Example 1

- Import dataset '5_GEM 2010 APS Croatia.sta' in Statistica and calculate:
- Frequency table of variable number 155: gender. Interpret it. Does it seem right?
- Descriptive statistics of variable number 156: age. Interpret it.

Note: description of the variables you can find in 4_variables description_GEM_2010.xlsx (sheet: APS Croatia):

Guidelines for making a solution:

Import 5_GEM 2010 APS Croatia.sta

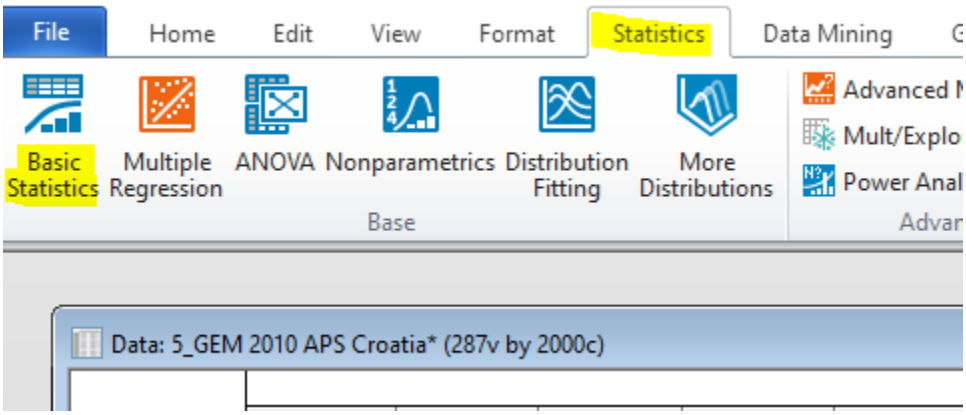
1. Make a frequency table of variable 155. Interpret it.
2. Make a descriptive statistics of variable 156. Interpret it.

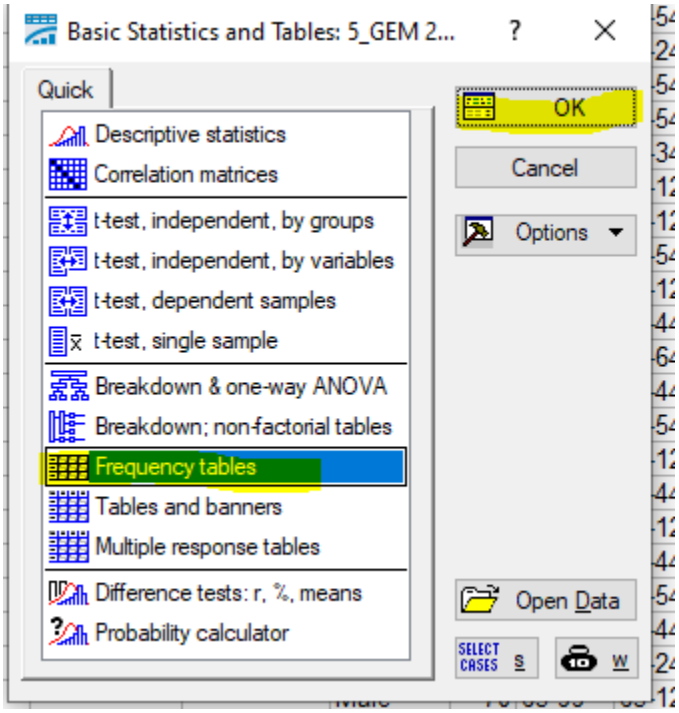
Solution:

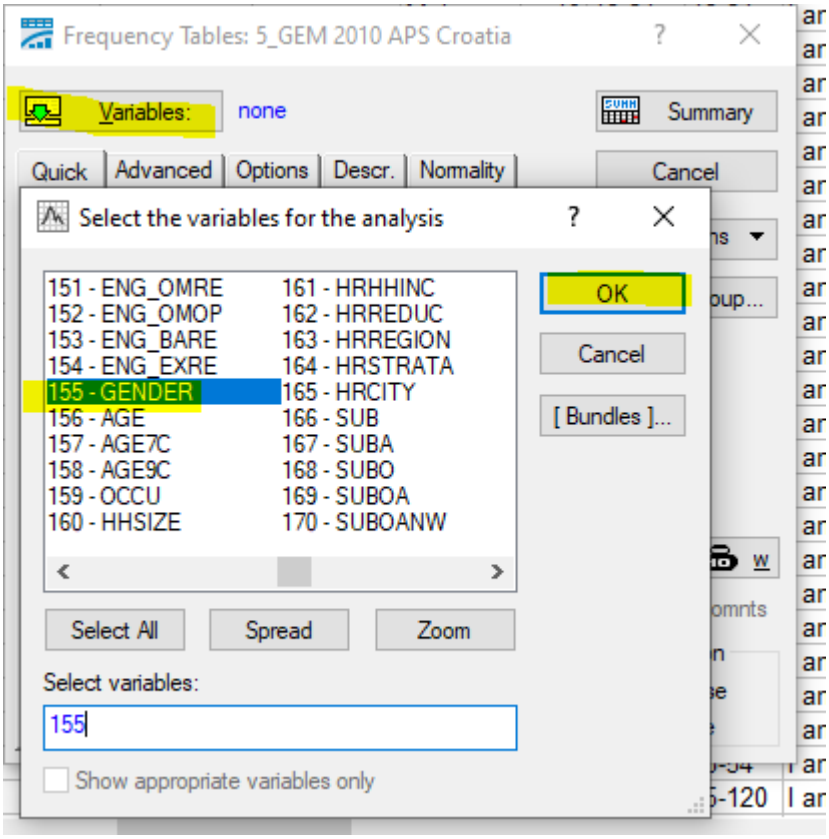
Import: File - Open - 5_GEM 2010 APS Croatia.sta

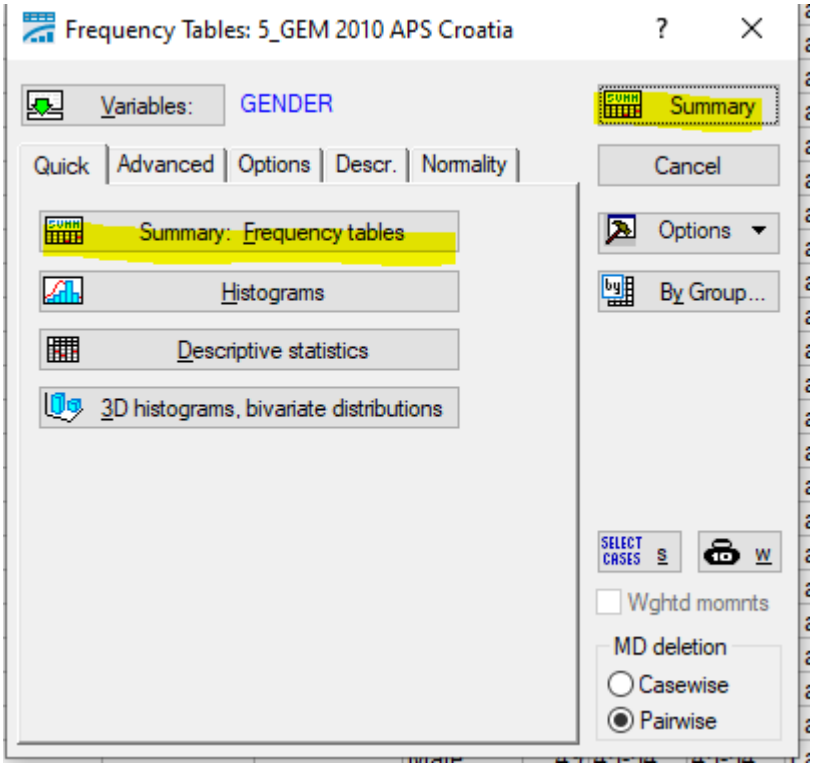
1. Make a frequency table of variable 155. Interpret it.

Frequency table: Statistics - Basic statistics and tables - Frequency tables - Variables: 155 Gender




Basic Statistics and Tables: 5_GEM 2...
Quick
Descriptive statistics
Correlation matrices
t-test, independent, by groups
t-test, independent, by variables
t-test, dependent samples
t-test, single sample
Breakdown & one-way ANOVA
Breakdown; non-factorial tables
Frequency tables
Tables and banners
Multiple response tables
Difference tests: r, %, means
Probability calculator
OK
Cancel
Options
Open Data


Frequency Tables: 5_GEM 2010 APS Croatia
Variables:
none
Summary
Quick
Advanced
Options
Descr.
Normality
Cancel
Select the variables for the analysis
151 - ENG_OMRE
152 - ENG_OMOP
153 - ENG_BARE
154 - ENG_EXRE
155 - GENDER
156 - AGE
157 - AGE7C
158 - AGE9C
159 - OCCU
160 - HHSIZE
161 - HRHHINC
162 - HRREDUC
163 - HRREGION
164 - HRSTRATA
165 - HRCITY
166 - SUB
167 - SUBA
168 - SUBO
169 - SUBOA
170 - SUBOANW
OK
Cancel
[Bundles]...
Select All
Spread
Zoom
Select variables:
155
Show appropriate variables only

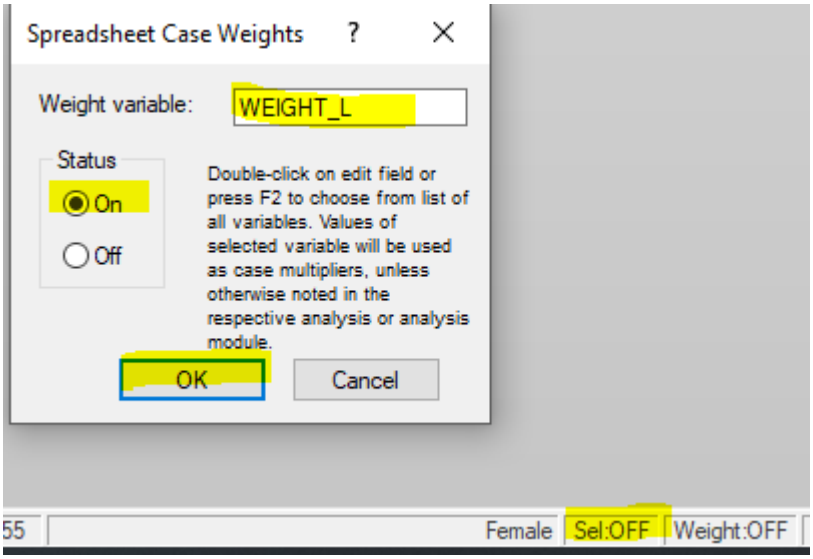


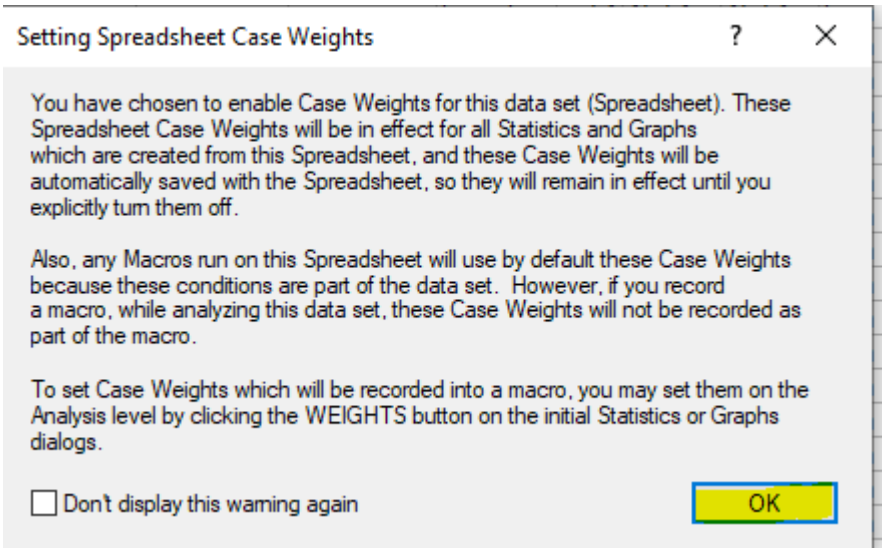
Category	Frequency table: GENDER: DEMA. What is your gender? (5_GEM 2010 APS Croatia.sta)			
	Count	Cumulative Count	Percent	Cumulative Percent
Male	775	775	38.75000	38.7500
Female	1225	2000	61.25000	100.0000
Missing	0	2000	0.00000	100.0000

