askia analyse
Significancy tests User guide

Cpc – Analyse 5.3.2.X

The aim of this document is to help you step by step to apply the significancy tests in askiaanalyse 5.3.2.X

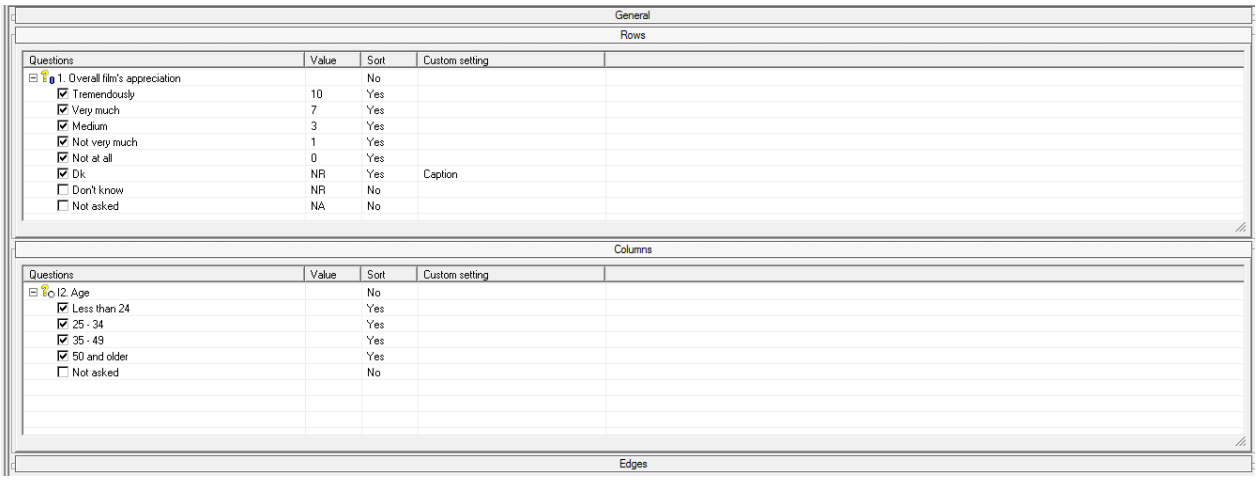
Analyse provides you Significancy tests to compare proportions and means.

Contents

1. Proportions comparison

To compare proportions between 2 or more samples, you need to run a crossed table.

You can select any closed question (single or multicoded or scale responses).
Insert a closed question in row **and** in column in a tab definition



And then select the significancy test.

1.1.Significance

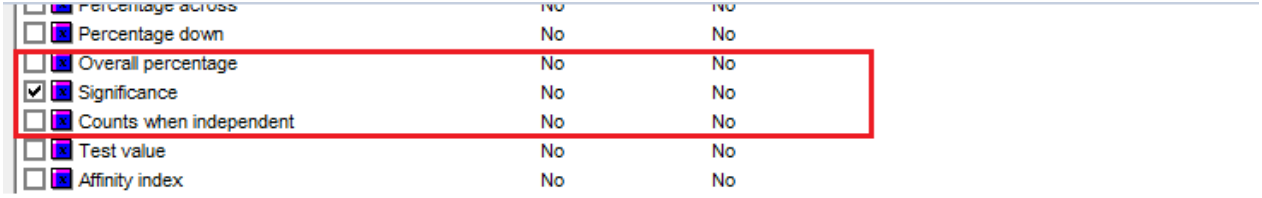
The significance test (for *closed* questions) is used in order to see if the proportion observed (p_i) in the sample *i* (n_i) is different from the p_N observed in the sample *N*.

We will display "–" or "+" signs to indicate where the difference is significative (see the figure below):

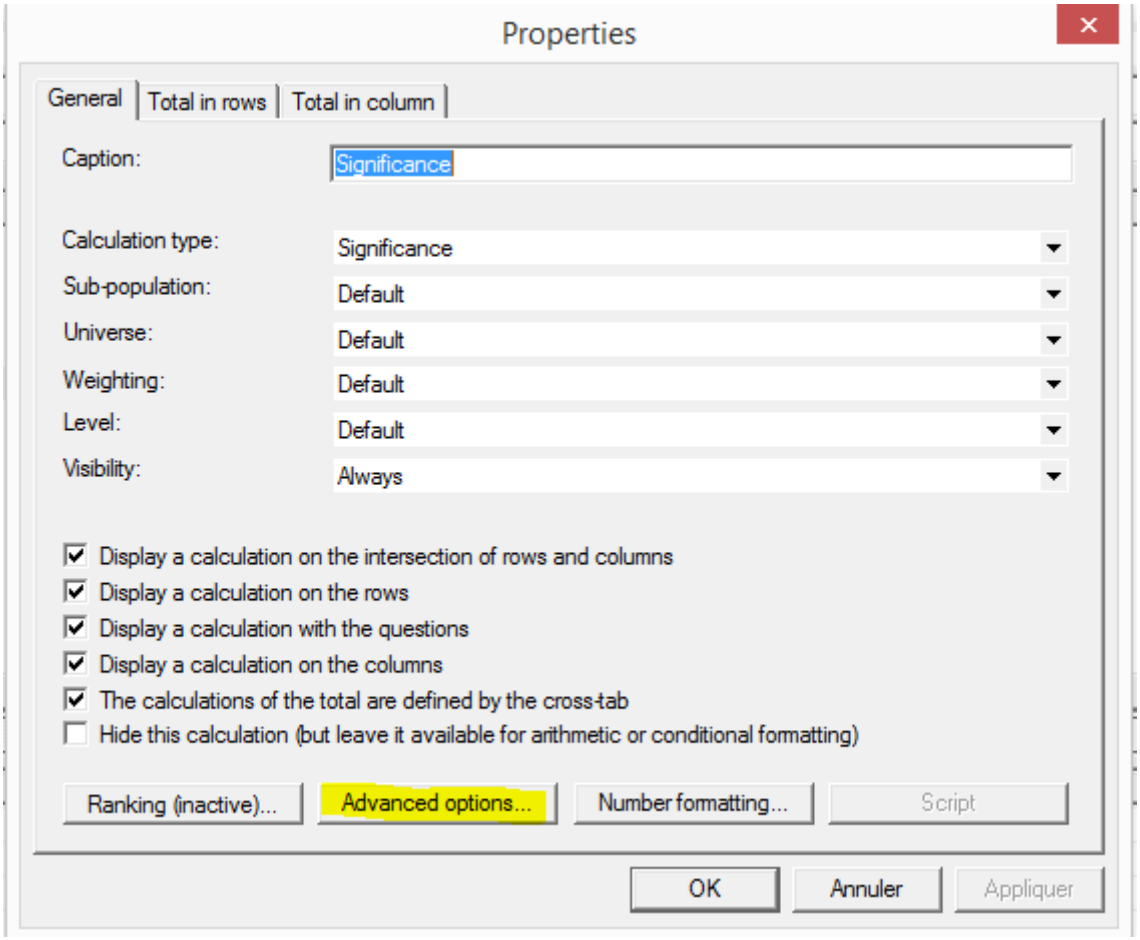
Base: 295 All interviews No weighting	Total	I2. Age			
		Less than 24	25 - 34	35 - 49	50 and older
Total	295 100,0%	42 14,2%	106 35,9%	103 34,9%	44 14,9%
i3. Socio-professional category					
Pupil, Student, Military	30 10,2%	2 4,8%	10 9,4%	9 8,7%	9 20,5%
Managerial Staff, Executive Manager, Liberal Profession	66 22,4%	19 45,2% +++	12 11,3% ---	25 24,3%	10 22,7%
Salaried Staff	60 20,3%	12 28,6%	16 15,1% -	20 19,4%	12 27,3%
Worker, Employee	62 21,0%	- - ---	28 26,4% +	30 29,1% ++	4 9,1% ---
Craftsman, Commercial, Artist	48 16,3%	6 14,3%	26 24,5% +++	7 6,8% ---	9 20,5%
Inactive, Jobless, Retired	29 9,8%	3 7,1%	14 13,2%	12 11,7%	- -
Khi2 test		Dep: 59,1 - Dof : 15 - Proba: 100,0%			

This test is used on independent samples.

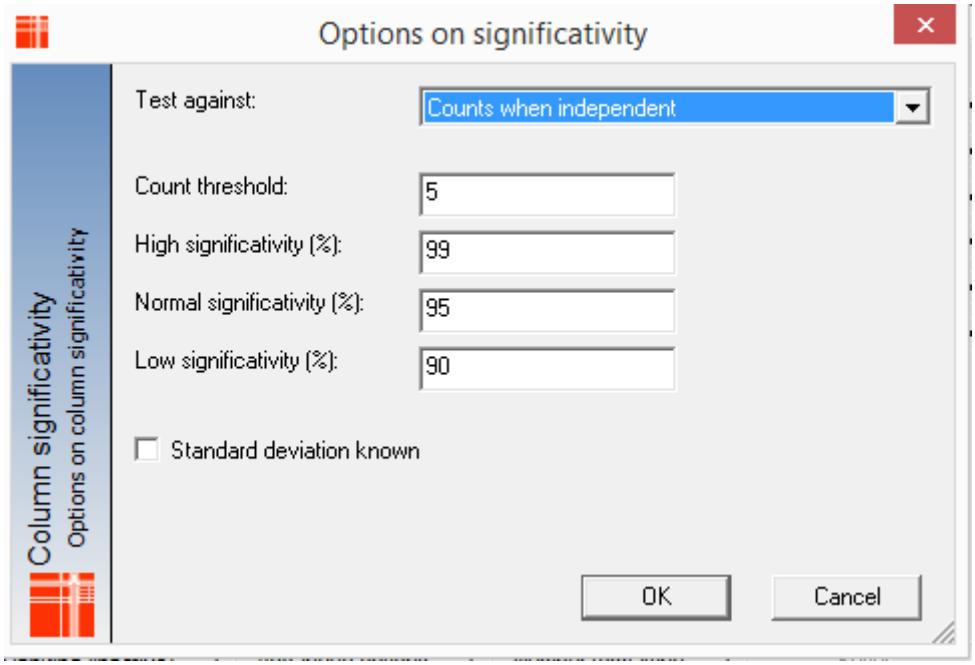
Select the calculation "Significance" in general section (*below Overall percentage*)



To define the parameters, click right on the significance calculation properties and click on Advanced options... :



And select the options as follows:



Test against	Select what comparison will be made, as follows: • **(A) Counts when independent**: comparison of cell with total ($\mathcal{X}^2$); ✓ You can test the p_i against the p observed in the total population *N* : • **(B) All other columns**: comparison of **the column profile** with the percentage obtained in the average other columns; ✓ You can test the p_i against the p observed in the *N* - N_i population : • **(C) All other rows**: comparison of the **row profile** with the percentage obtained in the other rows. ✓ You can test the p_i against the p observed in the *N* - N_i population :
Count threshold	The minimum count to be taken into account in a cell.
High significancy (%)	The percentage at which values are to be regarded as highly significant (e.g. 99%).
Normal *significativity* (%)	The percentage at which values are to be regarded as of normal significance (e.g. 95%)
Low High *significativity* (%)	The percentage at which values are to be regarded as of low significance (e.g. 90%)
Standard deviation known	(D)

In the results, significant values will be indicated by symbols:

- High threshold: +++ or ---
- Medium threshold: ++ or --
- Lower threshold: + or -

Note: it is possible to define a threshold to be 0, so that the test is not run at that threshold.

