Contingency Tables

The **Contingency Tables** procedure is designed to analyze and display frequency data contained in a two-way table. Such data is often collected as the result of a survey. Statistics are constructed to quantify the degree of association between the rows and columns, and tests are run to determine whether or not there is a statistically significant dependence between the row classification and the column classification. The frequencies are displayed both in tabular form and graphically as a barchart, mosaic plot, or skychart.

For data that has not yet been tabulated, use the *Crosstabulation* procedure, which creates similar output from raw response data.

Sample StatFolio: *contingency.sgp*

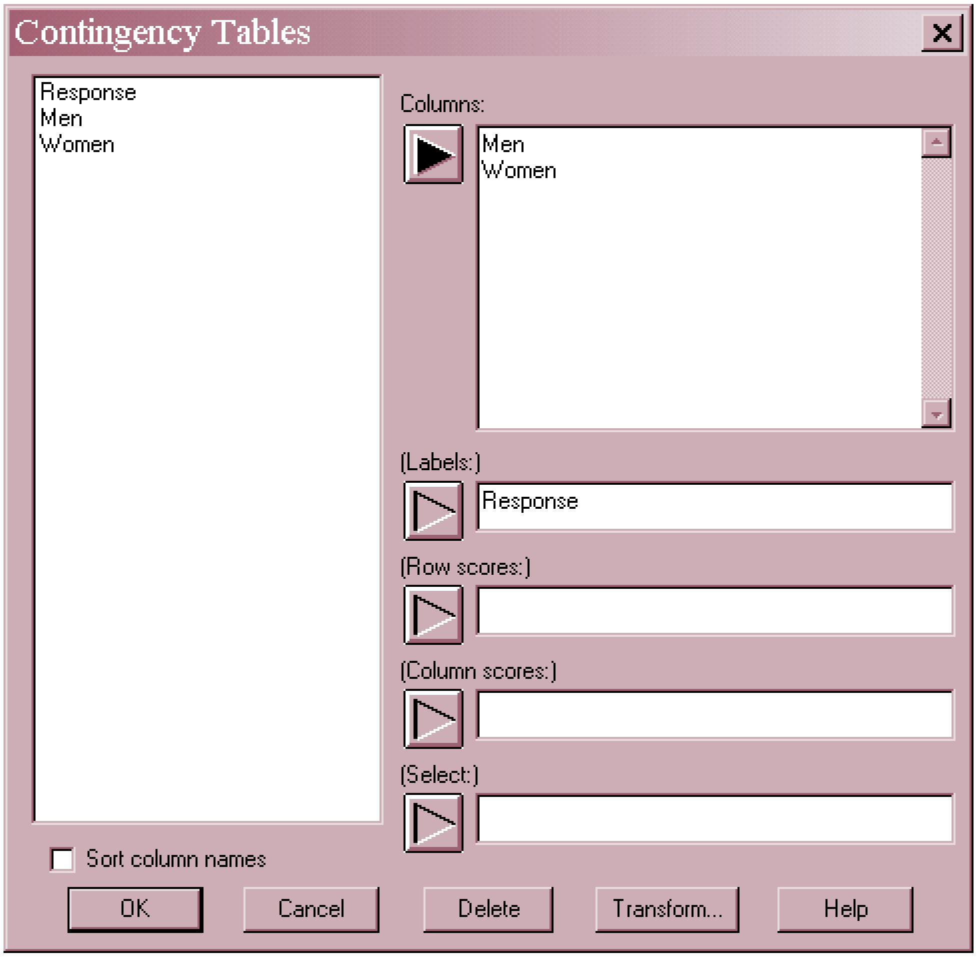
Sample data:

The file *opinion.sgd* contains the results of an opinion poll. *n* = 200 people, 107 men and 93 women, were asked to express an opinion about whether or not they agreed with a particular statement. The table below shows the results of that poll:

Response	*Men*	*Women*
Disagree strongly	5	17
Disagree	20	28
No opinion	12	3
Agree	50	35
Agree strongly	20	10

Data Input

The data input dialog box specifies the columns containing the data in the two-way table.



Contingency Tables

Response
Men
Women

Columns:
Men
Women

(Labels:) Response

(Row scores:)

(Column scores:)

(Select:)

☐ Sort column names

- **Columns:** two or more numeric columns corresponding to the columns of the table.

- **Labels:** optional labels to be assigned to each row of the table. Column labels are automatically generated from the column names.

- **Row scores:** optional numeric column with scores to be associated with each row. These scores are used when generating certain summary statistics and tests. If not specified, row scores will be constructed automatically using an algorithm based on their order and the row totals.

- **Column scores:** optional numeric column with scores to be associated with each column. If not specified, column scores will be constructed automatically using an algorithm based on their order and the column totals.

- **Select:** subset selection.

Analysis Summary

The *Analysis Summary* shows the number of rows and columns, as well the sum of the frequencies in all cells of the table.

```
Contingency Tables
Column variables:
    Men
    Women

Number of observations: 200
Number of rows: 5
Number of columns: 2
```

Frequency Table

The *Frequency Table* shows the frequency of occurrence of each pair of values in the row and column variables, together with other information as defined on the *Pane Options* dialog box.

