# askia analyse Significancy tests User guide



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.	Prop	portions comparison	3
	1.1.	Significance	4
	1.2.	Khi2 Value	9
	1.3.	Column Significativity	. 11
	1.4.	Z-Score	. 19
	1.5.	Paired preference test	. 20
2.	Mea	In comparison	. 22
	2.1.	Significance	. 23
	2.2.	Z-Score	. 26
	2.3.	Column Significativity	. 27

# **1. Proportions comparison**

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

You can select any closed question ( $^{\circ}$  single or  $^{\circ}$  multicoded or  $^{\circ}$  scale responses). Insert a closed question in row **and** in column in a tab definition

q				General	
r				Rows	
Questions	Value	Sort	Custom setting		-
🖃 🖥 👔 1. Overall film's appreciation		No			
✓ Tremendously	10	Yes			
Very much	7	Yes			
Medium	3	Yes			
Not very much	1	Yes			
✓ Not at all	0	Yes			
☑ Dk	NB	Yes	Caption		
Don't know	NB	No			
Not asked	NA	No			
1					1
				Columns	
		_			
Questions	Value	Sort	Lustom setting		-
E To 12 Age		No			
Less than 24		Yes			
25 - 34		Yes			
✓ 35 · 49		Yes			
✓ 50 and older		Yes			
Not asked		No			
1					
				Edges	-

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

This test is used on independent samples.

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

	NO	110	
Percentage down	No	No	
🔣 Overall percentage	No	No	
Significance	No	No	
Counts when independent	No	No	
Test value	No	No	
Affinity index	No	No	

	Advanced options	1
To define the parameters, click right on the significance calculation properties and click on	/ availood options	L-
To define the parameters, the nght on the significance calculation properties and the on		÷.,

Caption:	Significance					
Calculation type:	Significance 🗸					
Sub-population:	Default 👻					
Universe:	Default 🗸					
Weighting:	Default 🗸					
Level:	Default 🗸					
Visibility:	Always 🗸					
<ul> <li>Always</li> <li>Display a calculation on the intersection of rows and columns</li> <li>Display a calculation on the rows</li> <li>Display a calculation with the questions</li> <li>Display a calculation on the columns</li> <li>Display a calculation on the columns</li> <li>The calculations of the total are defined by the cross-tab</li> <li>Hide this calculation (but leave it available for arithmetic or conditional formatting)</li> </ul>						

And select the options as follows:

-1	Optio	ns on significativity ×
	Test against:	Counts when independent
	Count threshold:	5
civity	High significativity (%):	99
vity nificat	Normal significativity (%):	95
icati <sup>n</sup> nn sig	Low significativity (%):	90
Column signif	Standard deviation know	Wn OK Cancel

Test against	Select what comparison will be made, as follows:			
	<ul> <li>(A) Counts when independent: comparison of cell with total (X<sup>2</sup>);         <ul> <li>You can test the p<sub>i</sub> against the p observed in the total population N :</li> </ul> </li> <li>(B) All other columns: comparison of the column profile with the percentage obtained in the average other columns;         <ul> <li>You can test the p<sub>i</sub> against the p observed in the N - N<sub>i</sub> population :</li> </ul> </li> <li>(C) All other rows: comparison of the row profile with the percentage obtained in the other rows.         <ul> <li>You can test the p<sub>i</sub> against the p observed in the N - N<sub>i</sub> population :</li> </ul> </li> </ul>			
Count threshold	old 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 <u>Test Values</u> 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: $\chi^2$ )

The  $\chi^2\,$  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*.  $Total_i * Total_j$ 

$$x_{expected i} = \frac{10tut_i + 10t}{N}$$

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

And N: Total Sample size,

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

If the

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

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)		
$N_2 = N - N_1$				
$p_{obs(1,j)} = \frac{x_{obs1,j}}{N1}$				
	$p_2 = \frac{(\text{Total}(j) - \boldsymbol{x_{obs 2,j}})}{N_2}$	$p_2 = \frac{(\text{Total}(i) - \boldsymbol{x_{obs 2,j}})}{N_2}$		

