Chapter 520

Two Correlated Proportions (McNemar Test)

Introduction

This procedure computes confidence intervals and hypothesis tests for the comparison of the marginal frequencies of two factors (each with two levels) based on a 2-by-2 table of *n* pairs. Confidence limits can be obtained for the marginal probability difference, ratio, or odds ratio. Inequality tests are available for the marginal probability difference and ratio.

Experimental Design

A typical design for this scenario involves *N* pairs of individuals where a dichotomous measurement of one factor is measured on one of the individuals of the pair (case), and a second dichotomous measurement based on a second factor is measured on the second individual of the pair (control). Or similarly *N* individuals are measured twice, once for each of two dichotomous factors.

Comparing Two Correlated Proportions

Suppose you have two dichotomous measurements Y_1 and Y_2 on each of *N* subjects (where in many cases the 'subject' may be a pair of matched individuals). The proportions P_1 and P_2 represent the success probabilities. That is,

$$P_1 = \Pr(Y_1 = 1)$$

$$P_2 = \Pr(Y_2 = 1)$$

The data from this design can be summarized in the following 2-by-2 table:

	$Y_2 = 1$ (Yes, Present)	$Y_2 = 0$ (No, Absent)	Total
$Y_1 = 1$ (Yes, Present)	*A*	*B*	*A* + *B*
$Y_1 = 0$ (No, Absent)	*C*	*D*	*C* + *D*
Total	*A* + *C*	*B* + *D*	*N*

The marginal proportions P_1 and P_2 are estimated from these data using the formulae

$$p_1 = \frac{A+B}{N} \text{ and } p_2 = \frac{A+C}{N}$$

Three quantities which allow these proportions to be compared are

Quantity	Notation
Difference	$\Delta = P_1 - P_2$
Risk Ratio	$\phi = P_1 / P_2$
Odds Ratio	$\psi = \frac{O_1}{O_2}$

Although these three parameters are (non-linear) functions of each other, the choice of which is to be used should not be taken lightly. The associated tests and confidence intervals of each of these parameters can vary widely in power and coverage probability.

Difference

The proportion (risk) difference $\delta = P_1 - P_2$ is perhaps the most direct method of comparison between the two event probabilities. This parameter is easy to interpret and communicate. It gives the absolute impact of the treatment. However, there are subtle difficulties that can arise with its interpretation.

One interpretation difficulty occurs when the event of interest is rare. If a difference of 0.001 were reported for an event with a baseline probability of 0.40, we would probably dismiss this as being of little importance. That is, there usually is little interest in a treatment that decreases the probability from 0.400 to 0.399. However, if the baseline probably of a disease was 0.002 and 0.001 was the decrease in the disease probability, this would represent a reduction of 50%. Thus we see that interpretation depends on the baseline probability of the event.

A similar situation occurs when the amount of possible difference is considered. Consider two events, one with a baseline event rate of 0.40 and the other with a rate of 0.02. What is the maximum decrease that can occur? Obviously, the first event rate can be decreased by an absolute amount of 0.40 while the second can only be decreased by a maximum of 0.02.

So, although creating the simple difference is a useful method of comparison, care must be taken that it fits the situation.

Ratio

The proportion (risk) ratio $\phi = p_1 / p_2$ gives the relative change in risk in a treatment group (group 1) compared to a control group (group 2). This parameter is also direct and easy to interpret. To compare this with the difference, consider a treatment that reduces the risk of disease from 0.1437 to 0.0793. Which single number is most enlightening, the fact that the absolute risk of disease has been decreased by 0.0644, or the fact that risk of disease in the treatment group is only 55.18% of that in the control group? In many cases, the percentage (100 x risk ratio) communicates the impact of the treatment better than the absolute change.

Perhaps the biggest drawback of this parameter is that it cannot be calculated in one of the most common experimental designs: the case-control study. Another drawback, when compared to the odds ratio, is that the odds ratio occurs naturally in the likelihood equations and as a parameter in logistic regression, while the proportion ratio does not.

Odds Ratio

Chances are usually communicated as long-term proportions or probabilities. In betting, chances are often given as odds. For example, the odds of a horse winning a race might be set at 10-to-1 or 3-to-2. How do you translate from odds to probability? An odds of 3-to-2 means that the event will occur three out of five times. That is, an odds of 3-to-2 (1.5) translates to a probability of winning of 0.60.