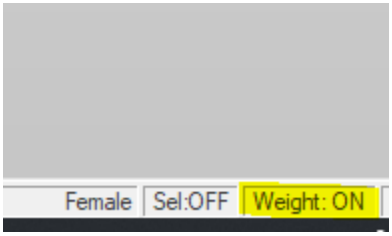
Interpretation: 38,75 % males and 61,25% females

weighing:

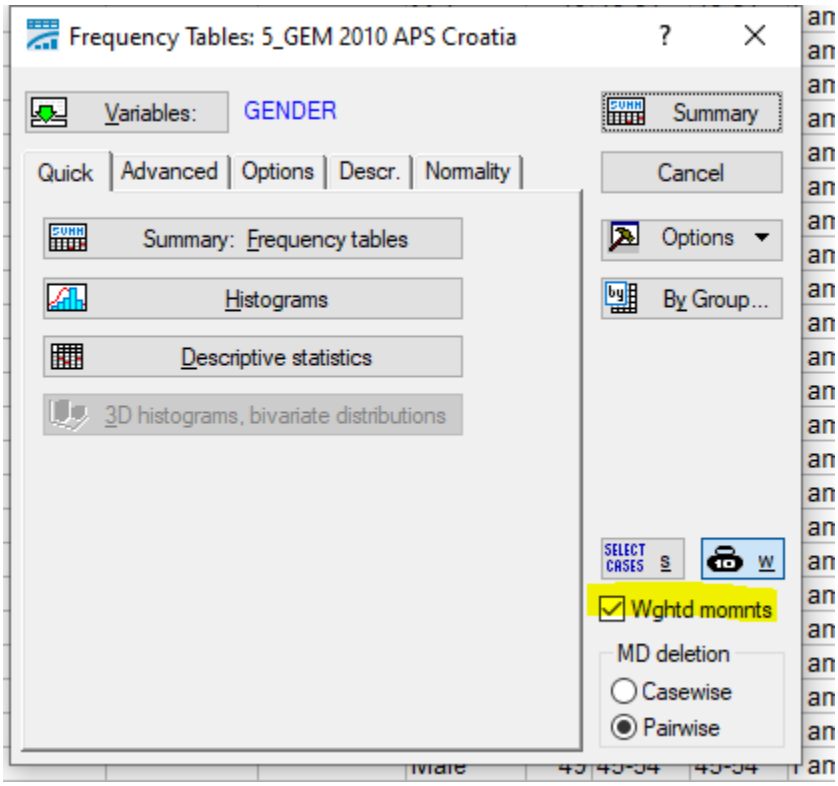
1. click on weight (bottom right corner), click 'On', weight variable: weight_l


Spreadsheet Case Weights
Weight variable:
WEIGHT_L
Status
On
Off
Double-click on edit field or press F2 to choose from list of all variables. Values of selected variable will be used as case multipliers, unless otherwise noted in the respective analysis or analysis module.
OK
Cancel
55
Female
Sel:OFF
Weight:OFF


Setting Spreadsheet Case Weights
You have chosen to enable Case Weights for this data set (Spreadsheet). These Spreadsheet Case Weights will be in effect for all Statistics and Graphs which are created from this Spreadsheet, and these Case Weights will be automatically saved with the Spreadsheet, so they will remain in effect until you explicitly turn them off.
Also, any Macros run on this Spreadsheet will use by default these Case Weights because these conditions are part of the data set. However, if you record a macro, while analyzing this data set, these Case Weights will not be recorded as part of the macro.
To set Case Weights which will be recorded into a macro, you may set them on the Analysis level by clicking the WEIGHTS button on the initial Statistics or Graphs dialogs.
Don't display this warning again
OK


Female
Sel:OFF
Weight: ON

2. in frequency table window click wghtd momnts

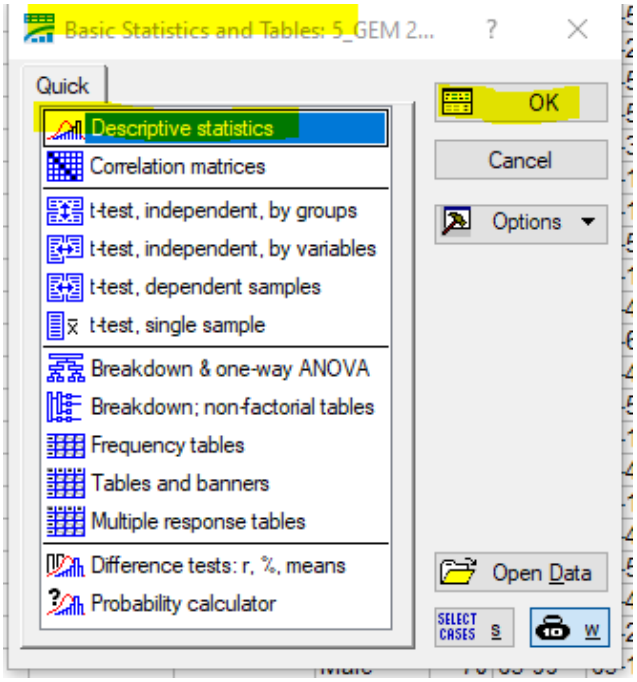


Category	Frequency table: GENDER: DEMA. What is your gender? (5_GEM 2010 APS Croatia.sta)			
	Count	Cumulative Count	Percent	Cumulative Percent
Male	800.6660	800.666	49.60756	49.6076
Female	813.3340	1614.000	50.39244	100.0000
Missing	0.0000	1614.000	0.00000	100.0000

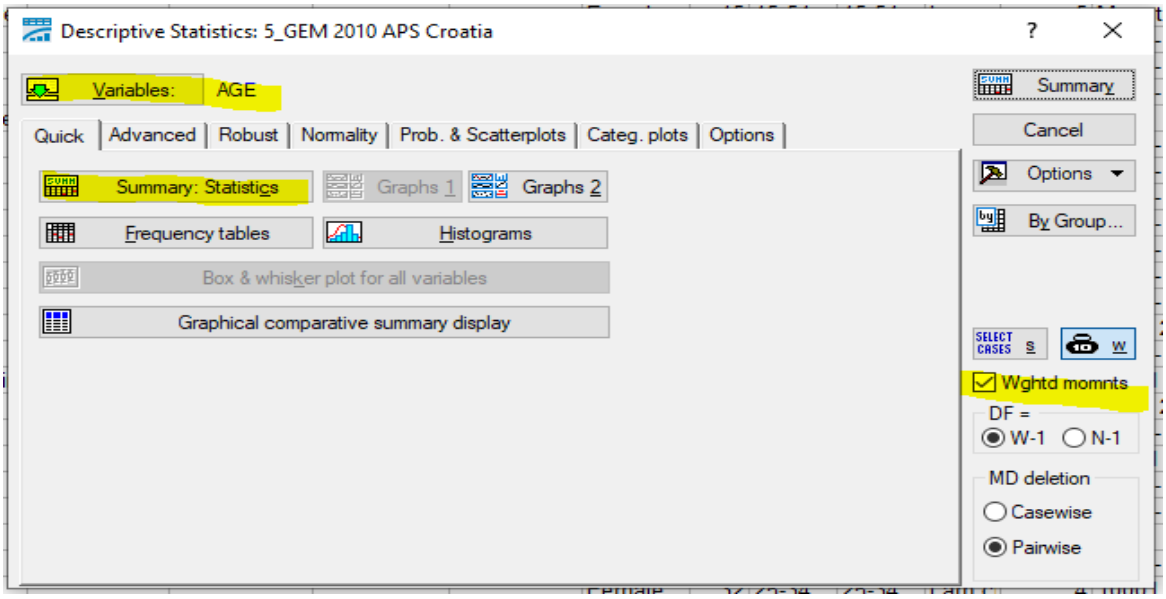
Interpretation: 49,6 % males and 50,4% females

2. Make a descriptive statistics of variable 156. Interpret it.

Descriptive statistics: Statistics - Basic statistics and tables - Descriptive statistics - Variables: 156 age



in descriptive statistics window click wghtd momnts



Variable	Descriptive Statistics (5_GEM 2010 APS Croatia.sta)				
	Sum of w	Mean	Minimum	Maximum	Std.Dev.
AGE	1610.850	40.33982	18.00000	64.00000	13.06047

Average age of interviewed population is 40.33 (working age population).

Example 2

- Import dataset '5_GEM 2010 APS Croatia.sta' in Statistica and calculate:
- Frequency table of variable number 188: frfail10. Interpret it.
- Descriptive statistics of variable number 193: tea10. Interpret it.

Note: description of the variables you can find in 4_variables description_GEM_2010.xlsx (sheet: APS Croatia):
Note: do not forget to turn-on weight_l and wghtd momnts

Guidelines for making a solution:

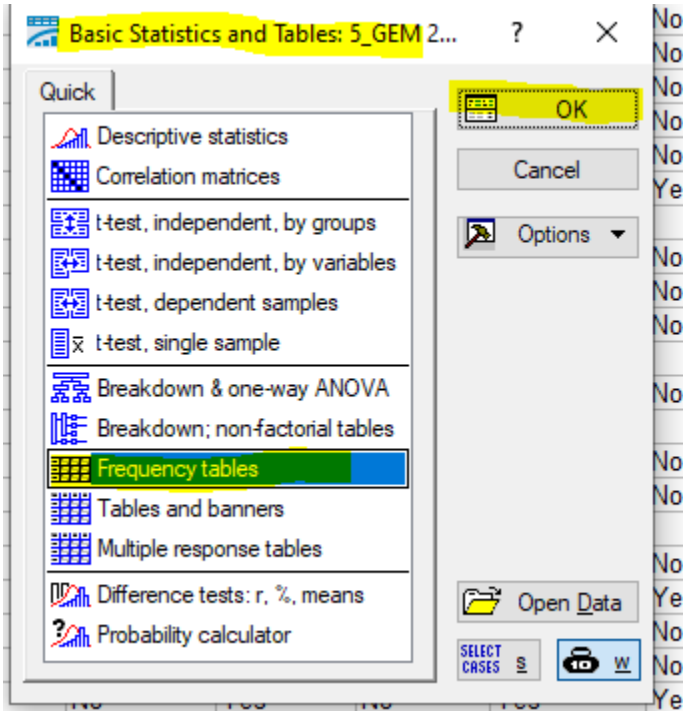
Import 5_GEM 2010 APS Croatia.sta

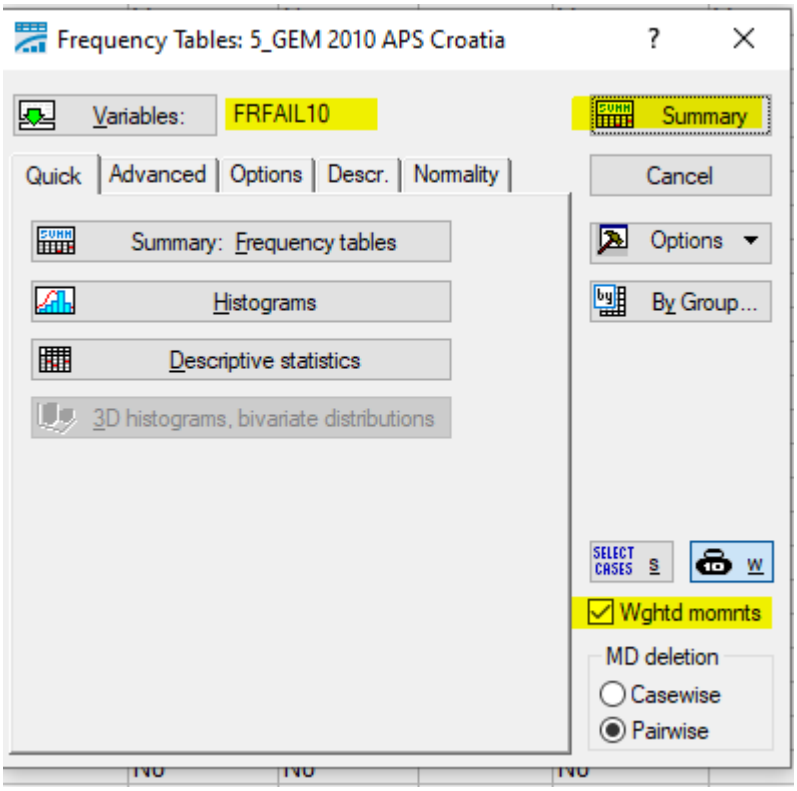
1. Make a frequency table of variable 188. Interpret it.
2. Make a descriptive statistics of variable 193. Interpret it.