#### D) If the standard deviation is known

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

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  $N(p_1 - p_2, sd)$  for

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

### 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:

Caption:	Mai 2 to at
Seption.	IN 12 Lesi
Calculation type:	Khi2 test
Sub-population:	Script (deviation)
Iniverse:	Script (standard error)
Shiveise.	Script (count)
Neighting:	Calculation anthmetic (number)
Level:	Calculation arithmetic (string)
Visibility:	Paired preference
<ul> <li>Display a calculation</li> <li>The calculations on</li> <li>Hide this calculation</li> </ul>	on on the intersection of rows and columns on on the rows on with the questions on on the columns f the total are defined by the cross-tab on (but leave it available for arithmetic or conditional formatting)

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

Khi <sup>2</sup>	??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 :

- P		Options on test
Khi2 options	Caption when dependent: Caption when independent: Percentage:	Dep: ??K?? - Dof : ??D?? - Proba: ??INVP??%           Indep: ??K?? - Dof : ??D?? - Proba: ??P??%           95           OK         Cancel

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:					

		Formattin	g		×
🗌 Use d	efault				
• Numb	er formatting				
C Script	formatting				
C Suffix	with script			Script	
C Script	then number forma	atting			_
Number	5				
Digits	after decimal point	t 1	-		
🗆 🗆 S	how % sign				
🗆 🗆 V	/rap text with parer	nthesis			
F F	eplace 0 by		Replace n/a by		-
F	eplace values <		by		
		Γ	ок	Cancel	
		L			

The  $\chi^2\,$  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*.  $x_{obs\,i} = \frac{Total_i * Total_j}{Total_i * Total_j}$ 

$$x_{expected i} = \frac{1}{N}$$

We calculated the  $\chi^2$  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	Total		i3. 9	Socio-profes	sional categ	ory	
All interviews No weighting							
NU Yregnung		Pupil, Student, Military	Managerial Staff, Executive Manager, Liberal Profession	Salaried Staff	Worker, Employee	Craftsman, Commercial , Artist	Inactive, Jobless, Retired
	A	В	С	D	E	F	G
Total	295	30	66	60	62	48	29
	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
1. Overall film's appreciation							
Tremendously	25	2	6	5	6	3	3
	9,7%	8,3%	10,2%	9,8%	10,7%	6,7%	13,0%
Very much	66	6	18	15	13	12	2
	25,6%	25,0%	30,5%	29,4%	23,2%	26,7%	8,7%
	g		G	G		g	
Medium	64	12	13	6	18	9	6
	24,8%	50,0%	22,0%	11,8%	32,1%	20,0%	26,1%
	D	A+CD+F+g			D		
Not very much	75	4	21	22	7	15	6
	29,1%	16,7%	35,6%	43,1%	12,5%	33,3%	26,1%
	E		bE+	ABE+		E	
				<b>4</b>		4	

To define where the letters in header will be displayed

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

Tab template: 🖉 Askia Crossed 🗨	D 🖬 🛍 🗸	<u>S</u> ettings
---------------------------------	---------	------------------

Totals					
	Show	At begining of tab	Caption		
Row		$\checkmark$	Total		
Question			Total		
Column			Total		
Edge		<b>v</b>	Total		
Captions					
Question cap	otion in tab:	[	Use default	•	
Question cap	otion in title:	Ĩ	Use default	•	
Question cap	otion in title for :	superpositions:	Use default	•	
Show ca	alculation capti	ons			
	captions on a	row on their own			
Edges o	n a column on	their own			
Empty ro	w after each g	uestion			
Empty ro	w after each re	esponse			
Empty c	olumn after eac	ch edge			
Show qu	lestion caption	s in rows			
Show qu	Jestion caption	s in columns			
Show re	sponse caption	s in edges hs in rows			
Change		ttor for cal size for			
Show a	column order le column order le	etter (for col signific etter in edge total c	column		
Show a	column order le	etter in total column	ı		
Show a	row order letter	r (for row significati	vity)		
Superpo	se responses t	or flat counts			

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

🔲 🖪 Counts when independent	No	No
🔲 🖪 Test value	No	No
🔄 💌 Affinity index	No	No
🗹 🔽 Column significativity	No	No
🗌 🖪 Z-score (closed)	No	No
🔲 🖪 Row significativity	No	No
🔲 🖪 Effective base	No	No

