples, 810-4
validation, 810-6
Intracluster correlation
cluster randomized - two means, 480-3
cluster randomized - non-inferiority, 235-4
cluster randomized - non-zero null, 233-4
cluster randomized design, 230-2
Introduction to power analysis, 6-1
Iterations
maximum, 4-7
Iterations tab, 4-7

K

Kappa
examples, 811-6
validation, 811-11
Kappa test for rater agreement, 811-1
Kenward and Roger method
mixed models, 571-11
Kolmogorov-Smirnov test, 670-2
Kruskal-Wallis
multiple comparisons - simulation, 580-1
simulation, 555-1
Kruskal-Wallis test
multiple comparisons - simulation, 580-3, 585-1
simulation, 585-3

L

Labels
histogram, 945-4
scatter plot, 940-2
Lachin and Foulkes
logrank test, 705-1
Lakatos
logrank, 715-1
Lan-DeMets
survival - group sequential, 710-1, 711-1
two means - group sequential, 475-1
two proportions - group sequential, 220-1
Latin square
ANOVA, 560-17
Latin square designs, 884-1
eamples, 884-4
Layout
histogram, 945-4
scatter plot, 940-2
Legend
histogram, 945-9
parameter, 4-10
scatter plot, 940-8
<table>
<thead>
<tr>
<th>Index-10</th>
</tr>
</thead>
<tbody>
<tr>
<td>Likelihood ratio test</td>
</tr>
<tr>
<td>two proportions, 200-7</td>
</tr>
<tr>
<td>Likert-scale</td>
</tr>
<tr>
<td>simulation, 410-21, 920-8</td>
</tr>
<tr>
<td>simulation, 920-21</td>
</tr>
<tr>
<td>Lilliefors' critical values</td>
</tr>
<tr>
<td>normality tests, 670-2</td>
</tr>
<tr>
<td>Linear model</td>
</tr>
<tr>
<td>ANOVA, 560-2</td>
</tr>
<tr>
<td>repeated measures ANOVA, 570-3</td>
</tr>
<tr>
<td>Linear regression, 855-1</td>
</tr>
<tr>
<td>confidence interval, 856-1</td>
</tr>
<tr>
<td>examples, 855-5</td>
</tr>
<tr>
<td>validation, 855-7</td>
</tr>
<tr>
<td>Lines</td>
</tr>
<tr>
<td>histogram, 945-10</td>
</tr>
<tr>
<td>scatter plot, 940-10</td>
</tr>
<tr>
<td>Loading a procedure, 2-3</td>
</tr>
<tr>
<td>Log base</td>
</tr>
<tr>
<td>scatter plot, 940-2</td>
</tr>
<tr>
<td>Log transformation</td>
</tr>
<tr>
<td>cross-over, 505-2</td>
</tr>
<tr>
<td>cross-over - equivalence, 525-2</td>
</tr>
<tr>
<td>cross-over - higher-order - equivalence, 545-4</td>
</tr>
<tr>
<td>cross-over - higher-order - non-zero null, 533-4</td>
</tr>
<tr>
<td>cross-over - higher-order - non-inferiority, 535-4</td>
</tr>
<tr>
<td>cross-over - non-inferiority, 515-2</td>
</tr>
<tr>
<td>cross-over - non-unity null, 513-2</td>
</tr>
<tr>
<td>mean ratio, 445-2</td>
</tr>
<tr>
<td>mean ratio - equivalence, 470-2</td>
</tr>
<tr>
<td>mean ratio - non-zero null, 453-2</td>
</tr>
<tr>
<td>mean ratio - non-inferiority, 455-2</td>
</tr>
<tr>
<td>Logarithm</td>
</tr>
<tr>
<td>scatter plot, 940-1</td>
</tr>
<tr>
<td>Logistic regression, 860-1</td>
</tr>
<tr>
<td>examples, 860-9</td>
</tr>
<tr>
<td>validation, 860-14</td>
</tr>
<tr>
<td>Logit</td>
</tr>
<tr>
<td>logistic regression, 860-2</td>
</tr>
<tr>
<td>Logrank</td>
</tr>
<tr>
<td>hazard rate parameterization, 711-2, 715-2</td>
</tr>
<tr>
<td>median survival time parameterization, 711-2, 715-2</td>
</tr>
<tr>
<td>mortality parameterization, 711-2, 715-2</td>
</tr>
<tr>
<td>proportion surviving parameterization, 711-2, 715-2</td>
</tr>
<tr>
<td>Log-rank test</td>
</tr>
<tr>
<td>group sequential test, 710-1, 711-1</td>
</tr>
<tr>
<td>Logrank procedure comparison, 715-3</td>
</tr>
<tr>
<td>Logrank test, 700-1, 715-1</td>
</tr>
<tr>
<td>Lachin and Foulkes, 705-1</td>
</tr>
<tr>
<td>non-inferiority, 706-1</td>
</tr>
<tr>
<td>Log-rank test</td>
</tr>
<tr>
<td>Group Sequential, 711-5</td>
</tr>
<tr>
<td>Logrank tests - non-inferiority</td>
</tr>
<tr>
<td>examples, 706-7</td>
</tr>
<tr>
<td>validation, 706-11</td>
</tr>
<tr>
<td>Logrank tests (Lakatos)</td>
</tr>
<tr>
<td>examples, 715-15</td>
</tr>
<tr>
<td>validation, 715-21</td>
</tr>
<tr>
<td>Longitudinal data models</td>
</tr>
<tr>
<td>mixed models, 571-2</td>
</tr>
<tr>
<td>Longitudinal models, 571-1</td>
</tr>
</tbody>
</table>

