Contingency table

Menu:	QCExpert	Testing	Contingency tables

This module tests the hypothesis about independence of two categorical variables *A*, *B* based on experimentally observed occurrence of combinations of particular levels of *A*, *B*. Number of levels of *A* is denoted *r*, number of levels of *B* is denoted *c*. The data are organized in the table of frequencies. Only the thick bordered box needs to be specified, the totals are calculated automatically.

	B levels			
A levels	B_1	…	B_c	Total
A_1	n_{11}	…	n_{1c}	$n_{1.}$
…	…	…	…	…
A_r	n_{r1}		n_{rc}	$n_{r.}$
Total	$n_{.1}$	…	$n_{.c}$	n

From this table we compute a test statistic *C*, which has a χ^2 distribution provided that *A* and *B* are independent.

$$C = n\sum_{i=1}^{r}\sum_{j=1}^{c}\frac{n_{ij}^2}{n_{i.}n_{.j}} - n$$

The test statistic *C* is then compared to the critical quantile of the χ^2 distribution $\chi^2_{(r\text{-}1).(c\text{-}1)}(1-\alpha)$. If $C > \chi^2$, the hypothesis H_0 about independence of *A* and *B* is rejected. Dependence of *A* and *B* means that probabilities and numbers of occurrences in columns of the frequency table for at least one level of A are affected by levels of B. If A and B are independent, then the probability p_{ij} of observing the event $A_j \wedge B_i$ is equal of the product of marginal probabilities $p_{i.}\ p_{.j}$. These probabilities may be computed from the total frequencies, $p_{i.} = n_{i.}/n$, $p_{.j} = n_{.j}/n$. The marginal probabilities sum to unity $\Sigma\, p_{i.} = \Sigma\, p_{.j} = 1$.

Data and parameters

The module expects data in form of frequencies shown in the thick-bordered part of the above table. The column names are taken from the header of the data table, row names may be in any column (preferably the first one) of the data table. In the dialog window, see Fig. 1, the columns of frequencies and column of row names. You may modify the task name and the significance level (usual value is $\alpha = 0.05$). It is recommended that $n_{ij} \geq 5$, to ensure reliability of the test. After having selected data, run the test by clicking on the OK button. Results are written into the Protocol window and the dialog window will close.

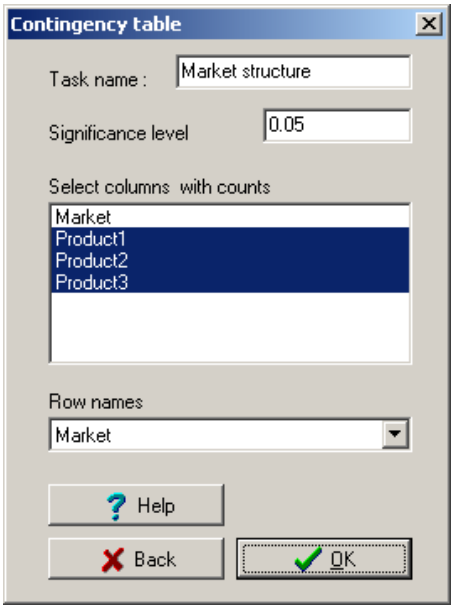

Fig. 1 Dialog panel for the contingency table

Example

The the following table we have volumes of products of three types 1, 2 3 sold to 4 different markets, Europe, Asia, USA, and Africa. We want to test if there is a difference in structure between products or between markets. Practically, we test if the proportion between products is the same for any market and simultaneously if the proportion between markets is the same for any product.

Market	Product1	Product2	Product3
Europe	114	25	301
Asia	68	20	225
USA	131	30	240
Africa	57	23	159

In the dialog window (Fig. 1) type the significance level (say, 0.05), select the columns with frequencies *Product1*, *Product2*, *Product3* and the column with market names, *Market*. Press *OK* to run the analysis.

Conclusion	Independence of variables is rejected
Significance level	0.05
Degrees of freedom	6
Chi2 statistic	17.11555756
Critical value	12.59158724
p-value	0.008867805134

Chi-squared statistic is 17.1, which is greater than the critical value 12.59, so the independence of the variables is rejected. That means, that the structure of the market is dependent on locality and product.

Protocol

Analysis of contingency table	Name of the module

Task name	Task name from the dialog window
Table of counts	
„row name“	Input frequencies (or counts) from data table. In the last column are the marginal frequencies (or sums of rows).
Theoretical	In brackets there are the theoretical frequencies based on assumption of independence.
Total	The column marginal frequencies, or sums of columns.
Table of ratios and probabilities	
„row name“	Empirical probabilities calculated from the given frequencies as $p_{ij} = n_{ij}/n$, in the last column are the marginal probabilities, or sums of rows.
Theoretical	Theoretical probabilities based on assumption of independence. calculated as $n_{i.}\, n_{.j}/n^2$, in the last column are theoretical row marginal probabilities.
Total	The column marginal frequencies, or sums of empirical probabilities in columns.
Conclusion	Verbal conclusion of the test
Significance level	Significance level of the test α ($1 - \alpha$ is called confidence). Recommended is $\alpha = 0.05$.
Degrees of freedom	$(r - 1)(c - 1)$.
Chi2 statistic	The value of the sample statistic C calculated from data
Critical value	Theoretical quantile of the χ^2 distribution $H0$ $\chi^2_{(r\text{-}1).(c\text{-}1)}(1-\alpha)$.
p-value	Calculated p-value of the test