Threshold	Only 1 sign: "+" or "-"	2 signs: "++" or "--"
High significancy (%)	0	0
Normal *significativity* (%)	0	95
Low High *significativity* (%)	90	0

The test allows comparison of Test Values with threshold values.

To take the decision, we compare the calculated Sigma to the significativity threshold:
If Sigma > test value, then there is a significative difference.

The sign will indicate if the percentage is significatively decreasing (-) or increasing (+).

A) Counts when independent (khi²: χ^2)

The χ^2 is calculated as follow:

$$\chi^2 = \frac{\sum(x_{obs\,i} - x_{expected\,i})2}{N}$$

where $x_{obs\,i}$ is the count observed in sample *i* and $x_{expected\,i}$ is the expected count in the global sample *N*.

$$x_{expected\,i} = \frac{Total_i * Total_j}{N}$$

We calculated the χ^2 with k-1 degrees of freedom and the probability that the variable is dependant.

And *N*: Total Sample size,

Then we compare the $\boldsymbol{p_{obs(i,j)}} = \frac{\boldsymbol{x_{obs\,i,j}}}{\boldsymbol{N}}$ and the $\boldsymbol{p_{ind(i,j)}} = \frac{\boldsymbol{x_{expected\,i,j}}}{\boldsymbol{N}}$

If the

$$Sigma = \frac{(\boldsymbol{p_{obs(i,j)}} - \boldsymbol{p_{ind(i,j)}})}{\sqrt{\frac{\boldsymbol{p_{ind(i,j)}} * (1.- \boldsymbol{p_{ind(i,j)}})}{N}}} > \text{test value}$$

We conclude that the proportions are significatively **different** from the others.

B) and C) all other columns/rows

All other columns(j)	**All other rows (i)**
N_1 = Total(j)	N_1 = Total(i)
$N_2 = N - N_1$	
$\boldsymbol{p_{obs(1,j)}} = \frac{\boldsymbol{x_{obs\,1,j}}}{\boldsymbol{N1}}$	
$p_2 = \frac{(\text{Total(j)} - \boldsymbol{x_{obs\,2,j}})}{N_2}$	$p_2 = \frac{(\text{Total(i)} - \boldsymbol{x_{obs\,2,j}})}{N_2}$

D) If the standard deviation is known

If the standard deviation is known, we will use a normal law $\mathrm{N}(0\ (p_1 - p_2 = 0), s'd)$ where f is a calculated estimator $f = \frac{(p_1 * N_1) + (p_2 * N_2)}{N_1 + N_2}$

for

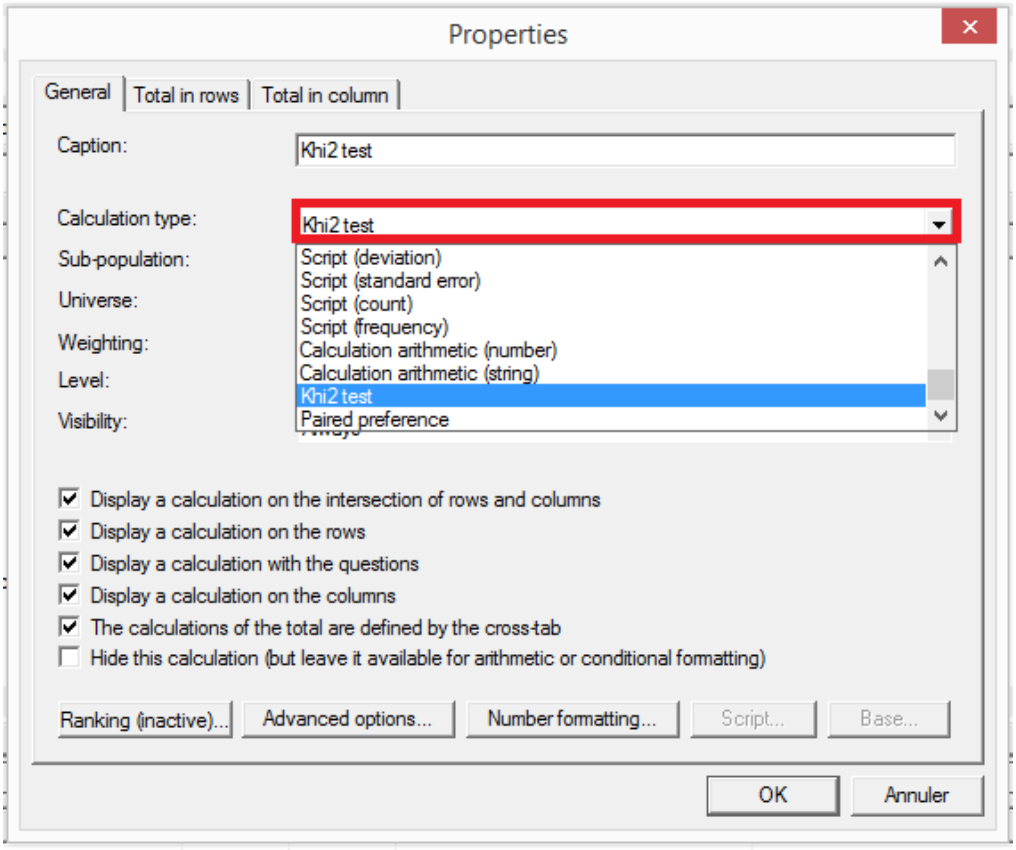
$$Sigma = \frac{f}{s'd = \sqrt{f * (1 - f) * (\frac{1}{N_1} + \frac{1}{N_2})}}$$

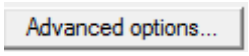
*If the standard deviation is unknown (by default in askia**analyse**), we will use a normal law* $\mathrm{N}(p_1 - p_2, sd)$ for

$$Sigma = \frac{p_1 - p_2}{sd = \sqrt{(p_1 * \frac{1 - p_1}{N_1}) + \left(p_2 * \frac{1 - p_2}{N_2}\right)}}$$

1.2. Khi2 Value

If you want to display the Khi2 value, the number of Degree of Freedom and the probability, you have to insert a new calculation and select Khi2 test as follow:

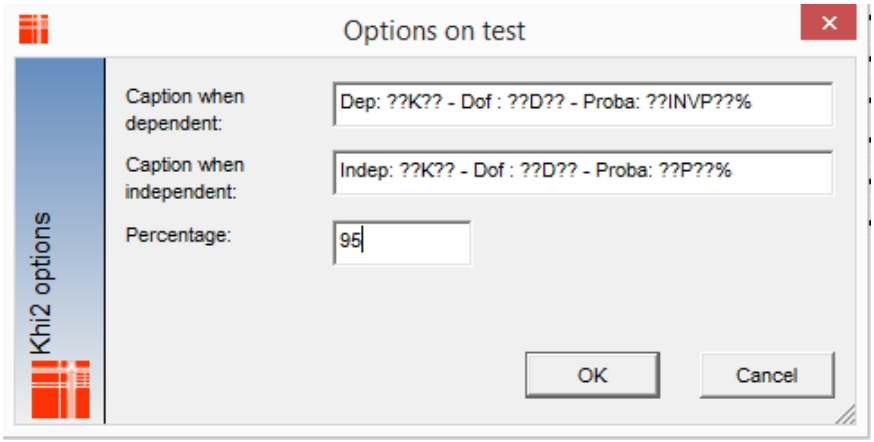


To customize the caption of the information displayed at the bottom of the table, click on 

And insert the following askiascript:

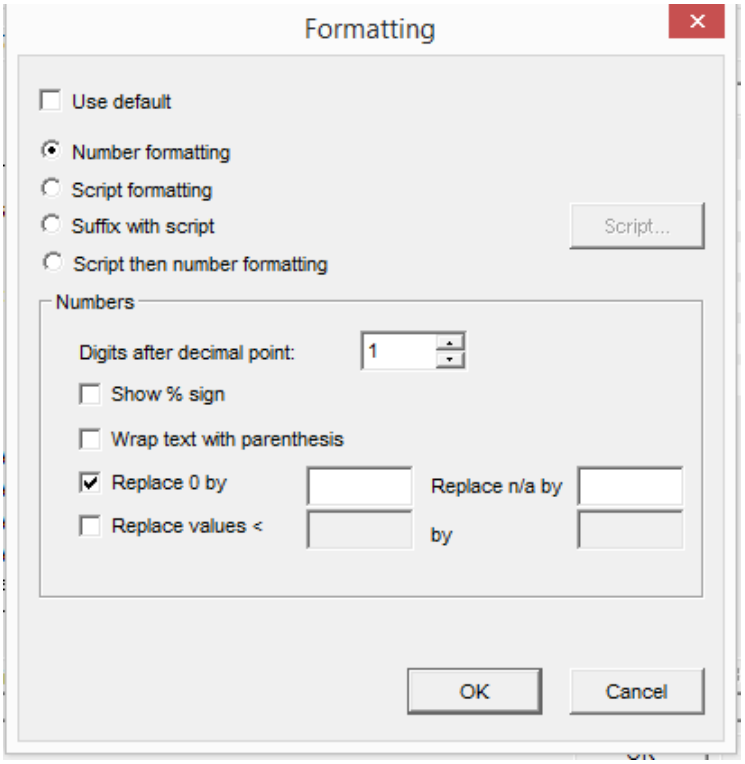
Khi²	??K??
The number of degree of Freedom (dof)	??D??
The probability that the variables are fully dependent:	??INVP??
The probability that the variables are fully independent:	??P??

For example :



You can also format the number as follow:

Click on Number formatting... in the calculation property, unclick the "Use Default", and specify the number of digits:



The χ^2 is calculated as follows:

$$\chi^2 = \frac{\sum(x_{obs\,i} - x_{expected\,i})2}{N}$$

where $x_{obs\,i}$ is the count observed in sample *i* and $x_{expected\,i}$ is the expected count in the global sample *N*.

$$x_{expected\,i} = \frac{Total_i * Total_j}{N}$$

We calculated the χ^2 with k-1 degrees of freedom and the probability that the variable is dependant.