Solution

1. Make a frequency table of variable 188. Interpret it.

Frequency table: Statistics - Basic statistics and tables - Frequency tables - Variables: 188



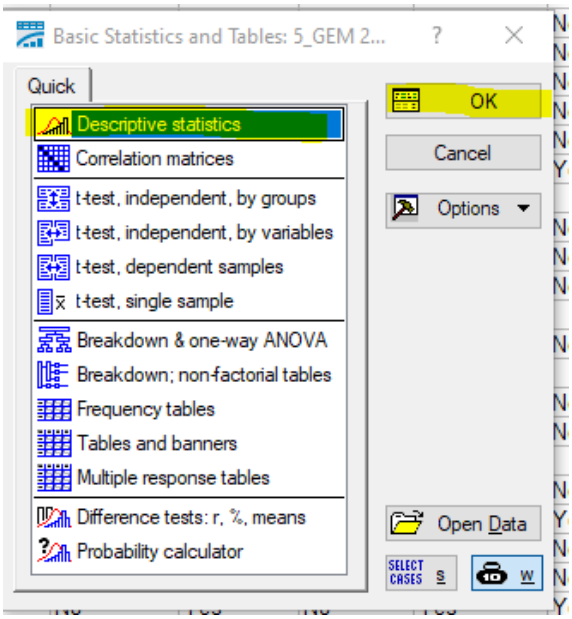


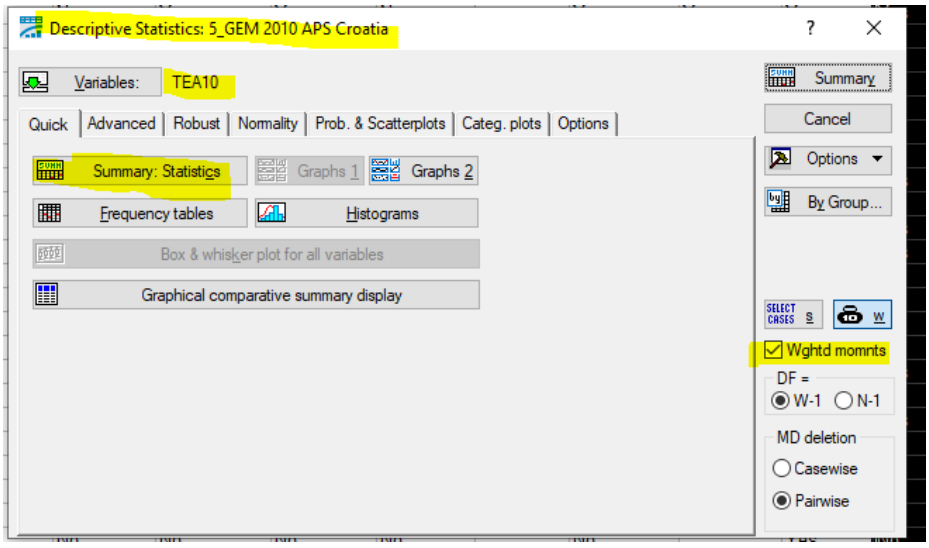
Category	Frequency table: FRFAIL10: FRFAIL adapted to make it fit for national level agg (5_GEM 2010 APS Croatia.sta)			
	Count	Cumulative Count	Percent	Cumulative Percent
No	979.0269	979.027	60.65842	60.6584
Yes	632.3383	1611.365	39.17833	99.8368
Missing	2.6348	1614.000	0.16324	100.0000

60,7% of the working age population have no fear of failure while 39,2% have fear of failure in starting a business.

2. Make a descriptive statistics of variable 193. Interpret it.

Descriptive statistics: Statistics - Basic statistics and tables - Descriptive statistics - Variables: 193





Variable	Descriptive Statistics (5_GEM 2010 APS Croatia.sta)				
	Sum of w	Mean	Minimum	Maximum	Std.Dev.
TEA10	1614.000	0.055188	0.00	1.000000	0.228418

There are 5,5 individuals on 100 individuals between 18 and 64 years of age who are entrepreneurially active.

Example 3

- Import dataset '6_GEM 2010 APS Master.sta' in Statistica and calculate:
- Descriptive statistics for TEA index, TEA male, TEA female for all countries in the data set (var.number 37, 38, 39)

Example 4

- We have read somewhere that the average age of entrepreneurial active person is 43 years. We want to check if this is true for Croatia.
- Use APS Croatia data set.

Guidelines for making a solution:

Import 5_GEM 2010 APS Croatia.sta

1. Set up hypotheses.
2. In the sample select entrepreneurially active people.
3. Test the hypotheses in Statistica.

Solution

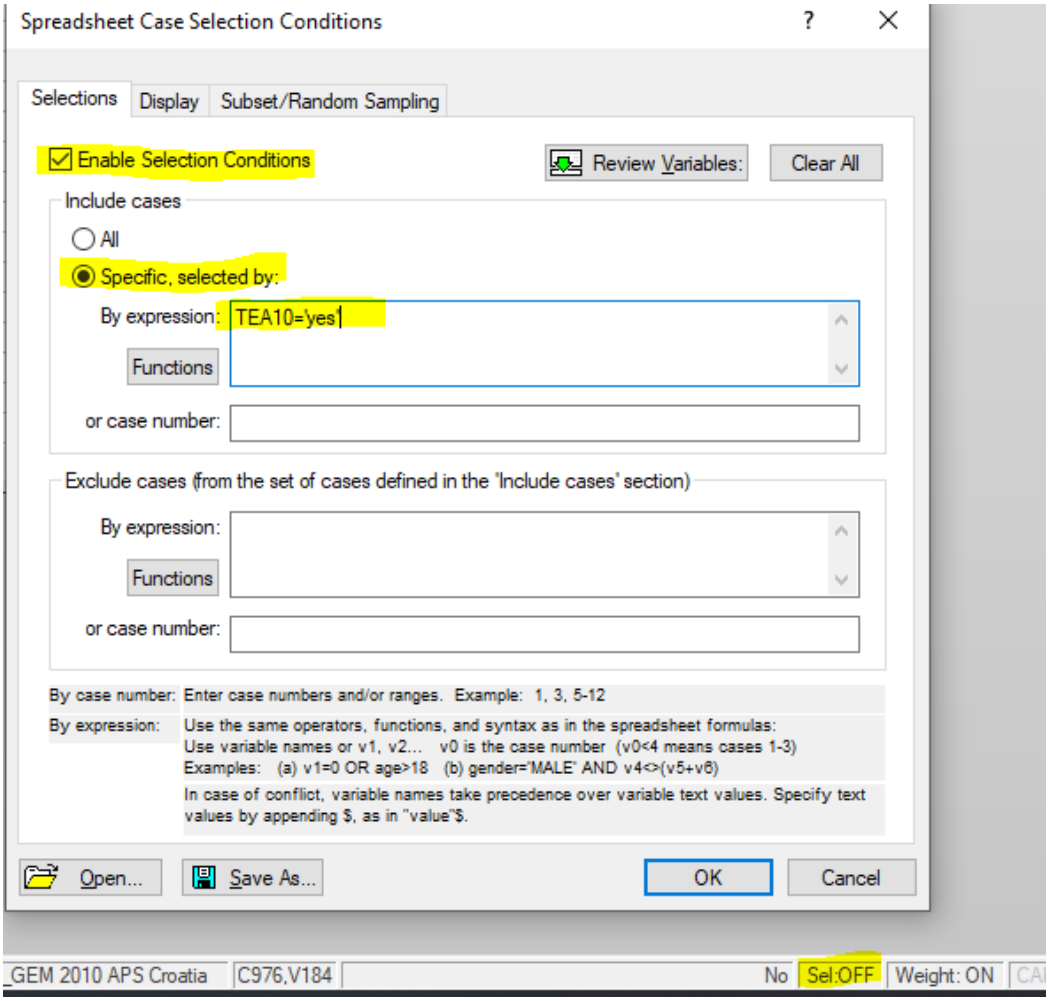
1. Set up hypotheses.

$H_0 : \mu = \mu_0 = 43$

$H_1 : \mu \neq \mu_0 = 43$

2. In the sample select entrepreneurially active people.

In the bottom right corner click 'Sel' and in the window that opens select 'Enable selection conditions' and write: TEA10='yes'

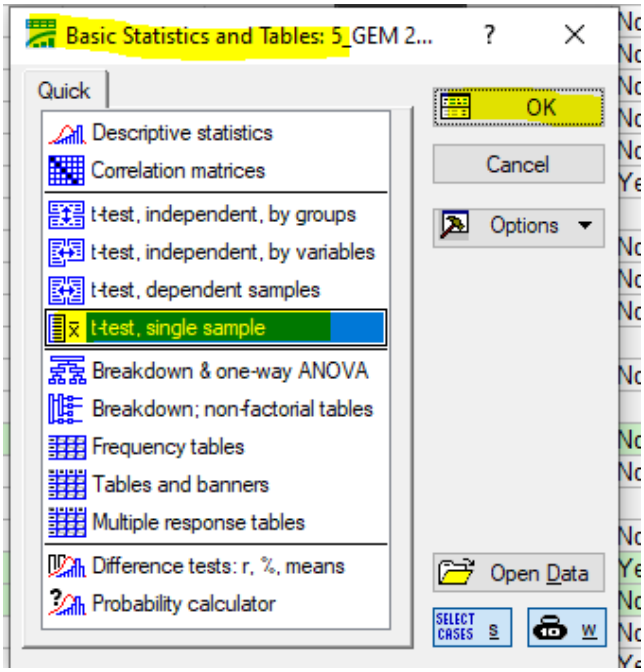

Spreadsheet Case Selection Conditions
Selections
Display
Subset/Random Sampling
Enable Selection Conditions
Review Variables:
Clear All
Include cases
All
Specific, selected by:
By expression:
TEA10='yes'
Functions
or case number:
Exclude cases (from the set of cases defined in the 'Include cases' section)
By expression:
Functions
or case number:
By case number: Enter case numbers and/or ranges. Example: 1, 3, 5-12
By expression: Use the same operators, functions, and syntax as in the spreadsheet formulas:
Use variable names or v1, v2... v0 is the case number (v0<4 means cases 1-3)
Examples: (a) v1=0 OR age>18 (b) gender='MALE' AND v4<>(v5+v6)
In case of conflict, variable names take precedence over variable text values. Specify text values by appending $, as in "value"$.
Open...
Save As...
OK
Cancel
_GEM 2010 APS Croatia
C976,V184
No
Sel:OFF
Weight: ON

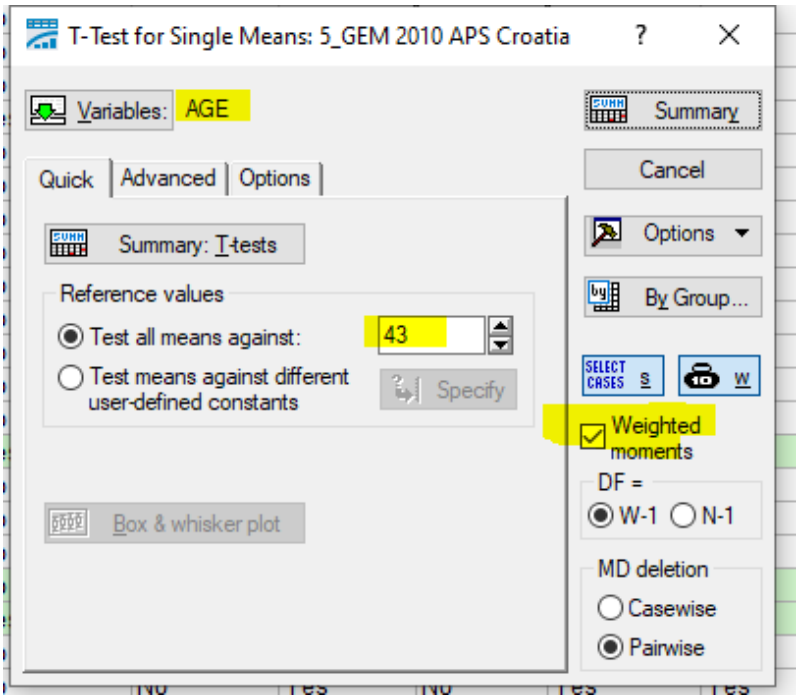

No
Sel: ON
Weight: ON

3. Test the hypotheses in Statistica.

Statistics - Basic statistics and tables - t-test single sample - Variables: 156 age - test all means against: 43;

click wghtd momnts





Variable	Test of means against reference constant (value) (5_GEM 2010 APS Croatia.sta) Include condition: tea10='yes'							
	Mean	Std.Dv.	Sum of w	Std.Err.	Reference Constant	t-value	df	p
AGE	38.56364	9.984640	87.81843	1.065466	43.00000	-4.16377	86.81843	0.000074

The results show that $p < \alpha$ (0,05) and we should reject H_0.

