Chapter 501

Contingency Tables (Crosstabs / Chi-Square Test)

Introduction

This procedure produces tables of counts and percentages for the joint distribution of two categorical variables. Such tables are known as contingency, cross-tabulation, or crosstab tables. When a breakdown of more than two variables is desired, you can specify up to eight grouping (break) variables in addition to the two table variables. A separate table is generated for each unique set of values of these grouping variables.

This procedure serves as a summary reporting tool and is often used to analyze survey data. It calculates most of the popular contingency-table statistics and tests such as chi-square, Fisher’s exact, and McNemar’s tests, as well as the Cochran-Armitage test for trend in proportions and the Kappa and weighted Kappa tests for inter-rater agreement.

This procedure also produces a broad set of association and correlation statistics for contingency tables: Phi, Cramer’s V, Pearson’s Contingency Coefficient, Tschuprow’s T, Lambda, Kendall’s Tau, and Gamma.

Types of Categorical Variables

Note that we will refer to two types of categorical variables: Table variables and Grouping variables. The values of the Table variables are used to define the rows and columns of a single contingency table. Two Table variables are used for each table, one variable defining the rows of the table and the other defining the columns. Grouping variables are used to split a data into subgroups. A separate table is generated for each unique set of values of the Grouping variables.

Note that if you only want to use one Table variable, you should use the Frequency Table procedure.
Technical Details

For the technical details that follow, we assume a contingency table of counts with \( R \) rows and \( C \) columns as in the table below. Let \( O_{ij} \) be the observed count for the \( i^{th} \) row (\( i = 1 \) to \( R \)) and \( j^{th} \) column (\( j = 1 \) to \( C \)). Let the row

<table>
<thead>
<tr>
<th>Row Variable</th>
<th>Column 1</th>
<th>( \cdots )</th>
<th>Column ( j )</th>
<th>( \cdots )</th>
<th>Column ( C )</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row 1</td>
<td>( O_{11} )</td>
<td>( \cdots )</td>
<td>( O_{1j} )</td>
<td>( \cdots )</td>
<td>( O_{1C} )</td>
<td>( n_1 )</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>Row ( i )</td>
<td>( O_{i1} )</td>
<td>( \cdots )</td>
<td>( O_{ij} )</td>
<td>( \cdots )</td>
<td>( O_{iC} )</td>
<td>( n_i )</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>Row ( R )</td>
<td>( O_{R1} )</td>
<td>( \cdots )</td>
<td>( O_{Rj} )</td>
<td>( \cdots )</td>
<td>( O_{RC} )</td>
<td>( n_R )</td>
</tr>
<tr>
<td>Total</td>
<td>( n_1 )</td>
<td>( \cdots )</td>
<td>( n_j )</td>
<td>( \cdots )</td>
<td>( n_C )</td>
<td>1</td>
</tr>
</tbody>
</table>

and column marginal totals be designated as \( n_i \) and \( n_j \), respectively, where

\[
\begin{align*}
n_i &= \sum_j O_{ij} \\
n_j &= \sum_i O_{ij}
\end{align*}
\]

Let the total number of counts in the table be \( N \), where

\[
\begin{align*}
N &= \sum_{i,j} O_{ij} \\
&= \sum_i n_i \\
&= \sum_j n_j
\end{align*}
\]

The table of associated proportions can then be written as

<table>
<thead>
<tr>
<th>Row Variable</th>
<th>Column 1</th>
<th>( \cdots )</th>
<th>Column ( j )</th>
<th>( \cdots )</th>
<th>Column ( C )</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Row 1</td>
<td>( p_{11} )</td>
<td>( \cdots )</td>
<td>( p_{1j} )</td>
<td>( \cdots )</td>
<td>( p_{1C} )</td>
<td>( p_1 )</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>Row ( i )</td>
<td>( p_{i1} )</td>
<td>( \cdots )</td>
<td>( p_{ij} )</td>
<td>( \cdots )</td>
<td>( p_{iC} )</td>
<td>( p_i )</td>
</tr>
<tr>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
<td>\vdots</td>
</tr>
<tr>
<td>Row ( R )</td>
<td>( p_{R1} )</td>
<td>( \cdots )</td>
<td>( p_{Rj} )</td>
<td>( \cdots )</td>
<td>( p_{RC} )</td>
<td>( p_R )</td>
</tr>
<tr>
<td>Total</td>
<td>( p_1 )</td>
<td>( \cdots )</td>
<td>( p_j )</td>
<td>( \cdots )</td>
<td>( p_C )</td>
<td>1</td>
</tr>
</tbody>
</table>
Contingency Tables (Crosstabs / Chi-Square Test)

where

\[ p_{ij} = \frac{O_{ij}}{N} \]
\[ p_i = \frac{n_i}{N} \]
\[ p_j = \frac{n_j}{N} \]

Finally, designate the expected counts and expected proportions for the \( i^{th} \) row and \( j^{th} \) column as \( E_{ij} \) and \( P_{eij} \), respectively, where

\[ E_{ij} = \frac{n_i n_j}{N} \]
\[ P_{eij} = \frac{E_{ij}}{N} = p_i p_j \]

In the sections that follow we will describe the various tests and statistics calculated by this procedure using the preceding notation.

**Table Statistics**

This section presents various statistics that can be output for each individual cell. These are useful for studying the independence between rows and columns. The statistics for the \( i^{th} \) row and \( j^{th} \) column are as follows.

**Count**
The cell count, \( O_{ij} \), is the number of observations for the cell.

**Row Percentage**
The percentage for column \( j \) within row \( i \), \( p_{ji|i} \), is calculated as

\[ p_{ji|i} = \frac{O_{ij}}{n_i} \]

**Column Percentage**
The percentage for row \( i \) within column \( j \), \( p_{i|j} \), is calculated as

\[ p_{i|j} = \frac{O_{ij}}{n_j} \]

**Table Percentage**
The overall percentage for the cell, \( p_{ij} \), is calculated as

\[ p_{ij} = \frac{O_{ij}}{N} \]

**Expected Counts Assuming Independence**
The expected count, \( E_{ij} \), is the count that would be obtained if the hypothesis of row-column independence were true. It is calculated as

\[ E_{ij} = \frac{n_i n_j}{N} \]
Chi-Square Contribution

The chi-square contribution, $CS_{ij}$, measures the amount that a cell contributes to the overall chi-square statistic for the table. This and the next two items let you determine which cells impact the chi-square statistic the most.

$$CS_{ij} = \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Deviation from Independence

The deviation statistic, $D_{ij}$, measures how much the observed count differs from the expected count.

$$D_{ij} = O_{ij} - E_{ij}$$

Std. Residual

The standardized residual, $SR_{ij}$, is equal to the deviation divided by the square root of the expected value:

$$SR_{ij} = \frac{O_{ij} - E_{ij}}{\sqrt{E_{ij}}}$$

Tests for Row-Column Independence

Pearson’s Chi-Square Test

Pearson’s chi-square statistic is used to test independence between the row and column variables. Independence means that knowing the value of the row variable does not change the probabilities of the column variable (and vice versa). Another way of looking at independence is to say that the row percentages (or column percentages) remain constant from row to row (or column to column).

This test requires large sample sizes to be accurate. An often quoted rule of thumb regarding sample size is that none of the expected cell values should be less than five. Although some users ignore the sample size requirement, you should also be very skeptical of the test if you have cells in your table with zero counts. For $2 \times 2$ tables, consider using Yates’ Continuity Correction or Fisher’s Exact Test for small samples.

Pearson’s chi-square test statistic follows an asymptotic chi-square distribution with $(R - 1)(C - 1)$ degrees of freedom when the row and column variables are independent. It is calculated as

$$\chi^2_P = \sum_i \sum_j \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

Yates' Continuity Corrected Chi-Square Test ($2 \times 2$ Tables)

Yates’ Continuity Corrected Chi-Square Test (or just Yates’ Continuity Correction) is similar to Pearson's chi-square test, but is adjusted for the continuity of the chi-square distribution. This test is particularly useful when you have small sample sizes. This test is only calculated for $2 \times 2$ tables.

Yates’ continuity corrected test statistic follows an asymptotic chi-square distribution with $(R - 1)(C - 1)$ degrees of freedom when the row and column variables are independent. It is calculated as

$$\chi^2_Y = \sum_i \sum_j \left( \max(0, |O_{ij} - E_{ij}| - 0.5) \right)^2 \frac{1}{E_{ij}}$$
Contingency Tables (Crosstabs / Chi-Square Test)

Likelihood Ratio Test
This test makes use of the fact that under the null hypothesis of independence, the likelihood ratio statistic follows an asymptotic chi-square distribution.

The likelihood ratio test statistic follows an asymptotic chi-square distribution with \((R - 1)(C - 1)\) degrees of freedom when the row and column variables are independent. It is calculated as

\[
\chi^2_{LR} = 2 \sum_i \sum_j O_{ij} \ln \left( \frac{O_{ij}}{E_{ij}} \right).
\]

Fisher’s Exact Test (2 × 2 Tables)
This test was designed to test the hypothesis that the two column percentages in a 2 × 2 table are equal. It is especially useful when sample sizes are small (even zero in some cells) and the chi-square test is not appropriate.

Using the hypergeometric distribution with fixed row and column totals, this test computes probabilities of all possible tables with the observed row and column totals. This test is often used when sample sizes are small, but it is appropriate for all sample sizes because Fisher’s exact test does not depend on any large-sample asymptotic distribution assumptions. This test is only calculated for 2 × 2 tables.

If we assume that \(P_H\) is the hypergeometric probability of any table with the observed row and column marginal totals, then Fisher’s Exact Test probabilities are calculated by summing over defined sets of tables depending on the hypothesis being tested (one-sided or two-sided).

Define the difference between conditional column proportions for row 1 from the observed table as \(D_O\), with

\[D_O = p_{1|1} - p_{1|2}\]

and the difference between conditional column proportions for row 1 from other possible tables with the observed row and column marginal totals as \(D\), with

\[D = p_{1|1} - p_{1|2}\]

The two-sided Fisher’s Exact Test P-value is calculated as

\[P - Value_{Two-Sided} = \sum_{\text{Tables where } |D| \geq |D_O|} P_H\]

The lower one-sided Fisher’s Exact Test P-value is calculated as

\[P - Value_{Lower} = \sum_{\text{Tables where } D \leq D_O} P_H\]

The upper one-sided Fisher’s Exact Test P-value is calculated as

\[P - Value_{Upper} = \sum_{\text{Tables where } D \geq D_O} P_H\]
Tests for Trend in Proportions (2 × k Tables)

When one variable is ordinal (e.g. “Low, Medium, High” or “1, 2, 3, 4, 5”) and the other has exactly two levels (e.g. “success”, “failure”), you can test the hypothesis that there is a linear trend in the proportion of successes (i.e. that the true proportion of successes increases (or decreases) across the levels of the ordinal variable). Three tests for linear trend in proportions are available in NCSS: the Cochran-Armitage Test, the Cochran-Armitage Test with Continuity Correction, and the Armitage Rank Correlation Test. Of these, the Cochran-Armitage Test is the most widely used.

Cochran-Armitage Test

The Cochran-Armitage test is described in Cochran (1954) and Armitage (1955). Though the formulas that follow appear different from those presented in the articles, the results are equivalent.

Suppose we have \( k \) independent binomial variates, \( y_i \), with response probabilities, \( p_i \), based on samples of size \( n_i \) at covariate (or dose) levels, \( x_i \), for \( i = 1, 2, ..., k \), where \( x_1 < x_2 < ... < x_k \). The scores \( x_i \), come from the row (or column) names of the ordinal variable. When the names are numeric (e.g. “1 2 3 4 etc.”) then the actual numeric values are used for the scores, allowing the user to input unequally spaced score values. When the names are not numeric, even though they may represent an ordinal scale (e.g. “Low, Medium, High”), then the scores are assigned automatically as evenly spaced integers from 1 to \( k \).