The odds of an event are calculated by dividing the event risk by the non-event risk. Thus, in our case of two populations, the odds are

$$O_1 = \frac{P_1}{1-P_1} \text{ and } O_2 = \frac{P_2}{1-P_2}$$

For example, if P_1 is 0.60, the odds are 0.60/0.40 = 1.5. In some cases, rather than representing the odds as a decimal amount, it is re-scaled into whole numbers. Thus, instead of saying the odds are 1.5-to-1, we may equivalently say they are 3-to-2.

In this context, the comparison of proportions may be done by comparing the odds through the ratio of the odds. The odds ratio of two events is

$$\psi = \frac{O_1}{O_2} = \frac{\frac{P_1}{1-P_1}}{\frac{P_2}{1-P_2}}$$

In the case of two correlated proportions, the odds ratio is calculated as

$$\psi = \frac{B}{C}$$

Until one is accustomed to working with odds, the odds ratio is usually more difficult to interpret than the proportion (risk) ratio, but it is still the parameter of choice for many researchers. Reasons for this include the fact that the odds ratio can be accurately estimated from case-control studies, while the risk ratio cannot. Also, the odds ratio is the basis of logistic regression (used to study the influence of risk factors). Furthermore, the odds ratio is the natural parameter in the conditional likelihood of the two-group, binomial-response design. Finally, when the baseline event-rates are rare, the odds ratio provides a close approximation to the risk ratio since, in this case, $1-P_1 \approx 1-P_2$, so that

$$\psi = \frac{\frac{P_1}{1-P_1}}{\frac{P_2}{1-P_2}} \approx \frac{P_1}{P_2} = \phi$$

One benefit of the log of the odds ratio is its desirable statistical properties, such as its continuous range from negative infinity to positive infinity.

Confidence Intervals

Several methods for computing confidence intervals for proportion difference, proportion ratio, and odds ratio have been proposed. We now show the methods that are available in **NCSS**.

Difference

Four methods are available for computing a confidence interval of the difference between the two proportions $\Delta = P_1 - P_2$. The lower (L) and upper (U) limits of these intervals are computed as follows. Note that $z = |z_{\alpha/2}|$ is the appropriate percentile from the standard normal distribution.

Newcombe (1998) conducted a comparative evaluation of ten confidence interval methods. He recommended that the modified Wilson score method be used instead of the Pearson Chi-square or the Yate's Corrected Chi-square.

Nam's Score

For details, see Nam (1997) or Tango (1998). The lower limit is the solution of

$$L = \inf\left\{\Delta_0 : \frac{\hat{\Delta} - \Delta_0}{\tilde{\sigma}_{\Delta_0}} < z\right\}$$

and the upper limit is the solution of

$$U = \sup\left\{\Delta_0 : \frac{\hat{\Delta} - \Delta_0}{\tilde{\sigma}_{\Delta_0}} > -z\right\}$$

where $\tilde{\sigma}_{\Delta_0}$ is given by

$$\tilde{\sigma}_{\Delta} = \frac{\tilde{p}_{21} + \tilde{p}_{12} - \Delta^2}{n}$$

$$\tilde{p}_{21} = \left\{\frac{-e + \sqrt{e^2 - 8f}}{4}\right\}$$

$$\tilde{p}_{12} = \tilde{p}_{21} - \Delta$$

$$e = -\hat{\Delta}(1 - \Delta) - 2(\hat{p}_{21} + \Delta)$$

$$f = \Delta(1 + \Delta)\hat{p}_{21}$$

Wilson's Score as modified by Newcombe

For further details, see Newcombe (1998c), page 2639. This is Newcombe's method 10.

$$L = \hat{\Delta} - \delta$$

$$U = \hat{\Delta} + \varepsilon$$

where

$$\delta = \sqrt{f_2^2 - 2\hat{\phi} f_2 g_3 + g_3^2}$$

$$\varepsilon = \sqrt{g_2^2 - 2\hat{\phi} g_2 f_3 + f_3^2}$$

$$f_2 = \frac{(A+B)}{N} - l_2$$

$$g_2 = u_2 - \frac{(A+B)}{N}$$

$$f_3 = \frac{(A+C)}{N} - l_3$$

$$g_3 = u_3 - \frac{(A+C)}{N}$$

and l_2 and u_2 are the roots of

$$\left| x - \frac{A+B}{N} \right| = z\sqrt{\frac{x(1-x)}{N}}$$

and l_3 and u_3 are the roots of

$$\left| x - \frac{A+C}{N} \right| = z\sqrt{\frac{x(1-x)}{N}}$$

$$\hat{\phi}$$

$$\hat{\phi} = \begin{cases} \dfrac{\max(\text{AD - BC - N/2,0})}{\sqrt{(\text{A + B})(C+D)(A+C)(B+D)}} & \text{if } AD > BC \\ \dfrac{\text{AD - BC}}{\sqrt{(\text{A + B})(C+D)(A+C)(B+D)}} & \text{otherwise} \end{cases}$$

Note that if the denominator of $\hat{\phi}$ is zero, $\hat{\phi}$ is set to zero.