We conclude that average age of entrepreneurial active person in Croatia is not equal 43. We can notice from the table that she/he is younger.

Example 5

- In the dataset '6_GEM 2010 APS Master' in the variable OPPORT10 (13) there are percentages of people for 60 countries who think that there will be a good business opportunities in the next 6 months.
- Make desctiptive statistics of the variable.
- Check if mean value is statistically different than 50%. What does it mean?

Guidelines for making a solution:

Import 6_GEM 2010 APS Master.sta

1. Make a descriptive statistics of variable 13.
2. Set up hypotheses.
3. Test the hypotheses in Statistica. Make a conclusion.

Solution

1. Make a descriptive statistics of variable 13: Statistics - Basic statistics and tables - Descriptive statistics - Variables: 13

Variable	Descriptive Statistics (6_GEM 2010 APS Master.sta)				
	Valid N	Mean	Minimum	Maximum	Std.Dev.
OPPORT10	60	43.10467	5.919197	81.35810	18.27157

43,1% of people in all GEM countries in 2010 believe that there will be a good business opportunities in the next 6 months.

2. Set up hypotheses.

$H_0 : \mu = \mu_0 = 50$

$H_1 : \mu \neq \mu_0 = 50$

3. Test the hypotheses in Statistica.

Variable	Test of means against reference constant (value) (6_GEM 2010 APS Master.sta)							
	Mean	Std.Dv.	N	Std.Err.	Reference Constant	t-value	df	p
OPPORT10	43.10467	18.27157	60	2.358849	50.00000	-2.92318	59	0.004907

$p<0,05$ - reject H_0

We can conclude that percentage of people who think there will be good business opportunities in the next 6 months is not equal 50%. There are less than 50% who think there will be a good business opportunities.

Example 6

- I believe that in Croatia there is at least 50% of the population who agree with the statement that you will often see stories in the public media about successful new businesses.
- Am I right? Check the data and let me know.
  (5_GEM 2010 APS Croatia.sta)

Example 7

- We want to test whether there is a difference in percentage of fear of failure (188) between TEA active and TEA non-active (193).
  (5_GEM 2010 APS Croatia.sta)

Guidelines for making a solution:

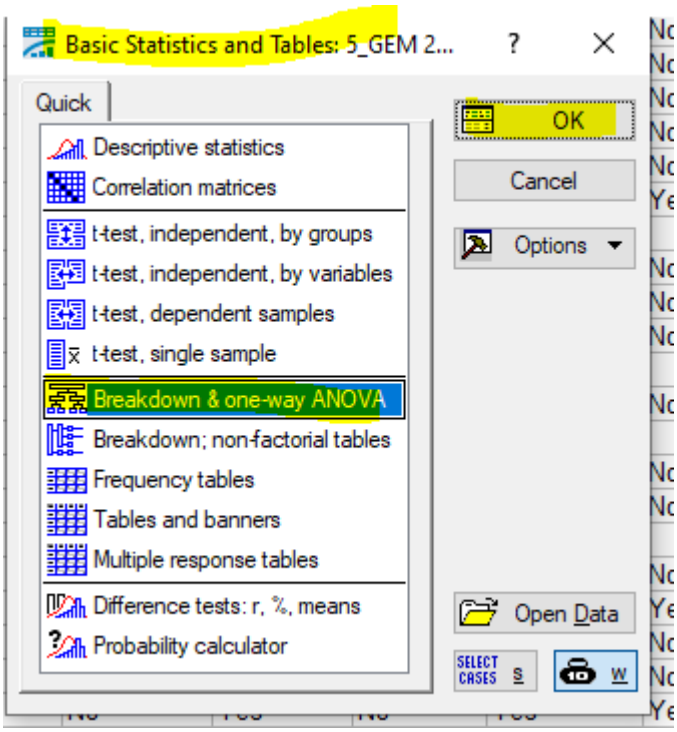
Import 5_GEM 2010 APS Croatia.sta

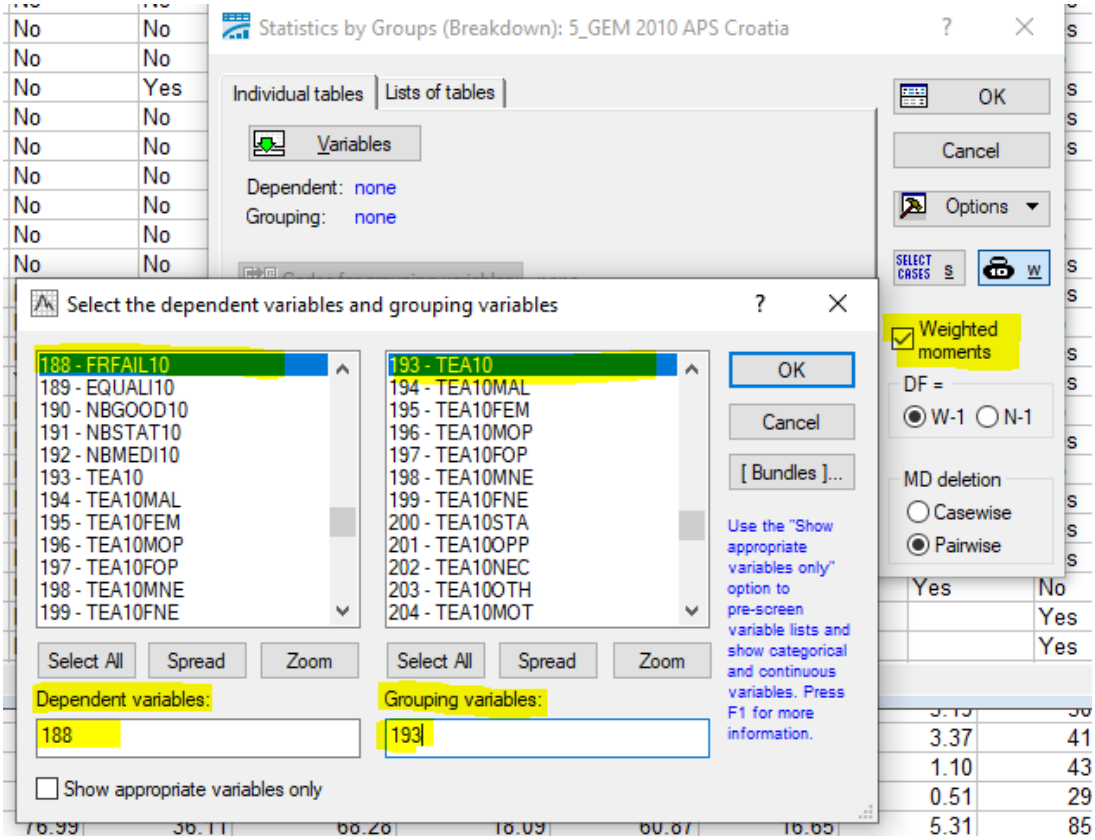
1. Calculate fear of failure for TEA active and TEA non-active
2. Set up hypotheses.
3. Test the hypotheses in Statistica. Make a conclusion.

Solution

1. Calculate fear of failure for TEA active and TEA non-active

Statistics - Basic statistics and tables - Breakdown and one-way ANOVA (wghtd momnts) - dependent: 188, independent: 193





TEA10	2-Way Tables of Descriptive Statistics (5_GEM 2010 APS Croatia.sta) Smallest N for any variable: 1611		
	FRFAIL10 Means	FRFAIL10 Sum of w	FRFAIL10 Std.Dev.
No	0.402687	1522.292	0.490600
Yes	0.217025	89.074	0.414554
All Grps	0.392424	1611.365	0.488442

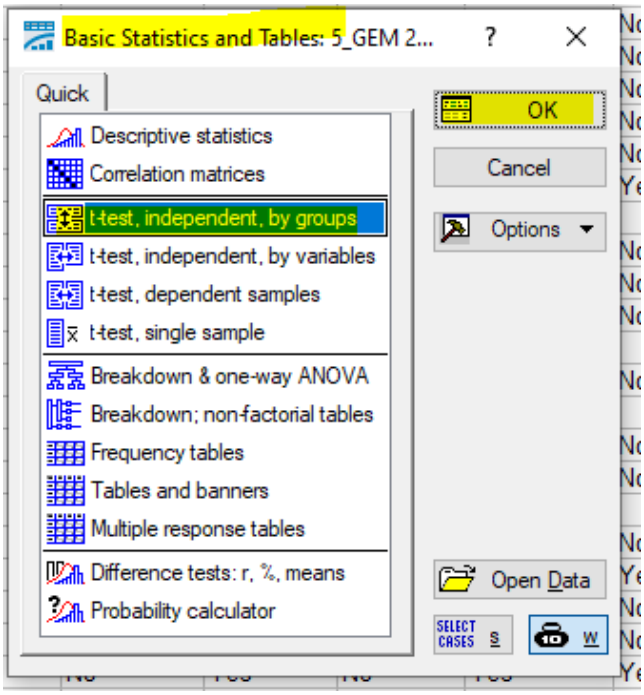
It can be seen that there are 21,7% TEA active have fear of failure while the same percentage for TEA non-active is 40,3%. We are going to test if the difference between the percentages is statistically significant.