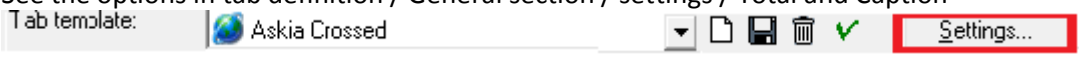
1.3. Column Significativity

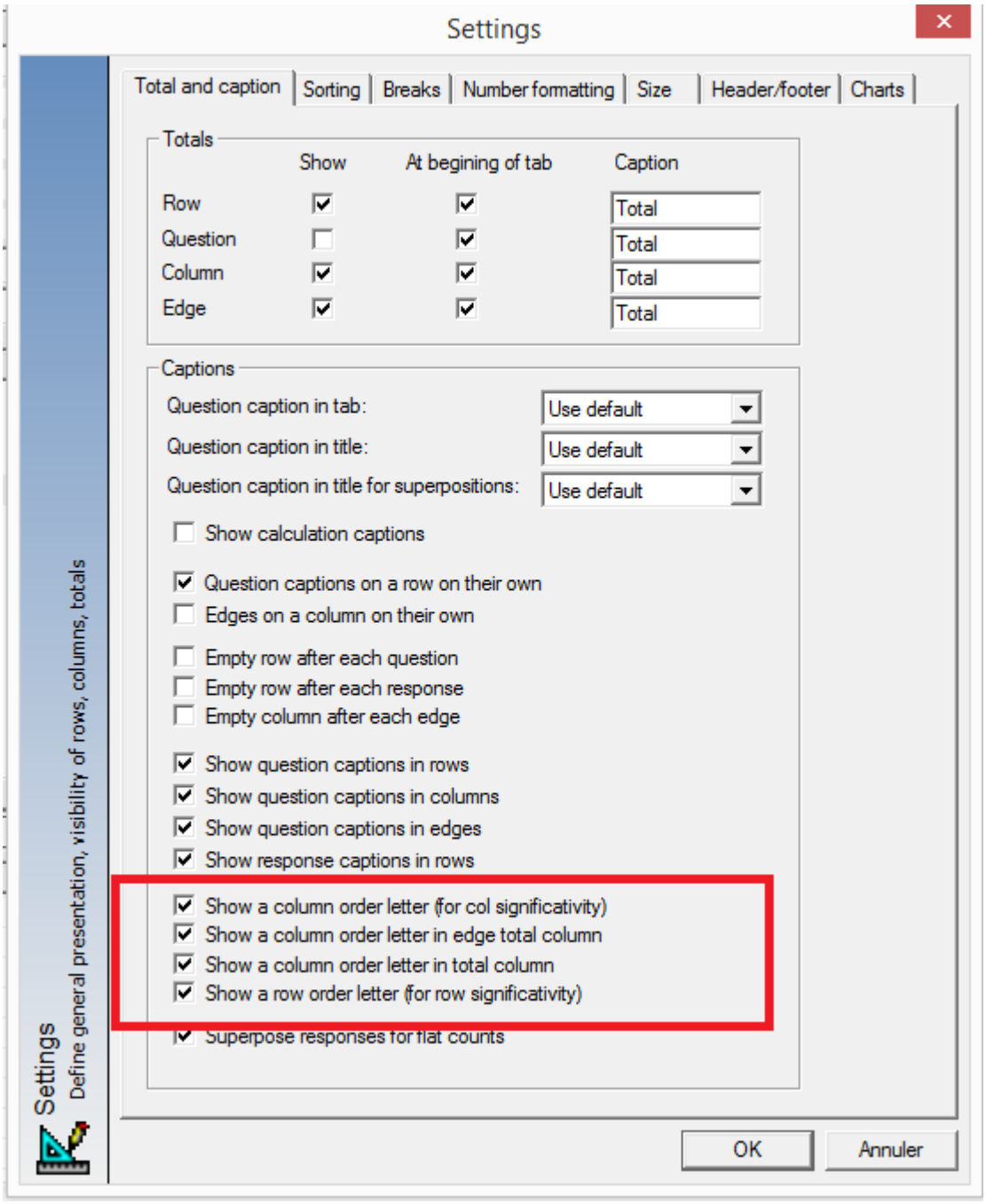
We use Column significativity if we want to compare proportions between 2 independent samples, and the output will display letters “a, A, A+”

Base: 295 All interviews No weighting	Total	i3. Socio-professional category					
		Pupil, Student, Military	Managerial Staff, Executive Manager, Liberal Profession	Salaried Staff	Worker, Employee	Craftsman, Commercial , Artist	Inactive, Jobless, Retired
	A	B	C	D	E	F	G
Total	295 100,0%	30 100,0%	66 100,0%	60 100,0%	62 100,0%	48 100,0%	29 100,0%
1. Overall film's appreciation							
Tremendously	25 9,7%	2 8,3%	6 10,2%	5 9,8%	6 10,7%	3 6,7%	3 13,0%
Very much	66 25,6% g	6 25,0%	18 30,5% G	15 29,4% G	13 23,2%	12 26,7% g	2 8,7%
Medium	64 24,8% D	12 50,0% A+CD+F+g	13 22,0%	6 11,8%	18 32,1% D	9 20,0%	6 26,1%
Not very much	75 29,1% E	4 16,7%	21 35,6% bE+	22 43,1% ABE+	7 12,5%	15 33,3% E	6 26,1%

To define where the letters in header will be displayed

See the options in tab definition / General section / settings / Total and Caption

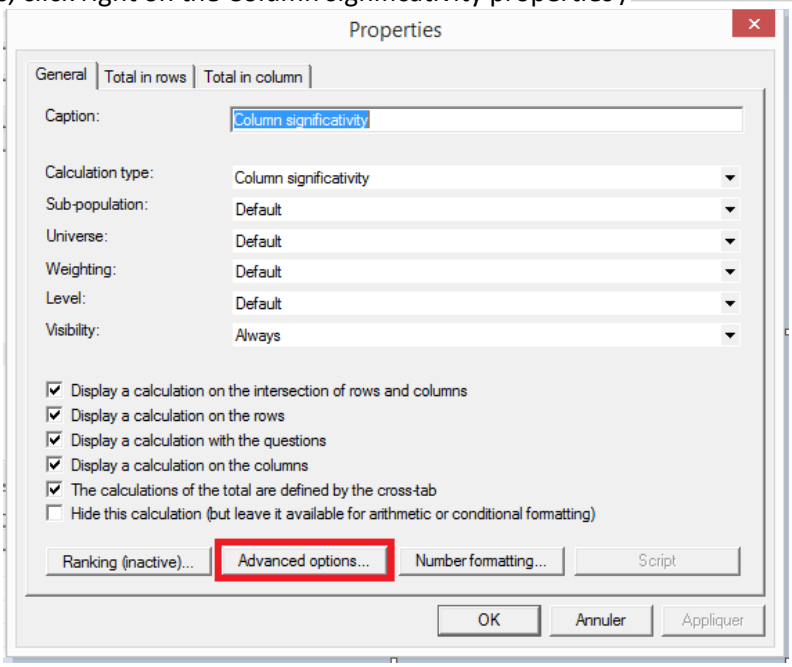




Once defined where the letters will be displayed, select the calculation “Column significativity”/”Row significativity” in general section.

Note that the column/row significativity *(for proportions)* is in the section of calculations for closed questions (*below Affinity index)*

To define the parameters, click right on the Column significativity properties / Advanced options... :



Let's imagine we want to compare the Age between the Social Category X Gender:

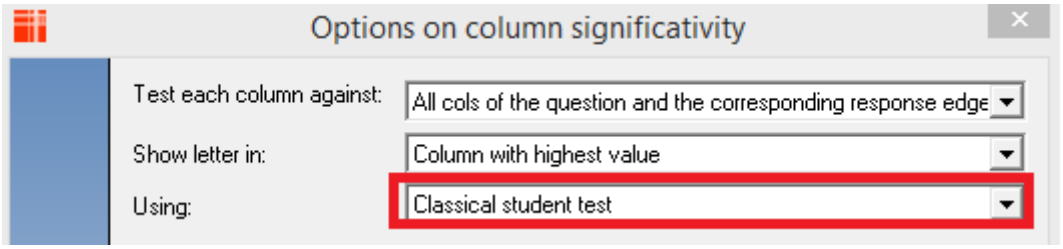
Base: 295 All interviews No weighting	Total	i1. Respondant's gender									
		Man					Woman				
		Total	I2. Age		i3. Socio-professional category		Total	I2. Age		i3. Socio-professional category	
			Less than 35	35 and older	CSP+	CSP -		Less than 35	35 and older	CSP+	CSP -
	A	B	C	D	E	F	G	H	I	J	K
Total	295 100,0%	162 100,0%	76 100,0%	86 100,0%	121 100,0%	41 100,0%	133 100,0%	72 100,0%	61 100,0%	115 100,0%	18 100,0%
1. Overall film's appreciation											
Tremendously	25 9,7%	19 13,6% G	10 14,3%	9 12,9%	14 13,0%	5 15,6%	6 5,1%	6 10,0%	- -	6 5,8%	- -
Very much	66 25,6% C	28 20,0% G	10 14,3%	18 25,7%	22 20,4% J	6 18,8%	38 32,2%	16 26,7%	22 37,9%	36 35,0%	2 13,3%
Medium	64 24,8% J	39 27,9%	21 30,0%	18 25,7%	30 27,8% J	9 28,1%	25 21,2%	13 21,7%	12 20,7%	16 15,5%	9 60,0%
Not very much	75 29,1%	42 30,0%	23 32,9%	19 27,1%	36 33,3%	6 18,8%	33 28,0%	15 25,0%	18 31,0%	29 28,2%	4 26,7%
Not at all	28 10,9%	12 8,6%	6 8,6%	6 8,6%	6 5,6% J	6 18,8%	16 13,6%	10 16,7%	6 10,3%	16 15,5%	- -
Nbre de réponses	258	140	70	70	108	32	118	60	58	103	15
M	3,79	3,89	3,66	4,13	3,89	3,91	3,68	3,77	3,59	3,78	3,00
SD	3,22	3,32	3,31	3,32	3,26	3,50	3,09	3,33	2,81	3,22	1,79
Sign. M											
Col Sign. M											

The available options in this dialog vary depending on the type of calculation being defined.
The following options are available:

Test each column against	Select what comparison will be made, as follows: **All columns:** all the columns of questions selected in column (independently of the question) will be compared with each other. For example, we will compare A/B, ….C/G, …H/K **All columns of the question** : All columns of the corresponding question will be compared with each other We will compare only between the Age categories (C/D) and between the CSP categories (E/F) **The total only**: Only the column Total (A) will be compared to the others columns. **All columns of the question and the corresponding responses edges:** all the columns of the same question only will be compared with each other and with the corresponding edge. For example , we will compare C/D or E/F or C/H **Previous column only:** all the columns of the same question or different questions will be compared to the previous column. **All columns of the question and the corresponding questions edges:** all the columns of the same question only will be compared with the corresponding edge. For example , we will compare C/H or D/I or E/J but we will not compare the C/D or the E/F **All columns of edge responses:** Only the columns within the edge response will be compared. For example, between B-C-D-E-F or between G-H-I-J-K **All columns of the current question and the first question:** All the columns of the first question will be compared to the all others columns. **Specify columns:** You can choose the columns to compare
Show letter in	Select the priority for the display of letters, as follows: • **Both columns:** the significance will appear in both columns • **First column:** the significance will appear in the first column • **Column with highest value:** the significance will appear in the column where the column profile is highest
Using	Specifies the type of test to be used: classical student test (A), student test using estimator (B), student test using efficiency coefficient (C), or student test using estimator and efficiency coefficient(D).
Test against total column	When this option is selected, the total column becomes a column like any other for the purposes of the calculation.
Display minus if under	When this option is selected, a minus sign is shown if the significancy goes down. • This can be used in conditional formatting when you test in conjunction with the column before: if there are two letters, the value has significantly gone down, if you have one, it has gone up.
Use student test when degrees of freedom <	If this option is selected, a student test will be used when the degrees of freedom are less than the amount stated in the adjacent box.
Use unweighted base	If this option is selected, col significativity will be carried out on weighted % and unweighted counts.