Define the following:

\[
N = \sum_{i=1}^{k} n_i \\
\bar{p} = \frac{1}{N} \sum_{i=1}^{k} y_i \\
\bar{q} = 1 - \bar{p} \\
\bar{x} = \frac{1}{N} \sum_{i=1}^{k} n_i x_i
\]

If we assume that the probability of response follows a linear trend on the logistic scale, then

\[
p_i = \frac{\exp(\alpha + \beta x_i)}{1 + \exp(\alpha + \beta x_i)}.
\]

The Cochran-Armitage test can be used to test the following hypotheses:

- One-Sided (Increasing Trend) \( H_0 : p_1 = p_2 = ... = p_k \) vs. \( H_1 : p_1 < p_2 < ... < p_k \)
- One-Sided (Decreasing Trend) \( H_0 : p_1 = p_2 = ... = p_k \) vs. \( H_1 : p_1 > p_2 > ... > p_k \)
- Two-Sided \( H_0 : p_1 = p_2 = ... = p_k \) vs. \( H_1 : p_1 < p_2 < ... < p_k \) or \( p_1 > p_2 > ... > p_k \)

Nam (1987) presents the following asymptotic test statistic for detecting a linear trend in proportions

\[
z = \frac{\sum_{i=1}^{k} y_i (x_i - \bar{x})}{\sqrt{pq \sum_{i=1}^{k} n_i (x_i - \bar{x})^2}}.
\]
A one-sided test rejects $H_0$ in favor of an increasing trend if $z \geq z_{1-\alpha}$, where $z_{1-\alpha}$ is the value that leaves $1 - \alpha$ in the upper tail of the standard normal distribution. A one-sided test rejects $H_0$ in favor of a decreasing trend if $z \leq z_{\alpha}$, where $z_{\alpha}$ is the value that leaves $\alpha$ in the lower tail of the standard normal distribution. A two-sided test rejects $H_0$ in favor of either an increasing or decreasing trend if $|z| \geq z_{1-\alpha/2}$.

### Cochran-Armitage Test with Continuity Correction

The Cochran-Armitage test with continuity correction is nearly the same as the uncorrected Cochran-Armitage test described earlier. In the continuity corrected test, a small continuity correction factor, $\Delta/2$, is added or subtracted from the numerator, depending on the direction of the test. If the scores, $x_i$, are equally-spaced then

$$\Delta = x_{i+1} - x_i \text{ for all } i < k$$

or the interval between adjacent scores. **NCSS** computes $\Delta$ for unequally-spaced scores as

$$\Delta = \frac{1}{k-1} \sum_{i=1}^{k-1} (x_{i+1} - x_i).$$

For the case of unequally-spaced covariates, Nam (1987) states, “For unequally spaced doses, no constant correction is adequate for all outcomes.” Therefore, we caution against the use of the continuity-corrected test statistic in the case of unequally-spaced covariates.

Using the same notation as that described for the Cochran-Armitage test, Nam (1987) presents the following continuity corrected asymptotic test statistic for detecting an increasing linear trend in proportions

$$z_{c.c.U} = \frac{\sum_{i=1}^{k} y_i (x_i - \bar{x}) - \frac{\Delta}{2}}{\sqrt{pq \left[ \sum_{i=1}^{k} n_i (x_i - \bar{x})^2 \right]^{1/2}}}.$$  

A one-sided test rejects $H_0$ in favor of an increasing trend if $z_{c.c.U} \geq z_{1-\alpha}$, where $z_{1-\alpha}$ is the value that leaves $1 - \alpha$ in the upper tail of the standard normal distribution.

The continuity-corrected test statistic for a decreasing trend is the same as that for an increasing trend, except that $\Delta/2$ is added in the numerator instead of subtracted

$$z_{c.c.L} = \frac{\sum_{i=1}^{k} y_i (x_i - \bar{x}) + \frac{\Delta}{2}}{\sqrt{pq \left[ \sum_{i=1}^{k} n_i (x_i - \bar{x})^2 \right]^{1/2}}}.$$  

A one-sided test rejects $H_0$ in favor of a decreasing trend if $z_{c.c.L} \leq z_{\alpha}$, where $z_{\alpha}$ is the value that leaves $\alpha$ in the lower tail of the standard normal distribution.

A two-sided test rejects $H_0$ in favor of either an increasing or decreasing trend if $z_{c.c.U} \geq z_{1-\alpha/2}$ or if $z_{c.c.L} \leq z_{\alpha/2}$. 

---

**NCSS Statistical Software**

Contingency Tables (Crosstabs / Chi-Square Test)

© NCSS, LLC. All Rights Reserved.
Armitage Rank Correlation Test
The Armitage Rank Correlation test is described in section 4 of Armitage (1955) (the test is referred to as Kendall’s Rank Correlation Test in the paper). The statistic, \( S \), is standardized to a normal z-value by dividing by the estimated standard error of \( S \) (which we label \( \sqrt{V} \) below). This z-value can be tested using the standard-normal distribution.

When there are two columns and we want to test for the presence of a trend in proportions down the rows, the calculations for this test are as follows:

\[
z = \frac{S}{\sqrt{V}}
\]

where

\[
S = A - B
\]

\[
V = \frac{n_1 n_2 \left( N^3 - \sum_{i=1}^{R} n_i^3 \right)}{3N(N - 1)}
\]

with

\[
A = \sum_{j=1}^{R-1} O_{j2} \sum_{i=j+1}^{R} O_{i1}
\]

\[
B = \sum_{j=1}^{R-1} O_{j1} \sum_{i=j+1}^{R} O_{i2}
\]

A one-sided test rejects \( H_0 \) in favor of an increasing trend if \( z \geq z_{1-\alpha} \), where \( z_{1-\alpha} \) is the value that leaves \( 1 - \alpha \) in the upper tail of the standard normal distribution. A one-sided test rejects \( H_0 \) in favor of a decreasing trend if \( z \leq z_{\alpha} \), where \( z_{\alpha} \) is the value that leaves \( \alpha \) in the lower tail of the standard normal distribution. A two-sided test rejects \( H_0 \) in favor of either an increasing or decreasing trend if \( |z| \geq z_{1-\alpha/2} \).

McNemar Test (\( k \times k \) Tables)
The McNemar test was first used to compare two proportions that are based on matched samples. Matched samples occur when individuals (or matched pairs) are given two different treatments, asked two different questions, or measured in the same way at two different points in time. Match pairs can be obtained by matching individuals on several other variables, by selecting two people from the same family (especially twins), or by dividing a piece of material in half.

The McNemar test has been extended so that the measured variable can have more than two possible outcomes. It is then called the McNemar test of symmetry. It tests for symmetry around the diagonal of the table. The diagonal elements of the table are ignored. The test is computed for square \( k \times k \) tables only.

The McNemar test statistic follows an asymptotic chi-square distribution with \( R(R - 1)/2 \) degrees of freedom. It is calculated as

\[
\chi^2_M = \frac{1}{2} \sum_{i} \sum_{j} \left( \frac{O_{ij} - O_{ji}}{O_{ij} + O_{ji}} \right)^2
\]
Kappa and Weighted Kappa Tests for Inter-Rater Agreement ($k \times k$ Tables)

Kappa is a measure of association (correlation or reliability) between two measurements on the same individual when the measurements are categorical. It tests if the counts along the diagonal are significantly large. Because Kappa is used when the same variable is measured twice, it is only appropriate for square tables where the row and column categories are the same. Kappa is often used to study the agreement of two raters such as judges or doctors, where each rater classifies each individual into one of $k$ categories.

**Rules-of-thumb for kappa:** values less than 0.40 indicate low association; values between 0.40 and 0.75 indicate medium association; and values greater than 0.75 indicate high association between the two raters.

Kappa and weighted kappa are only output for square $k \times k$ tables with identical row and column labels. If your data have entire rows or columns missing because they were never reported by the raters, you must add a row or column of zeros to make the table square (see Example 6).

The results of this section are based on Fleiss, Levin, and Paik (2003). The kappa procedure also outputs the Maximum Kappa and Maximum-Adjusted Kappa statistics.

**Kappa Estimation**

Define the overall proportion of observed agreement, $p_o$, as

$$p_o = \sum_i p_{ii}$$

and the overall chance-expected proportion of agreement, $p_e$, as

$$p_e = \sum_i p_i \cdot p_i$$

Kappa is calculated as from $p_o$ and $p_e$ as

$$\hat{\kappa} = \frac{p_o - p_e}{1 - p_e}$$

with asymptotic standard error

$$SE_{\kappa} = \frac{\sqrt{A + B - C}}{(1 - p_e)\sqrt{N}}$$

where

$$A = \sum_i p_{ii}[1 - (p_i + p_i)(1 - \hat{\kappa})]^2$$

$$B = (1 - \hat{\kappa})^2 \sum_i \sum_{j \neq i} p_{ij}(p_i + p_j)^2$$

$$C = [\hat{\kappa} - p_e(1 - \hat{\kappa})]^2$$

An approximate $100(1 - \alpha)\%$ confidence interval for $\kappa$ is

$$\hat{\kappa} - z_{\alpha/2}SE_{\kappa} \leq \kappa \leq \hat{\kappa} + z_{\alpha/2}SE_{\kappa}$$
Contingency Tables (Crosstabs / Chi-Square Test)

Kappa Hypothesis Test
To test the null hypothesis that $\kappa = 0$, the standard error of kappa under the null hypothesis is calculated as

$$SE_{\kappa_0} = \frac{1}{(1 - p_e)\sqrt{N}} \sqrt{p_e + p^2_e - \sum_i p_i p_i (p_i + p_i)}$$

and the kappa test statistic, $z_\kappa$, with asymptotic standard normal distribution is

$$z_\kappa = \frac{\hat{\kappa}}{SE_{\kappa_0}}$$

Weighted Kappa Estimation
Weighted kappa should only be used when the rater categories are ordered (e.g. "Low", "Medium", "High" or 1, 2, 3, 4). The procedure applies weights to quantify relative distances between categories. These weights can be calculated as either linear or quadratic in NCSS.

The linear weights are calculated as

$$w_{ij} = 1 - \frac{|i - j|}{R - 1}$$

with $R = C$. For a $4 \times 4$ table, the linear weight matrix would be

1.00 0.67 0.33 0.00  
0.67 1.00 0.67 0.33  
0.33 0.67 1.00 0.67  
0.00 0.33 0.67 1.00

The quadratic weights are calculated as

$$w_{ij} = 1 - \frac{(i - j)^2}{(R - 1)^2}$$

again with $R = C$. For a $4 \times 4$ table, the quadratic weight matrix would be

1.00 0.89 0.56 0.00  
0.89 1.00 0.89 0.56  
0.56 0.89 1.00 0.89  
0.00 0.56 0.89 1.00

Note that in both cases the weights for cells on the diagonal are equal to 1 and weights off the diagonal are between 0 and 1. Weighted kappa is the same as simple kappa when using a weight matrix with all diagonal weight elements equal to 1 and all off-diagonal weight elements equal to 0.