Frequency Table

	Men	Women	Row Total
Disagree strongly	5	17	22
	2.50%	8.50%	11.00%
Disagree	20	28	48
	10.00%	14.00%	24.00%
No opinion	12	3	15
	6.00%	1.50%	7.50%
Agree	50	35	85
	25.00%	17.50%	42.50%
Agree strongly	20	10	30
	10.00%	5.00%	15.00%
Column Total	107	93	200
	53.50%	46.50%	100.00%

Cell contents:
    Observed frequency
    Percentage of table

The sample data consists of r = 5 rows by c = 2 columns. Included in the table are:

- **Observed Frequency:** The cells in the main part of the table display O_{ij}, the entered frequency for row i and column j.

- **Percentage of Table:** Beneath the cell frequencies are the percentage that each cell represents of the entire table.

- **Row totals**: The rightmost column of the table contains the row totals R_i:

$$R_i = \sum_{j=1}^{c} O_{ij} \tag{1}$$

- **Column totals**: The bottom row of the table contains the column totals C_j:

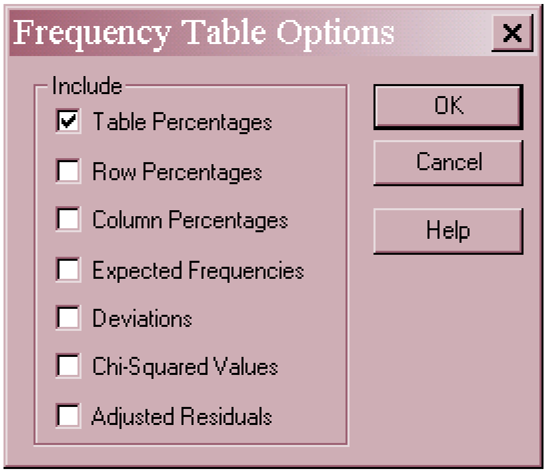
$$C_j = \sum_{i=1}^{r} O_{ij} \tag{2}$$

- **Table total**: the bottom right cell contains the sum of all the frequencies:

$$n = \sum_{i=1}^{r}\sum_{j=1}^{c} O_{ij} \tag{3}$$

For example, 5 men “Disagreed Strongly” with the statement posed to them.

Pane Options
Additional information may be added to each cell of the table using *Pane Options*:



Frequency Table Options

Include

☑ Table Percentages
☐ Row Percentages
☐ Column Percentages
☐ Expected Frequencies
☐ Deviations
☐ Chi-Squared Values
☐ Adjusted Residuals

- **Table Percentages:** the percentage each cell represents of the overall table, defined by

$$100\frac{O_{ij}}{n}\% \tag{4}$$

- **Row Percentages:** the percentage each cell represents of the total count in its row, defined by

$$100\frac{O_{ij}}{R_i}\% \tag{5}$$

- **Column Percentages:** the percentage each cell represents of the total count in its column, defined by

$$100\frac{O_{ij}}{C_j}\% \tag{6}$$

- **Expected frequency:** E_{ij}, the expected number of times row i would have appeared together with column j in the data file if the row and column classifications were independent:

$$E_{ij} = \frac{R_i C_j}{n} \tag{7}$$

- **Deviations**: the differences between the observed and expected frequencies:

$$O_{ij} - E_{ij} \tag{8}$$

- **Chi-Squared Values**: the contribution of each cell to a chi-squared statistic, used to test for independence between row and column classifications:

$$\frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} \tag{9}$$

- **Adjusted Residuals:** a form of standardized residual computed by dividing each cell deviation by an estimate of its standard error:

$$\varepsilon_{ij} = \frac{\left(O_{ij} - E_{ij}\right)}{\sqrt{E_{ij}\left(1 - \frac{R_i}{n}\right)\left(1 - \frac{C_j}{n}\right)}} \tag{10}$$

Example – Additional Information on Men Agreeing Strongly

Frequency Table

	Men	Women	Row Total
Agree strongly	20	10	30
	10.00%	5.00%	15.00%
	66.67%	33.33%	
	18.69%	10.75%	
	16.05	13.95	
	3.95	-3.95	
	0.97	1.12	
	1.57	-1.57	
Column Total	107	93	200
	53.50%	46.50%	100.00%

Cell contents:
Observed frequency
Percentage of table
Percentage of row
Percentage of column
Expected frequency
Observed - expected frequency
Contribution to chi-squared
Adjusted residual

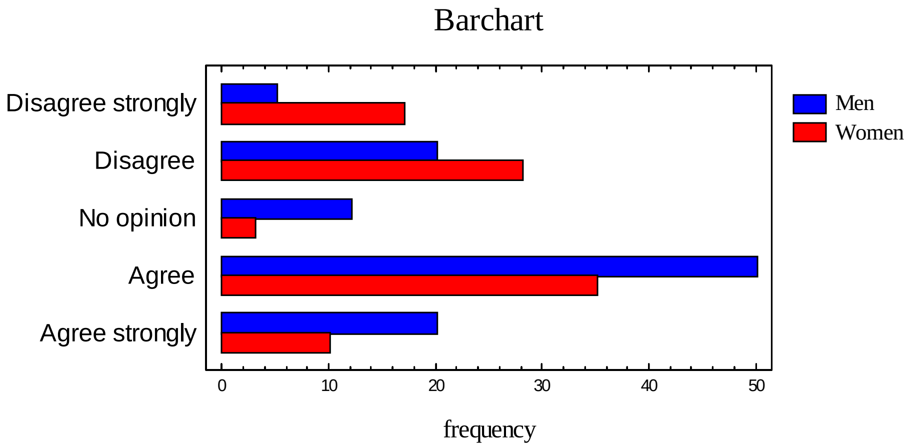
The 20 men who agreed strongly with the statement represent:

10.00% of the total of *n* = 200 respondents
66.67% of all 30 respondents who *Agreed Strongly*
18.69% of all 107 men

Were row and column classifications independent, the expected number of men who *Agreed Strongly* is 16.05, for a deviation of 13.95. In the computation of the Chi-squared test statistic, described below, this cell adds a total of 0.97 to that statistic. The adjusted residual indicates that the observed number of respondents in this cell is 1.57 standard deviations above its expected value.

Barchart

A common way to display the data in a two-way table is by using a multiple barchart.



The length of each bar in the above chart represents the number of respondents in each cell of the table.

Pane Options

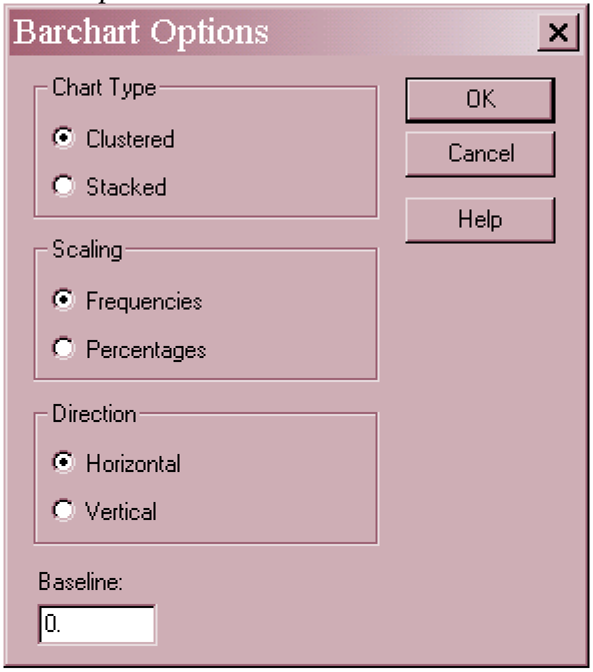


Chart Type: ● Clustered ○ Stacked

Scaling: ● Frequencies ○ Percentages

Direction: ● Horizontal ○ Vertical

Baseline: 0.

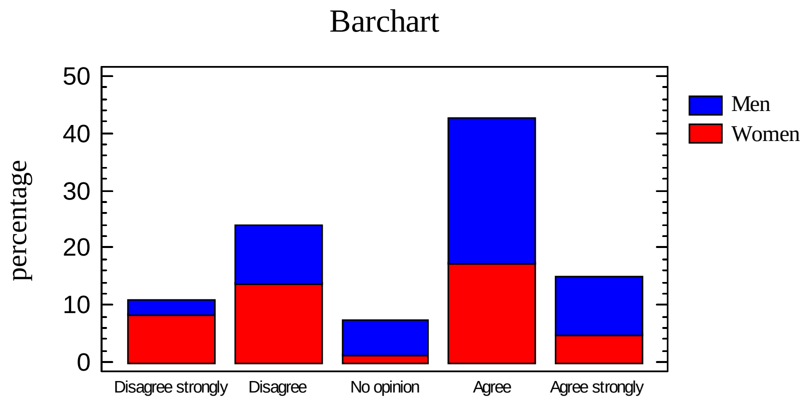
- **Chart Type:** The bars may be clustered side by side as shown in the example or stacked one upon the other.

- **Scaling**: whether the axis scale shows the frequencies O_{ij} or the percentages given by

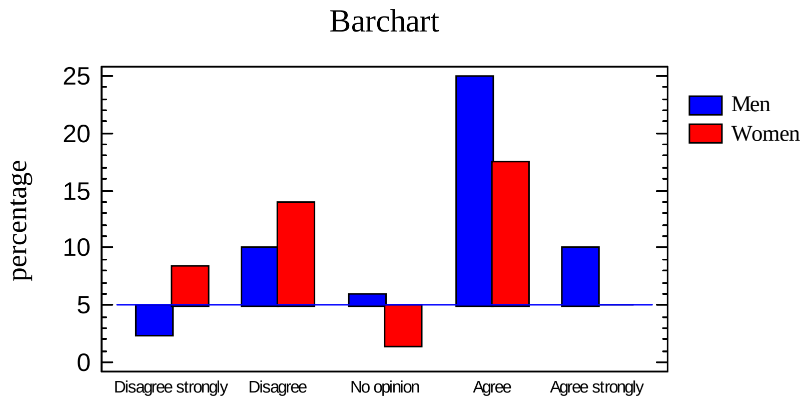
$$p_{ij} = 100\frac{O_{ij}}{n}\% \tag{11}$$

- **Direction**: whether the bars extend horizontally or vertically.
- **Baseline**: the value from which the bars extend.