Wald Z Method

For further details, see Newcombe (1998c), page 2638

$$L = \hat{\Delta} - z s_W$$

$$U = \hat{\Delta} + z s_W$$

where

$$\hat{\Delta} = p_1 - p_2 = (B - C)/N$$

$$s_W^2 = \frac{(A+D)(B+C) + 4BC}{N^3}$$

Wald Z Method with Continuity Correction

For details, see Newcombe (1998c), page 2638.

$$L = \hat{\Delta} - z s_W - \frac{1}{N}$$

$$U = \hat{\Delta} + z s_W + \frac{1}{N}$$

Ratio

Two methods are available for computing a confidence interval of the risk ratio $\phi = P_1 / P_2$. Note that $z = |z_{\alpha/2}|$ is the appropriate percentile from the standard normal distribution.

Nam and Blackwelder (2002) present two methods for computing confidence intervals for the risk ratio. These are presented here. Note that the score method is recommended.

Score (Nam and Blackwelder)

For details, see Nam and Blackwelder (2002), page 691. The lower limit is the solution of

$$z(\phi) = |z_{\alpha/2}|$$

and the upper limit is the solution of

$$z(\phi) = -|z_{\alpha/2}|$$

where

$$z(\phi) = \frac{\sqrt{N}(p_1 - \phi p_2)}{\sqrt{\phi(\tilde{p}_{12} + \tilde{p}_{21})}}$$

and

$$\tilde{p}_{12} = \frac{-p_1 + \phi^2(p_2 + 2p_{12}) + \sqrt{(p_1 - \phi p_2)^2 + 4\phi^2 p_{12} p_{12}}}{2\phi(\phi + 1)}$$

$$\tilde{p}_{21} = \phi\tilde{p}_{12} - (\phi - 1)(1 - p_{22})$$

Wald Z (Nam and Blackwelder)

For details, see Nam and Blackwelder (2002), page 692. The lower limit is the solution of

$$z_W(\phi) = |z_{\alpha/2}|$$

and the upper limit is the solution of

$$z_W(\phi) = -|z_{\alpha/2}|$$

where

$$z_W(\phi) = \frac{\sqrt{N}(\hat{p}_1 - \phi\hat{p}_2)}{\sqrt{\phi(\hat{p}_{12} + \hat{p}_{21})}}$$

Odds Ratio

Sahai and Khurshid (1995) present two methods for computing confidence intervals of the odds ratio $\psi = O_1/O_2$. Note that the maximum likelihood estimate of this is given by

$$\hat{\psi} = B/C$$

Exact Binomial

The lower limit is

$$\psi_L = \frac{B}{(C+1)F_{\alpha/2,2C+2,2B}}$$

and the upper limit

$$\psi_U = \frac{B+1}{CF_{\alpha/2,2B+2,2C}}$$

where F is the ordinate of the F distribution.

Maximum Likelihood

The lower limit is

$$\psi_L = \exp\left\{\ln(\hat{\psi}) - z_{\alpha/2}s_{\hat{\psi}}\right\}$$

and the upper limit

$$\psi_U = \exp\left\{\ln(\hat{\psi}) + z_{\alpha/2}s_{\hat{\psi}}\right\}$$

where

$$s_{\hat{\psi}} = \sqrt{\frac{1}{B} + \frac{1}{C}}$$

Hypothesis Tests

Difference

This module tests three statistical hypotheses about the difference in the two proportions:

1. $H_0: P_1 - P_2 = \Delta$ versus $H_a: P_1 - P_2 \neq \Delta$; this is a *two-tailed test*.
2. $H_{0L}: P_1 - P_2 \geq \Delta$ versus $H_{aL}: P_1 - P_2 < \Delta$; this is a *one-tailed test*.
3. $H_{0U}: P_1 - P_2 \leq \Delta$ versus $H_{aU}: P_1 - P_2 > \Delta$; this is a *one-tailed test*.

McNemar Test

Fleiss (1981) presents a test that is attributed to McNemar for testing the two-tailed null hypothesis. This is calculated as

$$\chi_1^2 = \frac{(B - C)^2}{B + C}$$

For this test, Δ must be equal to 0.