Using the cell weights, we can calculate the observed weighted proportion of agreement as

$$p_{ow} = \sum_i \sum_j w_{ij}p_{ij}$$
and the overall chance-expected weighted proportion of agreement, \( p_e \), as

\[
p_{ew} = \sum_i \sum_j w_{ij} p_i p_j
\]

Further define

\[
\bar{w}_i = \sum_j w_{ij} p_j
\]

\[
\bar{w}_j = \sum_i w_{ij} p_i
\]

Weighted kappa is calculated as

\[
\hat{k}_w = \frac{p_{ow} - p_{ew}}{1 - p_{ew}}
\]

with asymptotic standard error

\[
SE_{\hat{k}_w} = \frac{\sqrt{A - B}}{(1 - p_{ew})\sqrt{N}}
\]

where

\[
A = \sum_i \sum_j p_{ij} \left[ w_{ij} - (\bar{w}_i + \bar{w}_j)(1 - \hat{k}_w) \right]^2
\]

\[
B = \left[ \hat{k}_w - p_{ew}(1 - \hat{k}_w) \right]^2
\]

An approximate 100(1 - \( \alpha \))% confidence interval for \( \kappa_w \) is

\[
\hat{k}_w - z_{\alpha/2} SE_{\hat{k}_w} \leq \kappa_w \leq \hat{k}_w + z_{\alpha/2} SE_{\hat{k}_w}
\]

**Weighted Kappa Hypothesis Test**

To test the null hypothesis that \( \kappa_w = 0 \), the standard error of weighted kappa under the null hypothesis is calculated as

\[
SE_{\hat{k}_w_0} = \frac{1}{(1 - p_{ew})\sqrt{N}} \left( \sum_i \sum_j p_{ij} \left[ w_{ij} - (\bar{w}_i + \bar{w}_j) \right]^2 - p_{ew}^2 \right)
\]

and the weighted kappa test statistic, \( z_{\kappa_w} \), with asymptotic standard normal distribution is

\[
z_{\kappa_w} = \frac{\hat{k}_w}{SE_{\hat{k}_w_0}}
\]
**Maximum-Adjusted Kappa**

If we define the overall chance-expected proportion of agreement, $p_e$, as before with

$$p_e = \sum_i p_i \cdot p_i$$

and

$$p_{max} = \sum_i \min(p_i, p_i)$$

then the maximum kappa for a table with the observed marginal totals, $\hat{k}_{max}$, can be calculated as

$$\hat{k}_{max} = \frac{p_{max} - p_e}{1 - p_e}$$

The maximum-adjusted kappa statistic, $\hat{k}_{max-adj}$, is calculated as

$$\hat{k}_{max-adj} = \frac{\hat{k}}{\hat{k}_{max}}$$

---

**Association and Correlation Statistics**

**Phi**

A measure of association independent of the sample size. *Phi* ranges between 0 (no relationship) and 1 (perfect relationship). *Phi* was designed for $2 \times 2$ tables only. For larger tables, it has no upper limit and Cramer’s V should be used instead. The formula is

$$\phi = \sqrt{\frac{\chi^2}{N}}$$

**Cramer’s V**

A measure of association independent of sample size. This statistic is a modification of the Phi statistic so that it is appropriate for larger than $2 \times 2$ tables. V ranges between 0 (no relationship) and 1 (perfect relationship).

$$V = \sqrt{\frac{\phi^2}{\min(R, C)}}$$

**Pearson’s Contingency Coefficient**

A measure of association independent of sample size. It ranges between 0 (no relationship) and 1 (perfect relationship). For any particular table, the maximum possible depends on the size of the table (a $2 \times 2$ table has a maximum of 0.707), so it should only be used to compare tables with the same dimensions. The formula is

$$C = \sqrt{\frac{\chi^2}{\frac{\chi^2}{N} + N}}$$
**Tschuprow’s T**

A measure of association independent of sample size. This statistic is a modification of the Phi statistic so that it is appropriate for larger than 2 × 2 tables. T ranges between 0 (no relationship) and 1 (perfect relationship), but 1 is only attainable for square tables. The formula is

\[ T = \frac{\phi^2}{\sqrt{(R - 1)(C - 1)}} \]

**Lambda A - Rows dependent**

This is a measure of association for cross tabulations of nominal-level variables. It measures the percentage improvement in predictability of the dependent variable (row variable or column variable), given the value of the other variable (column variable or row variable). The formula is

\[ \lambda_a = \frac{\sum_i \max(O_{ij}) - \max(n_i)}{N - \max(n_i)} \]

**Lambda B - Columns dependent**

See Lambda A above. The formula is

\[ \lambda_a = \frac{\sum_j \max(O_{ij}) - \max(n_j)}{N - \max(n_j)} \]

**Symmetric Lambda**

This is a weighted average of the Lambda A and Lambda B above. The formula is

\[ \lambda = \frac{\sum_i \max(O_{ij}) + \sum_j \max(O_{ij}) - \max(n_i) - \max(n_j)}{2N - \max(n_i) - \max(n_j)} \]

**Kendall’s tau-B**

This is a measure of correlation between two ordinal-level (rankable) variables. It is most appropriate for square tables. To compute this statistic, you first compute two values, \( P \) and \( Q \), which represent the number of concordant and discordant pairs, respectively. The formula is

\[ \tau_b = \frac{P - Q}{N(N - 1)/2} \]

**Kendall’s tau-B (with correction for ties)**

This is the same as the above, except a correction is made for the case when ties are found in the data.

**Kendall’s tau-C**

This is used in the case where the number of rows does not match the number of columns. The formula is

\[ \tau_c = \frac{P - Q}{N^2(\min(R, C) - 1)/2\min(R, C)} \]
**Gamma**

This is another measure based on concordant \((P)\) and discordant \((Q)\) pairs. The formula is

\[
\gamma = \frac{P - Q}{P + Q}
\]

**Data Structure**

You may use either summarized or non-summarized data for this procedure. Typically, you will use data columns of categorical data. If you want to perform crosstab analysis on numeric data, the data must be grouped into categories before a table can be created. This is best accomplished by using an *If-Then* or *Recode* transformation. You can also use this procedure’s facility to categorize numeric data by checking “Create Other Row/Column Variables for Numeric Data” on the Variables tab.

The following are two example datasets that illustrate the type of data that can be analyzed using this procedure. The datasets are provided with the software. The first dataset “CrossTabs1” contains fictitious responses to a survey of 100 people in which respondents were asked about their weekly sugar intake and exercise. The data are in raw form. The second hypothetical dataset “McNemar” contains summarized responses from 23 individuals who were asked about their desire to purchase a certain home-improvement product before and after a sales demonstration.

**CrossTabs1 dataset (subset)**

<table>
<thead>
<tr>
<th>Sugar</th>
<th>Exercise</th>
</tr>
</thead>
<tbody>
<tr>
<td>High</td>
<td>Infrequent</td>
</tr>
<tr>
<td>Low</td>
<td>Frequent</td>
</tr>
<tr>
<td>High</td>
<td>Infrequent</td>
</tr>
<tr>
<td>High</td>
<td>Frequent</td>
</tr>
<tr>
<td>High</td>
<td>Infrequent</td>
</tr>
<tr>
<td>High</td>
<td>Frequent</td>
</tr>
<tr>
<td>High</td>
<td>Infrequent</td>
</tr>
<tr>
<td>High</td>
<td>Infrequent</td>
</tr>
<tr>
<td>High</td>
<td>Infrequent</td>
</tr>
<tr>
<td>Low</td>
<td>Frequent</td>
</tr>
</tbody>
</table>

**McNemar dataset**

<table>
<thead>
<tr>
<th>Before</th>
<th>After</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>Yes</td>
<td>10</td>
</tr>
<tr>
<td>No</td>
<td>No</td>
<td>6</td>
</tr>
<tr>
<td>Yes</td>
<td>Yes</td>
<td>4</td>
</tr>
<tr>
<td>Yes</td>
<td>No</td>
<td>3</td>
</tr>
</tbody>
</table>

The data below are a subset of the *Real Estate Sales* database provided with the software. This (computer-simulated) data gives information including the selling price, the number of bedrooms, the total square footage (finished and unfinished), and the size of the lots for 150 residential properties sold during the last four months in two states. Only the first 6 of 150 observations are displayed here. The variables “Price”, “TotalSqft”, and “LotSize” would need to be categorized before they could be displayed in a table.
Resale dataset (subset)

<table>
<thead>
<tr>
<th>State</th>
<th>Price</th>
<th>Bedrooms</th>
<th>TotalSqft</th>
<th>LotSize</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nev</td>
<td>260000</td>
<td>2</td>
<td>2042</td>
<td>10173</td>
</tr>
<tr>
<td>Nev</td>
<td>66900</td>
<td>3</td>
<td>1392</td>
<td>13069</td>
</tr>
<tr>
<td>Vir</td>
<td>127900</td>
<td>2</td>
<td>1792</td>
<td>7065</td>
</tr>
<tr>
<td>Nev</td>
<td>181900</td>
<td>3</td>
<td>2645</td>
<td>8484</td>
</tr>
<tr>
<td>Nev</td>
<td>262100</td>
<td>2</td>
<td>2613</td>
<td>8355</td>
</tr>
</tbody>
</table>

Missing Values

Missing values may be ignored or included in the table’s counts, percentages, statistics, and tests. This is controlled on the procedure’s Missing tab.

Summary Table Data Input

NCSS also allows you to input the contingency table data directly into the procedure without using the Data Window. To do this, select “Summary Table” for Type of Data Input.

Procedure Options

This section describes the options available in this procedure.

Variables Tab

Data Input

Type of Data Input

Select the source of the data to analyze. The choices are

- Database
  
  Data will be read from the Data Table on the Data Window using selected variables. Specify at least one Column variable and at least one Row variable to be used to create the contingency table. The unique values of these two variables will form the columns and rows of the table. If more than one variable is specified in either section, a separate table will be generated for each combination of variables.

Two types of variables may be specified to be used in rows and columns: Categorical Variables and Numeric Variables. Usually, you will enter categorical variables.

1. Categorical Variables
   
   Categorical or discrete variables may include text values (e.g. “Male, Female”) or index numbers (e.g. “1, 4, 7, 15” to represent 4 states). The numbers or categories may be ordinal (e.g. “Low, Medium, High” or “1, 2, 3, 4, 5” as in a Likert scale). In fact, some table statistics like the Armitage test for trend in proportions and weighted kappa assume that one or both of the table variables are ordinal.

2. Numeric Variables
   
   Since contingency tables display categorical data, all numeric variables with continuous data must be grouped into categories by the procedure using user-specified rules before the table is created. To enter this type of variable, you must first put a check by “Create Other Row (Column) Variables from Numeric Data” to display the numeric variable entry box. You can specify the groups by entering the numeric
boundaries directly (e.g. “Under 21, 21 to 55, and Over 55”) or by entering the number of intervals to create, the minimum boundary, and/or the interval width. You can only specify a single grouping rule that will be used for all numeric variables in either a row or column. If you have more than one numeric variable, then you should group the data directly on the dataset using a Recode Transformation and enter the resulting variables as a categorical row or column variables.

- **Summary Table**
  Summarized count data will be read directly from a spreadsheet on the window. Enter the row and column variable names, category labels, and cell counts directly into this summary table.

---

**Categorical Table Variables (Database Input)**

### Row (Column) Variable(s)
Specify one or more categorical variables for use in table rows (columns). Each unique value in each variable will result in a separate row in the table. The data values themselves may be text (e.g. “Low, Med, High”) or numeric (e.g. “1, 2, 3”), but the data as a whole should be categorical. If more than one variable is entered, a separate table will be created for each variable.

The data values in each variable will be sorted alpha-numerically before the table rows (columns) are created. If you want the values to be displayed in a different order, specify a custom value order for the data column(s) entered here using the Column Info Table on the Data Window.

### Create Other Row (Column) Variables from Numeric Data
Check this box to create tables with rows from numeric data. When checked, additional options will be displayed to specify how the numeric data will be classified into categorical variables.

If you choose to create row (column) variables from numeric data, you do not have to enter a categorical row (column) variable in the input box above (but you can). If both numeric and categorical row (column) variables are entered, a separate table and analysis will be calculated for each variable.

### Numeric Variable(s) to Categorize for Use in Table Rows (Columns)
Specify one or more variables that have only numeric values to be used in rows (columns) of the table. Numeric values from these variables will be combined into a set of categories using the categorization options that follow. If more than one variable is entered, a separate table will be created for each variable.

For example, suppose you want to tabulate a variable containing individual income values into four categories: “Below 10000”, “10000 to 40000”, “40000 to 80000”, and “Over 80000”. You could select the income variable here, set **Group Numeric Data into Categories Using** to “List of Interval Upper Limits” and set the **List** to “10000 40000 80000”.

