Companion formulae version 2020.1.5

This document describes the calculations for significance testing in tables and drill down.

Weighted data

Weighting the data affects the calculations. There are two sorts of weighting in CL:

- Respondent weighting, for example target or rim weighting. Each respondent is given a "weight" to adjust the figures to some known population or to compensate for sample imbalances.

- Quantity weighting. A value is used to weight the table so that the figures in the table are not respondents but represent the total of some value from the questionnaire, for example total area or volume.

Both types of weighting can be applied on the same table.

The calculations generally use the weighted figures.

Where appropriate the un-weighted or effective sample size is used. This is shown as e in the formulae.

If the **recommended format ESS** is used then the effective sample size used in calculations is:

$$e = \frac{(\sum w)^2}{\sum w^2}$$

Where w is the weight applied to each respondent included in the calculation.

If format NESS is used or the data is not weighted, then e is the un-weighted number of respondents in the calculation.

Significance markers

When marking significance levels under or against the figures the markers are:

- Plus or minus signs (asterisks in older versions). These show a comparison with the total column. The two columns that are compared are the test column and the total column **minus the test column**. The test for a column is therefore a test against **the rest of data**. The markers are used when the column is **higher (plus sign) or lower (minus sign)** than the rest of the data. One plus or minus sign shows the lower significance level (format SLA) and two plus or minus signs show the higher level (format SLB).

- Letters. The column identifiers are shown under the column labels. The columns are normally compared one at a time with all the other columns under the same heading (format SHG1 or SHG11). The letter markers are only placed on the column with the **higher** value. If all the column heading letters are lower case, then a lower case letter shows the lower significance level (format SLA) and an upper case letter shows the higher level (format SLB).

Important: the total column should not be allocated a letter unless the overlap formula is used. You should use SHG11 or SHG12 instead.

Additional higher levels of significance (format SLC) can be shown with an additional plus or minus after the other markers.

Mean score or average (format AVG)

Where:

> x is each value or score value.

$$\bar{x} = \frac{\sum wx}{\sum w}$$

Variance (format VAR)

This is calculated as:

$$S^2 = \frac{\sum wx^2 - \frac{(\sum wx)^2}{\sum w}}{\sum w - 1}$$

Standard deviation (format SDV)

$$S = \sqrt{S^2}$$

Standard error (format SER)

$$\text{SER} = \frac{S}{\sqrt{e}}$$

Error variance (format EVR)

$$\text{EVR} = \frac{S^2}{e}$$

Mean over standard error (format MSE)

Used when the expected mean is zero:

$$\text{MSE} = \frac{\bar{x}}{\text{ser}}$$

Means comparison t-test (TTV1, default)

The pooled variance for formats TTV1 and TTV3 is:

$$S^2 = \frac{\left(S_1^2(\sum w_1 - 1)\right) + (S_2^2(\sum w_2 - 1))}{w_1 + w_2 - \frac{w_1}{e_1} - \frac{w_2}{e_2}}$$

Note that:

$$\frac{w}{e} \equiv \frac{\sum w^2}{\sum w}$$

Which is 1 for unweighted data.

When comparing two column means with format TTV1;

$$t = \frac{\overline{x_1} - \overline{x_2}}{S\sqrt{\left(\frac{1}{e_1} + \frac{1}{e_2}\right)}}$$

Means comparison t-test (TTV2)

The variance for format TTV2 is:

$$S^2 = S_1^2 + S_2^2$$

When comparing two column means with format TTV2:

$$t = \frac{\overline{x_1} - \overline{x_2}}{S\sqrt{\left(\dfrac{1}{e_1} + \dfrac{1}{e_2}\right)}}$$

Means comparison t-test overlap formula (TTV3)

When comparing two column means with format TTV3 (overlap formula):

$$t = \frac{\overline{x_1} - \overline{x_2}}{S\sqrt{\left(\dfrac{1}{e_1} + \dfrac{1}{e_2} - \dfrac{2e_o}{e_1 e_2}\right)}}$$

Where e_o is the unweighted or effective sample size of the overlapping records and S is described under TTV1.

Proportions comparisons

When comparing two column percentages in the same row.

p_t is the combined proportion from both columns added together.

Proportions comparison Z test (SIG1)

When comparing proportions (percentages) with SIG1:

$$z = \frac{p_1 - p_2}{\sqrt{p_t(1 - p_t)\left(\dfrac{1}{e_1} + \dfrac{1}{e_2}\right)}}$$

Proportions comparison Z test (SIG2)

$$z = \frac{p_1 - p_2}{\sqrt{\dfrac{p_1(1 - p_1)}{e_1} + \dfrac{p_2(1 - p_2)}{e_2}}}$$

Proportions comparison t test (SIG3, recommended)

$$t = \frac{p_1 - p_2}{\sqrt{S^2 \left(\frac{1}{e_1} + \frac{1}{e_2}\right)}}$$

Where S^2 is:

$$\frac{p_t(1 - p_t)}{\left(1 - \frac{1}{e_1 + e_2}\right)}$$

Proportions comparison t test with continuity correction (SIG4)

$$t = \frac{p_1 - p_2 - cc}{\sqrt{S^2 \left(\frac{1}{e_1} + \frac{1}{e_2}\right)}}$$

Where S^2 is described under SIG3 and CC is:

$$\frac{\frac{1}{e_1} + \frac{1}{e_2}}{2}$$

Proportions comparison t test overlap formula (SIG5)

$$t = \frac{p_1 - p_2}{\sqrt{S^2 \left(\frac{1}{e_1} + \frac{1}{e_2} - \frac{2e_o}{e_1 e_2}\right)}}$$

Where S^2 is described under SIG3 and e_o is the unweighted or effective sample size of the overlapping records.

Proportions comparison t test overlap with continuity correction (SIG6)

$$t = \frac{p_1 - p_2 - cc}{\sqrt{S^2 \left(\frac{1}{e_1} + \frac{1}{e_2} - \frac{2e_o}{e_1 e_2}\right)}}$$

Where S^2 is described under SIG3, and e_o is the unweighted or effective sample size of the overlapping records, and CC is described under SIG4.

Degrees of freedom for TTV1, SIG3 and SIG5

$$\mathrm{DOF} = e_1 + e_2 - 2$$

Degrees of freedom for TTV3, SIG4 and SIG6

$$\mathrm{DOF} = e_1 + e_2 - e_o - 2$$