McNemar Test with Continuity Correction

Fleiss (1981) also presents a continuity-corrected version of McNemar test. This is calculated as

$$\chi_1^2 = \frac{(|B - C| - 1)^2}{B + C}$$

For this test, Δ must be equal to 0.

McNemar Test (Exact Binomial)

This test uses the asymptotic McNemar Test statistic and enumerates all possible outcomes using the binomial distribution to provide a p-value. For this test, Δ must be equal to 0.

Nam Score Test

Liu *et al.* (2002) recommend a likelihood score test which was originally proposed by Nam (1997). The tests are calculated as

$$z_L = \frac{\hat{\Delta} + \Delta}{\tilde{\sigma}_L} \text{ and } z_U = \frac{\hat{\Delta} - \Delta}{\tilde{\sigma}_U}$$

where

$$\tilde{\sigma}_L = \sigma_{-\Delta}$$

$$\tilde{\sigma}_U = \sigma_{\Delta}$$

and

$$\sigma_D = \frac{\tilde{p}_{21} + \tilde{p}_{12} - D^2}{N}$$

$$\tilde{p}_{21} = \left\{ \frac{-e + \sqrt{e^2 - 8f}}{4} \right\}$$

$$\tilde{p}_{12} = \tilde{p}_{21} - D$$

$$e = -\hat{\Delta}(1 - D) - 2(p_{21} + D)$$

$$f = D(1 + D)p_{21}$$

Wald Z Test

Liu *et al.* (2002) present a pair of large-sample, Wald-type *z* tests for testing the two one-tailed hypothesis about the difference $p_1 - p_2 = \Delta$. These are calculated as

$$z_L = \frac{\hat{\Delta} + \Delta - \frac{1}{2N}}{\hat{\sigma}} \text{ and } z_U = \frac{\hat{\Delta} - \Delta + \frac{1}{2N}}{\hat{\sigma}}$$

where

$$\hat{\sigma}^2 = \frac{p_{21} + p_{12} - \hat{\Delta}^2}{N}$$

$$\hat{\Delta} = p_1 - p_2$$

Ratio

This module tests three statistical hypotheses about the difference in the two proportions:

1. $H_0: P_1 / P_2 = \phi$ versus $H_a: P_1 / P_2 \neq \phi$; this is a *two-tailed test*.
2. $H_{0L}: P_1 / P_2 \geq \phi$ versus $H_{aL}: P_1 / P_2 > \phi$; this is a *one-tailed test*.
3. $H_{0U}: P_1 / P_2 \geq \phi$ versus $H_{aU}: P_1 / P_2 < \phi$; this is a *one-tailed test*.

Nam Test

For details, see Nam and Blackwelder (2002), page 691. The test statistic for testing a specific value of ϕ is

$$z(\phi) = \frac{\sqrt{N}\left(p_1 - \phi p_2\right)}{\sqrt{\phi\left(\tilde{p}_{12} + \tilde{p}_{21}\right)}}$$

where

$$\tilde{p}_{12} = \frac{-p_1 + \phi^2\left(p_2 + 2p_{12}\right) + \sqrt{\left(p_1 - \phi p_2\right)^2 + 4\phi^2 p_{12} p_{12}}}{2\phi(\phi + 1)}$$

$$\tilde{p}_{21} = \phi\tilde{p}_{12} - (\phi - 1)(1 - p_{22})$$

Data Structure

This procedure can summarize data from a database or summarized count values can be entered directly into the procedure panel.

Example 1 – Analysis of Two Correlated Proportions

This section presents an example of how to run an analysis on hypothetical data. In this example, two dichotomous variables where measured on each of fifty subjects; 30 subjects scored ‘yes’ on both variables, 9 subjects scored ‘no’ on both variables, 6 scored a ‘yes’ and then a ‘no’, and 5 scored a ‘no and then a ‘yes’.

Setup

To run this example, complete the following steps:

1 Specify the Two Correlated Proportions (McNemar Test) procedure options

- Find and open the **Two Correlated Proportions (McNemar Test)** procedure using the menus or the Procedure Navigator.
- The settings for this example are listed below and are stored in the **Example 1** settings template. To load this template, click **Open Example Template** in the Help Center or File menu.