Columns are assumed independent	When this option is selected, the individuals belonging to a sub-total will be considered different to those present in the category grouped in the same sub-total.
Count threshold	The minimum count to be taken into account in a cell.
Base threshold	The minimum base that must be met before column sig. testing is displayed. By default, the minimum base is 0. This option affects closed and numeric questions.
High significativity (%)	The percentage at which values are to be regarded as highly significant.
Normal significativity (%)	The percentage at which values are to be regarded as of normal significance.
Low significativity (%)	The percentage at which values are to be regarded as of low significance.
Display "A+"	Mark highly significant values with *A+.*
Display "A"	Mark values of medium significance with *A.*
Display "a"	Mark values of low significance with a.
Test cols (A:B,C:D-F)	*This option appears when you select **specify columns** in test **each column against**. It allows you to list specific columns to be compared against each other.* *Enter the letters or numbers of the columns you want to test. Separate each test with a comma, and use a colon to separate the columns to be compared within a test.* *Examples:* • *To test AvsB and DvsE only, enter A:B,D:E or 1:2,4:5* • *To test A against B,C,D,E,F: enter A:B-F* • *To test A,B,C against A,B,C: enter A-C:A-C or A-C* *You can display the list of columns being compared in your table (e.g. the footer), by entering the keyword ??*ColSig?? *in the appropriate field.*
Sig message (use ??sig??)	Defines a message which indicates at which level the columns have been tested during col significativity testing. The token ??sig?? can then be used to display this message. Select a message in the drop-down list, or enter your own. The items you can place in the message are: • p1 = high % • p2 = normal % • p3 = low % • invp1 = 100 - high % • invp2 = 100 - normal % • invp3 = 100 - low % • p1_1 = high proba (0-1) • p1_2 = normal proba (0-1) • p1_3 = low proba (0-1) • invp1_1 = 1 - high proba (0-1) • invp1_2 = 1- normal proba (0-1) • invp1_3 = 1- low proba (0-1) For example, if the message is "Columns are tested at ??p1??", p1 will be replaced by the high significativity value.

A) Classical Student Test



This test describes the Z-test using unpooled variance:

$$Z = \frac{(p_1 - p_2)}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

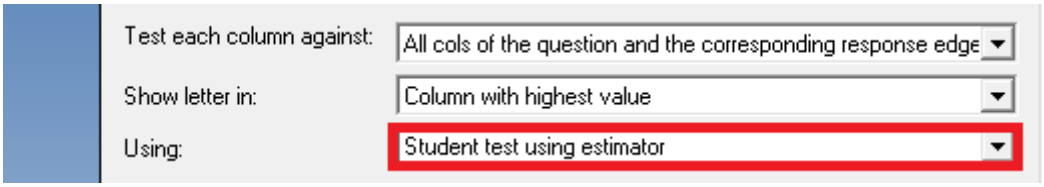
Where

p_1= Proportion 1 observed in the sample
p_2= Proportion 2 observed in the sample
n_1=Sample 1 size
n_2=Sample2 size

We compare the Z value > tα

- $t_\alpha 90\%$ = 1.65
- $t_\alpha 95\%$ = 1.96
- $t_\alpha 99\%$= 2.576, If Z > t_α then there is significative difference

B) Student Test using estimator



This test describes the Z-test using pooled variance

$$Z = \frac{(p_1 - p_2)}{\sqrt{dFo\,(1 - dFo) * (\frac{1}{n_1} + \frac{1}{n_2})}}$$

Where

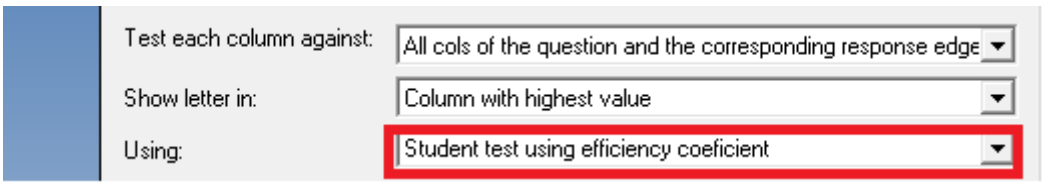
p_1= Proportion 1 observed in the sample
p_2= Proportion 2 observed in the sample
n_1=Sample 1 size
n_2=Sample2 size

dFo= $\frac{x_1+x_2}{n_1+n_2}$, And x_{ij} is the count observed in the cell *ij* and n_j is the sample size for the column j

We compare the Z value > tα

- $t_\alpha 90\%$ = 1.65
- $t_\alpha 95\%$ = 1.96
- $t_\alpha 99\%$= 2.576, If Z > t_α then there is significative difference

C) Student Test using efficiency coefficient



This test is used when we want to reduce the effect on the weighting.
Leslie Kish has analysed the effect of unequal weights in the accuracy of estimations through the 'Unequal Weighting Effect' (UWE). *(Kish L., Weighting for Unequal Pi, Journal of Official Statistics, Vol. 8, N°2, 1992, pp. 183-200.)*

$$Z = \frac{(p_1 - p_2)}{\sqrt{dFo\ (1 - dFo) * \left(\frac{\sum_{i=1}^{n_1} w_1^2}{n_1^2} + \frac{\sum_{i=1}^{n_2} w_2^2}{n_2^2}\right)}}$$

Where

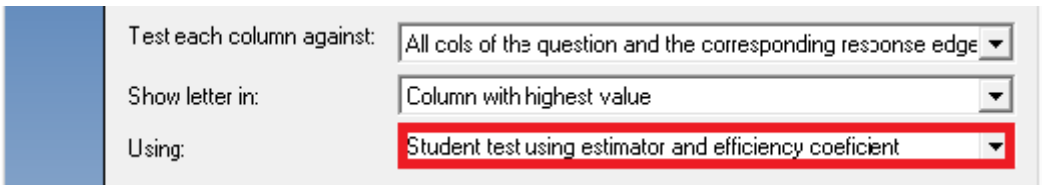
p_1= Proportion 1 observed in the sample
p_2= Proportion 2 observed in the sample
n_1=Sample 1 size
n_2=Sample2 size

dFo=$\frac{x_1+x_2}{n_1+n_2}$, and x_{ij} is the count observed in the cell *ij* and n_j is the sample size for the column j
w_1 = weight in the sample 1 per individual
w_2 = weight in the sample 2 per individual

We compare the Z value > tα

$t_\alpha 90\%$ = 1.65
$t_\alpha 95\%$ = 1.96
$t_\alpha 99\%$= 2.576, If Z > t_α then there is significative difference

D) Student Test using estimator and efficiency coefficient



This test is used when we want to reduce the effect on the weighting
Leslie Kish has analysed the effect of unequal weights in the accuracy of estimations through the 'Unequal Weighting Effect' (UWE). *(Kish L., Weighting for Unequal Pi, Journal of Official Statistics, Vol. 8, N°2, 1992, pp. 183-200.)*

$$Z = \frac{(p_1 - p_2)}{\sqrt{dFo * \frac{(1 - Fo)}{1 - \frac{1}{\sum_{i=1}^{n_1} w_1^2 + \sum_{i=1}^{n_2} w_2^2}} * \left(\frac{\sum_{i=1}^{n_1} w_1^2}{n_1^2} + \frac{\sum_{i=1}^{n_2} w_2^2}{n_2^2}\right)}}$$

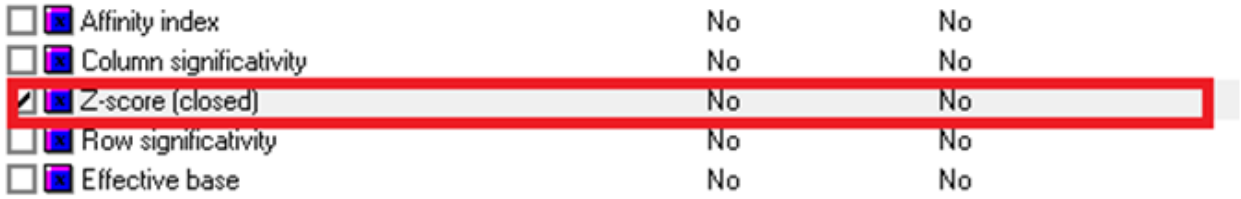
Where

p_1= Proportion 1 observed in the sample
p_2= Proportion 2 observed in the sample
n_1=Sample 1 size
n_2=Sample2 size
dF0= $\frac{x_1+x_2}{n_1+n_2}$, And x_{ij} is the count observed in the cell *ij* and n_j is the sample size for the column j
w_1 = weight in the sample 1 per individual
w_2 = weight in the sample 2 per individual
We compare the Z value > tα
$t_\alpha 90\%$ = 1.65
$t_\alpha 95\%$ = 1.96
$t_\alpha 99\%$= 2.576, If Z > t_α then there is significative difference

1.4. Z-Score

We use the Z-score to see if the proportion observed (p_i) in the sample *i* (n_i) is different from the p_N observed in the total sample *N.*

Select the Z-score (closed) in General menu in tab definition /calculations:



As output we will have a number displayed as an extra calculation:

Base: 295 All interviews No weighting	Total	i3. Socio-professional category					
		Pupil, Student, Military	Managerial Staff, Executive Manager, Liberal Profession	Salaried Staff	Worker, Employee	Craftsman, Commercial , Artist	Inactive, Jobless, Retired
	A	B	C	D	E	F	G
Total	295 100,0%	30 100,0%	66 100,0%	60 100,0%	62 100,0%	48 100,0%	29 100,0%
I2. Age							
Less than 24	42 14,2% E	2 6,7% 	19 28,8% AEF	12 20,0% E	- - --- 	6 12,5% E	3 10,3%
		-1,25	3,83	1,43	-3,60	-0,38	-0,63

If the Z-score is ≥ 1.96, there is a significative difference at 95% between the percentage observed in the sample i, $p_i$and the percentage observed in the total population N, p_N. The sign will indicate if the p_i is lower than the p_N.

