

# Cell Chi-Square Test

The cell chi-square test looks at each table cell and tests whether it is significantly different from its expected value in the overall table. For example, if it is thought that variations in political opinions might depend on the respondent's age, this test can be used to detect which cells contribute significantly to that dependence.

Unlike the chi-square test, which is carried out on a whole set of rows and columns, the cell chi-square test is carried out independently on each table cell. This is done by treating each cell as belonging to a two-by-two table, known as a contingency table, as follows:

	Base	In column	Not in column
Base	B	C	B-C
In row	R	V	R-V
Not in row	B-R	C-V	B-R-C+V

For each cell, the values B, C, R and V are taken from the table. The other values are calculated from these values.

The formula applied to this two-by-two table is the standard Pearson chi-square formula, with the Yates' correction for small samples when relevant as the p value associated with the Pearson chi-square test can be distorted if any cells in the table have very low expected counts (below 5).

Note: Although the significance level of the cell chi-square test is accurate for any given cell, the cell tests cannot be used instead of the chi-square test carried out on the overall table. Their purpose is simply to point to the parts of the table where dependencies between row and column categories may exist.

# Details and Restrictions of the Cell Chi-Square Test

## Cell chi-square validity

The following conditions must be met in order for the cell chi-square test to be valid:

- $0 < \text{column base} < \text{overall base}$
- $0 < \text{row base} < \text{overall base}$

This means the test is not valid for a cell where all or no respondents belong in the whole column or row for the cell.

## Significance level

You can select an explicit significance level of 10%, 5%, 2.5% 0.5% or 0.1%. Alternatively, an automatic selection can be requested, which sets the significance level to one of the above depending on the overall base as follows:

Level Applied	For overall base value "f" of
10%	$f \leq 300$
5%	$300 < f \leq 1000$
2.5%	$1000 < f \leq 4000$
0.5%	$4000 < f \leq 20000$
0.1%	$f > 20000$

## Cell chi-square usage

The cell chi-square test is not suitable for all tables. When you request the test on a table that is structurally unsuitable, IBM® SPSS® Data Collection Base Professional simply skips the test, leaves the cell chi-square and p value rows blank, and writes an explanation to the diagnostics data. It is up to you to make sure that the data in the table is generally suitable for testing, that the sample size is suitable, etc.

displays a message if you define a cell chi-square test on an unsuitable table or if you change a table that has a cell chi-square test defined so that it is no longer suitable for the test. When this happens, you can either adjust the table so that it conforms to the restrictions, or you can remove the test from the table.

Hierarchical data. This test is unsuitable for running on lower level data when you are working with hierarchical data a hierarchical view of the data. See the topic [Hierarchical Data](#) for more information.

Rows and columns For the cell chi-square test, the variables on the side and top axes must have at least two categories. does not perform the test on rows and columns in which all of the values are zero or on rows and columns that are formed from non-category elements, such as bases and means.

Nested and concatenated tables. When variables are concatenated in , a base is inserted at the start of each variable (variables can, in general, be different). With nesting, a base is needed for each "nest section" (they are almost always different). When a cell has the cell chi-square test applied to it, the base row and column used will be the nearest base in each dimension. This is consistent with the base used for percentages and other statistics in . When applied to the automatic selection option for significance level, this means that different parts of one table could have different significance levels applied.

Excluded Elements. The IncludeInBase=False property has no effect on the cell chi-square test. If a cell chi-square test is carried out on a table containing categories that are excluded from the base using IncludeInBase=False, the calculation includes rows and columns corresponding to the excluded categories.

# Statistical Formula for the Cell Chi-Square Test

For each cell, a two-by-two contingency table is established, with the actual cell from the cross table as the upper left cell, and the other cells derived by all three possible negations of the cross table cell:

	Total	Column	NOT column
Total	f	c	f-c
Row	r	x	r-x
NOT row	f-r	c-x	f-r-c+x

The two-by-two contingency table has one degree of freedom; the other three cells are determined when one of the four cells is known ( $x$ ).

The following table shows the notation used in this topic.

Notation	Description
$O_{ij}$	The observed values.
$E_{ij}$	The expected values given the null hypothesis.

The first index points to the two-by-two table rows and the second index to the columns:

$$O_{1,1} = x$$

$$O_{1,2} = r - x$$

$$O_{2,1} = c - x$$

$$O_{2,2} = f - r - c + x$$

$$E_{1,1} = r \cdot \frac{c}{f}$$

$$E_{1,2} = \frac{r \cdot (f - c)}{f}$$

$$E_{2,1} = \frac{(f - r) \cdot c}{f}$$

$$E_{2,2} = \frac{(f - r) \cdot (f - c)}{f}$$

## Null hypothesis

The deviation between the observed values  $O_{ij}$  and the expected values  $E_{ij}$  are not significant and are at random. That is the variables Row/NOT row and Column/NOT column are independent.

## $\chi^2$ statistic

$$\chi^2 = \sum_{i=1}^2 \sum_{j=1}^2 \frac{(O_{ij} - E_{ij})^2}{E_{ij}} = \frac{f \cdot (f \cdot x - c \cdot r)^2}{c \cdot r \cdot (f - c) \cdot (f - r)}$$

The null hypothesis is rejected if the p value for the  $\chi^2$  statistic is smaller than the specified significance level. The p value is calculated from the chi-square distribution with 1 degree of freedom.