Example – Stacked Horizontal Barchart Scaled by Percentage

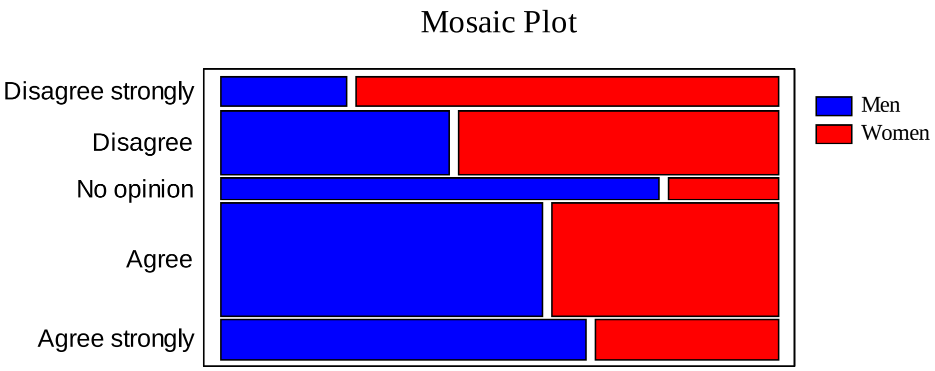


Example – Clustered Barchart with Nonzero Baseline of 5%



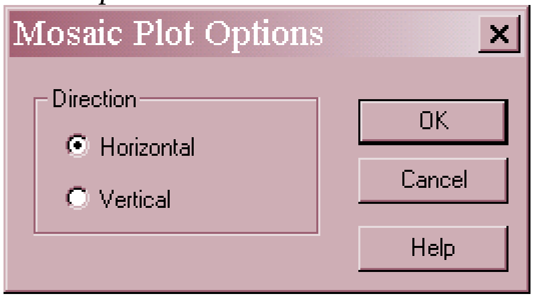
Mosaic Chart

An interesting variation of the Barchart occurs if the both the length and width of each bar are scaled to represent the frequency of the corresponding cell in the two-way table.



In this chart, the height of each row is proportional to its row total R_i. The width of the bars within each row is proportional to the frequency of each cell *within that row*. This results in bars whose *area* is proportional to the frequency in a particular cell. In the sample data, the largest bar corresponds to men who responded “*Agree*”.

Pane Options


Mosaic Plot Options

Direction

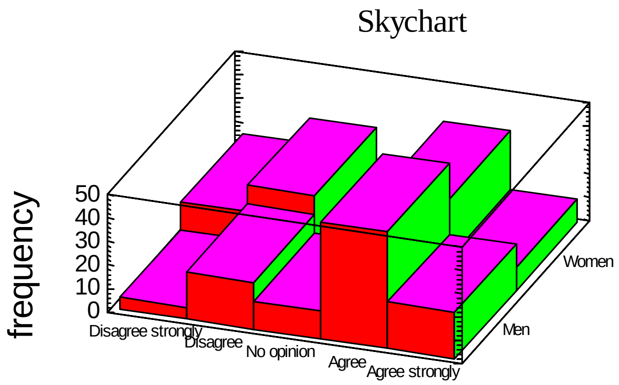
● Horizontal

○ Vertical

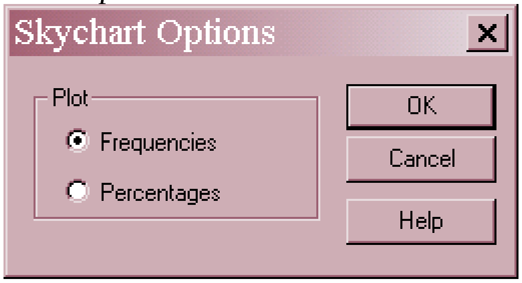
- **Direction:** the orientation of the bars.

SkyChart

The cell frequencies can also be represented using vertical bars.



Pane Options



Plot

● Frequencies

○ Percentages

- **Plot:** scaling for the vertical axis.

Tests of Independence

A common question asked of the data in two-way tables is whether or not the row and column classifications are independent. If rows and columns are independent, then the fact that an item falls in a particular row does not affect the probability of its falling in a given column. In the current example, independence would imply that both genders responded similarly to the statement posed to them.

STATGRAPHICS can perform any of 5 different tests, depending on the setting on the *Pane Options* dialog box. Each of the options tests the following hypotheses:

Null hypothesis: row and column classifications are independent
Alt. hypothesis: row and column classifications are not independent

Associated with each test is a P-Value. Small P-values (less than 0.05 if operating at the 5% significance level) lead to a rejection of the null hypothesis, implying a significant dependence between the row and column classifications.

Chi-Squared Test

The most common test for independence in a two-way table is the chi-squared test. This test compares the observed and expected frequencies by calculating:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}} \qquad (12)$$

STATGRAPHICS displays the results of this test and its corresponding P-value:

Tests of Independence

Test	*Statistic*	*Df*	*P-Value*
Chi-Squared	18.369	4	0.0010

The P-value is calculated by comparing the test statistic to a chi-squared distribution with *(r-1)(c-1)* degrees of freedom. The P-Value in the above table clearly shows that the gender and response on the test are *not* independent.