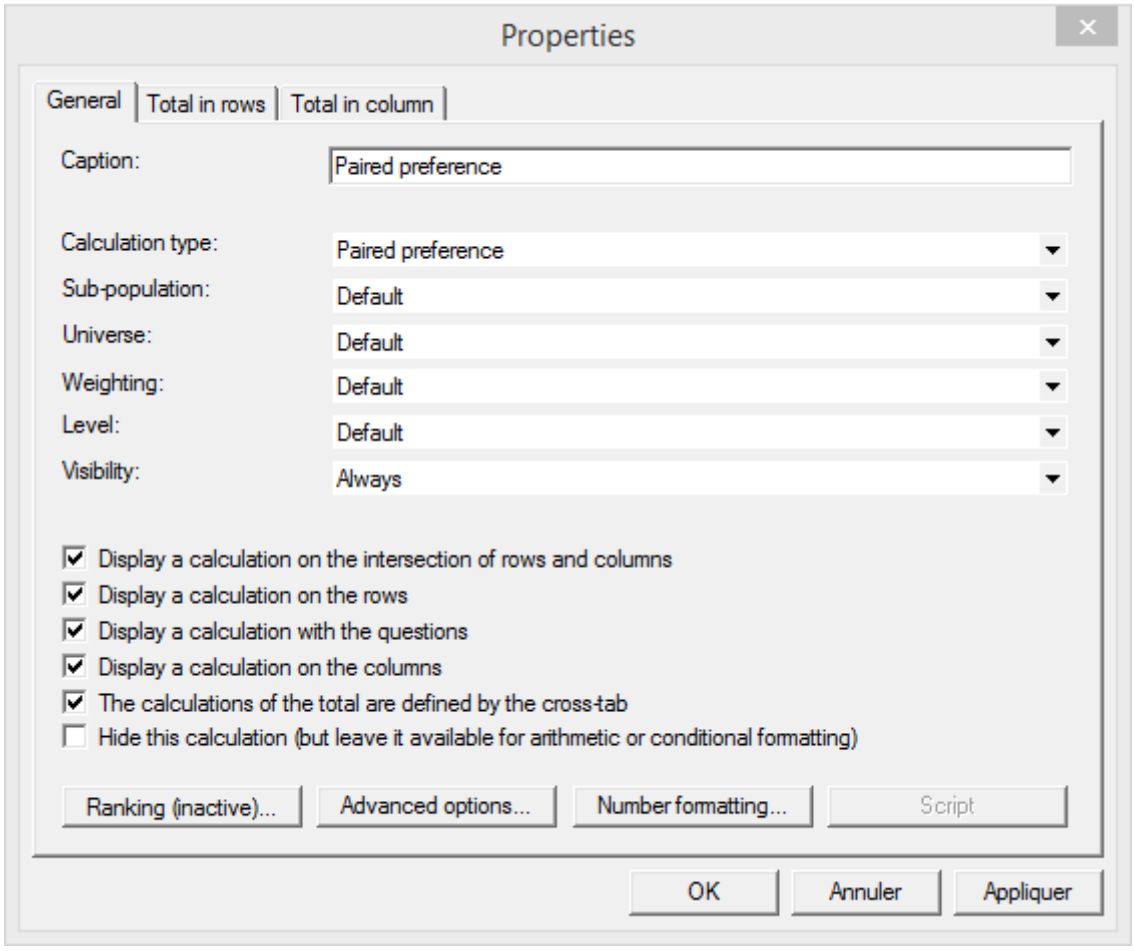
The z-score (for *closed* questions) will be calculated as follow:

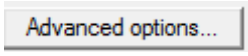
$$Z - score = \frac{(p_1 - p_N)}{\sqrt{\frac{p_N\,(1-p_N) * (\frac{N}{N_1} - 1)}{N-1}}}$$

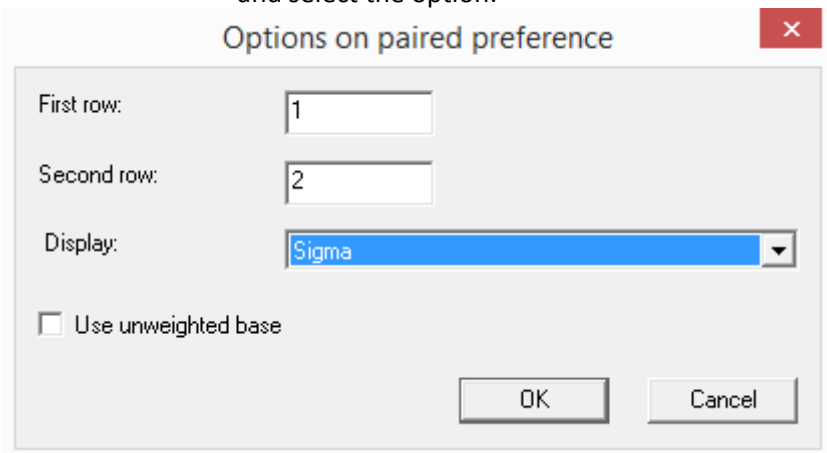
1.5.Paired preference test

We use a Paired preference test when we want to compare 2 rows (e.g. product R/ product T) for each column independently. We want to know if the product R is preferred at the Product T.

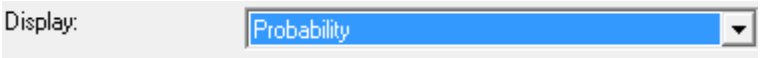
Insert a calculation in General Section of the tab definition:



Then click on  and select the option:



Note that you can display the sigma or the probability

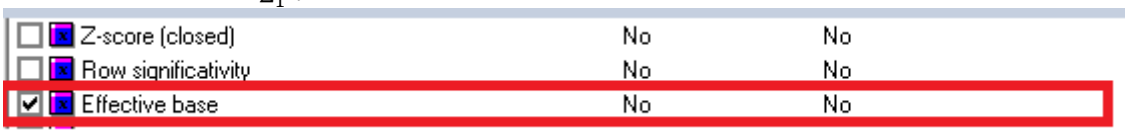


The formula is:

$$T = \frac{(p_1 - p_2)}{\sqrt{\frac{1}{e}(p_1 + p_2)}}$$

Where:

e = the effective base = $\frac{(\sum_1^j n)^2}{\sum_1^j n^2}$ for the column j

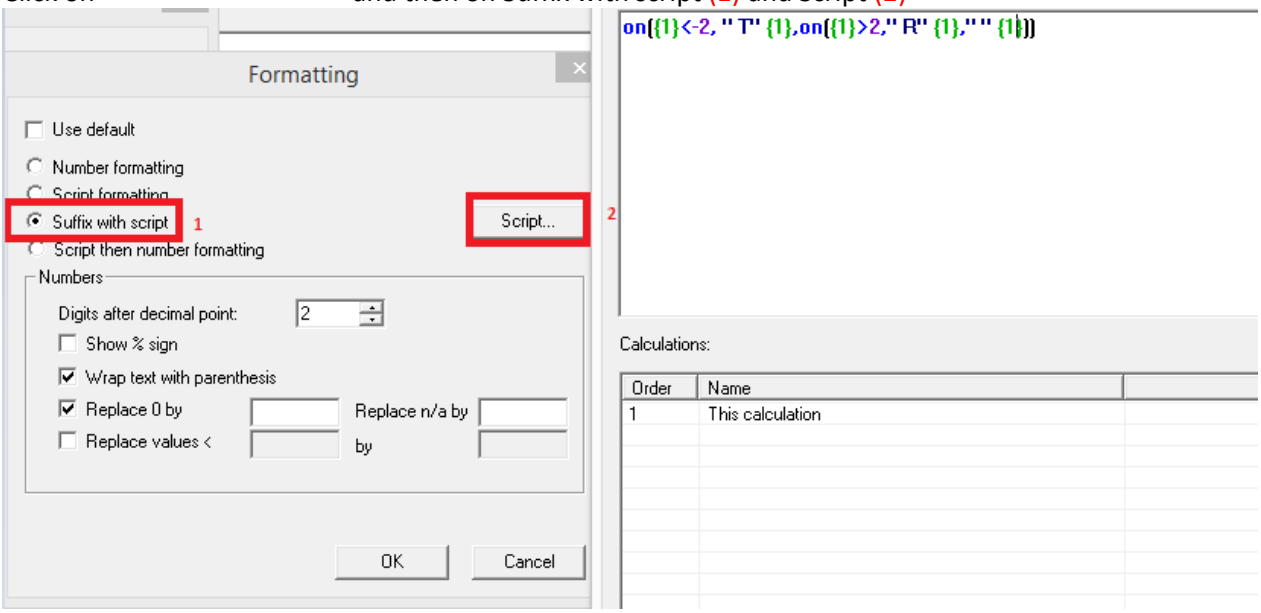


$p_1$is the percentage observed in the row 1
$p_2$is the percentage observed in the row 2

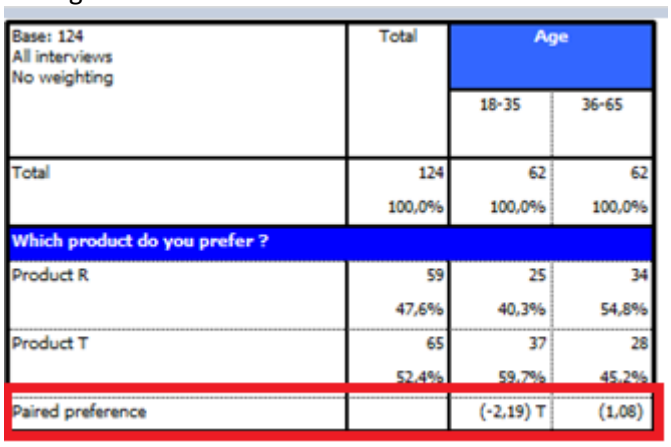
Note that this test is undefined if $p_1 - p_2 = 0$ and is $e < 2$,
We compare the T value to t distribution with (e -1) degrees of freedom.