**M**

| Macros, 930-1  |
| commands, 930-6  |
| examples, 930-18  |
| syntax, 930-2  |
| Major ticks  |
| histogram, 945-5  |
| scatter plot, 940-4  |
| Mann-Whitney test, 430-1  |
| assumptions, 430-4  |
| examples, 430-16  |
| group-sequential - simulation, 476-4, 477-4, 478-5  |
| non-zero null, 448-5  |
| non-inferiority, 450-5  |
| simulation, 440-5  |
| Mann-Whitney test - equivalence  |
| simulation, 465-5  |
| Mann-Whitney test - simulation  |
| equivalence, 465-1  |
| MANOVA, 605-1  |
| assumptions, 605-1  |
| examples, 605-12  |
| validation, 605-17  |
| Mantel Haenszel test  |
| two proportions, 200-7  |
| Mantel-Haenszel, 225-1  |
| Many proportions - trend  |
| examples, 255-12  |
| validation, 255-19  |
| Martinez-Iglewicz test, 670-3  |
| Matched case-control, 155-1  |
| post-marketing surveillance, 135-2  |
| Matched case-control - proportions  |
| examples, 155-6  |
| validation, 155-10  |
| Maximum  |
| histogram, 945-3  |
| scatter plot, 940-1  |
| Maximum iterations, 4-7  |
| Maximum likelihood  |
| mixed models, 571-8  |
| McNemar test, 150-1  |
| McNemar test - two correlated proportions  |
| examples, 150-7  |
| validation, 150-9  |
| Mean (one)  |
| confidence interval, 420-1  |
| control chart - simulation, 290-1  |
| exponential, 405-1  |
| Poisson, 412-1  |
| simulation, 410-1  |
| Mean (one) - confidence interval  |
| examples, 420-5  |
| validation, 420-7  |
Mean (one) - tolerance - confidence interval
examples, 421-6
validation, 421-8
Mean (one) - tolerance probability
confidence interval, 421-1
Mean absolute deviation
standard deviation estimator, 905-3
Mean ratio
equivalence, 470-1
inequality, 445-1
Means
introduction, 8-1
one-way - simulation, 555-1
Means - ratio - equivalence
cross-over, 525-1
Means - ratio - non-zero null - higher-order
cross-over, 533-1
Means - ratio - non-inferiority
cross-over, 515-1
Means - ratio - non-inferiority - higher-order
cross-over, 535-1
Means - ratio - non-unity null
cross-over, 513-1
Means (paired)
confidence interval, 496-1
simulation, 490-1
Means (paired) - confidence interval
examples, 496-5
validation, 496-6
Means (paired) - equivalence
simulation, 495-1
Means (paired) - tolerance - confidence interval
examples, 497-6
validation, 497-8
Means (paired) - tolerance probability
confidence interval, 497-1
Means (two)
cross-over, 500-1
equivalence, 460-1
exponential, 435-1
group sequential test, 475-1, 476-1, 477-1
interim analysis, 475-1, 476-1, 477-1
interim analysis – non-inferiority, 478-1
non-inferiority, 450-1
non-zero null, 448-1
Poisson, 437-1
ratio, 445-1
simulation, 440-1
T-test, 430-1
t-test - equivalence, 460-1
T-test - non-inferiority, 450-1
T-test - non-zero null, 448-1
Means (two) - cluster randomized design
examples, 480-4
validation, 480-6
Means (two) - confidence interval
examples, 471-5
validation, 471-7
Means (two) - equivalence
cross-over, 520-1
examples, 460-5
validation, 460-7
Means (two) - equivalence - higher-order
cross-over, 540-1, 545-1
Means (two) - equivalence - simulation
examples, 465-12
validation, 465-19
Means (two) - group sequential
examples, 475-10, 476-17, 477-15
validation, 475-17, 476-28, 477-26
Means (two) – non-inferiority – group sequential
examples, 478-17
validation, 478-28
Means (two) - non-zero null
cross-over, 508-1
examples, 448-10
validation, 448-13
Means (two) - non-zero null - higher-order
cross-over, 528-1
Means (two) - non-inferiority
cross-over, 510-1
examples, 450-10
validation, 450-13
Means (two) - non-inferiority - higher-order
cross-over, 530-1
Means (two) - ratio
cross-over, 505-1
equivalence, 470-1
non-zero null, 453-1
non-inferiority, 455-1
Means (two) - ratio - non-zero null
examples, 453-6
validation, 453-8
Means (two) - ratio - non-inferiority
examples, 455-6
validation, 455-8
Means (two) - simulation
equivalence, 465-1
t-test - equivalence, 465-1
Means (two) - tolerance - confidence interval
examples, 472-6
validation, 472-8
Means (two) - tolerance probability
confidence interval, 472-1
Means matrix
MANOVA, 605-4
repeated measures ANOVA, 570-7
Measurement error
randomized block ANOVA, 565-2
Median survival time conversion, 903-1
Median survival time parameterization
logrank, 711-2, 715-2
Menus
output window, 5-2
PASS home window, 3-2
procedure window, 4-3
spreadsheet, 925-2
Microarray data
- one sample t-test, 610-1
- paired T-test, 610-1
- two sample t-test, 615-1

Microarray one-sample or paired t-test
- examples, 610-12
- validation, 610-19

Microarray two-sample t-test
- examples, 615-12
- validation, 615-19

Miettinen-Nurminen confidence interval
- two proportions, 216-6

Miettinen-Nurminen test
- two proportions - equivalence, 215-7
- two proportions - non-inferiority, 210-8
- two proportions - offset, 205-7

Minimum
- histogram, 945-3
- scatter plot, 940-1

Minimum detectable difference
- mixed models, 571-14
- multiple comparisons, 575-11
- one-way ANOVA, 550-17
- t-test, 400-12
- two-sample t test, 430-14