If the expected value E_{ij} in any cell of the table is less than 5, a warning will be displayed. In such cases, the calculated Chi-squared statistic may not be well represented by a chi-squared distribution. This is particularly serious if any expected values are less than 2. When this occurs, consideration should be given to combining classes that do not contain much data.

Likelihood Ratio Test

An alternative to the chi-squared test is the likelihood ratio test. The test statistic for this test is given by

$$G^2 = 2\sum_{i=1}^{r}\sum_{j=1}^{c} O_{ij} \log\left(\frac{O_{ij}}{E_{ij}}\right) \tag{13}$$

Tests of Independence

Test	*Statistic*	*Df*	*P-Value*
Likelihood ratio	19.116	4	0.0007

This statistic is also compared to a chi-squared distribution with *(r-1)(c-1)* degrees of freedom.

Chi-Squared with Yates' Correction

In the case of 2 by 2 tables only, a modified version of the chi-squared test may be performed using Yates' correction for continuity:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{\left(\left|O_{ij} - E_{ij}\right| - 0.5\right)^2}{E_{ij}} \tag{14}$$

For example, suppose the results of the sample poll were condensed to a 2 by 2 table by combining classes as follows:

Response	*Men*	*Women*
Agree or agree strongly	70	45
Disagree or disagree strongly	25	45

The test could then be applied, with the following results.

Tests of Independence

Test	*Statistic*	*Df*	*P-Value*
Chi-squared with Yates' correction	10.038	1	0.0015

The test statistic is compared to a chi-squared distribution with 1 degree of freedom. Again, the result is highly significant.

Fisher's Exact Test

For 2 by 2 tables in which the sum of the counts *n* does not exceed 100, Fisher's exact test can be performed. Both two-sided tests (in which the direction of association is not specified in advance) and one-sided tests (in which the researcher expects an association in a particular direction) are performed. Typical output from this test is shown below:

Tests of Independence

Test	*One-sided P-Value*	*Two-sided P-Value*
Fisher's exact test	0.0138	0.0185

For a two-tailed test, the P-value is computed by summing the hypergeometric probability of every possible table with the same row and column totals as the observed table, and whose probability is less than or equal to that of the observed table. For a one-sided test, the P-value is computed by summing the probability of all such tables in which the frequency $O_{1,1}$ is less than or equal to that of the observed table. (Note: if necessary, rows and columns are rearranged so that $O_{1,1}$ is the smallest frequency in the table before the one-sided P-value is computed).

Linear Trend Test

If both rows and columns are ordinal, i.e., have a natural ordering, then a test can be run to determine whether or not a significant trend exists within the table. The test is performed by first defining scores for each category. For the sample data, scores could be defined as follows:

Response	*Score*
Disagree strongly	-2
Disagree	-1
No opinion	0
Agree	1
Agree strongly	2

Gender	*Score*
Men	1
Women	2

Given row scores $u_1 \le u_2 \le \ldots \le u_r$ and column scores $v_1 \le v_2 \le \ldots \le v_c$, the sum of crossproducts is calculated by

$$T = \sum_{i=1}^{r} \sum_{j=1}^{c} u_i v_j O_{ij} \qquad (15)$$

The result is then standardized to create a correlation r on a scale of –1 to 1. The larger the correlation, the stronger the linear relationship between the rows and columns. To test the hypothesis of independence, the test statistic

$$M^2 = (n-1)r^2 \qquad (16)$$

is compared to a chi-squared distribution with 1 degree of freedom.

The output of the test on the sample data is shown below:

Tests of Independence

Test	*Correlation*	*P-Value*
Linear trend test	-0.2481	0.0005

Scores

Row	*Score*	*Column*	*Score*
Disagree strongly	-2.0	Men	1.0
Disagree	-1.0	Women	2.0
No opinion	0.0		
Agree	1.0		
Agree strongly	2.0		

The table shows:

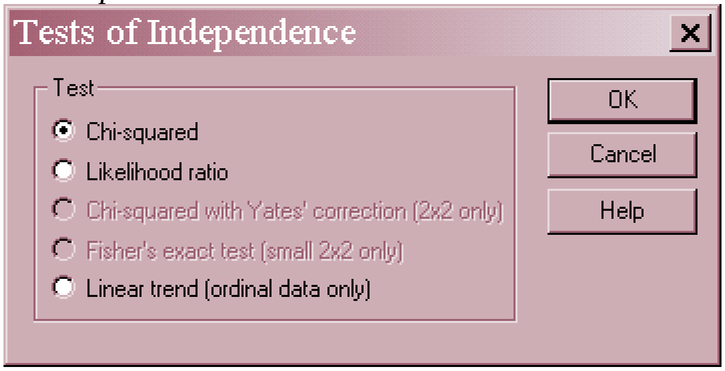
- **Correlation** – the calculated value of *r*. A negative correlation such as that observed above indicates a negative association between the scores. In the sample table, women, who have arbitrarily been given the higher score, tend to agree less with the statement.

- **P-Value** – tests the statistical significance of the correlation by comparing a normalized version of the test statistic to a Chi-squared distribution with one degree of freedom. Details may be found in Agresti (2002).

- **Scores** – the scores for each category.

The sample data shows a highly significant correlation.