### Group Numeric Data into Categories Using
Choose the method by which numeric data will be combined into categories for use in table rows or columns.

The choices are:

- **Number of Intervals, Minimum, and/or Width**
  This option allows you to specify the categories by entering any combination of the three parameters:

  - **Number of Intervals**
  - **Minimum**
  - **Width**

  All three are optional.

- **Number of Intervals**
  This is the number of intervals into which the values of the numeric variables are categorized. If not enough
intervals are specified to reach past the maximum data value, more will be added.

**Range**
Integer ≥ 2

**Minimum**
This value is used in conjunction with the Number of Intervals and Width values to construct a set of intervals into which the numeric variables are categorized. This is the minimum value of the first interval.

**Range**
This value must be less than the minimum data value.

**Width**
This value is used in conjunction with the Number of Intervals and Minimum values to construct a set of intervals into which the numeric variables are categorized. All intervals will have a width equal to this value.

A data value \( X \) is in this interval if

\[
\text{Lower Limit} < X \leq \text{Upper Limit}.
\]

- **List of Interval Upper Limits**
  This option allows you to specify the categories for the numeric variable by entering a list of interval boundaries directly, separated by blanks or commas. An interval of the form \( L1 < X \leq L2 \) is generated for each interval. The actual number of intervals is one more than the number of items specified here.

  For example, suppose you want to tabulate a variable containing individual income values into four categories: “Below 10000”, “10000 to 40000”, “40000 to 80000”, and “Over 80000”. You would set **List of Interval Upper Limits** to “10000 40000 80000”. Note that 10000 would be included in the “Below 10000” interval, but not the “10000 to 40000” interval. Also, 80000 would be included in the “40000 to 80000” interval, not the “Over 80000” interval.

**Frequency (Count) Variable (Database Input)**

**Frequency Variable**
Specify an optional frequency (count) variable. This data column contains integers that represent the number of observations (frequency) associated with each row of the dataset. If this option is left blank, each dataset row has a frequency of one. This variable lets you modify that frequency. This may be useful when your data are tabulated and you want to enter counts.

**Grouping (Break) Variables (Database Input)**

**Number of Grouping Variables**
Select the number of grouping (break) variables to include for the analysis. All reports and plots will be generated for each unique combination of the values of the grouping variables. You can select up to 8 grouping variables.

**Grouping Variable**
Select an optional categorical grouping (or break) variable. All tables, statistical reports, and plots will be generated for each unique value of this variable. If you specify more than one grouping variable, the tables, statistical reports, and plots will be generated for each unique combination of the values of the variables chosen.
Summary Table Data (Counts) (Summary Table Input)
Enter the row and column variable names, category labels, and cell counts directly into this summary table. Enter the variable names in the cells with a faint yellow background (A2 and B1). These are bolded automatically. Enter the category labels in the cells with a faint blue background (Column A, rows 3+ and Row 2, columns B+). Enter the individual cell counts in the cells with a white background.

Click the Reset Table button to reset the table and fill it with example values.

Missing Values Tab
This panel lets you specify up to five missing values (besides the default of blank). For example, “0”, “9”, or “NA” may be missing values in your dataset.

Missing Value Options
Missing Value Inclusion
Specify whether to include or exclude observations with missing values in the tables and/or reports.

Possible selections are:

• Delete All
  This option indicates that you want the missing values totally ignored.

• Include in Counts
  This option indicates that you want the number of missing values displayed, but you do not want them to influence any of the percentages.

• Include in All
  This option indicates that you want the missing values treated just like any other category. They will be included in all percentages and counts.

Label for Missing Values
Specify the label to be used to label missing values in the output.

Data Values to be Treated as “Missing”
Missing Value
Specify a value to be treated as a missing value by this procedure. This value is treated as a missing value in all active categorical variables. Up to 5 different missing values may be entered.

Reports Tab
This tab controls which tables and statistical reports are displayed in the output.

Select Reports
Data Summary Report
Check this option to display a report of the summarized data for each combination of row and column variables across all break variables.
Contingency Tables

Show Individual Tables
Check this option to display a separate table for each table statistic. After activating this option, you must specify which tables you would like to display.

The tables to choose from are:

- Counts
- Table Percentages
- Row Percentages
- Column Percentages
- Expected Counts Assuming Independence
- Chi-Square Contributions
- Deviations from Independence
- Standardized Residuals

Show Combined Table
Check this option to display a single table containing the selected statistics. After activating this option, you must specify which items you would like to display in the table.

The items to choose from are:

- Counts
- Table Percentages
- Row Percentages
- Column Percentages
- Expected Counts Assuming Independence
- Chi-Square Contributions
- Deviations from Independence
- Standardized Residuals

Table Statistics and Tests

Tests for Row-Column Independence
Check this option to output the “Tests for Row-Column Independence” report. These tests are used to test for independence between rows and columns of the table. Independence means that knowing the value of the row variable does not change the probabilities of the column variable (and vice versa). Another way of looking at independence is to say that the row percentages (or column percentages) remain constant from row to row (or column to column).

When this option is checked, you will have the option of choosing from 4 different independence tests:

- **Pearson’s Chi-Square Test**
  Check this option to output Pearson’s Chi-Square Test for row-column independence.
  
  This test requires large sample sizes to be accurate. An often quoted rule of thumb regarding sample size is that none of the expected cell values can be less than five. Although some users ignore the sample size requirement, you should be very skeptical of the test if you have cells in your table with zero counts. When these assumptions are violated, you should use Yates’ Continuity Corrected Test or Fisher’s Exact Test.

- **Yates’ Continuity Corrected Chi-Square Test [2 × 2 Tables]**
  Check this option to output Yates’ Continuity Corrected Chi-Square Test for row-column independence. This test is similar to Pearson's chi-square test, but is adjusted for the continuity of the chi-square distribution. This test is particularly useful when you have small sample sizes. This test is only calculated for $2 \times 2$ tables.
Contingency Tables (Crosstabs / Chi-Square Test)

- **Likelihood Ratio Test**
  Check this option to output the Likelihood Ratio Test for row-column independence. This test makes use of the fact that under the null hypothesis of independence, the likelihood ratio statistic follows an asymptotic chi-square distribution.

- **Fisher's Exact Test [2 × 2 Tables]**
  Check this option to output the Fisher's Exact Test for row-column independence. Using the hypergeometric distribution with fixed row and column totals, this test computes probabilities of all possible tables with the observed row and column totals. This test is often used when sample sizes are small, but it is appropriate for all sample sizes. This test is only calculated for 2 × 2 tables.

**Tests for Trend in Proportions [2 × k Tables]**
Check this option to output the various trend test reports. These tests are used to test for trend in proportions. These tests are only calculated for 2 × k tables. After selecting this option, you must select which trend tests to output. The options are

- **Cochran-Armitage Test**
  Check this option to output the Cochran-Armitage test for linear trend in proportions. The test may be used when you have exactly two rows or two columns in your table. This procedure tests the hypothesis that there is a linear trend in the proportion of successes. That is, the true proportion of successes increases (or decreases) as you move from row to row (or column to column). This test is only calculated for 2 × k tables.
  The Cochran-Armitage Test is the most widely-used test for trend in proportions.

- **Cochran-Armitage Test with Continuity Correction**
  Check this option to output the Continuity Corrected Cochran-Armitage test for linear trend in proportions. In this test, Z-values are adjusted by the factor Δ/2, where Δ is the average distance between scores. The test may be used when you have exactly two rows or two columns in your table. This procedure tests the hypothesis that there is a linear trend in the proportion of successes. That is, the true proportion of successes increases (or decreases) as you move from row to row (or column to column). This test is only calculated for 2 × k tables.

- **Armitage Rank Correlation Test**
  Check this option to output the Armitage rank correlation test for trend in proportions. The test may be used when you have exactly two rows or two columns in your table. This procedure tests the hypothesis that there is a trend in the proportion of successes. That is, the true proportion of successes increases (or decreases) as you move from row to row (or column to column). This test is only calculated for 2 × k tables.

**McNemar Test [k × k Tables]**
Check this option to output the McNemar Test. The McNemar test was first used to compare two proportions that are based on matched samples. Matched samples occur when individuals (or matched pairs) are given two different treatments, asked two different questions, or measured in the same way at two different points in time. Match pairs can be obtained by matching individuals on several other variables, by selecting two people from the same family (especially twins), or by dividing a piece of material in half.

The McNemar test has been extended so that the measured variable can have more than two possible outcomes. It is then called the McNemar test of symmetry. It tests for symmetry around the diagonal of the table. The diagonal elements of the table are ignored. This test is only calculated for square k × k tables.

**Kappa and Weighted Kappa Tests for Inter-Rater Agreement [k × k Tables]**
Check this option to output the Kappa Estimation and Hypothesis Tests reports. Kappa is a measure of association (correlation or reliability) between two measurements on the same individual when the measurements are categorical. It tests if the counts along the diagonal are significantly large. Because Kappa is used when the same
variable is measured twice, it is only appropriate for square tables where the row and columns have the same categories. Kappa is often used to study the agreement of two raters such as judges or doctors. Each rater classifies each individual into one of $k$ categories.

**Rules-of-thumb for kappa:** values less than 0.40 indicate low association; values between 0.40 and 0.75 indicate medium association; and values greater than 0.75 indicate high association between the two raters.

The items estimated and tested in the Kappa reports are

- **Kappa**
- **Weighted Kappa (With Linear and Quadratic Weights)**
- **Maximum Kappa**
- **Maximum-Adjusted Kappa**

This report is only output for square $k \times k$ tables with identical row and column categories.

**Weighted Kappa**

Weighted Kappa should only be used when the rater categories are ordered (e.g. “Low, Medium, High” or “1, 2, 3, 4”). The procedure applies weights to quantify relative distances between categories. These weights can be calculated as either linear or quadratic. Results from both are given in the report.

For $2 \times 2$ tables, Weighted Kappa is the same as simple Kappa.

**Confidence Level**

This confidence level is used for the kappa and weighted kappa confidence intervals. Typical confidence levels are 90%, 95%, and 99%, with 95% being the most common.

**Association and Correlation Statistics**

Check this option to output various categorical association and correlation statistics.

- **Phi**
  A measure of association independent of the sample size. Phi ranges between 0 (no relationship) and 1 (perfect relationship). Phi was designed for $2 \times 2$ tables only. For larger tables, it has no upper limit and Cramer’s V should be used instead.

- **Cramer’s V**
  A measure of association independent of sample size. This statistic is a modification of the Phi statistic so that it is appropriate for larger than $2 \times 2$ tables. Cramer’s V ranges between 0 (no relationship) and 1 (perfect relationship).

- **Pearson’s Contingency Coefficient**
  A measure of association independent of sample size. It ranges between 0 (no relationship) and 1 (perfect relationship). For any particular table, the maximum possible depends on the size of the table (a $2 \times 2$ table has a maximum of 0.707), so it should only be used to compare tables with the same dimensions.

- **Tschuprow’s T**
  A measure of association independent of sample size. This statistic is a modification of the Phi statistic so that it is appropriate for larger than $2 \times 2$ tables. T ranges between 0 (no relationship) and 1 (perfect relationship), but 1 is only attainable for square tables.

- **Lambda**
  This is a measure of association for cross tabulations of nominal-level variables. It measures the percentage improvement in predictability of the dependent variable (row variable or column variable), given the value of the other variable (column variable or row variable).
• **Kendall’s tau**
  This is a measure of correlation between two ordinal-level (rankable) variables. It is most appropriate for square tables.

• **Gamma**
  This is another measure based on concordant and discordant pairs. It is appropriate only when both row and column variables are ordinal.

---

**Alpha for Tests**

**Alpha**
Alpha is the significance level used in the hypothesis tests. A value of 0.05 is most commonly used, but 0.1, 0.025, 0.01, and other values are sometimes used. Typical values range from 0.001 to 0.20.

