What is the Chi-Square Test of Independence?

The Chi-Square Test of Independence is also known as Pearson's Chi-Square, Chi-Squared, or $\chi^2$. $\chi$ is the Greek letter Chi. The Chi-Square Test has two major fields of application: 1) goodness of fit test and 2) test of independence.

Firstly, the Chi-Square Test can test whether the distribution of a variable in a sample approximates an assumed theoretical distribution (e.g., normal distribution, Beta). [Please note that the Kolmogorov-Smirnov test is another test for the goodness of fit. The Kolmogorov-Smirnov test has a higher power, but can only be applied to continuous-level variables.]

Secondly, the Chi-Square Test can be used to test of independence between two variables. That means that it tests whether one variable is independent from another one. In other words, it tests whether or not a statistically significant relationship exists between a dependent and an independent variable. When used as test of independence, the Chi-Square Test is applied to a contingency table, or cross tabulation (sometimes called crosstabs for short).

Typical questions answered with the Chi-Square Test of Independence are as follows:

- **Medicine** - Are children more likely to get infected with virus A than adults?
- **Sociology** - Is there a difference between the marital status of men and woman in their early 30s?
- **Management** - Is customer segment A more likely to make an online purchase than segment B?
- **Economy** - Do white-collar employees have a brighter economical outlook than blue-collar workers?

As we can see from these questions and the decision tree, the Chi-Square Test of Independence works with nominal scales for both the dependent and independent variables. These example questions ask for answer choices on a nominal scale or a tick mark in a distinct category (e.g., male/female, infected/not infected, buy online/do not buy online).

In more academic terms, most quantities that are measured can be proven to have a distribution that approximates a Chi-Square distribution. Pearson's Chi Square Test of Independence is an approximate test. This means that the assumptions for the distribution of a variable are only approximately Chi-Square. This approximation improves with large sample sizes. However, it poses a problem with small
sample sizes, for which a typical cut-off point is a cell size below five expected occurrences.

Taking this into consideration, Fisher developed an exact test for contingency tables with small samples. Exact tests do not approximate a theoretical distribution, as in this case Chi-Square distribution. Fisher's exact test calculates all needed information from the sample using a hypergeocontinuous-level distribution.

What does this mean? Because it is an exact test, a significance value p calculated with Fisher's Exact Test will be correct; i.e., when ? =0.01 the test (in the long run) will actually reject a true null hypothesis in 1% of all tests conducted. For an approximate test such as Pearson's Chi-Square Test of Independence this is only asymptotically the case. Therefore the exact test has exactly the Type I Error (?-Error, false positives) it calculates as ?-value.

When applied to a research problem, however, this difference might simply have a smaller impact on the results. The rule of thumb is to use exact tests with sample sizes less than ten. Also both Fisher's exact test and Pearson's Chi-Square Test of Independence can be easily calculated with statistical software such as SPSS.

The Chi-Square Test of Independence is the simplest test to prove a causal relationship between an independent and one or more dependent variables. As the decision-tree for tests of independence shows, the Chi-Square Test can always be used.

**Chi-Square Test of Independence in SPSS**

In reference to our education example we want to find out whether or not there is a gender difference when we look at the results (pass or fail) of the exam.

The Chi-Square Test of Independence can be found in Analyze/Descriptive Statistics/Crosstabs…

This menu entry opens the crosstabs menu. Crosstabs is short for cross tabulation, which is sometimes referred to as contingency tables.

The first step is to add the variables to rows and columns by simply clicking on the variable name in the left list and adding it with a click on the arrow to either the row list or the column list.

The button Exact… opens the dialog for the Exact Tests. Exact tests are needed with small cell sizes below ten respondents per cell. SPSS has the choice between Monte-Carlo simulation and Fisher's Exact Test. Since our cells have a population greater or equal than ten we stick to the Asymptotic Test that is Pearson's Chi-Square Test of Independence.

The button Statistics… opens the dialog for the additional statics we want SPSS to compute. Since we want to run the Chi-Square Test of Independence we need to tick Chi-Square. We also want to include the contingency coefficient and the correlations which are the tests of interdependence between our two
variables.

The next step is to click on the button *Cells...* This brings up the dialog to specify the information each cell should contain. Per default, only the Observed Counts are selected; this would create a simple contingency table of our sample. However the output of the test, the directionality of the correlation, and the dependence between the variables are interpreted with greater ease when we look at the differences between observed and expected counts and percentages.

*Output, syntax, and interpretation can be found in our downloadable manual: Statistical Analysis: A Manual on Dissertation Statistics in SPSS (included in our member resources).* [Click here to download](#).