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 signifi	cativity properties / Adv	anced options
· · ·		Prope	erties	×
	General Total in rows Tot	tal in column		
	Caption:	Column significativity		
	Calculation type:	Column significativity		-
	Sub-population:	Default		-
	Universe:	Default		-
	Weighting:	Default		-
	Level:	Default		•
	Visibility:	Always		- C
	<ul> <li>Display a calculation on</li> <li>The calculations of the</li> <li>Hide this calculation (but Ranking (inactive))</li> </ul>	the intersection of rows an the rows th the questions the columns total are defined by the cro ut leave it available for arithr Advanced options	d columns sstab netic or conditional formatting) Number formatting	Specifiquer
				appindoer

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	12. /	Age	i3. Socio-pr cate	rofessional gory	Total	12.	Age	i3. Socio-pr cates	ofessional gory
			Less than 35	35 and older	CSP+	CSP -		Less than 35	35 and older	CSP+	CSP -
	A	В	С	D	E	F	G	Н	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%	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 M SD	258 3,79 3,22	140 3,89 3,32	70 3,66 3,31	70 4,13 3.32	108 3,89 3,26	32 3,91 3,50	118 3,68 3.09	60 3,77 3,33	58 3,59 2,81	103 3,78 3,22	15 3,00 1,79
 Sign, M Col Sign, M	5,66	5,52	5,51	5,52	5,20	2,20	5,05		2,01	5,66	

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: <b>All columns:</b> 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
	(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 F/F or C/H
	<b>Previous column only:</b> 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 <u>questions</u> edges: all the columns of the
	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
	Specify columns: You can choose the columns to compare
<b>a b b b b b b b b b b</b>	
Show letter in	Select the priority for the display of letters, as follows:
	<ul> <li>First column: the significance will appear in the first column</li> </ul>
	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	<ul> <li>When this option is selected, a minus sign is shown if the significancy goes down.</li> <li>This can be used in conditional formatting when you test in conjunction with the column</li> </ul>
	before: if there are two letters, the value has significantly gone down, if you have one, it has gone up.
Use student test	If this option is selected, a student test will be used when the degrees of freedom are less
when degrees of	than the amount stated in the adjacent box.
Use unweighted	If this antion is selected, selected is a significativity will be serviced out on weighted % and unweighted
	I TETHS ODTION IS SELECTED. COL SISTIMCATIVITA MILLAE CALLED ONE ON MEISINED 20 400 NUMERINED

Columns are	When this option is selected, the individuals belonging to a sub-total will be considered
assumed	different to those present in the category grouped in the same sub-total.
independent	
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	The percentage at which values are to be regarded as highly significant.
significativity (%)	
Normal	The percentage at which values are to be regarded as of normal significance.
significativity (%)	
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 sole (A:D C:D	This action supports when you colort modify columns in test each column anninet. It allows
r)	This option appears when you select <b>specify columns</b> in test <b>each column against</b> . It allows
F)	you to list specific columns to be compared against each other.
	Enter the latters or numbers of the columns you want to test. Congrate each test with a
	Enter the fetters of 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.
	Exumples:
	• To test A gagingt B C D E E: onter A:B E
	• To test A B C against A B C: enter $A = C + A = C$
	• To lest A, B, C uyumst A, B, C. enter A-C. A-C of A-C You can display the list of columns being compared in your table (e.g. the footer) by entering
	the keyword 22ColSig22 in the appropriate field
Sig message (use	Defines a message which indicates at which level the columns have been tested during col
27cig22)	significativity testing. The token 22 sig 22 can then he used to display this message
	Select a message in the dron-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 %
	<ul> <li>p1_1 = high proba ( 0-1)</li> </ul>
	<ul> <li>p1_2 = normal proba ( 0-1)</li> </ul>
	• p1_3 = low proba (0-1)
	<ul> <li>invp1_1 = 1 - high proba (0-1)</li> </ul>
	<ul> <li>invp1_2 = 1- normal proba (0-1)</li> </ul>
	• $invp1_3 = 1$ - low proba (0-1)
	For example, if the message is "Columns are tested at ??p1??", p1 will be replaced by the
	nign significativity value.