The output will display the sigma value or the probability. But you can customize the value displayed (e.g. Sigma value + preferred product)

Click on Number formatting... and then on Suffix with script (1) and Script (2)



You will obtain the following table:



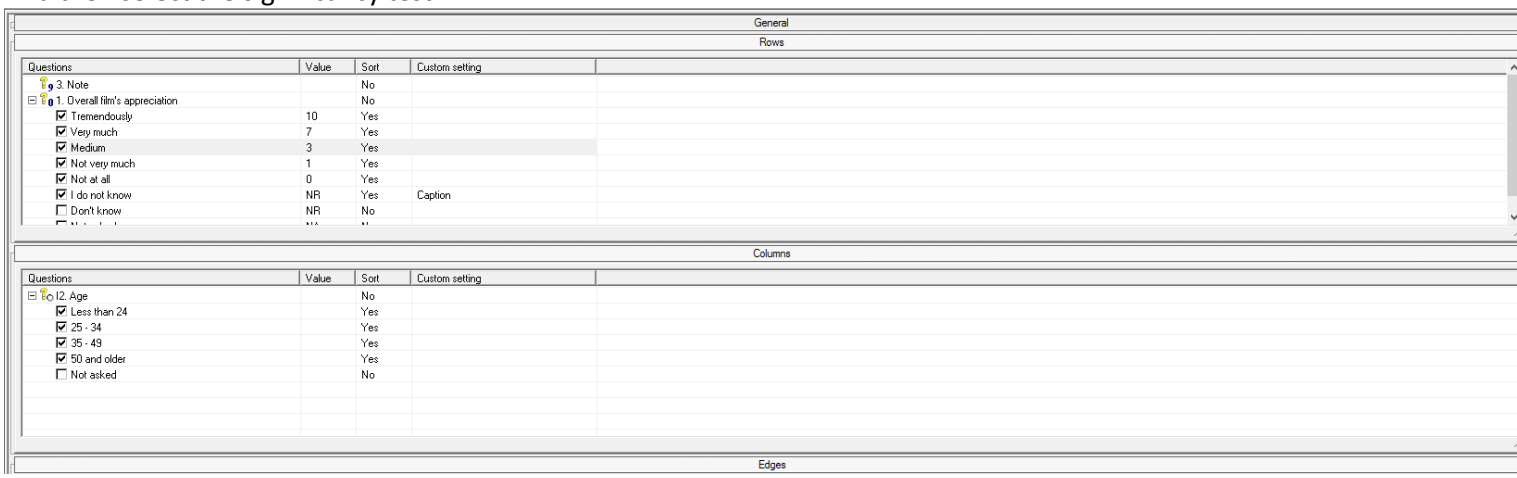
Base: 124 All interviews No weighting	Total	Age	
		18-35	36-65
Total	124 100,0%	62 100,0%	62 100,0%
Which product do you prefer ?			
Product R	59 47,6%	25 40,3%	34 54,8%
Product T	65 52,4%	37 59,7%	28 45,2%
Paired preference		(-2,19) T	(1,08)

2. Mean comparison

To apply it, you need to run a crossed table.

You can select any numeric question (numeric or scale responses)
Insert a **numeric question in row and** a **closed question in column** in a tab definition

And then select the significancy test.



2.1.Significance

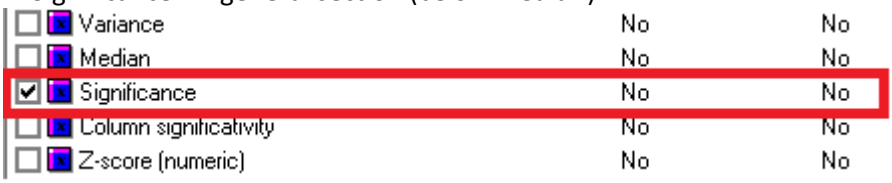
The significance test (for *numeric*) is used in order to see if the mean observed ($\bar{x}_i$) in the sample *i* (n_i) is different from the $\bar{x}_N$ observed in the sample *N*.

We will display "-" or "+" signs to indicate where the difference is significative (see the figure below):

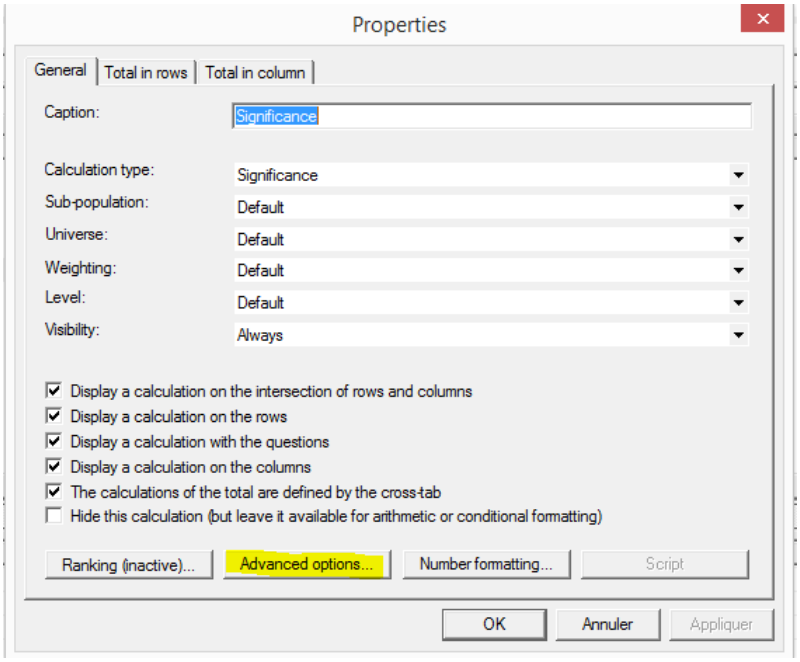
Base: 155 All interviews No weighting	Total	i3. Socio-professional category					
		Pupil, Student, Military	Managerial Staff, Executive Manager, Liberal Profession	Salaried Staff	Worker, Employee	Craftsman, Commercial, Artist	Inactive, Jobless, Retired
Total	155	20	37	26	37	24	11
	100,0	100,0	100,0	100,0	100,0	100,0	100,0
3. Note							
M	4,95	2,29	5,16	6,33	5,87	3,38	5,82
sd	3,00	2,86	3,58	2,68	1,90	2,06	2,25
nb	138	14	37	21	31	24	11
Significance		---		++	++	---	

This test is used on independent samples.

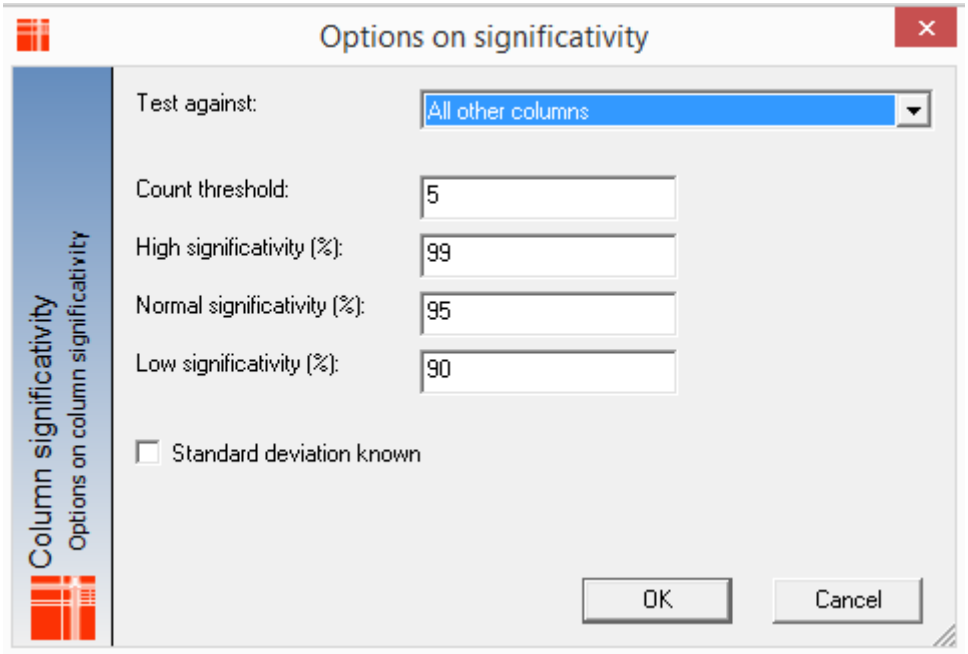
Select the calculation "Significance" in general section (*below median*)



To define the parameters, click right on the significance calculation properties and click on Advanced options...:



And select the options as follows:



Test against	Select what comparison will be made, as follows: • **(A) All other columns**: comparison of **the mean of the column** with the mean obtained in the other columns; ✓ You can test the $\bar{x}_i$ against the $\bar{x}$ observed in the $N - N_i$ population : • **(B) All other rows**: comparison of the **mean of the row** with the mean obtained in the other rows. ✓ You can test the $\bar{x}_i$ against the $\bar{x}$ observed in the $N - N_i$ population : • NB : The option "Count when independent" is useless (the A) test will be applied)
Count threshold	The minimum count to be taken into account in a cell.
High significancy (%)	The percentage at which values are to be regarded as highly significant (e.g. 99%).
Normal *significativity* (%)	The percentage at which values are to be regarded as of normal significance. (e.g. 95%)
Low High *significativity* (%)	The percentage at which values are to be regarded as of low significance. (e.g. 90%)
Standard deviation known	(D)

In the results, significant values will be indicated by symbols:

- High threshold: +++ or ---
- Medium threshold: ++ or --
- Lower threshold: + or -

Note: it is possible to define a threshold to be 0, so that the test is not run at that threshold.

Threshold	Only 1 sign: "+" or "-"	2 signs: "++" or "--"
High significancy (%)	0	0
Normal *significativity* (%)	0	95
Low High *significativity* (%)	90	0

The test allows comparison of Test Values with threshold values.

To take the decision, we compare the calculated Sigma=D to the significativity threshold:
If D > test value, then there is a significative difference.