Pane Options



Tests of Independence

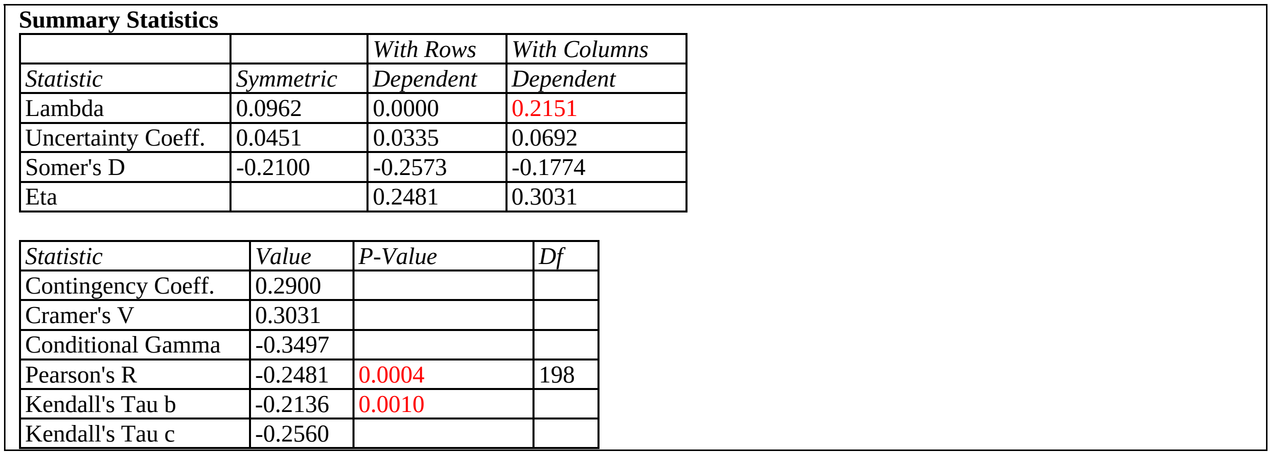
Test

● Chi-squared
○ Likelihood ratio
○ Chi-squared with Yates' correction (2x2 only)
○ Fisher's exact test (small 2x2 only)
○ Linear trend (ordinal data only)

- **Test** – the type of test to be performed.

Summary Statistics

Various statistics can also be computed to measure the degree of association between the rows and columns in the table.



Summary Statistics

		With Rows	*With Columns*
Statistic	*Symmetric*	*Dependent*	*Dependent*
Lambda	0.0962	0.0000	0.2151
Uncertainty Coeff.	0.0451	0.0335	0.0692
Somer's D	-0.2100	-0.2573	-0.1774
Eta		0.2481	0.3031

Statistic	*Value*	*P-Value*	*Df*
Contingency Coeff.	0.2900		
Cramer's V	0.3031		
Conditional Gamma	-0.3497		
Pearson's R	-0.2481	0.0004	198
Kendall's Tau b	-0.2136	0.0010	
Kendall's Tau c	-0.2560		

The following statistics are computed:

- **Lambda** - On a scale of 0 to 1, this statistic measures the relative improvement in predicting either rows or columns given knowledge of the other. In the above example, knowing the response helps predict the respondent’s gender, but knowing the respondent’s gender does not help predict the response (since the highest percentage of both men and women is in the “Agree” row).

- **Uncertainty coefficient** - On a scale of 0 to 1, this statistic measures the proportional reduction in the uncertainty about the value of the row or column variable given knowledge of the other. In the example, the reduction in uncertainty averages about 4.5%.

- **Somer’s D** - This statistic ranges from -1 to +1 and is based on the number of concordant and discordant pairs of observations. A concordant pair is one in which the two variables (row and column) have the same relative ranking (greater than or less than). A discordant pair is one in which the two variables have the opposite ranking. One variable (row or column) is considered to be the independent variable, while the other is considered to be the dependent variable. Both variables must be ordinal. A correction is made for ties on the independent variable. In this case, the association is negative.

- **Eta** - This statistic ranges between 0 and 1. When squared, eta represents the proportion of variation in the dependent variable that can be explained by knowledge of the independent variable. It is only appropriate when the dependent variable is of interval type and the independent variable is nominal or ordinal. When performing a crosstabulation of numeric data, Y_i equals the numeric value assigned to the dependent row or column. Otherwise, Y_i equals the row or column number. Since neither rows nor columns correspond to an interval variable in the current example, this statistic is not meaningful.

- **Contingency coefficient** - This statistic measures the degree of association between the values of the row and column variables on a scale of 0 to 1, based on the usual chi-square test statistic. It cannot in general attain the value 1 for all tables.

- **Cramer's V** - This statistic measures the degree of association between the values of the row and column variables on a scale of 0 to 1, based on the usual chi-square test statistic. Unlike the *contingency coefficient*, it can attain the value 1 for all tables.

- **Conditional gamma** - This statistic ranges from -1 to +1 and is based on the number of concordant and discordant pairs of observations. Both variables must be ordinal. No correction is made for ties.

- **Pearson's R** - This statistic measures the degree of association between the row and column variables using the ordinary correlation coefficient. It ranges between -1 and +1 and is relevant only if both variables are of interval type. Row and column values are assigned to each observation in a manner similar to that described for the *eta* statistic above. If $n > 2$ and $|R|$ is not equal to 1, a P value is computed to test the null hypothesis that the correlation equals 0.