### A) Classical Student Test

	Options on column significativity							
	Test each column against:	All cols of the question and the corresponding response $edge$						
	Show letter in:	Column with highest value						
	Using:	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  $n_1$ =Sample 1 size

 $p_2$  = Proportion 2 observed in the sample  $n_2$ =Sample2 size

We compare the Z value >  $t\alpha$ 

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

### B) Student Test using estimator

Test each column against:	All cols of the question and the corresponding response $edge$
Show letter in:	Column with highest value
Using:	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  $n_1$ =Sample 1 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

 $p_2$  = Proportion 2 observed in the sample  $n_2$ =Sample2 size

- We compare the Z value >  $t\alpha$ 
  - $t_{\alpha}90\% = 1.65$
  - $t_{\alpha}95\% = 1.96$
  - $t_{\alpha}$  99%= 2.576, If Z >  $t_{\alpha}$  then there is significative difference

### C) Student Test using efficiency coefficient

Test each column against:	All cols of the question and the corresponding response edge $\checkmark$
Show letter in:	Column with highest value
Using:	Student test using efficiency coeficient

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.)

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

### Where

 $p_1$ = Proportion 1 observed in the sample  $n_1$ =Sample 1 size

 $p_2$ = Proportion 2 observed in the sample  $n_2$ =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 >  $t\alpha$ 

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

### D) Student Test using estimator and efficiency coefficient

Test each column against:	All cols of the question and the corresponding response edge $\checkmark$
Show letter in:	Column with highest value
Using:	Student test using estimator and efficiency coeficient 🔹

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.*)

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

### 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 $w_1$  = weight in the sample 1 per individual $w_2$  = weight in the sample 2 per individualWe compare the Z value > ta $t_a 90\% = 1.65$  $t_a 95\% = 1.96$  $t_a 0.5$ 

 $t_{\alpha}$ 99%= 2.576, If Z >  $t_{\alpha}$  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:

🗌 💌 Affinity index	No	No	
🗌 🖪 Column significativity	No	No	
🖊 🔝 Z-score (closed)	No	No	
🔄 🖪 Row significativity	No	No	
🗌 🗖 Effective base	No	No	

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

Base: 295 All interviews No weighting	Total		i3. 9	sional category			
		Pupil, Student, Military	Managerial Staff, Executive Manager, Liberal Profession	Salaried Staff	Worker, Employee	Craftsman, Commercial , Artist	Inactive, Jobless, Retired
	A	В	С	D	E	F	G
Total	295	30	66	60	62	48	29
	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%	100,0%
I2. Age							
Less than 24	42	2	19	12		6	3
	14,2%	6,7%	28,8%	20,0%	- 	12,5%	10,3%
	E		AEF	E		E	
		-1,25	3,83	1,43	-3,60	-0,38	-0,63

If the Z-score is  $\geq$  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.

eneral Total in rows	Total in column			
Caption:	Paired preference			
Calculation type:	Paired preference			-
Sub-population:	Default			•
Universe:	Default			
Weighting:	Default			-
Level:	Default			-
Visibility:	Always			-
<ul> <li>Display a calculation</li> <li>The calculations of</li> <li>Hide this calculation</li> <li>Ranking (inactive)</li> </ul>	n on the intersection of rows and on n on the rows n with the questions n on the columns the total are defined by the cross-t n (but leave it available for arithmet Advanced options	ab ic or conditional forma Number formatting	t <b>ting)</b>	
		ок	Annuler	Applique

Insert a calculation in General Section of the tab definition:

Then click on	Advanced options	and select the option:
		Options on paired preference
	First row:	1
	Second row:	2
	Display:	Sigma
	🔲 Use unweighte	d base
		OK Cancel
Note that you	can display the sig	ma or the probability
Display:	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

Z-score (closed)	No	No
🔲 🔳 Row significativity	No	No
🗹 🔽 Effective base	No	No

 $p_1$  is the percentage observed in the row 1

 $p_{\rm 2} {\rm is}$  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 wi	ith script (1) and Script (2)
	on({1}<-2, " T" {1},on({1}>2," R" {1}," " {1}))
Formatting	
<ul> <li>□ Use default</li> <li>○ Number formatting</li> <li>○ Script formatting</li> <li>○ Script then number formatting</li> <li>○ Script then number formatting</li> <li>○ Numbers</li> <li>○ Digits after decimal point:</li> </ul>	2
✓ Wrap text with parenthesis	
Replace 0 by     Replace n/a by     Replace values < by	1 This calculation
OK Cancel	

You will obtain the following table:

Base: 124 All interviews No weighting	Total	Age		
		18-35	36-65	
Total	124	62		
	100,0%	100,0%	100,0%	
Which product do you prefer ?				
Product R	59	25	34	
	47,6%	40,3%	54,8%	
Product T	65	37	28	
	52,4%	59,7%	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 (<sup>1</sup>/<sub>9</sub> numeric or <sup>1</sup>/<sub>9</sub> scale responses) Insert a **numeric question in row and** a **closed question in column** in a tab definition