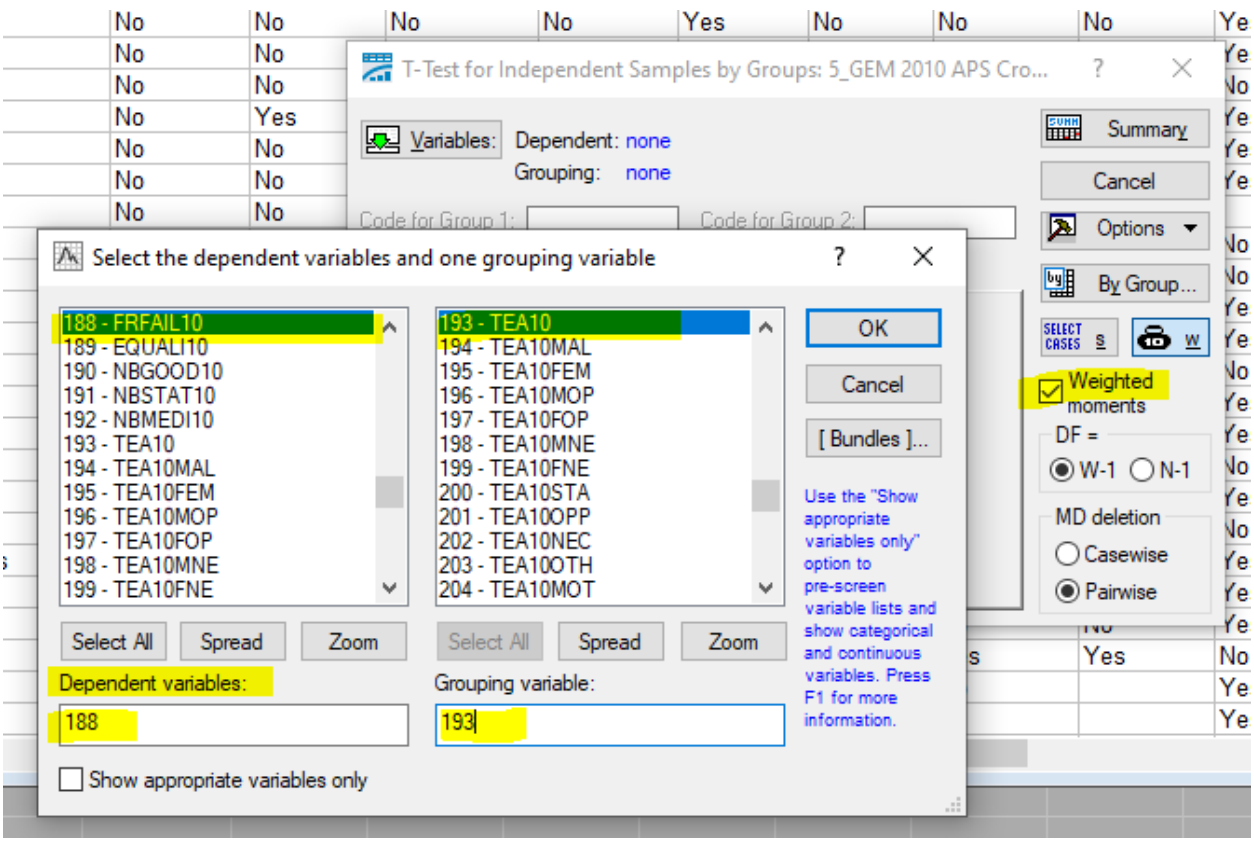
2. Set up hypotheses.

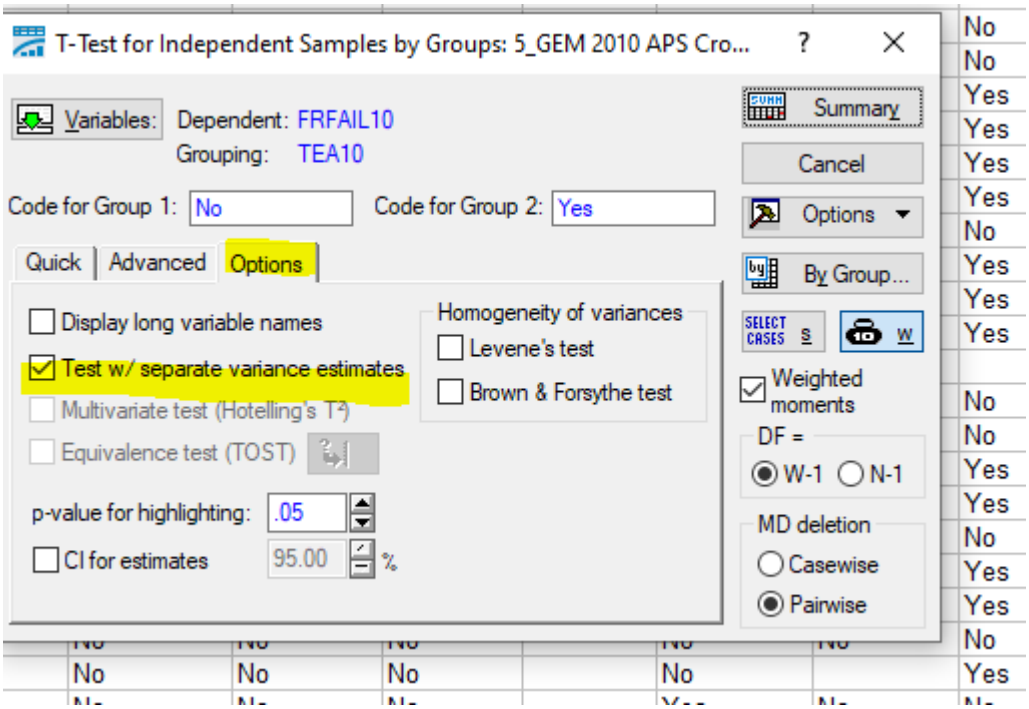
$$H_0 : p_{TEAactive} = p_{TEAnon-active}$$
$$H_1 : p_{TEAactive} \neq p_{TEAnon-active}$$

3. Test the hypotheses in Statistica. Make a conclusion.

Statistics - Basic statistics and tables - t-test, independent, by groups (wghtd momnts) - dependent: 188, independent: 193; options: test w/ separate variance estimates







Variable	T-tests; Grouping: TEA10: Involved in Total early-stage Entrepreneur (5_GEM 2010 APS Croatia.sta) Group 1: No Group 2: Yes													
	Mean No	Mean Yes	t-value	df	p	t separ. var.est.	df	p 2-sided	Sum of w No	Sum of w Yes	Std.Dev. No	Std.Dev. Yes	F-ratio Variances	p Variances
FRFAIL10	0.402687	0.217025	3.499021	1609.365	0.000480	4.063615	103.0602	0.000095	1522.292	89.07369	0.490600	0.414554	1.400533	0.043252

In the table of results, first look at the p variances:

if p variances < 0,05, it means that variances of two variables are not equal and we have to use t-test with separate variance estimates.

if p variances > 0,05, it means that variances of two variables are equal and we can use t-test with pooled variance estimates (first t-value in the table of results).

In our case p variances < 0,05:

0.043252 < 0,05

So, we look at the p 2 sided for t separ. var.est. = 0.000095 and compare it with 0,05

Since 0.000095 < 0,05, we reject H_0

We conclude that there is a statistically significant difference in fear of failure between TEA active and TEA non-active population. TEA active population have lower fear of failure compared to TEA non-active.

Example 8

- We would like to see whether there is a difference in total money required to start a business (SUMONTUS: 275) between women and men (gender: 155) who are entrepreneurial active (TEA10: 193).
  Do women start less expensive businesses?
  We know nothing about distribution of total money required.
  (5_GEM 2010 APS Croatia.sta)

Guidelines for making a solution:

Import 5_GEM 2010 APS Croatia.sta
1. Select entrepreneurially active people.
2. Check sample sizes.
3. Set up hypotheses.
4. Test the hypotheses in Statistica. Make a conclusion.

Solution

1. Select entrepreneurially active people.

weight

In the bottom right corner click 'Sel' and in the window that opens select 'Enable selection conditions' and write: TEA10='yes'

2. Check sample sizes

Statistics - Basic statistics and tables - Breakdown and one-way ANOVA (wghtd momnts) - dependent: 275, independent: 155

GENDER	2-Way Tables of Descriptive Statistics (5_GEM 2010 APS Croatia.sta) Smallest N for any variable: 60 Include condition: TEA10='yes'		
	SUMONTUS Means	SUMONTUS Sum of w	SUMONTUS Std.Dev.
Male	144794.2	40.63063	288519.5
Female	108711.7	16.90336	189280.7
All Grps	134193.2	57.53400	262121.1

There are 40,6 males and 16,9 females in the samples.

Since at least one sample is considered as small sample and we don't know if SUMONTUS is normally distributed, we will use nonparametric test - Mann Whitney test

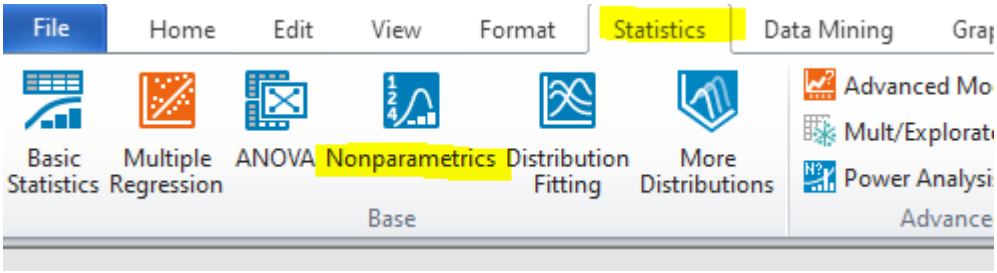
3. Set up hypotheses

H_0: distribution of total money required for females is equal to distribution of total money required for males

H_1: distributions are not equal

4. Test the hypotheses in Statistica. Make a conclusion.

Statistics - Nonparametrics - Comparing two independent samples (groups) - dependent: 275, independent: 155 - Mann Whitney U test



variable	Mann-Whitney U Test (w/ continuity correction) (5_GEM 2010 APS Croatia.sta) By variable GENDER Marked tests are significant at p									
	Rank Sum Female	Rank Sum Male	U	Z	p-value	Z adjusted	p-value	Valid N Female	Valid N Male	2*1sided exact p
SUMONTUS	787.0000	1841.000	462.0000	-1.47850	0.139276	-1.48248	0.138213	25	47	0.140060

you can use exact p

p= 0.140060 > 0,05, do not reject H_0

We conclude there is no statistically significant difference in distributions of total money required between females and males.

Women don't start less expensive businesses.

Example 9

- We would like to see whether there is a difference in entrepreneurial activity (TEA10: 37) between efficiency driven and innovation driven countries (CAT_GCR2: 4).
  Is it the same for opportunity (TEA10OPP: 42) and
  necessity (TEA10NEC: 43)?
  What do you expect?
  (6_GEM 2010 APS Master.sta)

Example 10

- Is there a correlation between FRFAIL (15) and SUSKILL (14)?
  Whether higher percentage of those who have knowledge means lower percentage of fear of failure.
  Test it for GEM countries.
  (6_GEM 2010 APS Master.sta)

Guidelines for making a solution:

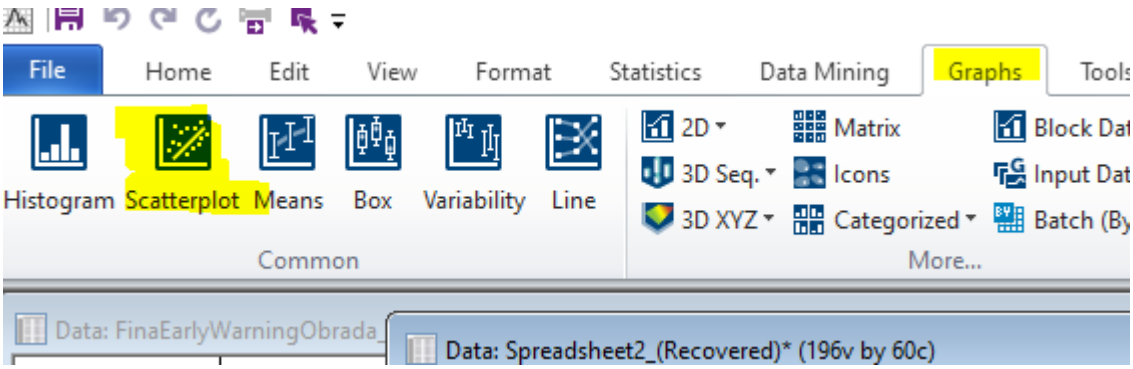
Import 6_GEM 2010 APS Master.sta

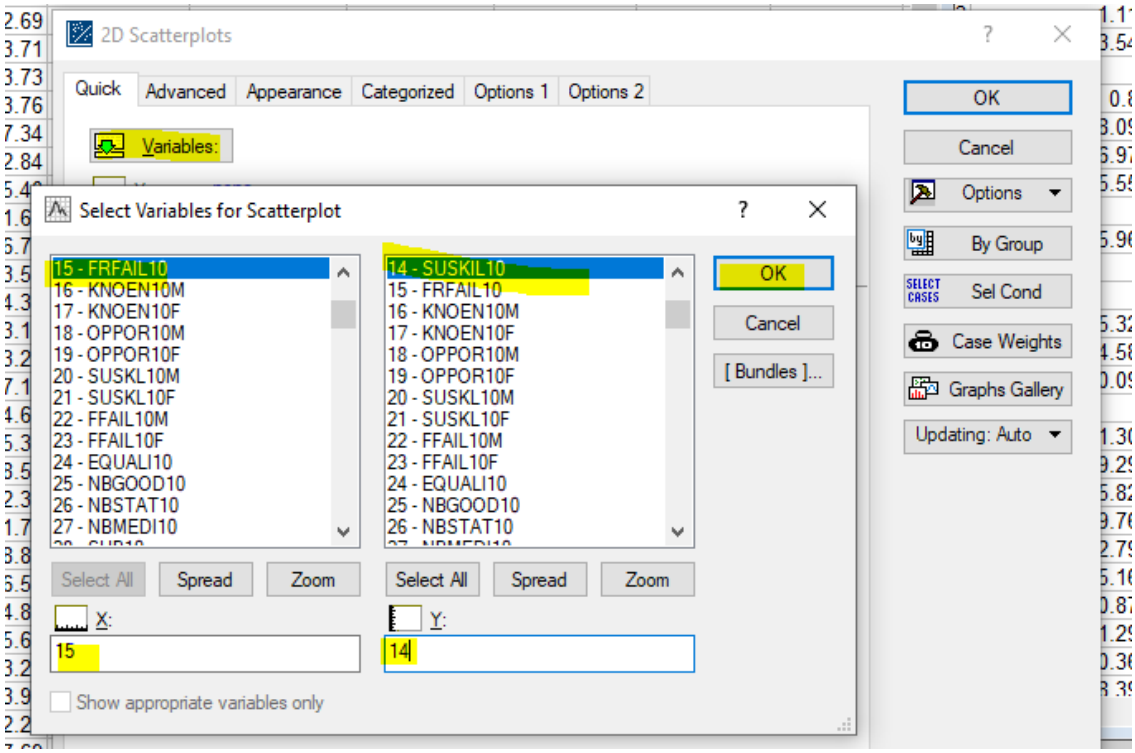
1. Draw scatterplots.
2. Test if our variables are normally distributed.
3. Set up hypotheses.
4. Test the hypotheses in Statistica. Make a conclusion.

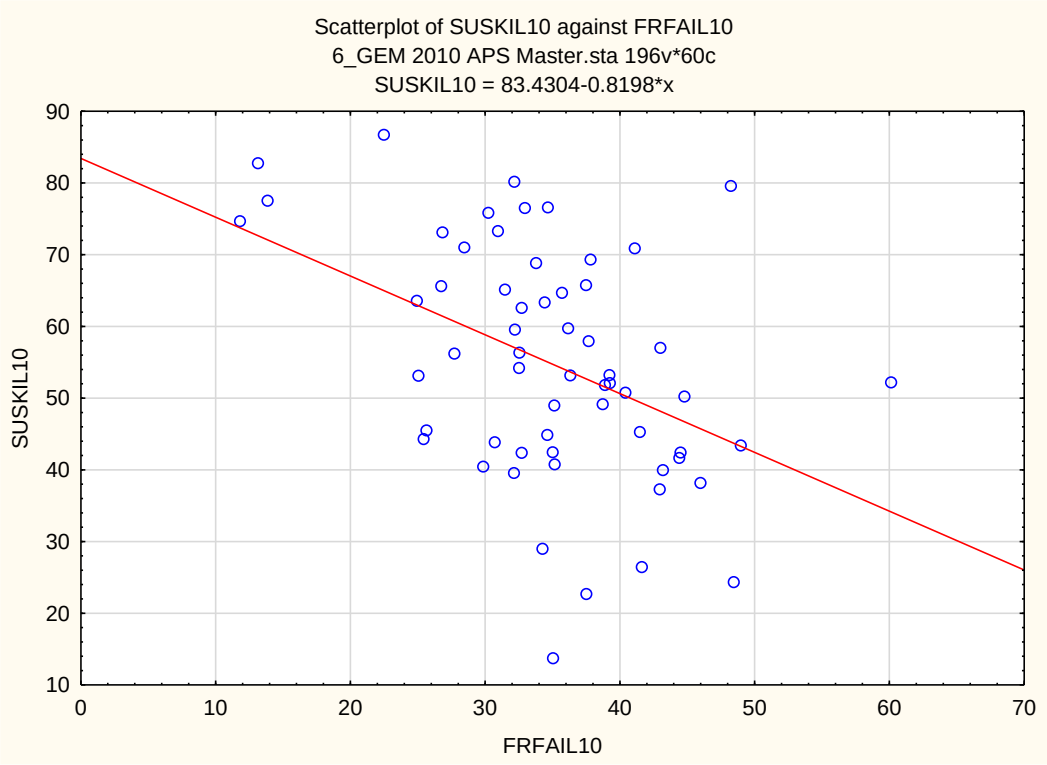
Solution

1. Draw a scatterplots.

Graphs - Scatterplot - X: 15, Y:14 (it can be vice versa)





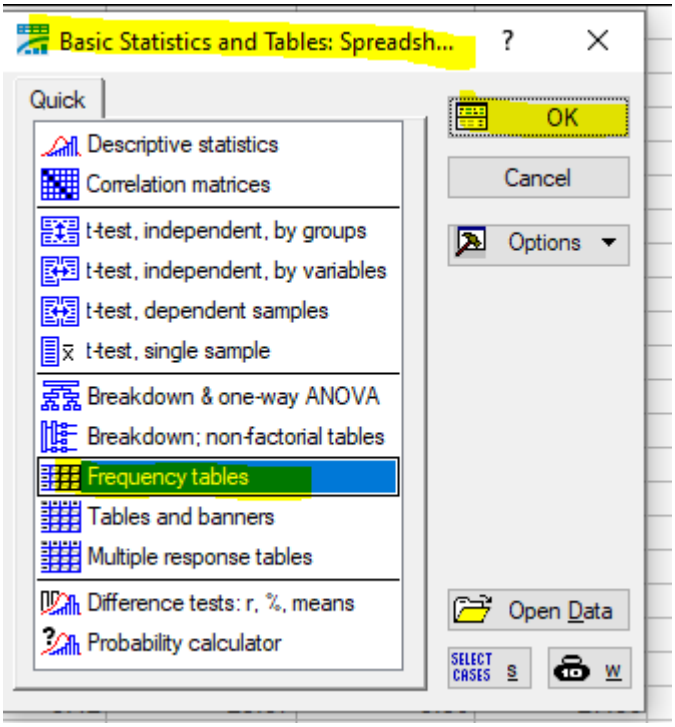


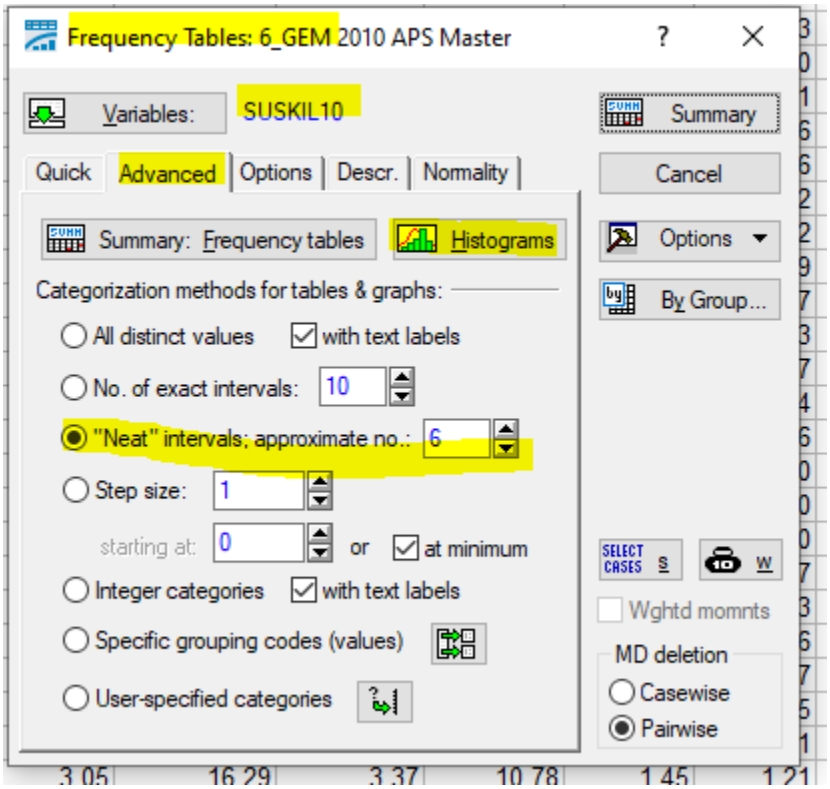
Negative correlation can be noticed - one variable increases with the decrease of the other variable.

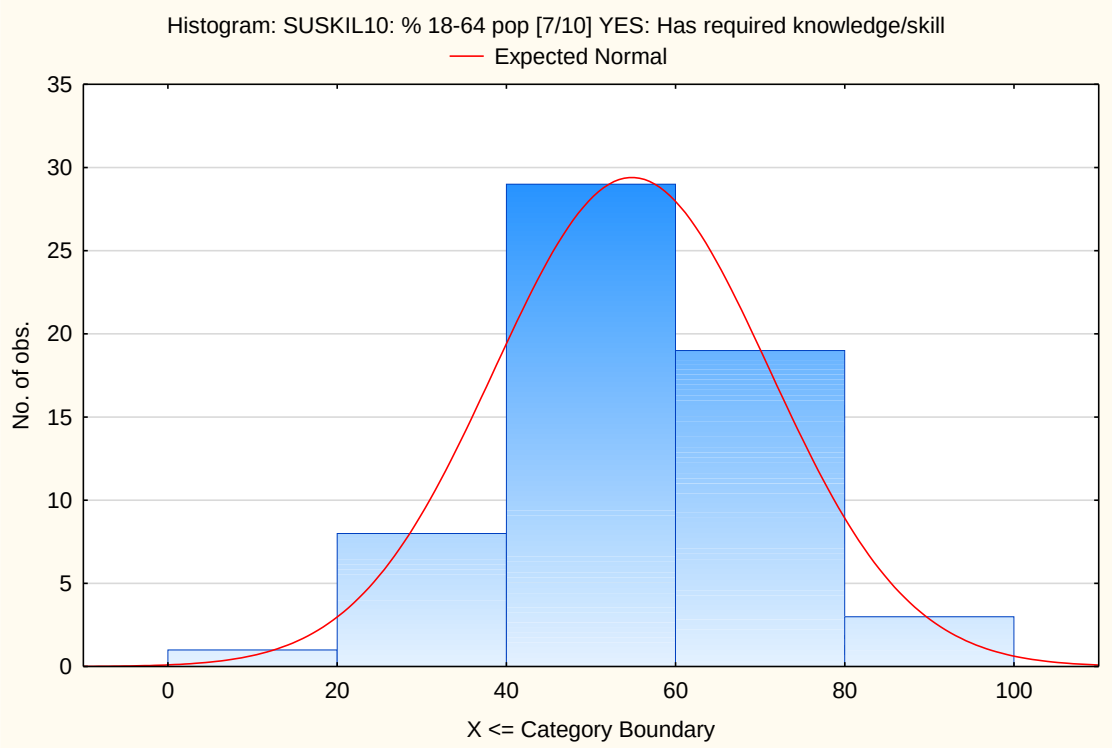
2. Test if our variables are normally distributed.