The sign will indicate if the mean in the sample 1 is significatively lower (-) or higher (+). Than the mean in the other columns/rows

The sigma value D is calculated (as follow)

$$D = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{sd_1^2}{N_1} + \frac{sd_2^2}{N_2}}}$$

Where D follows a normal mathematical expectation $N\ (\bar{x}_1 - \bar{x}_2 = 0, sd = \sqrt{\frac{sd_1^2}{N_1} + \frac{sd_2^2}{N_2}})$

$\boldsymbol{x_{obs1}}$ = is the count observed in the column/row for the sample $\boldsymbol{N_1}$
$\boldsymbol{x_{obs2}}$ = is the count observed in the sample $\boldsymbol{N_2 = N - N_1}$

$\bar{x}_1 = \frac{\sum_{i=1}^{N_1} x_{obsi}}{N_1}$ = mean 1 in the sample 1
$\bar{x}_2 = \frac{\sum_{i=1}^{N_2} x_{obsi}}{N_2}$ = mean 2 in the sample 2 (all others columns/rows)

$sd_1 = \sqrt{\sum_{i=1}^{N1}(x_{obs1} - \bar{x}_1)}$ is the standard deviation observed in the Sample 1
$sd_2 = \sqrt{\sum_{i=1}^{N2}(x_{obs2} - \bar{x}_2)}$ is the standard deviation observed in the Sample 2

We compare the abs(D)> tα

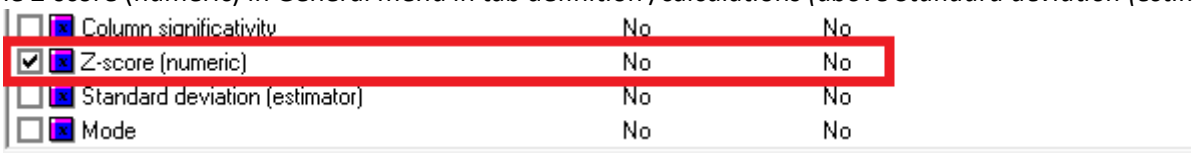
- $t_\alpha 90\%$ = 1.65
- $t_\alpha 95\%$ = 1.96
- $t_\alpha 99\% = 2.576$

if abs(D)> t_α , there is a significative difference between $\bar{x}_1$ and $\bar{x}_2$

2.2.Z-Score

We use the Z-score (numeric) to see if the mean observed ($\bar{x}_i$) in the sample *i* (n_i) is different from the mean observed ($\bar{x}_N$) observed in the total sample *N*.

Select the Z-score (numeric) in General menu in tab definition /calculations *(above Standard deviation (estimator)):*



As output we will have a number displayed as an extra calculation:

Base: 155 All interviews No weighting	Total	i3. Socio-professional category					
		Pupil, Student, Military	Managerial Staff, Executive Manager, Liberal Profession	Salaried Staff	Worker, Employee	Craftsman, Commercial , Artist	Inactive, Jobless, Retired
	A	B	C	D	E	F	G
Total	155	20	37	26	37	24	11
	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
3. Note							
Nb	138	14	37	21	31	24	11
M	4,95	2,29	5,16	6,33	5,87	3,38	5,82
SD	3,00	2,86	3,58	2,68	1,90	2,06	2,25
Sign. M		---		++	++	---	
Col Sign. M	BF		BF	ABF	ABF		BF
Z-score (numeric)		-0,89	0,07	0,46	0,31	-0,52	0,29

If the P(Z-score) > , there is a significative difference at 95% between the mean observed in the sample i, $\bar{x}_i$and the mean observed in the total population N, $\bar{x}_N$. The sign will indicate if the $\bar{x}_i$ is lower than the $\bar{x}_N$.

The z-score (for *numeric* questions) will be calculated as follow:

$$Z - score = \frac{(\bar{x}_1 - \bar{x}_N)}{sd_N}$$

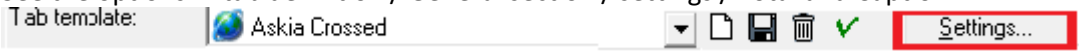
2.3.Column Significativity

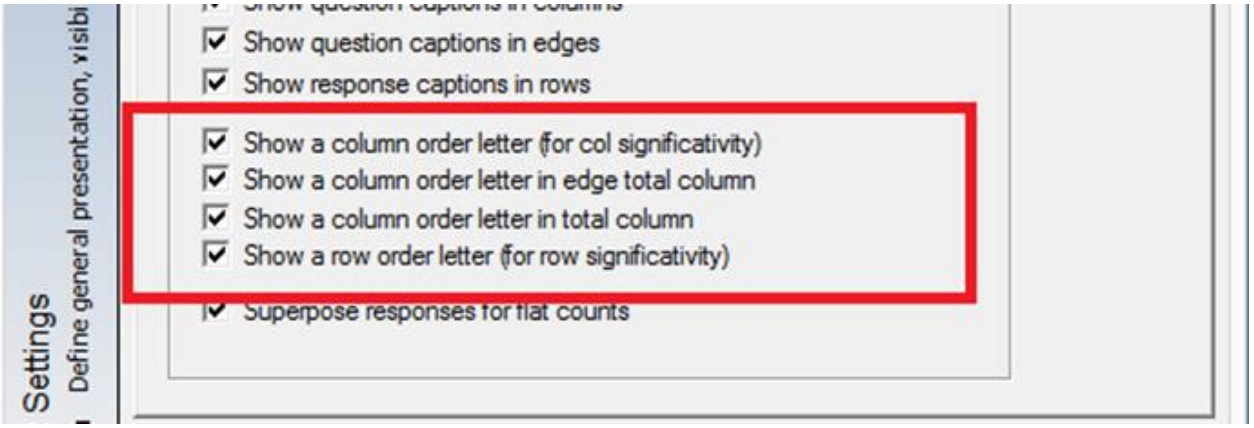
We use Column significativity (for *numeric*) if we want to compare means between 2 independent samples, and the output will display letters “a, A, A+”

Base: 155 All interviews No weighting	Total	i3. Socio-professional category					
		Pupil, Student, Military	Managerial Staff, Executive Manager, Liberal Profession	Salaried Staff	Worker, Employee	Craftsman, Commercial, Artist	Inactive, Jobless, Retired
	A	B	C	D	E	F	G
Total	155	20	37	26	37	24	11
3. Note							
Nb	138	14	37	21	31	24	11
M	4,95	2,29	5,16	6,33	5,87	3,38	5,82
SD	3,00	2,86	3,58	2,68	1,90	2,06	2,25
Col Sign. M	B+F+		B+F	AB+F+	AB+F+		B+F+

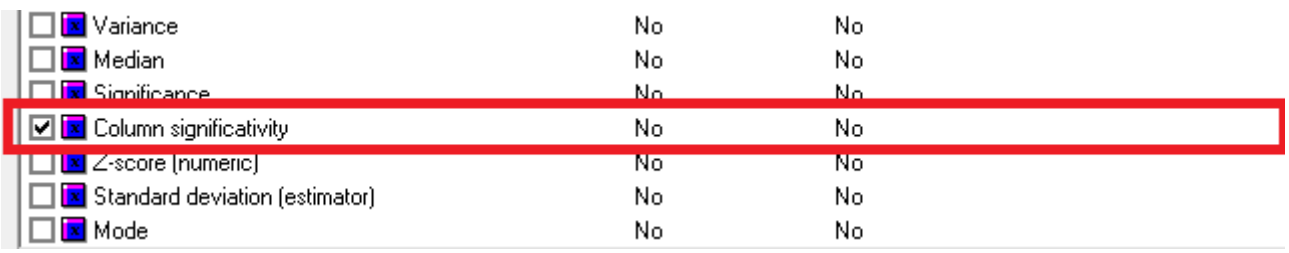
We can define where the letters in header will be displayed

See the options in tab definition / General section / settings / Total and Caption



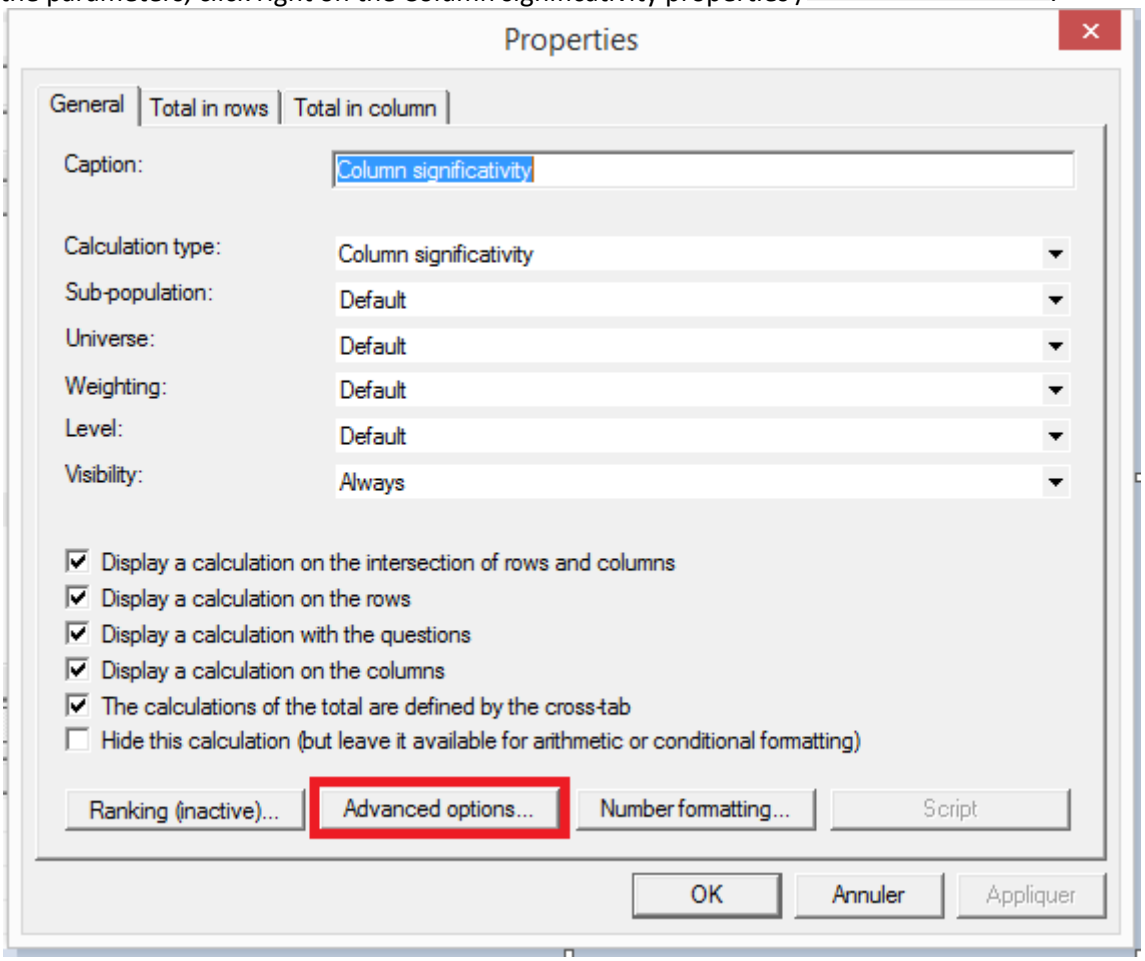


Once defined where the letters will be displayed, we can select the calculation “Column significativity”/”Row significativity” in general section.



Note that the column/row significativity *(for numerics)* is in the section of calculations for numeric question (above Z-score(numeric)).

To define the parameters, click right on the Column significativity properties / Advanced options... :



This dialog allows you to set advanced options for the selected calculation type. The available options in this dialog vary depending on the type of calculation being defined.