Minimum effective dose
- examples, 595-4
- validation, 595-7

Minimum Effective Dose
- Williams' Test, 595-1

Minimum system requirements, 1-1

Minor ticks
- histogram, 945-6
- scatter plot, 940-4

Mixed model
- defined, 571-2

Mixed models, 571-1
- differential evolution, 571-12
- examples, 571-36
- F test, 571-11
- Fisher scoring, 571-12
- fixed effects, 571-1
- G matrix, 571-5
- heterogeneous variances, 571-47, 571-49
- Kenward and Roger method, 571-11
- L matrix, 571-9
- likelihood formulas, 571-8
- maximum likelihood, 571-8
- minimum detectable difference, 571-14
- MIVQUE, 571-12
- Newton-Raphson, 571-12
- pairwise contrasts, 201-6, 431-7
- R matrix, 571-5
- restricted maximum likelihood, 571-9
- simulation steps, 571-17
- types, 571-1
- validation, 571-42

Mixture data
- simulation, 920-24

Mixture design
- D-optimal designs, 888-20

Monte Carlo, 8-3
- Monte Carlo simulation, 920-1
- Mortality conversion, 903-1
- Mortality parameterization
  - logrank, 711-2, 715-2
- MTBF
  - exponential mean (one), 405-1
- Multinomial
  - chi-square estimator, 900-3
- Multinomial distribution
  - simulation, 920-8
- Multiple comparisons, 575-1
  - Dunnett's test - simulation, 585-1
  - examples, 575-13
  - Games-Howell - simulation, 580-1
  - pair-wise - simulation, 580-1
  - validation, 575-19
  - vs control - simulation, 585-1
  - with a control, 575-2
  - with best, 575-5
- Multiple comparisons - simulation
  - examples, 580-13
  - power, 580-4
  - validation, 580-22
- Multiple comparisons - vs control - simulation
  - examples, 585-12
  - validation, 585-21
- Multiple contrasts
  - simulation, 590-1
- Multiple contrasts - simulation
  - examples, 590-14
  - validation, 590-24
- Multiple regression, 865-1
  - examples, 865-6
  - validation, 865-10
- Multiple testing adjustment
  - t-test - two groups, 615-2
  - t-test -one group, 610-4
- Multivariate analysis of variance, 605-1

