contingency table
A *multivariate frequency table*. The most common variant of the contingency table is the *bivariate* version, which lists the categories of one variable in the top row and those of the other in the left column. In addition it lists the marginal total of each category at the opposite end of the corresponding row/column and a grand total in the lower right corner. See figure CTT.

	X_1	$\cdots$	X_k	
Y_1	X_1Y_1		X_kY_1	$\sum_{Y_1}$
$\cdots$				
Y_h	X_1Y_h		X_kY_h	$\sum_{Y_h}$
	$\sum_{X_1}$	$\cdots$	$\sum_{X_k}$	$\sum_{X,Y}$

Figure **CTT:** contingency table of the size $k \times h$; the variable X has k categories and Y has h categories

The fields of the table list the *intersections* of the categories. For example, the field X_1Y_1 lists the number of items that belong to category X_1 *and* to category Y_1. The grand total $\sum_{X,Y}$ of the table is the sum of either $\sum_X$ or $\sum_Y$, i.e. the sum of all fields in the table.

Various *probabilities* can be derived from the table, for instance:

$$P(X_c) = \sum_{X_c} / \sum_{X,Y}$$

$$P(X_c \cap Y_d) = P(X_cY_d) = \frac{X_cY_d}{\sum_{X,Y}}$$

$$P(X_cY_d \mid X_c) = \frac{X_cY_d}{\sum_{X_c}}$$

The *expectation* of the contingency table is a table in which the variables are *independent*. Its fields are calculated as follows:

$$E(X_iY_j) = \frac{\sum_{X_i}\sum_{Y_j}}{\sum_{X,Y}}$$

Because the table rows and columns implement *conditional probabilities* and the expectation assumes that the variables X and Y are *independent*, the expectation will be close to the corresponding value in the table if, and only if, X and Y are weakly correlated ($\rightarrow$ *correlation*).

The accumulated sum of deviations from the expectation can be used to perform a χ^2 test ($\rightarrow$ *chi-square distribution*, *hypothesis test*) with $\nu = (k-1)\cdot(h-1)$ *degrees of freedom*. This test estimates ($\rightarrow$ *estimate*) the probability of the variables X and Y being independent, i.e. it is a "test of independence". The χ^2 test *score* is calculated as follows:

$$\chi^2 = \sum_{i=1}^{k}\sum_{j=1}^{h}\frac{[E(X_iY_j) - X_iY_j]^2}{E(X_iY_j)}$$

For this method to work, no field in the table should have a value less than 5.

Example: The following table lists people by behavioral priority (X) and gender (Y). Expectations are given in parentheses.

		Priority (X)		
		Assertiveness	Empathy	Total
Gender (Y)	Male	87 (82.3)	13 (17.7)	100
	Female	57 (61.7)	18 (13.3)	75
	Total	144	31	175

Given this table, for instance, the probability of

- being a woman is $P(Y = Female) = 75/175 \approx 0.43$
- preferring empathy given being male is $P(X = Empathy \mid Y = Male) = 13/100 \approx 0.13$,

The χ^2 score of the table is $\chi^2 \approx 3.59$ with one degree of freedom, giving a probability of $1 - F_{\chi^2(1)}(3.59) \approx 0.06$ for the variables being independent while the *observations* in the table arose by chance (F_{χ^2} being the *cumulative distribution function* of the χ^2-distribution).