CONTINGENCY TABLE

Supuestos
• Los datos estan seleccionados al azar(o son al menos representativos de) una gran población.
•Cada sujeto fue seleccionado independientemente
. No hay apareamiento ("matching")

He aquí un ejemplo
Subject with HIV infection were divided into two groups and given placebo or AZT. The result was recorded as disease progression or no progression (from New Eng. J. Med. 329:297-303, 1993).

Disease	progression	No progression	Total
AZT	76	399	475
Placebo	129	332	461
Total	205	731	936

Note these features of a contingency table, also called a cross tabulation table.

• The values in the table are the number of subjects. Table with averages, percentages or rates are not contingency tables.
• The two columns are mutually exclusive. A subject can be in one or the other, but not both.
• The two rows are also mutually exclusive.

Contingency tables are used to tabulate the results of four kinds of experiments:

• Experimental. Subjects are randomly divided into groups that become the rows of the table. Each group is given an alternative treatment. The columns of the table represent alternative outcomes. The example above is an experimental study.
• Prospective study. Subjects are recruited based on exposure to risk factors, and then are followed to observe the outcome. The experimenter assigns rows based on selection criteria and assigns columns based on the outcome [(n) the final result of sth, especially when none knows what it will happen until it actually happen).
• Cross-sectional study. A large group of subjects is recruited without regard to exposure or outcome. The investigators then divide the subjects into groups based on exposure to a risk factor and compare the prevalence of disease in the groups.
• Retrospective or case-control studies. Subjects are recruited based on the outcome (different columns). The investigators then look back at exposure to possible risk factors (or protective factors) which are shown as different rows. The investigator asks whether there is an association of the risk factor with the disease.

Contingency tables can tabulate the results of some basic science experiments. The rows represent alternative treatments, and the columns tabulate alternative outcomes.

Entering contingency tables into Prism
Prism is not able to cross-tabulate data to create a contingency table. You must enter the contingency table itself.

P value calculation
Prism offers three methods for calculating a P value from tables with two rows and two columns.
• Fisher's exact test. This test calculates a P value that is exactly correct.

• The chi-square test. The only advantage of the chi-square test is that it is easier to calculate by hand, and so is better known.

• Yates' continuity corrected chi-square test. This correction is designed to make the approximate results from a chi-square test more accurate with small samples. Statisticians disagree about whether to use it. If you always use Fisher's exact test, you don't have to think about Yates' correction.

If you have very large numbers (thousands) in your table, Prism will automatically perform the chi-square test even if you select Fisher's test, because the Fisher's test calculations are lengthy with large samples. With large samples, the chi-square test is very accurate and the Yates' continuity correction has negligible effect.

If your table has more than two rows or two columns, Prism always calculates the chi-square test. You have no choices. Extensions to Fisher's exact test have been developed for larger tables, but Prism doesn't offer them.

If your table has two columns and three or more rows, you may also select the chi-square test for trend. This calculation tests whether there is a linear trend between row number and the fraction of subjects in the left column. It only makes sense when the rows are arranged in a natural order (i.e. age, dose, time) and are equally spaced.

The P value answers this question: If there is no association between row and column variables in the overall population, what is the chance of observing so much association in an experiment of this size.

In the example, the P value from Fisher's test is less than 0.0001. If there was no association between AZT treatment and the fraction of patients whose disease progresses, there is less than a 0.01% of observing such a strong association (or stronger) in a study of this size.

Relative risk, Odds ratio, and Difference between proportions

There are three ways to quantify the results of a contingency table with two rows and two columns.

• Difference between proportions. In the example, disease progressed in 28% of the placebo-treated patients and in 16% of the AZT-treated subjects. The difference is 28% - 16% = 12%.

• Relative risk. The ratio is 16%/28%=0.57. A subject treated with AZT has 57% the chance of disease progression as a subject treated with placebo. The word "risk" is not always appropriate. Think of the relative risk as being simply the ratio of proportions.

• Odds ratio. This is a more difficult concept. There isn't much point in calculating an odds ratio for experimental or prospective studies. When analyzing case-control retrospective studies, however, you cannot meaningfully calculate the difference between proportions or the relative risk. The odds ratio is used to summarize the results of these kinds of studies. See a biostatistics or epidemiology book for details.

If your data came from an experimental, cross-sectional or prospective study, you can pick any or all of these. The odds ratio is rarely used with these kind of data.

If your data came from a retrospective case-control study, you should pick only the odds ratio.

Confidence interval
You can calculate the difference between proportions, relative risk, or odds ratio for the particular subjects you happened to sample. You can be 95% sure that the confidence interval contains the true value in the overall population.

Prism approximates the confidence interval of the relative risk and odds ratio using the methods of Woolf and Katz respectively. If any of the four values in the contingency table are zero, Prism adds 0.5 to all values before calculating the relative risk or odds ratio and the confidence intervals.