### And then select the significancy test.

1				General
-				Rows
Questions	Value	Sort	Custom setting	
To 3. Note		No		
🖃 🍞 🔒 1. Overall film's appreciation		No		
Tremendously	10	Yes		
Very much	7	Yes		
Medium	3	Yes		
Not very much	1	Yes		
✓ Not at all	0	Yes		
I do not know	NB	Yes	Caption	
🗖 Don't know	NB	No		
1 <b>H</b> arris	***			
1				Lolumns
Questions	Value	Sort	Custom setting	
E 1012. Age		No		
Less than 24		Yes		
		Yes		
		Yes		
☑ 50 and older		Yes		
Not asked		No		
1				
e de la companya de la				Edges

## **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	Total i3. Socio-professional category						
All interviews No weighting		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							
м	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)

🗖 🖪 Variance	No	No
🗖 🗖 Median	No	No
🗹 🔳 Significance	No	No
🔝 🔼 Column significativity	No	No
🔲 🔳 Z-score (numeric)	No	No

To define the parameters, click right on the significance calculation properties and click on

Advanced options...

Caption:       Significance         Calculation type:       Significance         Sub-population:       Default         Universe:       Default         Weighting:       Default         Level:       Default         Visibility:       Always         Image: Display a calculation on the intersection of rows and columns         Image: Display a calculation on the questions         Image: Display a calculation on the columns         Image: Display	eneral Total in rows	Total in column	
Calculation type:       Significance <ul> <li>Sub-population:</li> <li>Default</li> <li>Always</li> </ul> <li>Display a calculation on the intersection of rows and columns</li> <li>Display a calculation on the columns</li> <li>Display a calculation on the columns</li> <li>Display a calculation on the columns</li> <li>The calculation (but leave it available for arithmetic or conditional formatting)</li> <li>Ranking (inactive)</li> <li>Advanced options</li> <li>Number formatting</li>	Caption:	Significance	
Calculation type:       Significance <ul> <li>Sub-population:</li> <li>Default</li> <li>Di</li></ul>		par-management	
Sub-population:       Default          Universe:       Default          Weighting:       Default          Level:       Default          Visibility:       Always          Image: Population on the intersection of rows and columns          Display a calculation on the rows          Display a calculation on the questions          Display a calculation on the columns          The calculations of the total are defined by the cross-tab          Hide this calculation (but leave it available for arithmetic or conditional formatting)          Ranking (inactive)       Advanced options       Number formatting       Script	Calculation type:	Significance	-
Universe:       Default          Weighting:       Default          Level:       Default          Visibility:       Always          ✓       Display a calculation on the intersection of rows and columns          ✓       Display a calculation on the rows          ✓       Display a calculation on the questions          ✓       Display a calculation on the columns          ✓       The calculations of the total are defined by the cross tab          Hide this calculation (but leave it available for arithmetic or conditional formatting)          Ranking (inactive)       Advanced options       Number formatting       Script	Sub-population:	Default	-
Weighting:       Default          Level:       Default          Visibility:       Always          Image: Second Sec	Universe:	Default	-
Level:       Default          Visibility:       Always          Image: Display a calculation on the intersection of rows and columns          Display a calculation on the rows          Display a calculation on the rows          Display a calculation with the questions          Display a calculation on the columns          The calculations of the total are defined by the cross+ab          Hide this calculation (but leave it available for arithmetic or conditional formatting)          Ranking (inactive)       Advanced options       Number formatting	Weighting:	Default	-
Visibility:       Always       Image: Always         Display a calculation on the intersection of rows and columns       Display a calculation on the rows         Display a calculation on the rows       Display a calculation with the questions         Display a calculation on the columns       Display a calculation on the columns         Display a calculation on the columns       The calculations of the total are defined by the cross-tab         Hide this calculation (but leave it available for arithmetic or conditional formatting)       Script	Level:	Default	-
<ul> <li>Display a calculation on the intersection of rows and columns</li> <li>Display a calculation on the rows</li> <li>Display a calculation with the questions</li> <li>Display a calculation on the columns</li> <li>The calculations of the total are defined by the cross tab</li> <li>Hide this calculation (but leave it available for arithmetic or conditional formatting)</li> <li>Ranking (inactive)</li> <li>Advanced options</li> <li>Number formatting</li> </ul>	Visibility:	Always	-
	<ul> <li>Display a calculati</li> <li>The calculations of Hide this calculation</li> <li>Ranking (inactive)</li> </ul>	on on the intersection of rows and columns on on the rows on with the questions on on the columns of the total are defined by the cross tab on (but leave it available for arithmetic or conditional formatting) 	