Option	Value
Data Tab	
Type of Data Input	**Enter Table of Counts**
Variable 1 Heading	**Var1**
Variable 1 Label of 1st Value	**Yes**
Variable 1 Label of 2nd Value	**No**
Variable 2 Heading	**Var2**
Variable 2 Label of 1st Value	**Yes**
Variable 2 Label of 2nd Value	**No**
N11	**30**
N12	**6**
N21	**5**
N22	**9**
Summary Reports Tab	
Counts and Proportions	**Checked**
Proportions Analysis	**Checked**
Difference Reports Tab	
Nam RMLE Score	**Checked**
Wilson Score	**Checked**
Test Direction	**Two-Sided (H0: P1 - P2 = D0 vs. Ha: P1 - P2 ≠ D0)**
H0 Difference (D0)	**0.0**
McNemar Test	**Checked**

2 Run the procedure

- Click the **Run** button to perform the calculations and generate the output.

Counts and Proportions Sections

Counts and Proportions

Var1	Var2 Yes Count	Var2 No Count	Total Count
Yes	30	6	36
No	5	9	14
Total	35	15	50

p1 = (36/50) = 0.7200
p2 = (35/50) = 0.7000

Proportions Analysis

Statistic	Value
Variable 1 Event Rate (p1)	0.7200
Variable 2 Event Rate (p2)	0.7000
Proportion Matching	0.7800
Proportion Not Matching	0.2200
Absolute Risk Difference |p1 - p2|	0.0200
Number Needed to Treat 1/|p1 - p2|	50.0000
Relative Risk Reduction |p1 - p2|/p2	0.0286
Relative Risk p1/p2	1.0286
Odds Ratio o1/o2	1.2000

These reports document the values that were input and give various statistics of these values.

Confidence Interval of the Difference (P1 - P2)

Confidence Intervals of the Difference (P1 - P2)

Confidence Interval Name	p1	p2	Difference p1 - p2	Lower 95% C.L. of P1 - P2	Upper 95% C.L. of P1 - P2	Confidence Interval Width
Nam RMLE Score*	0.7200	0.7000	0.0200	-0.1197	0.1606	0.2802
Wilson Score	0.7200	0.7000	0.0200	-0.1145	0.1537	0.2683

This report provides large sample confidence intervals of the difference.

McNemar Inequality Test

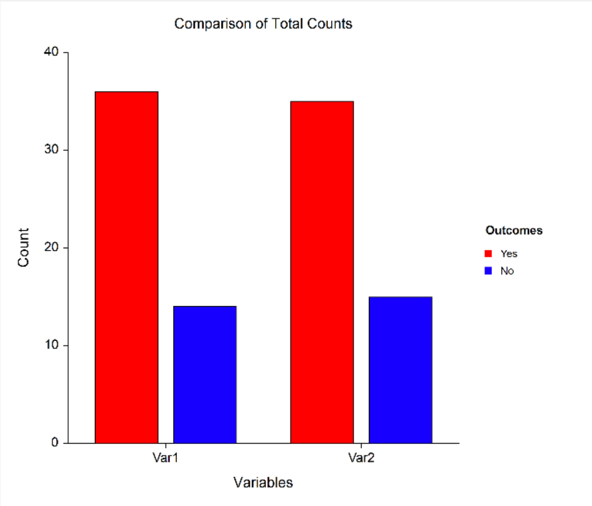
Two-Sided Hypothesis Tests of the Difference (P1 - P2)
H0: P1 - P2 = 0 vs. Ha: P1 - P2 ≠ 0

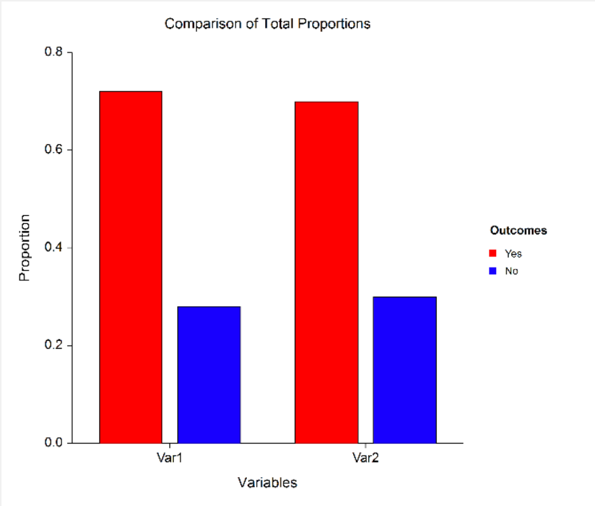
Test Statistic Name	p1	p2	Difference p1 - p2	Test Statistic Value	Prob Level	Reject H0 at α = 0.05?
McNemar	0.7200	0.7000	0.0200	0.091	0.7630	No

This report provides the McNemar test. The p-value of the test is the Prob Level.

Plots Section

Plots Section





These reports show the marginal totals and proportions of the two variables.