Navigating the help system, 1-4
Navigation pane
- output window, 5-9
Negative binomial distribution
- probability calculator, 915-5
Nested factors
- design generator, 889-1
New template, 4-3
Noncentrality
- one-way ANOVA, 550-3
- Noncentrality parameter
  - one-way ANOVA, 550-4
Non-inferiority
- correlated proportions, 160-1
  - logrank, 705-13, 706-1
  - means (two), 450-1
  - means (two) - ratio, 455-1
one proportion, 105-1
paired t-test, 415-1
survival, 706-1
two proportions, 210-1
Non-inferiority - two correlated proportions
examples, 160-10
validation, 160-13
Non-inferiority hypothesis, 6-7
Non-Inferiority test (two means)
simulation, 440-17
Nonparametric
t-test, 400-7, 485-7
Nonparametric tests
Wilcoxon test, 400-1
Non-zero null
means (two), 448-1
means (two) - ratio, 453-1
one proportion, 103-1
paired t-test, 413-1
Normal density
histogram, 945-3
Normal distribution
probability calculator, 915-5
simulation, 920-9, 920-18
Normality - simulation
elements, 670-8
validation, 670-11
Normality tests, 670-1
Anderson-Darling test, 670-2
D’Agostino kurtosis, 670-2
D’Agostino omnibus, 670-3
D’Agostino skewness, 670-4
Kolmogorov-Smirnov, 670-2
Lilliefors’ critical values, 670-2
Martinez-Iglewicz, 670-3
range, 670-4
Shapiro-Wilk, 670-4
Notes
histogram, 945-10
scatter plot, 940-10
Nuisance parameter, 8-2
Nuisance parameters, 6-6
Null hypothesis, 6-1, 8-1
Number bins
histogram, 945-2
O
O’Brien-Fleming
survival - group sequential, 710-1, 711-1
two means - group sequential, 475-1
two proportions - group sequential, 220-1
Odds ratio
logistic regression, 860-2
matched case-control, 155-1
McNemar test, 150-3
one proportion, 100-6
proportions, 7-3
two proportions, 200-3
Odds ratio estimator, 910-1
elements, 910-2
One proportion
equivalence, 110-1
inequality, 100-1
non-inferiority, 105-1
non-zero null, 103-1
superiority, 103-1
One-sample t-test
microarray data, 610-1
One-way ANOVA, 550-1
elements, 550-10
validation, 555-19
One-way ANOVA - simulation
elements, 555-10
validation, 555-16
Open example template, 4-3
Open template, 4-3
Operating system, 1-1
Ordered categorical variables (two), 253-1
elements, 253-5
validation, 253-10
Orthogonal arrays, 887-1
Orthogonal sets of Latin squares, 884-2
Outliers
multiple comparisons - simulation, 580-23
one-way ANOVA - simulation, 555-18
Outliers (two means)
simulation, 440-19
Output, 2-5
Output Gallery, 5-2
Output window, 5-1
bottom bar, 5-10
edit menu, 5-5
file menu, 5-2
help menu, 5-7
navigation pane, 5-9
toolbar, 5-8
view menu, 5-4
window menu, 5-7
P
P value, 6-4
Painting order
histogram, 945-11
scatter plot, 940-11
Paired design
microarray data, 610-1
Paired designs
non-zero null, 413-1
non-inferiority, 415-1
Paired distributions
simulating, 490-2
Paired means
simulation, 490-1
Paired means - equivalence
simulation, 495-1
Paired proportions
  equivalence, 165-1
  non-inferiority, 160-1
Paired t-test, 400-1, 485-1
  assumptions, 400-3, 485-3
  microarray data, 610-1
  non-zero null, 413-1
  non-inferiority, 415-1
  superiority, 413-1
Pairwise comparisons
  multiple comparisons, 575-7
Pair-wise comparisons
  simulation, 580-1
Pairwise contrasts
  mixed models, 201-6
Pairwise contrasts
  mixed models, 431-7
Panel, 4-1
Parameters
  abbreviations, 4-10
  axis, 4-10
  entering, 4-6
  legend, 4-10
PASS help system, 1-3
PASS home window, 3-1
  Category Tree, 3-4
  Find, 3-5
  Menus, 3-2
  Procedure list, 3-5
  Search, 3-5
  toolbar, 3-3
Patient entry
  logrank, 705-3
Peto-Peto test
  Group Sequential, 711-5
Phase I trials
  definition, 120-1
Phase II clinical trials
  single-stage one proportion, 120-1
  three-stage one proportion, 130-1
  two-stage one proportion, 125-1
Phase II trials
  definition, 120-1
Pillai-Bartlett trace, 605-1
  MANOVA, 605-3
  repeated measures ANOVA, 570-1, 570-6
Plackett-Burman designs, 886-1
Planned comparisons
  ANCOVA, 551-1
  one-way ANOVA, 550-1
Plot setup tab, 4-9
Pocock
  survival - group sequential, 710-1, 711-1
  two means - group sequential, 475-1
  two proportions - group sequential, 220-1
Poisson distribution
  probability calculator, 915-5
  simulation, 920-9
Poisson incidence
  post-marketing surveillance, 135-1
Poisson mean (one), 412-1
  examples, 412-4
  validation, 412-9
Poisson means (two), 437-1
  examples, 437-5
  validation, 437-10
Poisson regression, 870-1
  examples, 870-7
  validation, 870-9
Population size
  t-test, 400-8, 485-8
Post-marketing surveillance, 135-1
  examples, 135-6
  validation, 135-8
Power
  calculating, 6-4
  introduction, 6-1
  means, 8-1
  multiple comparisons - simulation, 580-4
Prevalence
  correlated proportions, 160-2, 165-2
Printing documentation, 1-7
Printing output, 5-4, 5-8
Probability calculator, 915-1
  Beta distribution, 915-1
  Binomial distribution, 915-2
  Bivariate normal distribution, 915-2
  Chi-square distribution, 915-2
  Correlation coefficient distribution, 915-3
  F distribution, 915-3
  Gamma distribution, 915-4
  Hotelling’s T2 distribution, 915-4
  Hypergeometric distribution, 915-4
  Negative binomial distribution, 915-5
  Normal distribution, 915-5
  Poisson distribution, 915-5
  Student’s t distribution, 915-6
  Studentized range distribution, 915-6
  Weibull distribution, 915-6
Procedure menus
  procedure window, 4-4
Procedure options, 4-1
Procedure window, 4-1
  file menu, 4-3
  help menu, 4-5
  procedure menus, 4-4
  run menu, 4-4
  tabs, 4-6
  toolbar, 4-5
  tools menu, 4-4
  view menu, 4-4
  window menu, 4-4
Producer’s risk
  exponential mean (one), 405-2
Program updates, 1-2
Proportion (one)
  binomial model, 100-2, 270-2
  confidence interval, 115-1
  continuity correction, 100-5
  equivalence, 110-1
  exact binomial test, 100-4
  examples, 100-13
hypergeometric model, 100-2
inequality, 100-1
non-inferiority, 105-1
odds ratio, 100-6
saw-tooth power function, 100-16
single-stage phase II trials, 120-1
superiority, 103-1
two-stage phase II trials, 125-1
validation, 100-20
Z test, 100-4
Proportion (one) - confidence interval
examples, 115-6
validation, 115-9
Proportion (one) - equivalence
test, 110-6
validation, 110-17
Proportion (one) - non-inferiority
test, 105-10
validation, 105-16
Proportion (one) - non-zero null
test, 103-10
validation, 103-14
Proportion surviving parameterization
logrank, 711-2, 715-2
Proportional trend test, 255-1
Proportions
comparing, 7-1
difference, 7-2
interpretation, 7-3
introduction, 7-1
logistic regression, 860-1
odds ratio, 7-3
odds ratio estimator, 910-1
paired (equivalence), 165-1
paired (non-inferiority), 160-1
ratio, 7-2
Proportions (many)
trend, 255-1
Proportions (many) - trend
test, 255-12
validation, 255-19
Proportions (multiple)
Chi-square test, 250-1
Proportions (two)
Chi-square test, 200-5, 223-6
cluster randomized - equivalence, 240-1
cluster randomized - non-inferiority, 235-1
cluster randomized – non-zero null, 233-1
cluster randomized design, 230-1
certainty intervals, 216-1
equivalence, 215-1
Fisher's exact, 200-4
group sequential test, 220-1, 221-1
group sequential test – non-inferiority, 223-1, 478-1
independent, 200-1
independent - equivalence, 215-1
independent - non-inferiority & superiority, 210-1
inequality, 200-1
inequality - offset, 205-1
interim analysis, 220-1, 221-1
interim analysis – non-inferiority, 223-1
interim analysis – non-zero null, 222-1
matched case control, 155-1
McNemar test, 150-1
non-inferiority & superiority, 210-1
stratified, 225-1
Proportions (two) - cluster - equivalence
test, 240-8
validation, 240-12
Proportions (two) - cluster - non-inferiority
test, 235-7
validation, 235-11
Proportions (two) - cluster - non-zero null
test, 233-7
validation, 233-11
Proportions (two) - cluster randomized design
test, 230-9
validation, 230-17
Proportions (two) - confidence interval
test, 216-26
validation, 216-28
Proportions (two) - equivalence
test, 215-17
validation, 215-23
Proportions (two) - group sequential
test, 220-9, 221-18
validation, 220-17, 221-29
Proportions (two) - inequality
test, 200-14
validation, 200-19
Proportions (two) – non-inferiority – group sequential
test, 223-22
validation, 223-33
Proportions (two) - non-inferiority
test, 210-17
validation, 210-24
Proportions (two) – non-zero null – group sequential
test, 222-22
validation, 222-33
Proportions (two) - offset
test, 205-18
validation, 205-24
Proportions (two) - stratified design
test, 225-8
validation, 225-11
Proportions (two) correlated
equivalence, 165-1
non-inferiority, 160-1
Proportions estimator, 910-1
validation, 910-2
Index-16