---

**Report Options Tab**

The following options control the format of the reports.

**Report Options**

**Variable Names**
Specify whether to use variable names, variable labels, or both to label output reports. In this discussion, the variables are the columns of the data table.

• **Names**
  Variable names are the column headings that appear on the data table. They may be modified by clicking the Column Info button on the Data window or by clicking the right mouse button while the mouse is pointing to the column heading.

• **Labels**
  This refers to the optional labels that may be specified for each column. Clicking the Column Info button on the Data window allows you to enter them.

• **Both**
  Both the variable names and labels are displayed.

**Comments**
1. Most reports are formatted to receive about 12 characters for variable names.
2. Variable Names cannot contain blanks or math symbols (like + - * / . ), but variable labels can.

**Value Labels**
Value Labels are used to make reports more legible by assigning meaningful labels to numbers and codes.
The options are

• **Data Values**
  All data are displayed in their original format, regardless of whether a value label has been set or not.
Value Labels
All values of variables that have a value label variable designated are converted to their corresponding value label when they are output. This does not modify their value during computation.

Both
Both data value and value label are displayed.

Example
A variable named GENDER (used as a grouping variable) contains 1’s and 2’s. By specifying a value label for GENDER, the report can display “Male” instead of 1 and “Female” instead of 2. This option specifies whether (and how) to use the value labels.

Table Formatting

Column Justification
Specify whether data columns in the contingency tables will be left or right justified.

Column Widths
Specify how the widths of columns in the contingency tables will be determined.

The options are

- Autosize to Minimum Widths
  Each data column is individually resized to the smallest width required to display the data in the column. This usually results in columns with different widths. This option produces the most compact table possible, displaying the most data per page.

- Autosize to Equal Minimum Width
  The smallest width of each data column is calculated and then all columns are resized to the width of the widest column. This results in the most compact table possible where all data columns have the same width. This is the default setting.

- Custom (User-Specified)
  Specify the widths (in inches) of the columns directly instead of having the software calculate them for you.

Custom Widths (Single Value or List)
Enter one or more values for the widths (in inches) of columns in the contingency tables. This option is only displayed if Column Widths is set to “Custom (User-Specified)”.

- Single Value
  If you enter a single value, that value will be used as the width for all data columns in the table.

- List of Values
  Enter a list of values separated by spaces corresponding to the widths of each column. The first value is used for the width of the first data column, the second for the width of the second data column, and so forth. Extra values will be ignored. If you enter fewer values than the number of columns, the last value in your list will be used for the remaining columns.

  Type the word “Autosize” for any column to cause the program to calculate it's width for you. For example, enter “1 Autosize 0.7” to make column 1 be 1 inch wide, column 2 be sized by the program, and column 3 be 0.7 inches wide.
Wrap Column Headings onto Two Lines
Check this option to make column headings wrap onto two lines. Use this option to condense your table when your data are spaced too far apart because of long column headings.

Decimal Places

Item Decimal Places
These decimal options allow the user to specify the number of decimal places for items in the output. Your choice here will not affect calculations; it will only affect the format of the output.

- Auto
  If one of the “Auto” options is selected, the ending zero digits are not shown. For example, if “Auto (0 to 7)” is chosen,
  
  0.0500 is displayed as 0.05
  1.314583689 is displayed as 1.314584
  
The output formatting system is not designed to accommodate “Auto (0 to 13)”, and if chosen, this will likely lead to lines that run on to a second line. This option is included, however, for the rare case when a very large number of decimals is needed.

Omit Percent Sign after Percentages
The program normally adds a percent sign, %, after each percentage. Checking this option will cause this percent sign to be omitted.

Plots Tab
The options on this panel allow you to select and control the appearance of the plots output by this procedure.

Select Plots

Show Plots
Check this option to display a separate plot for each table statistic. After activating this option, you must specify which plots you would like to display.

The plots to choose from are:

- Counts
- Table Percentages
- Row Percentages
- Column Percentages
- Expected Counts Assuming Independence
- Chi-Square Contributions
- Deviations from Independence
- Standardized Residuals

Click the plot format button to change the plot display settings.

Show Break as Title
Specify whether to display the values of the break variables as the second title line on the plots.
Example 1 – 2 × 2 Contingency Table and Statistics from Raw Categorical Data

The data for this example are found in the “CrossTabs1” dataset. This dataset contains fictitious survey data from 100 individuals asked about their sugar intake and exercise. Notice that we have entered custom value orders for the columns in the dataset so that the values will appear in the correct order. We use a 2 × 2 contingency table for this example so that all of the tests for row-column independence will be displayed.

You may follow along here by making the appropriate entries or load the completed template Example 1 by clicking on Open Example Template from the File menu of the Contingency Tables (Crosstabs / Chi-Square Test) window.

1. Open the CrossTabs1 dataset.
   - From the File menu of the NCSS Data window, select Open Example Data.
   - Click on the file CrossTabs1.NCSS.
   - Click Open.

2. Open the Contingency Tables (Crosstabs / Chi-Square Test) window.
   - On the menus, select Analysis, then Descriptive Statistics, then Contingency Tables (Crosstabs / Chi-Square Test). The procedure will be displayed.
   - On the menus, select File, then New Template. This will fill the procedure with the default template.

3. Specify the variables.
   - Select the Variables tab.
   - Double-click in the Row Variable(s) text box. This will bring up the variable selection window.
   - Select Sugar from the list of variables and then click OK. “Sugar” will appear in the Row Variable(s) box.
   - Double-click in the Column Variable(s) text box. This will bring up the variable selection window.
   - Select Exercise from the list of variables and then click OK. “Exercise” will appear in the Column Variable(s) box.

4. Specify the reports.
   - Select the Reports tab.
   - Leave Show Individual Tables and Counts checked.
   - Check Show Combined Table and leave the selected table items checked.
   - Leave Tests for Row-Column Independence and the selected tests checked.
   - Check Association and Correlation Statistics.

5. Specify the plots.
   - Select the Plots tab.
   - Check Show Plots and leave Counts checked.

6. Run the procedure.
   - From the Run menu, select Run Procedure. Alternatively, just click the green Run button.

The following reports and plots will be displayed in the Output window.
### Counts Table

<table>
<thead>
<tr>
<th>Sugar</th>
<th>Exercise</th>
<th>Infrequent</th>
<th>Frequent</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td></td>
<td>19</td>
<td>28</td>
<td>47</td>
</tr>
<tr>
<td>High</td>
<td></td>
<td>37</td>
<td>16</td>
<td>53</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>56</td>
<td>44</td>
<td>100</td>
</tr>
</tbody>
</table>

### Combined Table

<table>
<thead>
<tr>
<th>Sugar</th>
<th>Exercise</th>
<th>Infrequent</th>
<th>Frequent</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td>Count</td>
<td>19</td>
<td>28</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>% of Total</td>
<td>19.00%</td>
<td>28.00%</td>
<td>47.00%</td>
</tr>
<tr>
<td></td>
<td>% within Row</td>
<td>40.43%</td>
<td>59.57%</td>
<td>100.00%</td>
</tr>
<tr>
<td></td>
<td>% within Column</td>
<td>33.93%</td>
<td>63.64%</td>
<td>47.00%</td>
</tr>
<tr>
<td>High</td>
<td>Count</td>
<td>37</td>
<td>16</td>
<td>53</td>
</tr>
<tr>
<td></td>
<td>% of Total</td>
<td>37.00%</td>
<td>16.00%</td>
<td>53.00%</td>
</tr>
<tr>
<td></td>
<td>% within Row</td>
<td>69.81%</td>
<td>30.19%</td>
<td>100.00%</td>
</tr>
<tr>
<td></td>
<td>% within Column</td>
<td>66.07%</td>
<td>36.36%</td>
<td>53.00%</td>
</tr>
<tr>
<td>Total</td>
<td>Count</td>
<td>56</td>
<td>44</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>% of Total</td>
<td>56.00%</td>
<td>44.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td></td>
<td>% within Row</td>
<td>56.00%</td>
<td>44.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td></td>
<td>% within Column</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

### Tests for Row-Column Independence
(Sugar by Exercise)
H0: "Sugar" and "Exercise" are independent.
H1: "Sugar" and "Exercise" are associated (not independent).

<table>
<thead>
<tr>
<th>Test</th>
<th>Type</th>
<th>Chi-Square Value</th>
<th>DF</th>
<th>Prob at α = 0.05?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson's Chi-Square</td>
<td>2-Sided</td>
<td>8.7299</td>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>Yates' Cont. Correction</td>
<td>2-Sided</td>
<td>7.5780</td>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>2-Sided</td>
<td>8.8440</td>
<td>1</td>
<td>Yes</td>
</tr>
<tr>
<td>Fisher's Exact</td>
<td>2-Sided</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fisher's Exact (Lower)</td>
<td>1-Sided</td>
<td></td>
<td></td>
<td>Yes</td>
</tr>
<tr>
<td>Fisher's Exact (Upper)</td>
<td>1-Sided</td>
<td></td>
<td></td>
<td>No</td>
</tr>
</tbody>
</table>

### Association and Correlation Statistics
(Sugar by Exercise)

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phi</td>
<td>0.2955</td>
</tr>
<tr>
<td>Cramer's V</td>
<td>0.2955</td>
</tr>
<tr>
<td>Pearson's Contingency Coefficient</td>
<td>0.2834</td>
</tr>
<tr>
<td>Tschuprow's T</td>
<td>0.2955</td>
</tr>
<tr>
<td>Lambda A .. Rows dependent</td>
<td>0.2553</td>
</tr>
<tr>
<td>Lambda B .. Columns dependent</td>
<td>0.2045</td>
</tr>
<tr>
<td>Symmetric Lambda</td>
<td>0.2308</td>
</tr>
<tr>
<td>Kendall's tau-B</td>
<td>-0.1479</td>
</tr>
<tr>
<td>Kendall's tau-B (with correction for ties)</td>
<td>-0.2955</td>
</tr>
<tr>
<td>Kendall's tau-C</td>
<td>-0.2928</td>
</tr>
<tr>
<td>Gamma</td>
<td>-0.5463</td>
</tr>
</tbody>
</table>
This report presents the individual contingency table of counts, a combined table with counts and percentages, the results of the various row-column independence tests, and various association and correlation statistics. A plot of the counts is also displayed. The Pearson’s chi-square test results indicate that for these hypothetical data there is an association between a person’s sugar intake and exercise frequency (p-value = 0.00313).
Example 2 – 3 × 4 Contingency Table and Statistics from Summarized Categorical Data

The data for this example are found in the “CrossTabs2” dataset. Notice that we have entered custom value orders for the column labeled Region so that the values will appear in the correct order.

You may follow along here by making the appropriate entries or load the completed template Example 2a by clicking on Open Example Template from the File menu of the Contingency Tables (Crosstabs / Chi-Square Test) window.

1 **Open the CrossTabs2 dataset.**
   - From the File menu of the NCSS Data window, select Open Example Data.
   - Click on the file CrossTabs2.NCSS.
   - Click Open.

2 **Open the Contingency Tables (Crosstabs / Chi-Square Test) window.**
   - On the menus, select Analysis, then Descriptive Statistics, then Contingency Tables (Crosstabs / Chi-Square Test). The procedure will be displayed.
   - On the menus, select File, then New Template. This will fill the procedure with the default template.

3 **Specify the variables.**
   - Select the Variables tab.
   - Double-click in the Row Variable(s) text box. This will bring up the variable selection window.
   - Select Region from the list of variables and then click OK. “Region” will appear in the Row Variable(s) box.
   - Double-click in the Column Variable(s) text box. This will bring up the variable selection window.
   - Select Choice from the list of variables and then click OK. “Choice” will appear in the Column Variable(s) box.
   - Double-click in the Frequency Variable text box. This will bring up the variable selection window.
   - Select Count from the list of variables and then click OK. “Count” will appear in the Frequency Variable box.