And select the options as follows:

	Options on significativity									
	Test against:	All other columns	•							
	Count threshold:	5								
tivity	High significativity (%):	99								
vity mifical	Normal significativity (%):	95								
ficati mn sig	Low significativity (%):	90								
Column signi	Standard deviation know	wn OK Canc	el							

Test against	Select what comparison will be made, as follows:						
	<ul> <li>(A) All other columns: comparison of the mean of the column with the mean obtained in the other columns;         <ul> <li>✓ You can test the x             i against the x             observed in the N - N             i population :</li> </ul> </li> <li>(B) All other rows: comparison of the mean of the row with the mean obtained in the other rows.         <ul> <li>✓ You can test the x             i against the x             observed in the N - N             i population :</li> <li>✓ You can test the x             i against the x             observed in the N - N             i population :</li> <li>NB : The option "Count when independent" is useless (the A) test will be applied)</li> </ul> </li> </ul>						
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 <u>Test Values</u> 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 ( $\overline{x}_1 - \overline{x}_2 = 0$ ,  $sd = \sqrt{\frac{sd_1^2}{N_1} + \frac{sd_2^2}{N_2}}$ )

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

 $\overline{x}_{1} = \frac{\sum_{i=1}^{N_{1}} x_{obsi}}{N_{1}} = \text{mean 1 in the sample 1}$   $\overline{x}_{2} = \frac{\sum_{i=1}^{N_{2}} x_{obsi}}{N_{2}} = \text{mean 2 in the sample 2 (all others columns/rows)}$ 

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

We compare the  $abs(D) > t\alpha$ 

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

if  $abs(D) > t_{\alpha}$ , 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)):

Column significativity	No	No
🗹 🔀 Z-score (numeric)	No	No
🔝 🔝 Standard deviation (estimator)	No	No
🔲 🖪 Mode	No	No

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	В	С	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										
ΝЬ	138	14	37	21	31	24	11			
м	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	В	С	D	E	F	G			
Total	155	20	37	26	37	24	11			
3. Note										
ΝЬ	138	14	37	21	31	24	11			
м	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

	<ul> <li>✓ Show question captions in edges</li> <li>✓ Show response captions in rows</li> </ul>
ſ	<ul> <li>Show a column order letter (for col significativity)</li> <li>Show a column order letter in edge total column</li> <li>Show a column order letter in total column</li> <li>Show a row order letter (for row significativity)</li> </ul>
	✓ Superpose responses for flat counts

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

	🔲 🔳 Variance	No	No
	🔲 🔳 Median	No	No
	Significance	No	No
	🗹 🖪 Column significativity	No	No
Т	🛄 🔯 Z-score (numeric)	No	No
	🔲 💌 Standard deviation (estimator)	No	No
	🔲 🔳 Mode	No	No

Advanced options

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

To define th	e parameters, click righ	t on the Column signifi	icativity properties /	<u>.</u>								
		Prop	erties	×								
	General Total in rows Total in column											
-	Caption: Column significativity											
	Calculation type:	Column significativity		-								
	Sub-population:	Default		-								
	Universe:	Default	Default									
	Weighting:	Default	Default Default									
	Level:	Default										
	Visibility:	Always	Always									
-	<ul> <li>Display a calculation on the intersection of rows and columns</li> <li>Display a calculation on the rows</li> <li>Display a calculation with the questions</li> <li>Display a calculation on the columns</li> <li>Display a calculation on the columns</li> <li>The calculations of the total are defined by the cross-tab</li> <li>Hide this calculation (but leave it available for arithmetic or conditional formatting)</li> <li>Ranking (inactive)</li> <li>Advanced options</li> <li>Number formatting</li> </ul>											
				P 1								
			OK Annuler Ap	pliquer								
			7									

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	lotal	Total						1	1. Respond	lant's geno	ler					
					Man							Woman				
		Total	iotal 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	В	С	D	E	F	G	н	I	1	K	L	М	N	0	
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	
м	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	1	G	ABCFGL	G		G	J		3	NC	IJKNO	J	IJN	

