χ^2 test of independence (Chi-square test of/for independence)

Definition: We use χ^2 test of independence if we want to **analyze relationships between two qualitative statistical attributes**. This test is widely applied in market research since the base data for the test usually come from questionnaires.

Examples:
Analyze the relationships between gender and monthly income.
Analyze the relationships betwwen product quality and supplier.
Analyze the relationships between the preference of certain goods and country.
etc.

The procedure of the χ^2 test of independence lies in **comparing empirical (observed) frequencies (real data) and theoretical (expected) frequencies.**

χ^2 test of independence is another important application of assymetrical χ^2 distribution so the test statistic is as follows:

$$\chi^2 = \sum_{i=1}^{m}\sum_{j=1}^{k}\frac{((a_i b_j) - (a_i b_j)_0)^2}{(a_i b_j)_0}$$

where,
$(a_i b_j)$ are empirical (observed) frequencies,
$(a_i b_j)_0$ are theoretical (expected) frequencies,
m is the number of rows of the appropriate contingency table,
k is the number of columns of the appropriate contingency table.

Expected frequencies are calculated using a formula:

$$\left(a_i b_j\right)_0 = \frac{a_i \cdot b_j}{n}$$

where,
a_i is the sum of values (observed frequencies) in each row in the contingency table,
b_j is the sum of values (observed frequencies) in each column in the contingency table,
n is the number of observations.

Ctitical value should be calculated by using CHINV function in MS Excel (alpha; (m-1)(k-1)) and used to state wheter we accept the null hypothesis (i.e. there is no relationship between the two qualitative statistical attributes) or reject the null hypothesis and accept the alternate hypothesis (there is a relationship between the two qualitative statistical attributes).

Note: If we detect a relationship between the two qualitative statistical attributes, we might want to know how strong this relationship is. In that case, we can calculate several coeffcients:

Cramer's V:

$$V = \sqrt{\frac{\chi^2}{n(\min((m,k)-1)}}$$

Pearson's C:

$$C = \sqrt{\frac{\chi^2}{n+\chi^2}}$$

or Chuprov's T^2:

$$T^2 = \frac{\chi^2}{n(\sqrt{(m-1)(k-1)}}$$

where,
χ^2 is the test statistic of the test of independence,
n is the number of observations,
m is the number of rows of the appropriate contingency table,
k is the number of columns of the appropriate contingency table.

The scale for determining the strength of the relationships is as follows:
0-0,3->>weak relationship
0,3-0,5->>medium relationship
0,5-0,7(0,8)->>strong relationship
more than 0,8(0,9)->>very strong relationship

Note: If we **analyze relationships between two quantitative statistical attributes**, we shall perform **regression and correlation analysis**.