4 **Specify the reports.**
   - Select the Reports tab.
   - Uncheck Show Individual Tables.
   - Check Show Combined Table and leave the selected table items checked.
   - Leave Tests for Row-Column Independence and the selected tests checked.

5 **Specify the plots.**
   - Select the Plots tab.
   - Check Show Plots and leave Counts checked.

6 **Run the procedure.**
   - From the Run menu, select Run Procedure. Alternatively, just click the green Run button.

The following reports and plots will be displayed in the Output window.
## Combined Table

<table>
<thead>
<tr>
<th>Region</th>
<th>A</th>
<th>B</th>
<th>C</th>
<th>D</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>East</td>
<td>56</td>
<td>14</td>
<td>12</td>
<td>22</td>
<td>104</td>
</tr>
<tr>
<td>% of Total</td>
<td>20.59%</td>
<td>5.15%</td>
<td>4.41%</td>
<td>8.09%</td>
<td>38.24%</td>
</tr>
<tr>
<td>% within Row</td>
<td>53.85%</td>
<td>13.46%</td>
<td>11.54%</td>
<td>21.15%</td>
<td>100.00%</td>
</tr>
<tr>
<td>% within Column</td>
<td>62.22%</td>
<td>32.56%</td>
<td>26.09%</td>
<td>23.66%</td>
<td>38.24%</td>
</tr>
<tr>
<td>West</td>
<td>15</td>
<td>27</td>
<td>8</td>
<td>24</td>
<td>74</td>
</tr>
<tr>
<td>% of Total</td>
<td>5.51%</td>
<td>9.93%</td>
<td>2.94%</td>
<td>8.82%</td>
<td>27.21%</td>
</tr>
<tr>
<td>% within Row</td>
<td>20.27%</td>
<td>36.49%</td>
<td>10.81%</td>
<td>32.43%</td>
<td>100.00%</td>
</tr>
<tr>
<td>% within Column</td>
<td>16.67%</td>
<td>62.79%</td>
<td>17.39%</td>
<td>25.81%</td>
<td>27.21%</td>
</tr>
<tr>
<td>South</td>
<td>19</td>
<td>2</td>
<td>26</td>
<td>47</td>
<td>94</td>
</tr>
<tr>
<td>% of Total</td>
<td>6.99%</td>
<td>0.74%</td>
<td>9.56%</td>
<td>17.28%</td>
<td>34.56%</td>
</tr>
<tr>
<td>% within Row</td>
<td>20.21%</td>
<td>2.13%</td>
<td>27.66%</td>
<td>50.00%</td>
<td>100.00%</td>
</tr>
<tr>
<td>% within Column</td>
<td>21.11%</td>
<td>4.65%</td>
<td>56.52%</td>
<td>50.54%</td>
<td>34.56%</td>
</tr>
<tr>
<td>Total</td>
<td>90</td>
<td>43</td>
<td>46</td>
<td>93</td>
<td>272</td>
</tr>
<tr>
<td>% of Total</td>
<td>33.09%</td>
<td>15.81%</td>
<td>16.91%</td>
<td>34.19%</td>
<td>100.00%</td>
</tr>
<tr>
<td>% within Row</td>
<td>33.09%</td>
<td>15.81%</td>
<td>16.91%</td>
<td>34.19%</td>
<td>100.00%</td>
</tr>
<tr>
<td>% within Column</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

## Tests for Row-Column Independence

**(Region by Choice)**

<table>
<thead>
<tr>
<th>Test</th>
<th>Chi-Square Value</th>
<th>DF</th>
<th>Prob Level</th>
<th>Reject H0 at α = 0.05?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson's Chi-Square†</td>
<td>75.3662</td>
<td>6</td>
<td>0.00000</td>
<td>Yes</td>
</tr>
<tr>
<td>Yates' Cont. Correction*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>75.0616</td>
<td>6</td>
<td>0.00000</td>
<td>Yes</td>
</tr>
<tr>
<td>Fisher's Exact*</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

† WARNING: At least one cell had a value less than 5.
* Test computed only for 2×2 tables.

## Plots Section

**(Region by Choice)**

This report presents the results from the summarized data. The Pearson’s chi-square test results indicate that the row and column variables are not independent \((p\text{-value} = 0.00000)\), but there is a sample size warning that should be considered. Note that Fisher’s Exact Test and Yates’ Continuity Correction are not reported because this is not a \(2 \times 2\) table.
An alternate way to enter this summarized data is to set **Type of Data Input** to **Summary Table**. You may follow along here by making the appropriate modifications to the current settings or load the completed template **Example 2b** by clicking on Open Example Template from the File menu of the Contingency Tables (Crosstabs / Chi-Square Test) window.

7 **Modify the Data Input Type.**
   - Select the **Variables tab**.
   - For **Type of Data Input**, select **Summary Table**.
   - Enter the variable names, labels, and counts into the spreadsheet.

8 **Run the procedure again.**
   - From the Run menu, select **Run Procedure**. Alternatively, just click the green Run button.

The output will be exactly the same as that displayed above.
Example 3 – 6 × 4 Contingency Table and Statistics from Raw Numeric Data

The real estate data for this example are found in the “Resale” dataset. We’ll use the crosstabs procedure to create a table with city as the row variable and price groups as the column variable. The software will summarize the continuous price variable for us using a list of price group boundaries. We’ll use the column labels and value labels in the dataset to make the data easier to interpret in the reports.

You may follow along here by making the appropriate entries or load the completed template Example 3 by clicking on Open Example Template from the File menu of the Contingency Tables (Crosstabs / Chi-Square Test) window.

1 Open the Resale dataset.
   - From the File menu of the NCSS Data window, select Open Example Data.
   - Click on the file Resale.NCSS.
   - Click Open.

2 Open the Contingency Tables (Crosstabs / Chi-Square Test) window.
   - On the menus, select Analysis, then Descriptive Statistics, then Contingency Tables (Crosstabs / Chi-Square Test). The procedure will be displayed.
   - On the menus, select File, then New Template. This will fill the procedure with the default template.

3 Specify the variables.
   - Select the Variables tab.
   - Double-click in the Row Variable(s) text box. This will bring up the variable selection window.
   - Select City from the list of variables and then click OK. “City” will appear in the Row Variable(s) box.
   - Clear the value in the Column Variable(s) text box so that the box is empty.
   - Check Create Other Column Variables from Numeric Data.
   - Double-click in the Numeric Variable(s) to Categorize for Use in Table Columns text box. This will bring up the variable selection window.
   - Select Price from the list of variables and then click OK. “Price” will appear in the Numeric Variable(s) to Categorize for Use in Table Columns box.
   - For Group Numeric Data into Categories Using select List of Interval Upper Limits.
   - For List enter 100000 200000 300000.

4 Specify the reports.
   - Select the Reports tab.
   - Leave Show Individual Tables and Counts checked.
   - Check Row Percentages under Show Individual Tables.

5 Specify the format.
   - Select the Report Options tab.
   - For Variable Names select Labels.
   - For Value Labels select Value Labels.

6 Run the procedure.
   - From the Run menu, select Run Procedure. Alternatively, just click the green Run button.

The following reports and plots will be displayed in the Output window.
### Counts Table

<table>
<thead>
<tr>
<th>Community</th>
<th>Up To 100000</th>
<th>100000 To 200000</th>
<th>200000 To 300000</th>
<th>Over 300000</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silverville</td>
<td>5</td>
<td>12</td>
<td>4</td>
<td>6</td>
<td>27</td>
</tr>
<tr>
<td>Los Wages</td>
<td>18</td>
<td>16</td>
<td>9</td>
<td>6</td>
<td>49</td>
</tr>
<tr>
<td>Red Gulch</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>Politicville</td>
<td>5</td>
<td>11</td>
<td>10</td>
<td>1</td>
<td>27</td>
</tr>
<tr>
<td>Senate City</td>
<td>6</td>
<td>12</td>
<td>4</td>
<td>2</td>
<td>24</td>
</tr>
<tr>
<td>Congresstown</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>Total</td>
<td>41</td>
<td>58</td>
<td>32</td>
<td>19</td>
<td>150</td>
</tr>
</tbody>
</table>

### Row Percentages Table

<table>
<thead>
<tr>
<th>Community</th>
<th>Up To 100000</th>
<th>100000 To 200000</th>
<th>200000 To 300000</th>
<th>Over 300000</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Silverville</td>
<td>18.52%</td>
<td>44.44%</td>
<td>14.81%</td>
<td>22.22%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Los Wages</td>
<td>36.73%</td>
<td>32.65%</td>
<td>18.37%</td>
<td>12.24%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Red Gulch</td>
<td>41.67%</td>
<td>25.00%</td>
<td>25.00%</td>
<td>8.33%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Politicville</td>
<td>18.52%</td>
<td>40.74%</td>
<td>37.04%</td>
<td>3.70%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Senate City</td>
<td>25.00%</td>
<td>50.00%</td>
<td>16.67%</td>
<td>8.33%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Congresstown</td>
<td>18.18%</td>
<td>36.36%</td>
<td>18.18%</td>
<td>27.27%</td>
<td>100.00%</td>
</tr>
<tr>
<td>Total</td>
<td>27.33%</td>
<td>38.67%</td>
<td>21.33%</td>
<td>12.67%</td>
<td>100.00%</td>
</tr>
</tbody>
</table>

### Tests for Row-Column Independence (Community by Sales Price)

<table>
<thead>
<tr>
<th>Test</th>
<th>Type</th>
<th>Chi-Square Value</th>
<th>DF</th>
<th>Prob Level</th>
<th>Reject H0 at α = 0.05?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pearson's Chi-Square†</td>
<td>2-Sided</td>
<td>16.8045</td>
<td>15</td>
<td>0.33069</td>
<td>No</td>
</tr>
<tr>
<td>Yates' Cont. Correction*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Likelihood Ratio</td>
<td>2-Sided</td>
<td>16.2412</td>
<td>15</td>
<td>0.36620</td>
<td>No</td>
</tr>
<tr>
<td>Fisher's Exact*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

† WARNING: At least one cell had an expected value less than 5.
* Test computed only for 2×2 tables.

This report presents the results from the data with the continuous price variable grouped into 4 categories. The Pearson’s chi-square test results indicate that there is not enough evidence to conclude that the row and column variables are associated (p-value = 0.33069). There is an expected value warning that should be considered. Note that Fisher’s Exact Test and Yates’ Continuity Correction are not reported because this is not a 2 × 2 table.
Example 4 – Tests for Trend in Proportions (Validation using Armitage (1955))

The data for this example come from Table 1 of Armitage (1955) and are stored in the “Armitage” dataset. The dataset contains counts of tonsil sizes (+, ++, ++++) from 1398 children aged 0-15 years along with and indicator or whether each child is a carrier or non-carrier of the bacteria Streptococcus pyogenes. On page 378, Armitage (1955) calculates the Cochran-Armitage chi-square test statistic for the alternative hypothesis of any trend to be 7.19 on 1 df with a p-value of 0.007. On page 383, Armitage (1955) calculates the Rank Correlation Test chi-square test statistic for the alternative hypothesis of any trend to be 6.83 on 1 df with a p-value of 0.009.

You may follow along here by making the appropriate entries or load the completed template Example 4 by clicking on Open Example Template from the File menu of the Contingency Tables (Crosstabs / Chi-Square Test) window.

1 **Open the Armitage dataset.**
   - From the File menu of the NCSS Data window, select **Open Example Data**.
   - Click on the file **Armitage.NCSS**.
   - Click **Open**.

2 **Open the Contingency Tables (Crosstabs / Chi-Square Test) window.**
   - On the menus, select **Analysis**, then **Descriptive Statistics**, then **Contingency Tables (Crosstabs / Chi-Square Test)**. The procedure will be displayed.
   - On the menus, select **File**, then **New Template**. This will fill the procedure with the default template.