Q
Qualitative factors
  D-optimal designs, 888-5
Quitting PASS, 4-3

R
Random assignment, 880-1
Random effects models, 571-1
  mixed models, 571-2
Random factor
  ANOVA, 560-5
Random numbers, 920-1
  multiple comparisons - simulation, 580-5
  simulation, 585-5, 590-5
Random sorting, 880-3
Random sorting using max % deviation, 880-5
Randomization
  complete, 880-3
  Efron’s biased coin, 880-3
  Latin square designs, 884-2
  random sorting, 880-3
  random sorting using max % deviation, 880-5
  Smith, 880-4
  Wei’s urn, 880-4
Randomization lists, 880-1
  examples, 880-9
Randomized block ANOVA, 561-1
  examples, 565-9
  validation, 565-11
Randomized block design
  design generator, 889-5
Range test, 670-4
Rater agreement - kappa, 811-1
Rating data
  ROC curve (one), 260-4
  ROC curves (two), 265-3
Ratio
  proportions, 7-2
  two means, 445-1
  two means - equivalence, 470-1
Ratio (two means)
  non- zero null, 453-1
  non-inferiority, 455-1
Ratio of two means
  examples, 445-6
  validation, 445-8
Ratio of two means - equivalence
  examples, 470-7
  validation, 470-9
Ratio -two means
cross-over, 505-1
Regression
  confidence interval, 856-1
  Cox, 850-1
  linear, 855-1
  logistic, 860-1
  multiple, 865-1
  Poisson, 870-1
Regression - confidence interval
  examples, 856-7
  validation, 856-10
Rejection region, 6-3
Repeated measures
  mixed models, 571-1
  two means, 431-1
  two proportions, 201-1
Repeated measures - two means
  examples, 431-13
  validation, 431-19
Repeated measures - two proportions
  examples, 201-16
  validation, 201-24
Repeated measures ANOVA, 570-1
  examples, 570-29
  validation, 570-49
Repeated measures design
  design generator, 889-6
Replication
two-level designs, 881-3
Reports tab, 4-8
Requirements, 1-1
Resetting a template, 4-3
Response surface designs, 885-1
  examples, 885-3
Restricted maximum likelihood
  mixed models, 571-9
Risk ratio
  Blackwelder, 205-26
ROC curve (one), 260-1
  examples, 260-8
  validation, 260-12
ROC curves (two), 265-1
  examples, 265-8
  validation, 265-13
R-squared
  logistic regression, 860-7
  multiple regression, 865-1
RTF files, 5-1
Ruler, 5-4
Run menu
  procedure window, 4-4
Running PASS, 2-1

S
Sample size
  introduction, 6-1
Save template, 4-3
Saw-tooth power function
  one proportion, 100-16
Scale
  histogram, 945-4
  scatter plot, 940-1
Scaling
  histogram, 945-2
Scaling factors
D-optimal designs, 888-2

Scatter plot
alignment, 940-2
axis, 940-1
axis line, 940-3
background color, 940-11
bars, 940-11
boundaries, 940-1
color, 940-11
comments, 940-10
connecting points, 940-10
corners, 940-1
data, 940-11
data bars, 940-11
data ticks, 940-10
decimals, 940-4
example data, 940-11
font, 940-2
grid lines, 940-5
labels, 940-2
layout, 940-2
log base, 940-2
logarithm, 940-1
major ticks, 940-4
maximum, 940-1
minimum, 940-1
minor ticks, 940-4
notes, 940-10
painting order, 940-11
scale, 940-1
settings, 940-1
size, 940-11
symbols, 940-6, 940-10
text, 940-10
ticks, 940-3, 940-10
titles, 940-7
trend lines, 940-10

Score test
two proportions - equivalence, 215-7
two proportions - non-inferiority, 210-8
two proportions - offset, 205-7

Screening designs, 886-1
examples, 886-3

Search
PASS home window, 3-5

Sensitivities (paired)
examples, 276-7
validation, 276-11

Sensitivities (two)
examples, 275-9
validation, 275-13

Sensitivity
correlated proportions, 160-2, 165-2
one-sample, 270-1
paired, 276-1
ROC curve (one), 260-1
two-sample, 275-1

Sensitivity (one)
examples, 270-6
validation, 270-11

Serial number, 1-2

Settings
histogram, 945-1
Shapiro-Wilk test, 670-4
Sign test
simulation, 410-3
Sign test - paired means
simulation, 490-4
Sign test - paired means - equivalence
simulation, 495-5

Significance boundaries, 221-1, 222-1, 223-1, 476-1, 477-1, 478-1

Significance level, 6-3
adjusting, 8-5
Significance level - simulation
multiple comparisons, 580-1, 585-2
multiple contrasts, 590-1