Statistics - Basic statistics and tables - Frequency tables - Variables: 14

Advanced: 'Neat intervals', approximate no.: 6

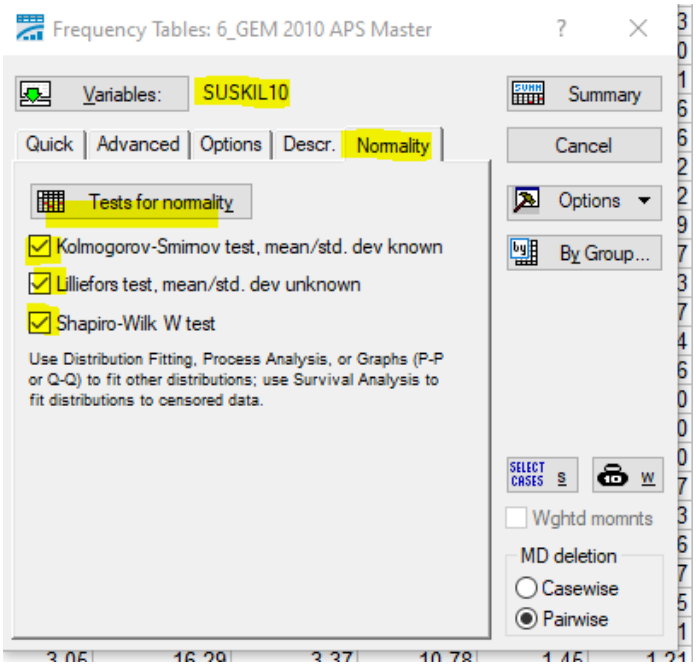






It looks normal.

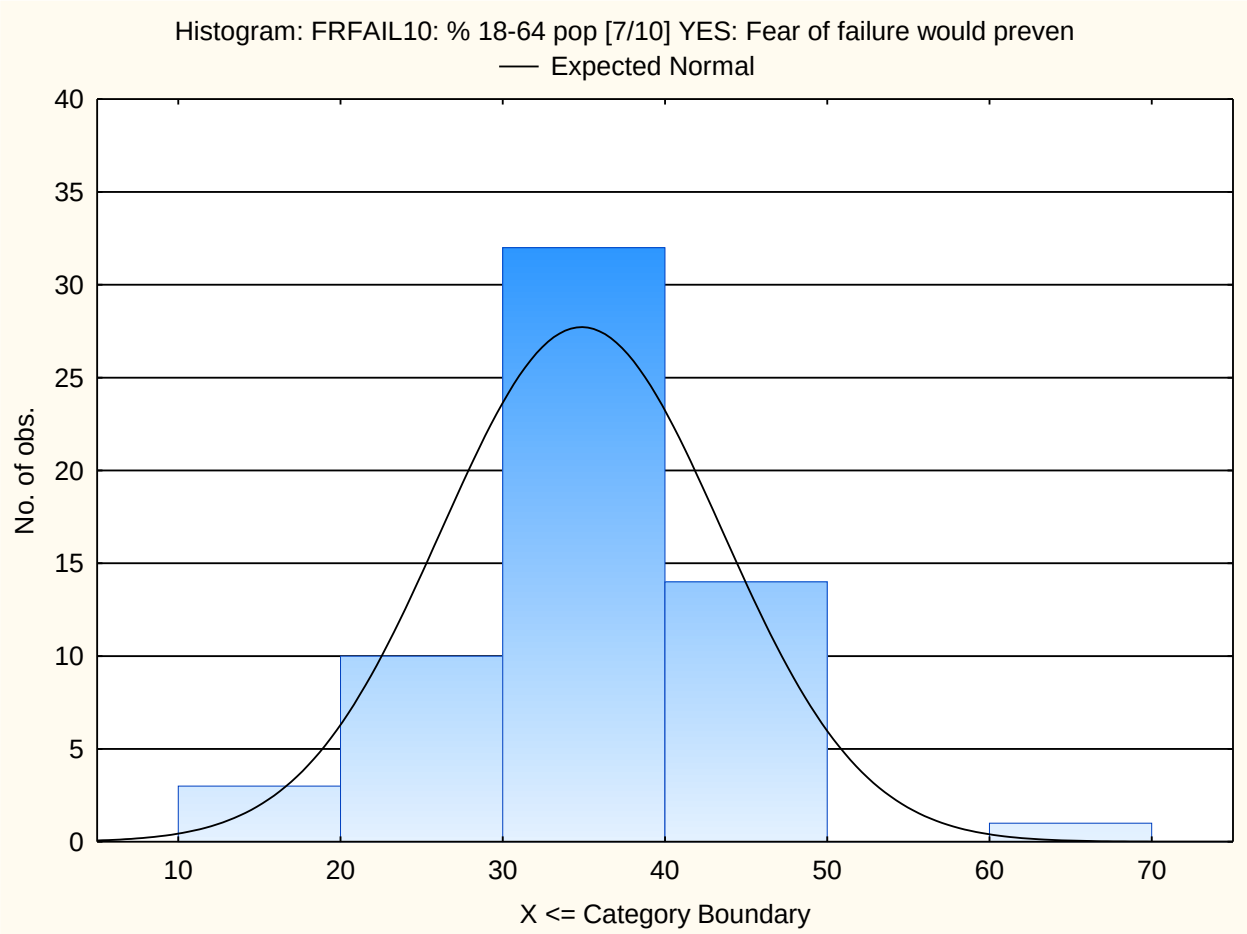
Test for normality:



Variable	Tests of Normality (6_GEM 2010 APS Master.sta)					
	N	max D	K-S p	Lilliefors p	W	p
SUSKIL10: % 18-64 pop [7/10] YES: Has required kno	60	0.066903	p > .20	p > .20	0.981766	0.507652

All p values are > 0,05 which means that our variable is normally distributed (H_0: variable is normal; H_1: variable is not normal).

We do the same for variable 15:



Variable	Tests of Normality (6_GEM 2010 APS Master.sta)					
	N	max D	K-S p	Lilliefors p	W	p
FRFAIL10: % 18-64 pop [7/10] YES: Fear of failure	60	0.092886	p > .20	p > .20	0.968331	0.120841

Since our variables are normally distributed and we can notice linear relationship between variables, we can apply Pearson correlation.

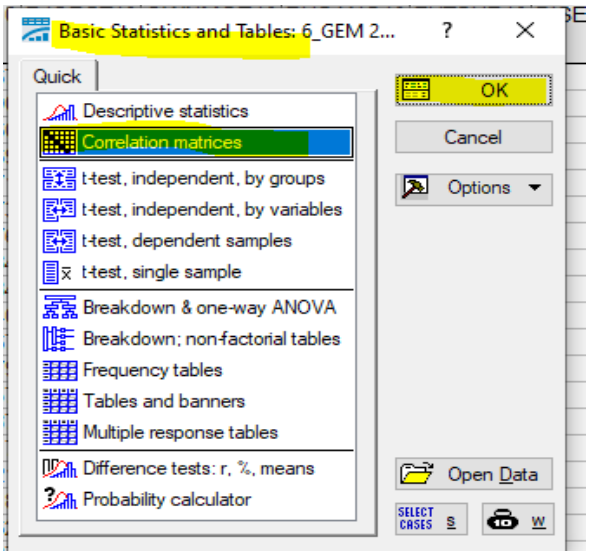
3. Set up hypotheses.

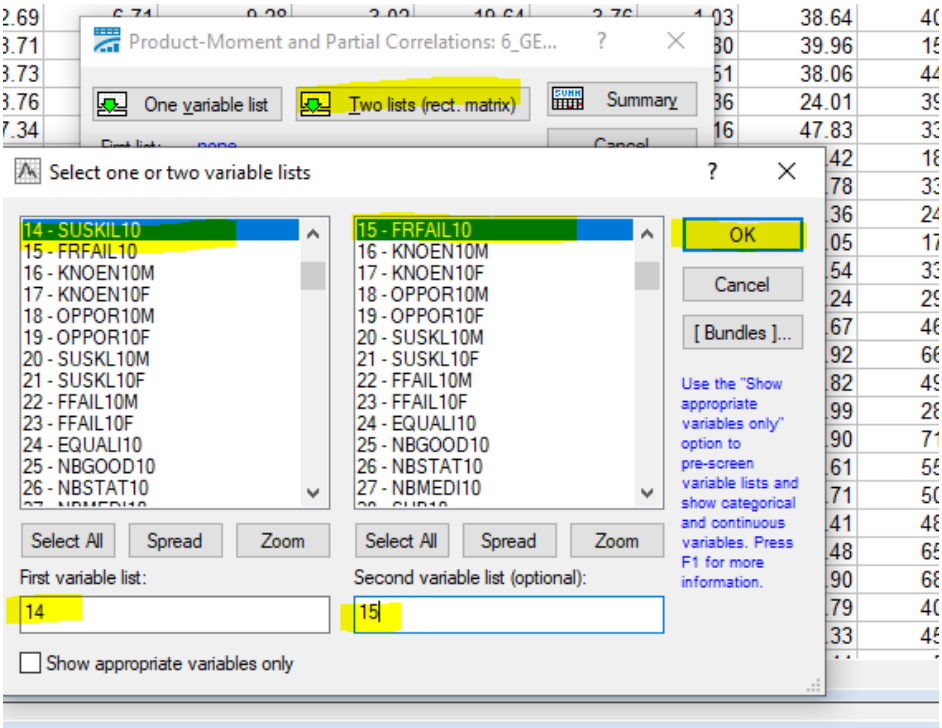
$H_0 : r = 0$

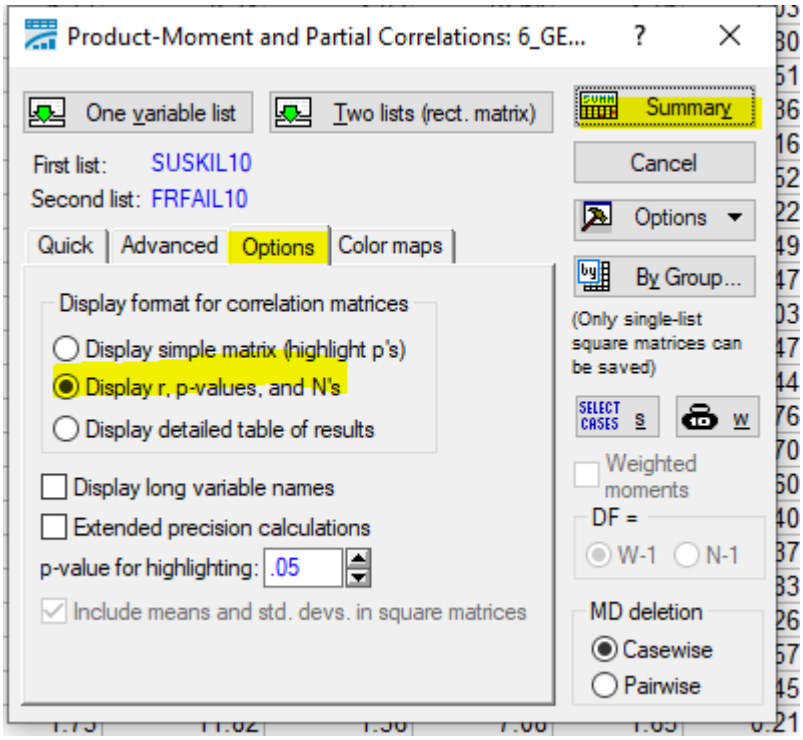
$H_1 : r \neq 0$

4. Test the hypotheses in Statistica. Make a conclusion.

Statistics - Basic statistics and tables - Correlation matrices - Display r, p-values and N's - Two lists - First variable list: 14; Second variable list:15 - Summary







Variable	Correlations (6_GEM 2010 APS Master.sta) Marked correlations are significant at p < .05000 N=60 (Casewise deletion of missing data)
	FRFAIL10
SUSKIL10	-.4348
	p=.001

p<0,05, we reject H_0

We conclude that there is a statistically significant correlation between two variables - higher knowledge means lower fear of failure and vise versa.