3 **Specify the variables.**
   - Select the **Variables tab**.
   - Double-click in the **Row Variable(s)** text box. This will bring up the variable selection window.
   - Select **Strep** from the list of variables and then click **OK**. “Strep” will appear in the **Row Variable(s)** box.
   - Double-click in the **Column Variable(s)** text box. This will bring up the variable selection window.
   - Select **Tonsils** from the list of variables and then click **OK**. “Tonsils” will appear in the **Column Variable(s)** box.
   - Double-click in the **Frequency Variable** text box. This will bring up the variable selection window.
   - Select **Count** from the list of variables and then click **OK**. “Count” will appear in the **Frequency Variable** box.

4 **Specify the reports.**
   - Select the **Reports tab**.
   - Uncheck **Show Individual Tables**.
   - Check **Show Combined Table** and leave only **Counts** and **Column Percentages** checked.
   - Uncheck **Tests for Row-Column Independence**.
   - Check **Tests for Trend in Proportions** and all three corresponding trend tests.

5 **Specify the report options.**
   - Select the **Report Options tab**.
   - For **Variable Names**, select **Labels**.

6 **Run the procedure.**
   - From the Run menu, select **Run Procedure**. Alternatively, just click the green Run button.

The following reports and plots will be displayed in the Output window.
### Output

#### Combined Table

<table>
<thead>
<tr>
<th>Streptococcus pyogenes</th>
<th>Tonsil Size</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>Total</td>
<td></td>
</tr>
<tr>
<td>Non-carriers</td>
<td>497</td>
<td>560</td>
<td>269</td>
<td>1326</td>
<td></td>
</tr>
<tr>
<td></td>
<td>96.32%</td>
<td>95.08%</td>
<td>91.81%</td>
<td>94.85%</td>
<td></td>
</tr>
<tr>
<td>Carriers</td>
<td>19</td>
<td>29</td>
<td>24</td>
<td>72</td>
<td></td>
</tr>
<tr>
<td></td>
<td>3.68%</td>
<td>4.92%</td>
<td>8.19%</td>
<td>5.15%</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>516</td>
<td>589</td>
<td>293</td>
<td>1398</td>
<td></td>
</tr>
<tr>
<td></td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td></td>
</tr>
</tbody>
</table>

#### Cochran-Armitage Trend Test

**(Streptococcus pyogenes by Tonsil Size)**

H0: $p(1) = p(2) = p(3) = \ldots = p(k)$

<table>
<thead>
<tr>
<th>Alternative Hypothesis*</th>
<th>Numerator†</th>
<th>Standard Error</th>
<th>$Z$</th>
<th>Prob Level</th>
<th>Reject H0 at $\alpha = 0.05$?</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1: Increasing Trend</td>
<td>16.48498</td>
<td>6.1467</td>
<td>2.6819</td>
<td>0.00366</td>
<td>Yes</td>
</tr>
<tr>
<td>H1: Decreasing Trend</td>
<td>16.48498</td>
<td>6.1467</td>
<td>2.6819</td>
<td>0.99634</td>
<td>No</td>
</tr>
<tr>
<td>H1: Any Trend</td>
<td>16.48498</td>
<td>6.1467</td>
<td>2.6819</td>
<td>0.00732</td>
<td>Yes</td>
</tr>
</tbody>
</table>

* Trend is based on % within Column for Streptococcus pyogenes = “Carriers”.

#### Cochran-Armitage Trend Test with Continuity Correction

**(Streptococcus pyogenes by Tonsil Size)**

H0: $p(1) = p(2) = p(3) = \ldots = p(k)$

<table>
<thead>
<tr>
<th>Alternative Hypothesis*</th>
<th>Numerator†</th>
<th>Standard Error</th>
<th>$Z$</th>
<th>Prob Level</th>
<th>Reject H0 at $\alpha = 0.05$?</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1: Increasing Trend</td>
<td>15.98498</td>
<td>6.1467</td>
<td>2.6006</td>
<td>0.00465</td>
<td>Yes</td>
</tr>
<tr>
<td>H1: Decreasing Trend</td>
<td>15.98498</td>
<td>6.1467</td>
<td>2.6006</td>
<td>0.99535</td>
<td>No</td>
</tr>
<tr>
<td>H1: Any Trend</td>
<td>15.98498</td>
<td>6.1467</td>
<td>2.6006</td>
<td>0.00931</td>
<td>Yes</td>
</tr>
</tbody>
</table>

* Trend is based on % within Column for Streptococcus pyogenes = “Carriers”.

† Continuity Correction Factor ($\Delta/2$) = 0.5

#### Armitage Rank Correlation Trend Test

**(Streptococcus pyogenes by Tonsil Size)**

H0: $p(1) = p(2) = p(3) = \ldots = p(k)$

<table>
<thead>
<tr>
<th>Alternative Hypothesis*</th>
<th>Numerator</th>
<th>Standard Error</th>
<th>$Z$</th>
<th>Prob Level</th>
<th>Reject H0 at $\alpha = 0.05$?</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1: Increasing Trend</td>
<td>16229</td>
<td>6208.3460</td>
<td>2.6141</td>
<td>0.00447</td>
<td>Yes</td>
</tr>
<tr>
<td>H1: Decreasing Trend</td>
<td>16229</td>
<td>6208.3460</td>
<td>2.6141</td>
<td>0.99553</td>
<td>No</td>
</tr>
<tr>
<td>H1: Any Trend</td>
<td>16229</td>
<td>6208.3460</td>
<td>2.6141</td>
<td>0.00895</td>
<td>Yes</td>
</tr>
</tbody>
</table>

* Trend is based on % within Column for Streptococcus pyogenes = “Carriers”.

The reported alternative hypotheses correspond to the trend in proportions for the second row (Strep = “Carriers”). The two-sided Cochran-Armitage test confirms that the carrier rate (% within Column for Strep = “Carriers”) does, in fact, change with the tonsil size ($Z = 2.6819$ and p-value = 0.00732). The continuity corrected test (p-value = 0.00931) and Armitage rank correlation test ($Z = 2.6141$ and p-value = 0.00895) show similar results. These test results match exactly those given in Armitage (1955) if we note that $Z^2 = \text{Chi-Square on 1 df}$ such that $2.6819^2 = 7.1926$ and $2.6141^2 = 6.8335$. 

---

501-34 © NCSS, LLC. All Rights Reserved.
Example 5 – McNemar Test

The data for this example are found in the “McNemar” dataset. This hypothetical data contains summarized responses from 23 individuals who were asked about their desire to purchase a certain home-improvement product before and after a sales demonstration.

You may follow along here by making the appropriate entries or load the completed template Example 5 by clicking on Open Example Template from the File menu of the Contingency Tables (Crosstabs / Chi-Square Test) window.

1 Open the McNemar dataset.
   - From the File menu of the NCSS Data window, select Open Example Data.
   - Click on the file McNemar.NCSS.
   - Click Open.

2 Open the Contingency Tables (Crosstabs / Chi-Square Test) window.
   - On the menus, select Analysis, then Descriptive Statistics, then Contingency Tables (Crosstabs / Chi-Square Test). The procedure will be displayed.
   - On the menus, select File, then New Template. This will fill the procedure with the default template.

3 Specify the variables.
   - Select the Variables tab.
   - Double-click in the Row Variable(s) text box. This will bring up the variable selection window.
   - Select Before from the list of variables and then click OK. “Before” will appear in the Row Variable(s) box.
   - Double-click in the Column Variable(s) text box. This will bring up the variable selection window.
   - Select After from the list of variables and then click OK. “After” will appear in the Column Variable(s) box.
   - Double-click in the Frequency Variable text box. This will bring up the variable selection window.
   - Select Count from the list of variables and then click OK. “Count” will appear in the Frequency Variable box.

4 Specify the reports.
   - Select the Reports tab.
   - Leave Show Individual Tables and Counts checked.
   - Uncheck Tests for Row-Column Independence.
   - Check McNemar Test.

5 Run the procedure.
   - From the Run menu, select Run Procedure. Alternatively, just click the green Run button.

The following reports and plots will be displayed in the Output window.
Output

Counts Table

<table>
<thead>
<tr>
<th>Before</th>
<th>After</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>No</td>
<td>6</td>
<td>10</td>
</tr>
<tr>
<td>Yes</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Total</td>
<td>9</td>
<td>14</td>
</tr>
</tbody>
</table>

McNemar Test
(Before by After)
H0: P12 = P21
H1: P12 ≠ P21

<table>
<thead>
<tr>
<th>Test</th>
<th>Type</th>
<th>Chi-Square Value</th>
<th>DF</th>
<th>Prob Level</th>
<th>Reject H0 at α = 0.05?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asymptotic Chi-Square</td>
<td>2-Sided</td>
<td>3.7692</td>
<td>1</td>
<td>0.05220</td>
<td>No</td>
</tr>
<tr>
<td>Binomial Exact</td>
<td>2-Sided</td>
<td></td>
<td></td>
<td>0.09229</td>
<td>No</td>
</tr>
</tbody>
</table>

Both the Asymptotic Chi-Square (p-value = 0.05220) and Binomial Exact (p-value = 0.09229) McNemar Tests indicate that there is not enough evidence to reject the null hypothesis.
Example 6 – Kappa Test for Inter-Rater Agreement from Summarized Data (Validation using Fleiss, Levin, and Paik (2003))

Fleiss, Levin, and Paik (2003) present a hypothetical example on pages 598-608 in which 100 subjects are diagnosed independently by two raters and placed into 1 of 3 categories: Psychotic, Neurotic, and Organic. The summarized data are contained in the “KappaFleiss” dataset. In this example dataset, please note that a custom value order has been entered so that the table categories appear in the same order as in the book.

They compute a kappa value of 0.68, a null standard error of 0.076, and a z-value of 8.95 for testing the null hypothesis that kappa = 0. They also compute the asymptotic standard error for computing confidence intervals to be 0.087.

You may follow along here by making the appropriate entries or load the completed template Example 6a by clicking on Open Example Template from the File menu of the Contingency Tables (Crosstabs / Chi-Square Test) window.

1 Open the KappaFleiss dataset.
   • From the File menu of the NCSS Data window, select Open Example Data.
   • Click on the file KappaFleiss.NCSS.
   • Click Open.

2 Open the Contingency Tables (Crosstabs / Chi-Square Test) window.
   • On the menus, select Analysis, then Descriptive Statistics, then Contingency Tables (Crosstabs / Chi-Square Test). The Cross Tabulation procedure will be displayed.
   • On the menus, select File, then New Template. This will fill the procedure with the default template.

3 Specify the variables.
   • Select the Variables tab.
   • Double-click in the Row Variable(s) text box. This will bring up the variable selection window.
   • Select Rater_A from the list of variables and then click OK. “Rater_A” will appear in the Row Variable(s) box.
   • Double-click in the Column Variable(s) text box. This will bring up the variable selection window.
   • Select Rater_B from the list of variables and then click OK. “Rater_B” will appear in the Column Variable(s) box.
   • Double-click in the Frequency Variable text box. This will bring up the variable selection window.
   • Select Count from the list of variables and then click OK. “Count” will appear in the Frequency Variable box.

4 Specify the reports.
   • Select the Reports tab.
   • Leave Show Individual Tables and Counts checked.
   • Uncheck Tests for Row-Column Independence.
   • Check Kappa and Weighted Kappa Tests for Inter-Rater Agreement.

5 Run the procedure.
   • From the Run menu, select Run Procedure. Alternatively, just click the green Run button.

The following reports and plots will be displayed in the Output window.