Simon
two-stage phase II trials, 125-1

Simple linear regression
confidence interval, 856-1

Simulation, 8-3, 920-1
Beta distribution, 920-3
Binomial distribution, 920-5
Cauchy distribution, 920-5
Constant distribution, 920-6
contaminated normal, 920-19
Exponential distribution, 920-6
F distribution, 920-7
Gamma distribution, 920-7
Likert-scale, 920-8, 920-21
Multinomial distribution, 920-8
multiple comparisons, 580-1, 585-2, 585-4
multiple contrasts, 590-1, 590-4
Normal distribution, 920-9, 920-18
normality tests, 670-1
one mean, 410-1
one mean – control chart, 290-1
one-way ANOVA, 555-1
paired means, 490-1
paired means - equivalence, 495-1
Poission distribution, 920-9
random number generation, 580-5, 585-5, 590-5
size, 8-3
skewed distribution, 920-10
Student's T distribution, 920-10
syntax, 920-13
T distribution, 920-10
Tukey's lambda distribution, 920-10
two means, 440-1
two means - equivalence, 465-1
Uniform distribution, 920-11
variability – control chart, 295-1
Weibull distribution, 920-12

Simulation steps
mixed models, 571-17

Single-stage design
one proportion, 120-1
Single-stage phase II trials  
examples, 120-3  
validation, 120-3  
Size  
  histogram, 945-11  
  scatter plot, 940-11  
Skewed data  
  one-way ANOVA - simulation, 555-20  
  simulation, 410-12  
Skewed data (two means)  
  simulation, 440-20  
Skewness distribution  
  simulation, 920-10  
Slope  
  linear regression, 855-1  
Slope - simple linear regression  
  confidence interval, 856-1  
Smith's randomization, 880-4  
Specificity  
  correlated proportions, 160-2, 165-2  
  one-sample, 270-1  
  ROC curve (one), 260-1  
Specificity (one)  
  examples, 270-6  
  validation, 270-11  
Spending functions  
  survival - group sequential, 710-3  
  two means - group sequential, 475-2, 476-5, 477-4  
  two proportions - group sequential, 220-2, 221-7  
  two proportions - group sequential – non-zero null, 222-10, 223-10  
Split plot analysis  
  mixed models, 571-1  
Spreadsheet, 925-1  
  edit menu, 925-2  
  file menu, 925-2  
  help menu, 925-3  
  menus, 925-2  
Standard deviation, 8-2  
  estimator, 905-1  
  interpretation, 905-1  
Standard deviation (one)  
  confidence interval, 640-1  
Standard deviation (one) - confidence interval  
  examples, 640-5  
  validation, 640-7  
Standard deviation (one) - relative error  
  confidence interval, 642-1  
Standard deviation (one) - relative error - confidence interval  
  examples, 642-4  
  validation, 642-6  
Standard deviation (one) - tolerance - confidence interval  
  examples, 641-6  
  validation, 641-8  
Standard deviation (one) - tolerance probability  
  confidence interval, 641-1  
Standard deviation estimator  
  examples, 905-6  
  Standard deviation of means estimator, 912-1  
Standard deviation test  
  one, 650-1  
  two, 655-1  
Starting PASS, 1-2, 2-1  
Stratified designs  
  two proportions, 225-1  
Student's t distribution  
  probability calculator, 915-6  
Studentized range distribution  
  probability calculator, 915-6  
Student's T distribution  
  simulation, 920-10  
Superiority  
  one proportion, 103-1  
  paired t-test, 413-1  
  two proportions, 210-1  
Superiority hypothesis, 6-8  
Survival  
  logrank, 700-1, 715-1  
  logrank - Lachin and Foulkes, 705-1  
  non-inferiority, 706-1  
Survival - group sequential  
  examples, 710-11, 711-27  
  validation, 710-18, 711-40  
Survival - logrank  
  examples, 700-6  
  validation, 700-9  
Survival - logrank - Lachin and Foulkes  
  examples, 705-8  
  validation, 705-12  
Survival Parameter Conversion Tool, 903-1  
Symbols  
  scatter plot, 940-6, 940-10  
Syntax  
  macros, 930-2  
System requirements, 1-1