Let's imagine we want to compare the Overall Note for the movie between the Social Categories X the Gender:

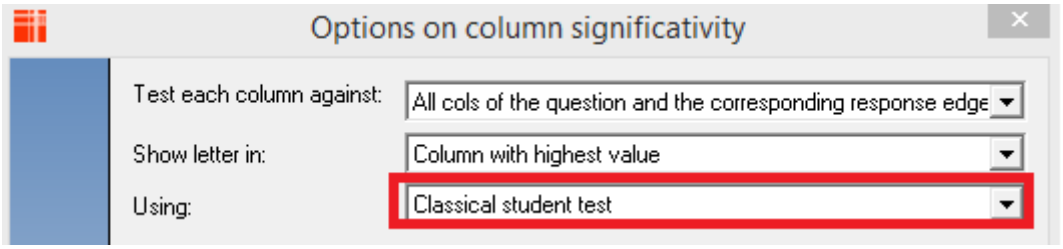
Base: 155 All interviews No weighting	Total	i1. Respondant's gender													
		Man							Woman						
		Total	i3. Socio-professional category						Total	i3. Socio-professional category					
			Pupil, Student, Military	Managerial Staff, Executive Manager, Liberal Profession	Salaried Staff	Worker, Employee	Craftsman, Commercial , Artist	Inactive, Jobless, Retired		Pupil, Student, Military	Managerial Staff, Executive Manager, Liberal Profession	Salaried Staff	Worker, Employee	Craftsman, Commercial , Artist	Inactive, Jobless, Retired
	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
Total	155	86	11	21	14	19	12	9	69	9	16	12	18	12	2
3. Note															
Nb	138	77	8	21	11	16	12	9	61	6	16	10	15	12	2
M	4,95	5,61	4,00	6,00	7,82	5,75	3,50	6,00	4,11	-	4,06	4,70	6,00	3,25	5,00
SD	3,00	3,08	2,74	3,63	2,37	1,89	2,69	2,45	2,69	-	3,21	1,95	1,90	1,09	-
Col Sign. M	JN	GI	J	G	ABCFGL	G		G	J		J	JN	IJKNO	J	IJN

The following options are available:

Test each column against	Select what comparison will be made, as follows: **All columns:** all the columns of questions selected in column (independently of the question) will be compared with each other. For example , we will compare A/B, ….C/G, …H/K **All columns of the question** : All columns of the corresponding question will be compared with each other We will compare only between the Age categories (C/D) and between the CSP categories (E/F) **The total only**: Only the column Total (A) will be compared to the others columns. **All columns of the question and the corresponding responses edges:** all the columns of the same question only will be compared with each other and with the corresponding edge. For example , we will compare C/D or E/F or C/H **Previous column only:** all the columns of the same question or different questions will be compared to the previous column. **All columns of the question and the corresponding questions edges:** all the columns of the same question only will be compared with the corresponding edge. For example , we will compare C/H or D/I or E/J but we will not compare the C/D or the E/F **All columns of edge responses:** Only the columns within the edge response will be compared. For example, between B-C-D-E-F or between G-H-I-J-K **All columns of the current question and the first question:** All the columns of the first question will be compared to the all others columns. **Specify columns:** You can choose the columns to compare
Show letter in	Select the priority for the display of letters, as follows: • **Both columns:** the significance will appear in both columns • **First column:** the significance will appear in the first column • **Column with highest value:** the significance will appear in the column where the column profile is highest
Using	Specifies the type of test to be used: classical student test (A), student test using estimator (B), student test using efficiency coefficient (C), or student test using estimator and efficiency coefficient(D).
Test against total column	When this option is selected, the total column becomes a column like any other for the purposes of the calculation.
Display minus if under	When this option is selected, a minus sign is shown if the significancy goes down. • This can be used in conditional formatting when you test in conjunction with the column before: if there are two letters, the value has significantly gone down, if you have one, it has gone up.
Use student test when degrees of freedom <	If this option is selected, a student test will be used when the degrees of freedom are less than the amount stated in the adjacent box.
Use unweighted base	If this option is selected, col significativity will be carried out on weighted mean and unweighted counts.
Columns are assumed independent	When this option is selected, the individuals belonging to a sub-total will be considered different to those present in the category grouped in the same sub-total.
Count threshold	The minimum count to be taken into account in a cell.

Base threshold	The minimum base that must be met before column sig. testing is displayed. By default, the minimum base is 0. This option affects numeric questions.
High significativity (%)	The percentage at which values are to be regarded as highly significant.
Normal significativity (%)	The percentage at which values are to be regarded as of normal significance.
Low significativity (%)	The percentage at which values are to be regarded as of low significance.
Display "A+"	Mark highly significant values with *A+.*
Display "A"	Mark values of medium significance with *A.*
Display "a"	Mark values of low significance with a.
Test cols (A:B,C:D-F)	This option appears when you select **specify columns** in test **each column against**. It allows you to list specific columns to be compared against each other. Enter the letters or numbers of the columns you want to test. Separate each test with a comma, and use a colon to separate the columns to be compared within a test. Examples: • To test AvsB and DvsE only, enter A:B,D:E or 1:2,4:5 • To test A against B,C,D,E,F: enter A:B-F • To test A,B,C against A,B,C: enter A-C:A-C or A-C You can display the list of columns being compared in your table (e.g. the footer), by entering the keyword ??*ColSig??* in the appropriate field.
Sig message (use ??sig??)	Defines a message which indicates at which level the columns have been tested during col significativity testing. The token ??sig?? can then be used to display this message. Select a message in the drop-down list, or enter your own. The items you can place in the message are: • p1 = high % • p2 = normal % • p3 = low % • invp1 = 100 - high % • invp2 = 100 - normal % • invp3 = 100 - low % • p1_1 = high proba (0-1) • p1_2 = normal proba (0-1) • p1_3 = low proba (0-1) • invp1_1 = 1 - high proba (0-1) • invp1_2 = 1- normal proba (0-1) • invp1_3 = 1- low proba (0-1) For example, if the message is "Columns are tested at ??p1??", p1 will be replaced by the high significativity value.

A) Classical Student Test



This test describes the t-test using unpooled variance:

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{sd_1^2}{n_1} + \frac{sd_2^2}{n_2}}}$$

Where

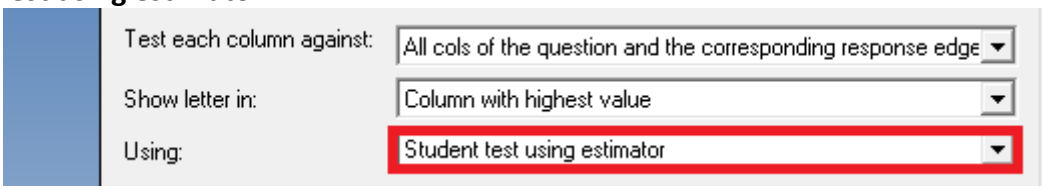
$\bar{x}_1$= Mean 1 observed in the sample
n_1=Sample 1 size
sd_1^2=Variance 1

$\bar{x}_2$= Mean 2 observed in the sample
n_2=Sample2 size
sd_2^2=Variance 2

We compare the T > tα

- $t_\alpha 90\%$ = 1.65
- $t_\alpha 95\%$ = 1.96
- $t_\alpha 99\%$= 2.576, If T > t_α then there is significative difference

B) Student Test using estimator



This test describes the t-test using pooled variance

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\frac{sd_1^2(n_1 - 1) + sd_2^2(n_2 - 1)}{n_1 + n_2 - 2}\right) \mathrm{X} \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

Where

$\bar{x}_1$= Mean 1 observed in the sample
n_1=Sample 1 size
sd_1^2=Variance 1

$\bar{x}_2$= Mean 2 observed in the sample
n_2=Sample2 size
sd_2^2=Variance 2

We compare the T value > tα

- $t_\alpha 90\%$ = 1.65
- $t_\alpha 95\%$ = 1.96
- $t_\alpha 99\%$= 2.576, If T > t_α then there is significative difference

C) Wilcoxon Rank test

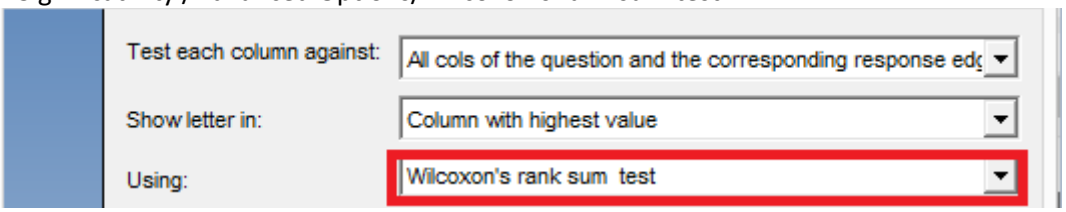
Note that this test is available only in the askiaanalyse 5.3.3.0. and +

We use the Wilcoxon Rank Test when we want to compare 2 means between 2 paired samples (e.g. before / after) and when the differences between pairs are severely non-normal distributed.

The output will display letters “a, A, A+”.

Base: 490 All interviews No weighting	LQ2	
	Before	After
	A	B
Total	49 100,0%	49 100,0%
Q2		
Nb	49	49
M	7,92	3,65
SD	1,34	1,84
Wilcoxon		A

Select Column Significativity /Advanced Options/Wilcoxon’s rank sum test:



Step 1: We calculate the difference between the 2 data sets, for each person i

$$diff_i = xi_a - xi_b$$

Step 2: Then we rank each absolute$|diff_i|$, excluding the $diff_i = 0$

Step 3: we allocate the sign of each Rank($diff_i \neq 0$)
Where N is the number of $diff_i \neq 0$

Step 4: We calculate the sum of positive Rank (W_+) and the sum of negative Rank (W_-)

$$W = Min(W_+, W_-)$$

If N ≥ 10

$$Z = \frac{W - \left(\frac{N(N+1)}{4}\right)}{\sqrt{\frac{N(N+0{,}5)(N+1)}{12}}}$$

If $Z > Z_{critical}$ then there is significative difference between the 2 means

Critical Values of ±z

Level of Significance for a				
Directional Test				
.05	.025	.01	.005	.0005
Non-Directional Test				
--	.05	.02	.01	.001
$z_{critical}$				
1.645	1.960	2.326	2.576	3.291

If N < 10

W is compared to a critical value from a reference table.

If $W > W_{critical}$ then there is significative difference between the 2 means.

Critical Values of ±W for Small Samples:

	Level of Significance for a			
	Directional Test			
	.05	.025	.01	.005
	Non-Directional Test			
N	--	.05	.02	.01
5	15	--	--	--
6	17	21	--	--
7	22	24	28	--
8	26	30	34	36
9	29	35	39	43