- **Kendall's Tau b** - This statistic ranges from -1 to +1 and is based on the number of concordant and discordant pairs of observations, where 1 corresponds to complete concordance and -1 corresponds to complete disagreement. Both variables must be ordinal. A correction is made for tied pairs.

- **Kendall's Tau c** - This statistic is similar to *Kendall's tau b* except in its handling of ties.

Note that not all statistics are relevant for all types of data.

Odds Ratios

A useful way of looking at 2 by 2 tables when one factor corresponds to the occurrence or nonoccurrence of an event is through the odds ratio or relative risk of the event. For example, Agresti (2002) presents the following data from a study of the effectiveness of aspirin in preventing heart attacks:

Treatment	*Heart Attack*	*No Heart Attack*
Placebo	189	10,845
Aspirin	104	10,933

The table shows the findings from a study of n = 22,071 individuals. For this table, STATGRAPHICS generates the following output:

Odds Ratios and Relative Risk

Odds Ratios

Numerator	*Denominator*	*Odds Ratio*	*95% LCL*	*95% UCL*
Placebo	Aspirin	1.83205	1.44004	2.33078

The odds of an event are defined as the probability of an event divided by the probability that the event does not occur. In the sample data, the odds of having a heart attack as a function of the treatment that was given are estimated to be:

$$\textit{Placebo}: \frac{n_{11}}{n_{12}} = \frac{189}{10{,}845} = 0.01743 \tag{17}$$

$$\textit{Aspirin}: \frac{n_{21}}{n_{22}} = \frac{104}{10{,}933} = 0.00951 \tag{18}$$

The *odds ratio* is the ratio of these 2 numbers:

$$\hat{\theta} = \frac{n_{11}n_{22}}{n_{12}n_{21}} = 1.832 \tag{19}$$

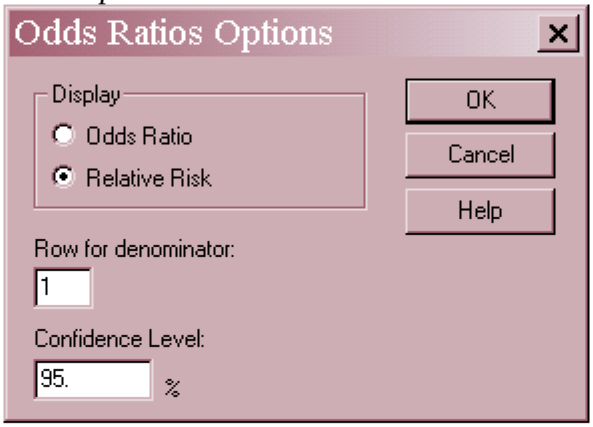
This implies that the odds of a heart attack are about 83% higher for those taking the placebo rather than the aspirin.

STATGRAPHICS also displays a confidence interval for the odds ratio, calculated from the inverse logarithms of:

$$\log\hat{\theta} \pm z_{\alpha/2}\sqrt{\frac{1}{n_{11}} + \frac{1}{n_{12}} + \frac{1}{n_{21}} + \frac{1}{n_{22}}} \tag{20}$$

In the sample data, since the 95% confidence interval is completely above 1.0, we can state with 95% confidence that the odds of a heart attack for those taking the placebo is greater than the odds for those taking aspirin.

Pane Options



Odds Ratios Options

Display

○ Odds Ratio

● Relative Risk

Row for denominator: 1

Confidence Level: 95. %

- **Display**: whether to display the odds ratio or the relative risk of an event.

- **Row for denominator**: of the 2 rows, which row should be used as the denominator in the ratio.

- **Confidence level:** the percentage to be used for the confidence interval.

Example: Relative Risk

Rather than using the odds ratio, the likelihood of an event can be compared using the estimated risks instead. For example, the estimated *risk* or probability of a heart attack for the two groups is:

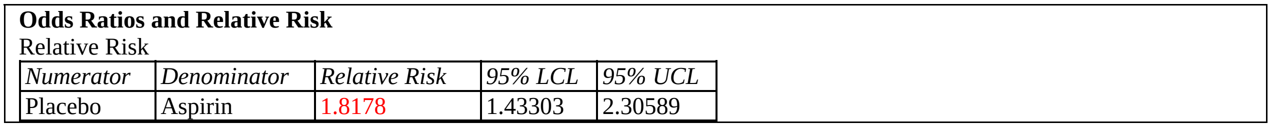
$$\textit{Placebo}:\ \hat{p}_1 = \frac{n_{11}}{n_{11}+n_{12}} = \frac{189}{11{,}034} = 0.01713 \tag{21}$$

$$\textit{Aspirin}:\ \hat{p}_2 = \frac{n_{21}}{n_{21}+n_{22}} = \frac{104}{11{,}037} = 0.00942 \tag{22}$$

The ratio of these two quantities is called the *relative risk*:

$$r = \frac{\hat{p}_1}{\hat{p}_2} = \frac{0.01713}{0.00942} = 1.82 \tag{23}$$

STATGRAPHICS displays this estimate together with a confidence interval:



Odds Ratios and Relative Risk

Relative Risk

Numerator	*Denominator*	*Relative Risk*	*95% LCL*	*95% UCL*
Placebo	Aspirin	1.8178	1.43303	2.30589

The confidence interval is based on:

$$\log r \pm z_{\alpha/2}\sqrt{\frac{1-\hat{p}_1}{\hat{p}_1(n_{11}+n_{12})}+\frac{1-\hat{p}_2}{\hat{p}_2(n_{21}+n_{22})}} \tag{24}$$

Calculations

Lambda

Rows dependent: $$\lambda = \frac{\left(\sum_{j=1}^{c} O_{*j} - R_*\right)}{n - R_*} \qquad (25)$$

Columns dependent: $$\lambda = \frac{\left(\sum_{i=1}^{r} O_{i*} - C_*\right)}{n - C_*} \qquad (26)$$

Symmetric: $$\lambda = \frac{\left(\sum_{j=1}^{c} O_{*j} + \sum_{i=1}^{r} O_{i*} - R_* - C_*\right)}{2n - R_* - C_*} \qquad (27)$$

where

O_{i*} = largest frequency in row i

O_{*j} = largest frequency in column j

R_* = largest row total

C_* = largest column total

Uncertainty Coefficient

Rows dependent: $$U = \frac{U(X) + U(Y) - U(XY)}{U(X)} \qquad (28)$$

Columns dependent: $$U = \frac{U(X) + U(Y) - U(XY)}{U(Y)} \qquad (29)$$

Symmetric: $$U = 2\left[\frac{U(X) + U(Y) - U(XY)}{U(X) + U(Y)}\right] \qquad (30)$$

where

$$U(X) = -\sum_{i=1}^{r} \frac{R_i}{n} \log\left(\frac{R_i}{n}\right) \qquad (31)$$

$$U(Y) = -\sum_{j=1}^{c} \frac{C_j}{n} \log\left(\frac{C_j}{n}\right) \qquad (32)$$

$$U(XY) = -\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{O_{ij}}{n}\log\left(\frac{O_{ij}}{n}\right) \text{ for } O_{ij} > 0 \qquad (33)$$

Somer's D

Letting P be the number of concordant pairs and Q the number of discordant pairs:

Rows dependent: $$D = \frac{2(P-Q)}{n^2 - \sum_{j=1}^{c} C_j^2} \qquad (34)$$

Columns dependent: $$D = \frac{2(P-Q)}{n^2 - \sum_{i=1}^{r} R_i^2} \qquad (35)$$

Symmetric: $$D = \frac{4(P-Q)}{\left(n^2 - \sum_{i=1}^{r} R_i^2\right) + \left(n^2 - \sum_{j=1}^{c} C_j^2\right)} \qquad (36)$$

Eta

$$\eta = \sqrt{1 - \frac{SS_W}{SS_T}} \qquad (37)$$

where SS_T is the total corrected sum of squares for Y when each observation i, i=1,2,…n, is assigned a value Y_i, and $\mathrm{SS_W}$ is the sum of squares within categories of the independent variable.

Contingency Coefficient

$$C = \sqrt{\frac{\chi^2}{\chi^2 + n}} \qquad (38)$$

Cramer's V

For a 2-by-2 table:

$$V = \sqrt{\frac{\chi^2}{n}} \qquad (39)$$

using the uncorrected value of χ^2. For other tables:

$$V = \sqrt{\frac{\chi^2}{n[\min(r,c)-1]}} \tag{40}$$

Conditional Gamma

$$\gamma = \frac{(P-Q)}{P+Q} \tag{41}$$

Pearson's R

The P-value is calculated by computing the probability of exceeding

$$t = \left| R\sqrt{\frac{n-2}{1-R^2}} \right| \tag{42}$$

using Student's t distribution with *n* - 2 degrees of freedom.

Kendall's Tau b

$$\tau = \frac{2(P-Q)}{\sqrt{\left(n^2 - \sum_{i=1}^{r} R_i^2\right)\left(n^2 - \sum_{j=1}^{c} C_j^2\right)}} \tag{43}$$

If *n* >10, a P-value is computed by calculating the two-tailed probability of a standard normal random variable exceeding

$$z = \frac{(P-Q)}{\sqrt{d}} \tag{44}$$

where

$$d = \frac{1}{18}\left[\left(n^2-n\right)\left(2n+5\right) - \sum_{j=1}^{c}\left(C_j^2 - C_j\right)\left(2C_j+5\right) - \sum_{i=1}^{r}\left(R_i^2 - R_i\right)\left(2R_i+5\right)\right]$$

$$+ \frac{\left[\sum_{j=1}^{c}\left(C_j^2 - C_j\right)\left(C_j-2\right)\right]\left[\sum_{i=1}^{r}\left(R_i^2 - R_i\right)\left(R_i-2\right)\right]}{9\left(n^2-n\right)\left(n-2\right)}$$

$$+\frac{\left[\sum_{j=1}^{c}\left(C_j^2-C_j\right)\sum_{i=1}^{r}\left(R_i^2-R_i\right)\right]}{2\left(n^2-n\right)} \tag{45}$$

Kendall's Tau c

$$\tau=\frac{2\min(r,c)(P-Q)}{\left[\min(r,c)-1\right]n^2} \tag{46}$$