T  
T distribution  
  simulation, 920-10  
T test  
  two proportions - equivalence, 215-7  
  two proportions - non-inferiority, 210-8  
T2  
  Hotelling’s, 600-1  
Tabs  
  data, 4-6  
  iterations, 4-7  
  plot setup, 4-9  
  reports, 4-8  
Tabs on the procedure window, 4-6  
Taguchi designs, 887-1  
  examples, 887-4  
Tarone-Ware test  
  Group Sequential, 711-5  
Technical support, 1-9
Templates, 4-1
  automatic, 4-2
  creating a new, 4-3
  default, 4-2
  loading, 4-3
  loading examples, 4-3
  opening, 4-3
  opening examples, 4-3
  saving, 4-3
Test statistics, 6-6
Text
  histogram, 945-10
  scatter plot, 940-10
Three-stage design
  one proportion, 130-1
Three-stage phase II trials
  examples, 130-8
  validation, 130-8
Ticks
  histogram, 945-5
  scatter plot, 940-3, 940-10
Time averaged difference
  binary data, 201-1
  normal data, 431-1
  power for, 201-2, 431-2
  two means, 431-1
  two proportions, 201-1
Titles
  histogram, 945-8
  scatter plot, 940-7
Toolbar
  output window, 5-8
  PASS home window, 3-3
  procedure window, 4-5
Tools menu
  procedure window, 4-4
Trend in proportions, 255-1
Trend lines
  scatter plot, 940-10
Trimmed t-test (two means)
  simulation, 440-4
Trimmed t-test (two means) - equivalence
  simulation, 465-4
T-test
  cross-over, 500-3
  microarray data, 610-1, 615-1
  one group - multiple testing adjustment, 610-4
  one mean, 400-1
  one mean - simulation, 410-1
  paired, 400-1, 485-1
  paired - equivalence - simulation, 495-1
  paired - simulation, 490-1
  two groups - multiple testing adjustment, 615-2
  two means, 430-1
  two means - equivalence, 460-1
  two means - non-inferiority, 450-1
  two means - non-zero null, 448-1
  two means - simulation, 440-1
  two proportions, 200-8, 223-6
T-test - equivalence
  cross-over, 520-3
T-test - simulation
  two means - equivalence, 465-1
T-test (one mean)
  assumptions, 400-3
  examples, 400-8
  non-inferiority, 415-1
  non-zero null, 413-1
  superiority, 413-1
  validation, 400-17
T-test (one mean) - non-zero null
  examples, 413-8
  validation, 413-10
T-test (one mean) - non-inferiority
  examples, 415-8
  validation, 415-11
T-test (one mean) - simulation
  examples, 410-10
  validation, 410-16
T-test (paired means) - equivalence - simulation
  examples, 495-11
  validation, 495-17
T-test (paired means) - simulation
  examples, 490-11
  validation, 490-16
T-test (paired)
  examples, 400-13, 485-8
  validation, 485-10
T-test (two means)
  assumptions, 430-3
  examples, 430-10
  validation, 430-18
T-test (two means) - simulation
  examples, 440-12
  validation, 440-16
Tukey-Kramer
  simulation, 580-1
Tukey-Kramer test
  multiple comparisons, 575-7
  simulation, 580-3
Tukey's lambda distribution
  simulation, 920-10
Two correlated proportions - equivalence
  examples, 165-11
  validation, 165-14
Two correlated proportions - non-inferiority
  examples, 160-10
  validation, 160-13
Two means
  cluster randomized design, 480-1
  cross-over, 500-1
Two means - cluster randomized design
  examples, 480-4
  validation, 480-6
Two means - equivalence
  cross-over, 520-1
  examples, 460-5
  validation, 460-7
Two means - equivalence - higher-order
  cross-over, 540-1, 545-1
Two means - equivalence - simulation
  examples, 465-12
  validation, 465-19
Two means - group sequential
  examples, 475-10, 476-17, 477-15
  validation, 475-17, 476-28, 477-26
Two means – non- inferiority – group sequential
  examples, 478-17
  validation, 478-28
Two means - non- zero null
  cross-over, 508-1
  examples, 448-10
  validation, 448-13
Two means - non- zero null - higher-order
  cross-over, 528-1
Two means - non-inferiority
  cross-over, 510-1
  examples, 450-10
  validation, 450-13
Two means - non-inferiority - higher-order
  cross-over, 530-1
Two means - ratio
  cross-over, 505-1
  examples, 445-6
  validation, 445-8
Two means - ratio - equivalence
  cross-over, 525-1
  examples, 470-7
  validation, 470-9
Two means - ratio - non- zero null - higher-order
  cross-over, 533-1
Two means - ratio - non-inferiority
  cross-over, 515-1
Two means - ratio - non-inferiority - higher-order
  cross-over, 535-1
Two means - ratio - non-unity null
  cross-over, 513-1
Two means (ratio) - non- zero null
  examples, 453-6
  validation, 453-8
Two means (ratio) - non-inferiority
  examples, 455-6
  validation, 455-8
Two ordered categorical variables, 253-1
  examples, 253-5
  validation, 253-10
Two proportions
  cluster randomized - equivalence, 240-1
  cluster randomized - non-inferiority, 235-1
  cluster randomized - non-zero null, 233-1
  cluster randomized design, 230-1
  equivalence, 215-1
  inequality, 200-1
  inequality - offset, 205-1
  matched case-control, 155-1
  McNemar test, 150-1
  non-inferiority & superiority, 210-1
Two proportions - cluster - equivalence
  examples, 240-8
  validation, 240-12
Two proportions - cluster - non-inferiority
  examples, 235-7
  validation, 235-11
Two proportions - cluster - non-zero null
  examples, 233-7
  validation, 233-11
Two proportions - cluster randomized design
  examples, 230-9
  validation, 230-17
Two proportions - equivalence
  examples, 215-17
  validation, 215-23
Two proportions - group sequential
  examples, 220-9, 221-18
  validation, 220-17, 221-29
Two proportions - inequality
  examples, 200-14
  validation, 200-19
Two proportions – non- inferiority – group sequential
  examples, 223-22
  validation, 223-33
Two proportions - non-inferiority
  examples, 210-17
  validation, 210-24
Two proportions – non-zero null – group sequential
  examples, 222-22
  validation, 222-33
Two proportions - offset
  examples, 205-18
  validation, 205-24
Two proportions - stratified design
  examples, 225-8
  validation, 225-11
Two-channel arrays, 610-1
Two-level designs, 881-1
  examples, 881-6
Two-level factorial designs, 881-1
Two-sample t-test, 430-1
  equivalence, 460-1
  microarray data, 615-1
  non-inferiority, 450-1
  non-zero null, 448-1
Two-sample t-test - simulation
  equivalence, 465-1
Two-stage design
  one proportion, 125-1
Two-stage phase II trials
  examples, 125-8
  validation, 125-10
Type-I error, 6-2
Type-II error, 6-2