501-37
Counts Table

<table>
<thead>
<tr>
<th>Rater_A</th>
<th>Psychotic</th>
<th>Neurotic</th>
<th>Organic</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Psychotic</td>
<td>75</td>
<td>1</td>
<td>4</td>
<td>80</td>
</tr>
<tr>
<td>Neurotic</td>
<td>5</td>
<td>4</td>
<td>1</td>
<td>10</td>
</tr>
<tr>
<td>Organic</td>
<td>0</td>
<td>0</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>80</td>
<td>5</td>
<td>15</td>
<td>100</td>
</tr>
</tbody>
</table>

Kappa Estimation (Rater_A by Rater_B)

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
<th>Asymptotic Std. Error</th>
<th>95% Lower Conf. Limit</th>
<th>95% Upper Conf. Limit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kappa</td>
<td>0.6765</td>
<td>0.0877</td>
<td>0.5046</td>
<td>0.8484</td>
</tr>
<tr>
<td>Weighted Kappa (Linear)</td>
<td>0.7222</td>
<td>0.0843</td>
<td>0.5570</td>
<td>0.8874</td>
</tr>
<tr>
<td>Weighted Kappa (Quadratic)</td>
<td>0.7553</td>
<td>0.0867</td>
<td>0.5854</td>
<td>0.9253</td>
</tr>
<tr>
<td>Maximum-Adjusted Kappa*</td>
<td>0.7931</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Kappa Hypothesis Tests (Rater_A by Rater_B)

<table>
<thead>
<tr>
<th>Test</th>
<th>Value</th>
<th>Asymptotic Std. Error under H0</th>
<th>Z</th>
<th>One-Sided Prob Level</th>
<th>Two-Sided Prob Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Kappa</td>
<td>0.6765</td>
<td>0.0762</td>
<td>8.8791</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>Weighted Kappa (Linear)</td>
<td>0.7222</td>
<td>0.0879</td>
<td>8.2201</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
<tr>
<td>Weighted Kappa (Quadratic)</td>
<td>0.7553</td>
<td>0.0989</td>
<td>7.6335</td>
<td>0.00000</td>
<td>0.00000</td>
</tr>
</tbody>
</table>

The results from NCSS match Fleiss, Levin, and Paik (2003) with some small differences due to rounding (most pronounced in the z-value). The authors used rounded values in hand calculations to arrive at their results. NCSS uses full precision in all calculations.

The weighted kappa results are completely disregarded here because the rating categories have no inherent order. Weighted kappa should be ignored in this case.

An alternate way to enter this summarized data is to set Type of Data Input to Summary Table. You may follow along here by making the appropriate modifications to the current settings or load the completed template Example 6b by clicking on Open Example Template from the File menu of the Contingency Tables (Crosstabs / Chi-Square Test) window.

6 Modify the Data Input Type.
- Select the Variables tab.
- For Type of Data Input, select Summary Table.
- Enter the variable names, labels, and counts into the spreadsheet.

7 Run the procedure again.
- From the Run menu, select Run Procedure. Alternatively, just click the green Run button.

The output will be exactly the same as that displayed above.
Example 7 – Weighted Kappa Test from Raw Data with Missing Cell Combinations

The data for this example are contained in the “WeightedKappa” dataset. This dataset contains independent ratings by 2 raters on 12 individuals. Each individual was scored on a scale of 1 to 5. The goal is to determine how closely the raters’ scores agree. Weighted kappa is appropriate in this case because the categories are ordinal, meaning that the categories have magnitude with natural ordering.

The problem with the data in this dataset, however, is that there are no cases where Rater 1 scored an individual as “3” and there are no cases where Rater 2 scored individual as “2”. This results in a contingency table that is square, but does not have identical row and column categories. The second part of the example will show you how to modify the data by adding zeros appropriately so that kappa can be computed without having to go through the added step of summarizing the data first.

You may follow along here by making the appropriate entries or load the completed template Example 7a by clicking on Open Example Template from the File menu of the Contingency Tables (Crosstabs / Chi-Square Test) window.

1. Open the WeightedKappa dataset.
   - From the File menu of the NCSS Data window, select Open Example Data.
   - Click on the file WeightedKappa.NCSS.
   - Click Open.

2. Open the Contingency Tables (Crosstabs / Chi-Square Test) window.
   - On the menus, select Analysis, then Descriptive Statistics, then Contingency Tables (Crosstabs / Chi-Square Test). The procedure will be displayed.
   - On the menus, select File, then New Template. This will fill the procedure with the default template.

3. Specify the variables.
   - Select the Variables tab.
   - Double-click in the Row Variable(s) text box. This will bring up the variable selection window.
   - Select Rater_1 from the list of variables and then click OK. “Rater_1” will appear in the Row Variable(s) box.
   - Double-click in the Column Variable(s) text box. This will bring up the variable selection window.
   - Select Rater_2 from the list of variables and then click OK. “Rater_2” will appear in the Column Variable(s) box.

4. Specify the reports.
   - Select the Reports tab.
   - Leave Show Individual Tables and Counts checked.
   - Uncheck Tests for Row-Column Independence.
   - Check Kappa and Weighted Kappa Tests for Inter-Rater Agreement.

5. Specify the format.
   - Select the Report Options tab.
   - For Variable Names select Labels.

6. Run the procedure.
   - From the Run menu, select Run Procedure. Alternatively, just click the green Run button.

The following reports and plots will be displayed in the Output window.
### Count Table

<table>
<thead>
<tr>
<th></th>
<th>Rater 1</th>
<th>Rater 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>3</td>
<td>4</td>
</tr>
</tbody>
</table>

The number of rows with at least one missing value is 1.

### Kappa Estimation
(Rater 1 by Rater 2)

Not Calculated: Kappa statistics are only calculated for square k×k tables with identical row and column categories.

### Kappa Hypothesis Tests
(Rater 1 by Rater 2)

Not Calculated: Kappa tests are only calculated for square k×k tables with identical row and column categories.

This report indicates that it is not possible to calculate the kappa test statistic for this table since the table categories are not identical for rows and columns. There is no row category “3” for Rater 1 and no column category “2” for Rater 2. This is a result of the fact that some categories were not applied by each rater. We could re-enter this summarized data into the data table, including missing rows and columns with assigned counts of “0”, but there is an easier way. Simply add a new variable “Count” to the dataset and give each row a count value of “1”. In the row immediately after the last data row, enter the values for each rater that were never observed (i.e. in the first empty row at the end of the dataset, enter “3” under Rater_1 and “2” under Rater_2) and assign those a count value of “0”. In the example dataset we created new variables RaterMod_1, RaterMod_2, and Count to illustrate this principle, but in your dataset you do not necessarily have to create new rater variables; you can just add the rows with count = 0 starting at the first empty row. We’ll now run the analysis using the modified variables to get the kappa results.

You may follow along here by making the appropriate modifications to the current settings or load the completed template Example 7b by clicking on Open Example Template from the File menu of the Contingency Tables (Crosstabs / Chi-Square Test) window.

### 7 Modify the variables.

- Select the Variables tab.
- Double-click in the Row Variable(s) text box. This will bring up the variable selection window.
- Select RaterMod_1 from the list of variables and then click OK. “RaterMod_1” will appear in the Row Variable(s) box.
- Double-click in the Column Variable(s) text box. This will bring up the variable selection window.
- Select RaterMod_2 from the list of variables and then click OK. “RaterMod_2” will appear in the Column Variable(s) box.
- Double-click in the Frequency Variable text box. This will bring up the variable selection window.
- Select Count from the list of variables and then click OK. “Count” will appear in the Frequency Variable box.

### 8 Run the procedure again.

- From the Run menu, select Run Procedure. Alternatively, just click the green Run button.

The following reports and plots will be displayed in the Output window.
The counts table now has identical row and column categories with the exact same counts as the previous table. A row and column of zeros have been added to make the table suitable for the calculation of the kappa and weighted kappa statistics. It is appropriate to consider the weighted kappa statistic in this case because the data are comprised of ordered scores. A weighted kappa value of 0.7561 indicates moderate-to-high agreement between the raters. If we did not take into account the ordinal nature of the data and looked at the simple kappa statistic, we would conclude a much lower association of 0.5000. This demonstrates the importance of using weighted kappa when it is appropriate.
Example 8 – Data Summary Report

The data summary report was designed for situations in which you want to transfer a summarized table to another program. This format creates a vertical listing of the counts in a format that is easy to copy and paste into another NCSS dataset or into other programs. The data for this example are found in the “Resale” dataset.

You may follow along here by making the appropriate entries or load the completed template Example 8 by clicking on Open Example Template from the File menu of the Contingency Tables (Crosstabs / Chi-Square Test) window.

1 Open the Resale dataset.
   • From the File menu of the NCSS Data window, select Open Example Data.
   • Click on the file Resale.NCSS.
   • Click Open.

2 Open the Contingency Tables (Crosstabs / Chi-Square Test) window.
   • On the menus, select Analysis, then Descriptive Statistics, then Contingency Tables (Crosstabs / Chi-Square Test). The procedure will be displayed.
   • On the menus, select File, then New Template. This will fill the procedure with the default template.

3 Specify the variables.
   • Select the Variables tab.
   • Double-click in the Row Variable(s) text box. This will bring up the variable selection window.
   • Select City from the list of variables and then click OK. “City” will appear in the Row Variable(s) box.
   • Clear the value in the Column Variable(s) text box so that the box is empty.
   • Check Create Other Column Variables from Numeric Data.
   • Double-click in the Numeric Variable(s) to Categorize for Use in Table Columns text box. This will bring up the variable selection window.
   • Select Price from the list of variables and then click OK. “Price” will appear in the Numeric Variable(s) to Categorize for Use in Table Columns box.
   • For Group Numeric Data into Categories Using select List of Interval Upper Limits.
   • For List enter 100000 200000 300000.

4 Specify the reports.
   • Select the Reports tab.
   • Check Data Summary Report.
   • Leave Show Individual Tables and Counts checked.
   • Uncheck Tests for Row-Column Independence.

5 Run the procedure.
   • From the Run menu, select Run Procedure. Alternatively, just click the green Run button.

The following reports and plots will be displayed in the Output window.
Data Summary Report

<table>
<thead>
<tr>
<th>Price</th>
<th>City</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Up To 100000</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Up To 100000</td>
<td>2</td>
<td>10</td>
</tr>
<tr>
<td>Up To 100000</td>
<td>3</td>
<td>5</td>
</tr>
<tr>
<td>Up To 100000</td>
<td>4</td>
<td>5</td>
</tr>
<tr>
<td>Up To 100000</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Up To 100000</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>100000 To 200000</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>100000 To 200000</td>
<td>2</td>
<td>16</td>
</tr>
<tr>
<td>100000 To 200000</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>100000 To 200000</td>
<td>4</td>
<td>11</td>
</tr>
<tr>
<td>100000 To 200000</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>100000 To 200000</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>200000 To 300000</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>200000 To 300000</td>
<td>2</td>
<td>9</td>
</tr>
<tr>
<td>200000 To 300000</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>200000 To 300000</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>200000 To 300000</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>200000 To 300000</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>Over 300000</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Over 300000</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>Over 300000</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>Over 300000</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>Over 300000</td>
<td>5</td>
<td>2</td>
</tr>
<tr>
<td>Over 300000</td>
<td>6</td>
<td>3</td>
</tr>
</tbody>
</table>

Counts Table

<table>
<thead>
<tr>
<th>City</th>
<th>Up To 100000</th>
<th>100000 To 200000</th>
<th>200000 To 300000</th>
<th>Over 300000</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5</td>
<td>12</td>
<td>4</td>
<td>6</td>
<td>27</td>
</tr>
<tr>
<td>2</td>
<td>18</td>
<td>16</td>
<td>9</td>
<td>6</td>
<td>49</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>3</td>
<td>3</td>
<td>1</td>
<td>12</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>11</td>
<td>10</td>
<td>1</td>
<td>27</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>12</td>
<td>4</td>
<td>2</td>
<td>24</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>11</td>
</tr>
<tr>
<td>Total</td>
<td>41</td>
<td>58</td>
<td>32</td>
<td>19</td>
<td>150</td>
</tr>
</tbody>
</table>

This report gives the count (frequency) for each unique combination of the table and grouping variables, taken together. In this example, there are no grouping variables.