

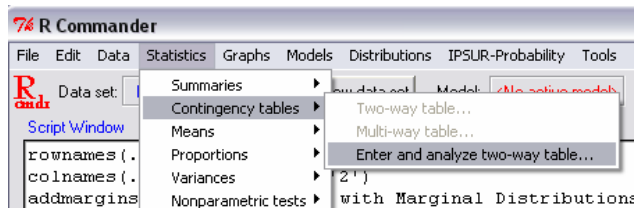
Contingency Tables

Contingency tables are used for organizing categorical variables and testing hypothesis with the chi-squared test for independence. The following example shows how to enter contingency table frequency information into R and testing for statistical significance.

Example: The following table shows results of HIV testing. What is the probability that you can test positive for HIV, but not actually have the disease; also, what is the probability that you test positive and are infected? Using a 0.05 level of significance, is there a relationship between test results and presence of the disease?

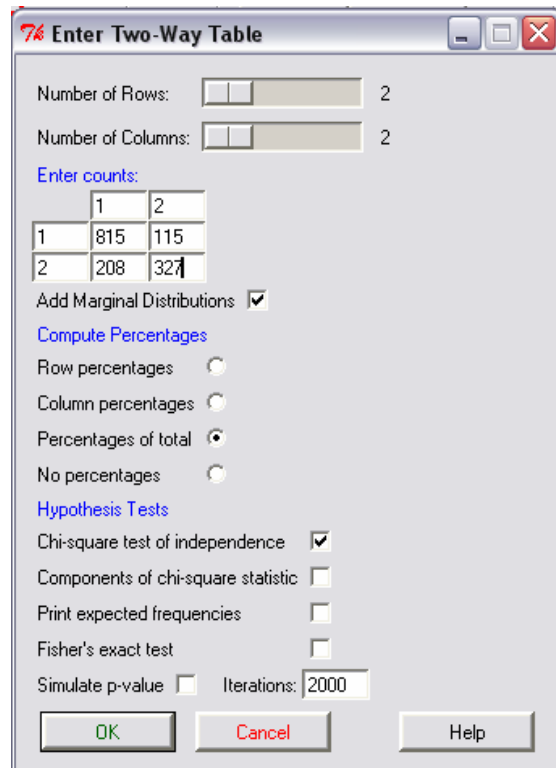
	Present D^+	Absent D^-	Total
Positive T^+	815	115	930
Negative T^-	208	327	535
Total	1023	442	1465

- 1) This can be done quickly by using the contingency table function in IPSUR – **Statistics / Contingency tables / Enter and analyze two-way table...**



Enter the data into the table as it appears in the example, and select “Percentage of total” for **Computer Percentages** – this will compute overall probabilities for each of the four cases.

For the **Hypothesis Tests** section, check next to “Chi-square test of independence.” This will test whether there is correlation between the two variables (the null hypothesis is that there is **no relationship** between the two).



2) The below results frequency and relative frequency distributions are shown in the output window:

	1	2	Sum
1	815	115	930
2	208	327	535
Sum	1023	442	1465

	1	2	Total
1	55.6	7.8	63.5
2	14.2	22.3	36.5
Total	69.8	30.2	100.0

Pearson's Chi-squared test
data: .Table
X-squared = 383.2047, df = 1, p-value < 2.2e-16

Interpret:

- 3) To determine if there is a relationship between the test results and disease presence, use the bottom portion of the test output (**Pearson's Chi-squared test**). Since:
p-value = 2.2e-16
and we are using a 0.05 level of significance, the null hypothesis (there is no relationship between the two variables) is rejected.