Uniform distribution
  simulation, 920-11
Updates, 1-2
Validation

ANCOVA, 551-14
Chi-square test, 250-9
Cochran-Armitage test, 255-19
correlation (one) - confidence interval, 801-7
Cox regression, 850-6
cross-over - higher-order - equivalence, 540-12
cross-over - higher-order - non-zero null, 528-11
cross-over - higher-order - non-inferiority, 530-11
cross-over - ratio - equivalence, 525-9
cross-over - ratio - higher-order - equivalence, 545-12
cross-over - ratio - higher-order - non-zero null, 533-12
cross-over - ratio - higher-order - non-inferiority, 535-12
cross-over - ratio - non-inferiority, 515-8
cross-over (two means), 500-10
cross-over (two means) - equivalence, 520-10
cross-over (two means) - non-unity null, 513-8
cross-over (two means) - non-zero null, 508-10
cross-over (two means) - non-inferiority, 510-10
cross-over (two means) - ratio, 505-8
equivalence - two correlated proportions, 165-14
exponential mean (one), 405-10
exponential means (two), 435-7
fixed effects ANOVA, 560-19
Hotelling’s T2, 600-11
intra-class correlation, 810-6
kappa, 811-11
linear regression, 855-7
logistic regression, 860-14
logrank tests - non-inferiority, 706-11
logrank tests (Lakatos), 715-21
MANOVA, 605-17
many proportions - trend, 255-19
matched case-control - proportions, 155-10
McNemar test - two correlated proportions, 150-9
mean ratio - non-zero null, 453-8
mean ratio - non-inferiority, 455-8
microarray one-sample or paired t-test, 610-19
microarray two-sample t-test, 615-19
minimum effective dose, 595-7
mixed models, 571-42
multiple comparisons, 575-19
multiple comparisons - simulation, 580-22
multiple comparisons - vs control - simulation, 585-21
multiple contrasts - simulation, 590-24
multiple regression, 865-10
non-inferiority - two correlated proportions, 160-13
normality - simulation, 670-11
one coefficient alpha, 815-7
one correlation, 800-7
one mean - confidence interval, 420-7
one mean - tolerance - confidence interval, 421-8
one proportion, 100-20
one proportion - confidence interval, 115-9
one proportion - equivalence, 110-17
one proportion - non-inferiority, 105-16
one proportion - non-zero null, 103-14
one sensitivity, 270-11
one specificity, 270-11
one standard deviation - confidence interval, 640-7
one standard deviation - relative error - confidence interval, 642-6
one standard deviation - tolerance - confidence interval, 641-8
one variance - confidence interval, 651-6
one variance - relative error - confidence interval, 653-6
one variance - tolerance - confidence interval, 652-8
one-way ANOVA, 550-19
one-way ANOVA - simulation, 555-16
paired means - confidence interval, 496-6
paired means - tolerance - confidence interval, 497-8
paired sensitivities, 276-11
Poisson mean (one), 412-9
Poisson means (two), 437-10
Poisson regression, 870-9
post-marketing surveillance, 135-8
randomized block ANOVA, 565-11
ratio of two means, 445-8
ratio of two means - equivalence, 470-9
regression - confidence interval, 856-10
repeated measures - two means, 431-19
repeated measures - two proportions, 201-24
repeated measures ANOVA, 570-49
ROC curve (one), 260-12
ROC curves (two), 265-13
single-stage phase II trials, 120-3
survival - group sequential, 710-18, 711-40
survival - logrank, 700-9
survival - logrank - Lachin and Foulkes, 705-12
three-stage phase II trials, 130-8
t-test (one mean), 400-17
t-test (one mean) - non-zero null, 413-10
t-test (one mean) - non-inferiority, 415-11
t-test (one mean) - simulation, 410-16
t-test (paired means) - equivalence - simulation, 495-17
t-test (paired means) - simulation, 490-16
t-test (paired), 485-10
t-test (two means), 430-18
t-test (two means) - simulation, 440-16
two coefficient alphas, 820-10
two correlated proportions - equivalence, 165-14
two correlated proportions - non-inferiority, 160-13
two correlations, 805-8
two means - cluster randomized, 480-6
two means - confidence interval, 471-7
two means - equivalence, 460-7
two means - equivalence - simulation, 465-19
two means - group sequential, 475-17, 476-28, 477-26
two means - non-zero null, 448-13
two means - non-inferiority, 450-13
two means – non-inferiority – group sequential, 478-26
four means - ratio, 445-8
two means - ratio - equivalence, 470-9
two means - tolerance - confidence interval, 472-8
two ordered categorical variables, 253-10
two proportions - cluster - equivalence, 240-12
two proportions - cluster - non-inferiority, 235-11
two proportions - cluster - non-zero null, 233-11
two proportions - cluster randomized, 230-17
two proportions - confidence interval, 216-28
two proportions - equivalence, 215-23
two proportions - group sequential, 220-17, 221-29
two proportions - inequality, 200-19
two proportions – non-inferiority – group sequential, 223-33
two proportions - non-inferiority, 210-24
two proportions – non-zero null – group sequential, 222-33
two proportions - offset, 205-24
two proportions - stratified design, 225-11
two sensitivities, 275-13
two-stage phase II trials, 125-10
variance (one), 650-7
variance ratio - confidence interval, 656-8
variance ratio - relative error - confidence interval, 657-6
variances (two), 655-8
Williams’ test, 595-1
Walter’s confidence intervals
two proportions, 216-11
Wei’s urn randomization, 880-4
Weibull distribution
probability calculator, 915-6
simulation, 920-12
Welch test
multiple contrasts - simulation, 590-4
Welch's test - simulation
equivalence, 465-1
Welch's t-test
non-zero null, 448-4
non-inferiority, 450-4
simulation, 440-3
Wilcoxon test, 400-1, 400-7, 413-7, 415-7, 448-9, 450-9, 485-7
assumptions, 400-3, 485-3
non-inferiority, 415-1
simulation, 410-3
superiority, 413-1
Wilcoxon test - paired means
simulation, 490-4
Wilcoxon test - paired means - equivalence
simulation, 495-4
Wilks’ Lambda, 605-1
MANOVA, 605-2
repeated measures ANOVA, 570-1, 570-6
Williams test
two proportions, 216-7
Window menu
output window, 5-4
procedure window, 4-4
Winsorized test (two means) - equivalence
simulation, 465-4
Within standard deviation
repeated measures ANOVA, 570-15
Within-subjects design
repeated measures ANOVA, 570-3
Z

Z test
  one proportion, 100-4

  one proportion - equivalence, 110-6
  two proportions - equivalence, 215-5
  two proportions - non-inferiority, 210-6
  two proportions - offset, 205-6