Test each column	Select what comparison will be made, as follows:				
against	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 colu					
same question only will be compared with each other and with the corresponding					
example , we will compare C/D or E/F or C/H					
Previous column only: all the columns of the same question or different question					
	compared to the previous column.				
All columns of the question and the corresponding <u>questions</u> edges: all the col					
	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 eage responses: Only the columns within the edge response will be					
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					
	Speen y commist. Fou can enouse the commisto compare				
Show letter in	Select the priority for the display of letters, as follows:				
	• <b>Both columns:</b> 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	When this option is selected, the total column becomes a column like any other for the				
column	purposes of the calculation.				
Display minus if	When this option is selected, a minus sign is shown if the significancy goes down.				
under	<ul> <li>This can be used in conditional formatting when you test in conjunction with the</li> </ul>				
	column before: if there are two letters, the value has significantly gone down, if you				
	have one, it has gone up.				
Use student test	If this option is selected, a student test will be used when the degrees of freedom are less				
when degrees of	ees of than the amount stated in the adjacent box.				
freedom <					
Use unweighted	If this option is selected, col significativity will be carried out on weighted mean and				
base	unweighted counts.				
Columns are	When this option is selected, the individuals belonging to a sub-total will be considered				
assumed	different to those present in the category grouped in the same sub-total.				
independent					
Count threshold	The minimum count to be taken into account in a cell.				

# The following options are available:

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	The percentage at which values are to be regarded as highly significant.				
significativity (%)					
Normal	The percentage at which values are to be regarded as of normal significance.				
significativity (%)	ificativity (%)				
Low significativity (%)	<b>Itivity</b> 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	This option appears when you select <b>specify columns</b> in test <b>each column against</b> . It allows				
(A:B,C:D-F)	you to list specific columns to be compared against each other.				
	<ul> <li>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.</li> <li>Examples: <ul> <li>To test AvsB and DvsE only, enter A:B,D:E or 1:2,4:5</li> <li>To test A against B,C,D,E,F: enter A:B-F</li> <li>To test A,B,C against A,B,C: enter A-C:A-C or A-C</li> </ul> </li> <li>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.</li> </ul>				
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: <ul> <li>p1 = high %</li> <li>p2 = normal %</li> <li>p3 = low %</li> <li>invp1 = 100 - high %</li> <li>invp2 = 100 - normal %</li> <li>invp3 = 100 - low %</li> <li>p1_1 = high proba ( 0-1)</li> <li>p1_2 = normal proba ( 0-1)</li> <li>invp1_1 = 1 - high proba ( 0-1)</li> <li>invp1_2 = 1 - normal proba ( 0-1)</li> <li>invp1_3 = 1 - low proba ( 0-1)</li> </ul> For example, if the message is "Columns are tested at ??p1??", p1 will be replaced by the bigh 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

 $\overline{x}_1$ = Mean 1 observed in the sample  $n_1$ =Sample 1 size  $sd_1^2$ =Variance 1 We compare the T > t $\alpha$ 

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

### B) Student Test using estimator

Test each column against:	All cols of the question and the corresponding response edge $\checkmark$
Show letter in:	Column with highest value
Using:	Student test using estimator 🔹

This test describes the t-test using pooled variance

$$T = \frac{x_1 - x_2}{\sqrt{\left(\frac{sd_1^2(n_1 - 1) + sd_2^2(n_2 - 1)}{n_1 + n_2 - 2}\right) X\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

### Where

 $\overline{x}_1$  = Mean 1 observed in the sample  $n_1$  = Sample 1 size  $sd_1^2$  = Variance 1 We compare the T value > t $\alpha$ 

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

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

 $\overline{x}_2$  = Mean 2 observed in the sample

 $n_2$ =Sample2 size

 $sd_2^2$ =Variance 2

### 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 <u>2 paired samples</u> (e.g. before / after) and when the differences between pairs are <u>severely non-normal distributed</u>.

The output will display letters "a, A, 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( $dif f_i \neq 0$ ) Where N is the number of  $dif f_i \neq 0$ 

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

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

If N  $\geq$  10

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

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 <sub>critical</sub>						
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						
Ν		.05	.02	.01			
5	15						
6	17	21					
7	22	24	28				
8	26	30	34	36			
9	29	35	39	43			