Example 11

- Is there a relationship between Knoent10 (12) and TEA10 (37).
  Test it separatey for factor driven countries and efficiency driven countries (CAT_GCR2).
  (6_GEM 2010 APS Master.sta)

Guidelines for making a solution:
Import 6_GEM 2010 APS Master.sta
1. Check for sample size
2. Select factor driven countries
3. Set up hypotheses
4. Calculate Spearman correlation coefficient
5. Make a conclusion
6. Do the same for efficiency driven counties

Solution:

1. Check for sample size

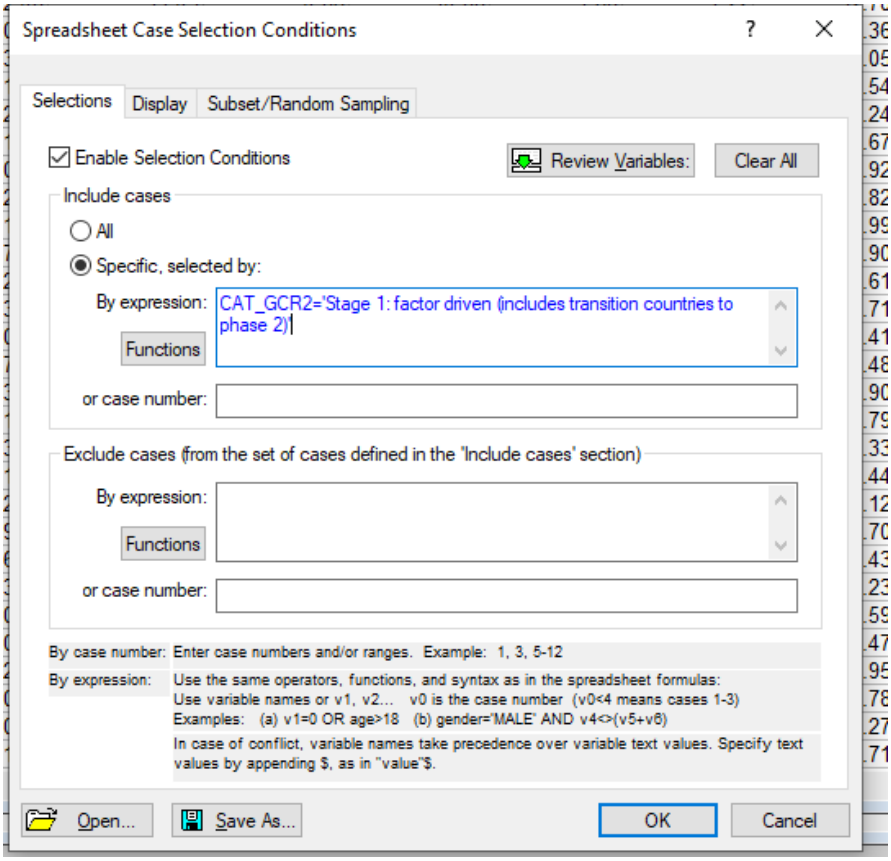
Make a frequency table of variable cat_gcr2: Statistics - Basic statistics and tables - Frequency tables - Variables: 4

CAT_GCR2: COUNTRY GROUP GCR REPORT 2009-2010 - 3 CAT (6_GEM 2010 APS Master)

	Frequency table: CAT_GCR2: COUNTRY GROUP GCR REPORT			
Category	Count	Cumulative Count	Percent	Cumulative Percent
Stage 1: factor driven (includes transition countries to phase 2)	13	13	21.66667	21.6667
Stage 2: efficiency driven (includes transition countries to phase 3)	24	37	40.00000	61.6667
Stage 3: innovation driven	23	60	38.33333	100.0000
Missing	0	60	0.00000	100.0000

2. Select factor driven countries

In the bottom right corner click 'Sel' and in the window that opens select 'Enable selection conditions' and write: cat_gcr2='Stage 1: factor driven (includes transition countries to phase 2)'



3. Set up hypotheses

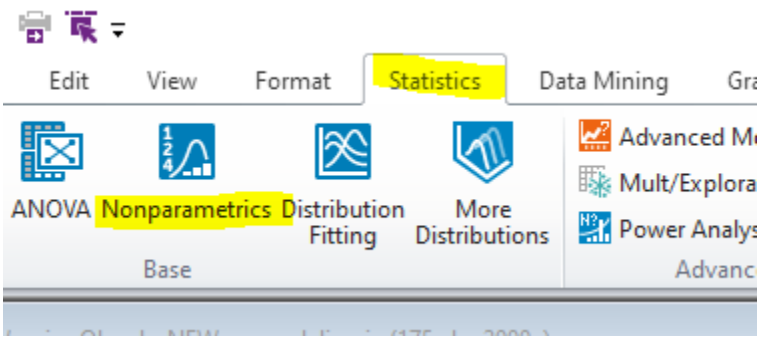
Since we have a small sample, it is recommended that we use Spearman correlation coefficient.

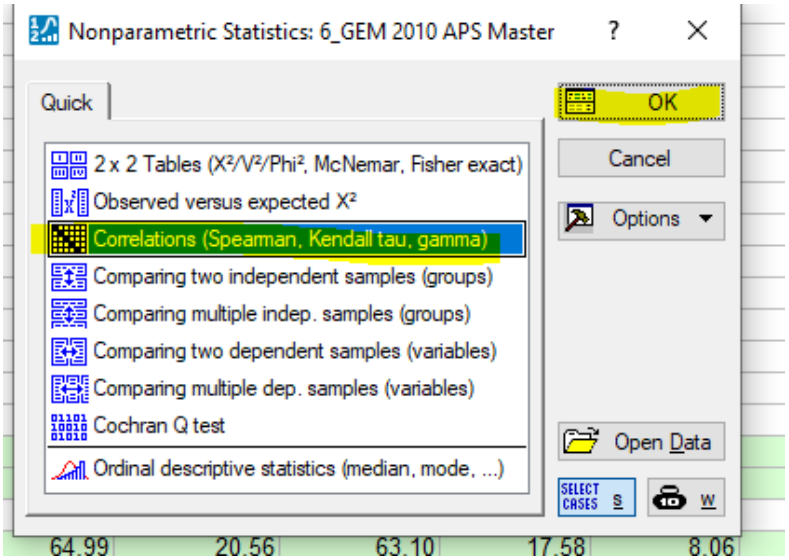
H_0: no correlation

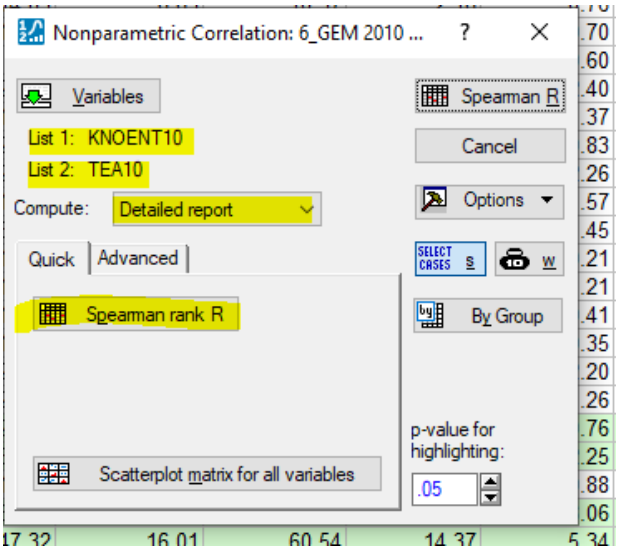
H_1: correlation exists

4. Calculate Spearman correlation coefficient

Statistics - Nonparametrics - Correlations (Spearman, Kendall tau, gamma) - Compute: Detailed report - Variables - List 1: 12, List 2: 37 - Spearman rank R







Pair of Variables	Spearman Rank Order Correlations (6_GEM 2010 APS Master.sta) MD pairwise deleted Marked correlations are significant at p			
	Valid N	Spearman R	t(N-2)	p-value
KNOENT10 & TEA10	13	0.675824	3.041064	0.011225

$p<0,05$ - reject H_0

5. Make a conclusion

There is a significant positive correlation between two variables - the higher percentage of those who know somebody who started a business the higher entrepreneurial activity.

6. Do the same for efficiency driven counties

Category	Frequency table: CAT_GCR2: COUNTRY GROUP GCR REPORT 2009-2010 - 3 CAT (6_GEM 2010 APS Master.sta) Include condition: cat_gcr2='Stage 2: efficiency driven (includes transition countries to phase 3)'			
	Count	Cumulative Count	Percent	Cumulative Percent
Stage 2: efficiency driven (includes transition countries to phase 3)	24	24	100.0000	100.0000
Missing	0	24	0.0000	100.0000

H_0: no correlation

H_1: correlation exists

Pair of Variables	Spearman Rank Order Correlations (6_GEM 2010 APS Master.sta) MD pairwise deleted Marked correlations are significant at p			
	Valid N	Spearman R	t(N-2)	p-value
KNOENT10 & TEA10	24	0.480000	2.566375	0.017606

It works for the efficiency driven countries as well.

Example 12

- We would like to examine whether knowledge and skills to start a business (SUSKIL10: 187) depend on gender (gender: 155), in other words, is there a difference in perception of having knowledge and skills for business between women and men.
  (5_GEM 2010 APS Croatia.sta)

Guidelines for making a solution:
Import 5_GEM 2010 APS Croatia.sta
weight variable: weight_l
1. Set up hypotheses
2. Calculate crosstable
3. Calculate χ^2
4. Make a conclusion

Solution:

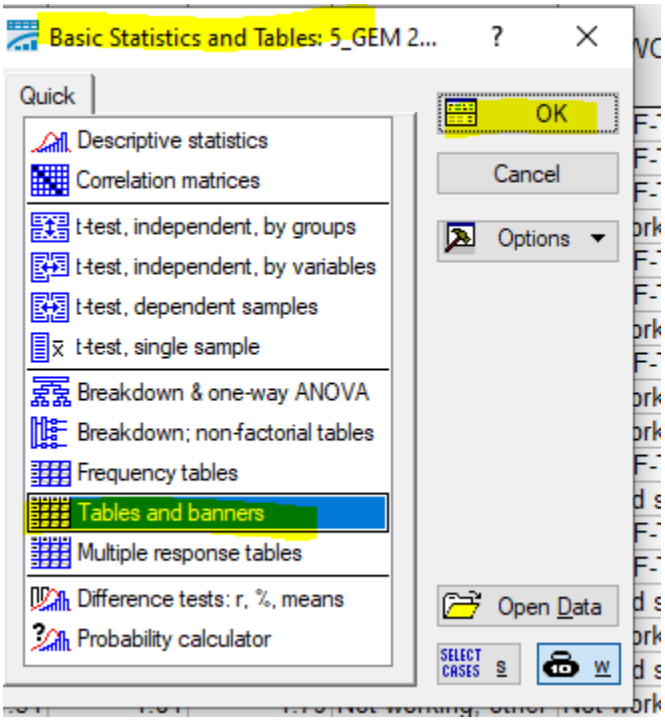
1. Set up hypotheses

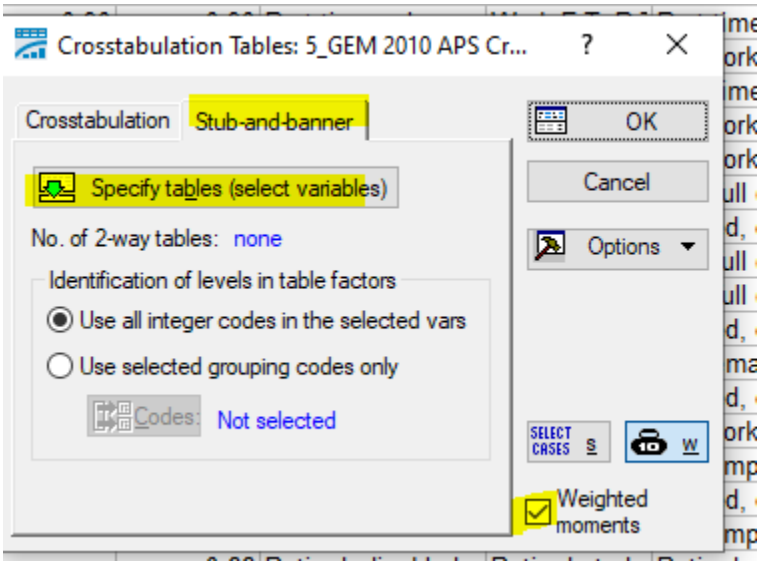
Since both variables are categorical - Gender: male, female; Suskil: yes, no, we can use χ^2 test.

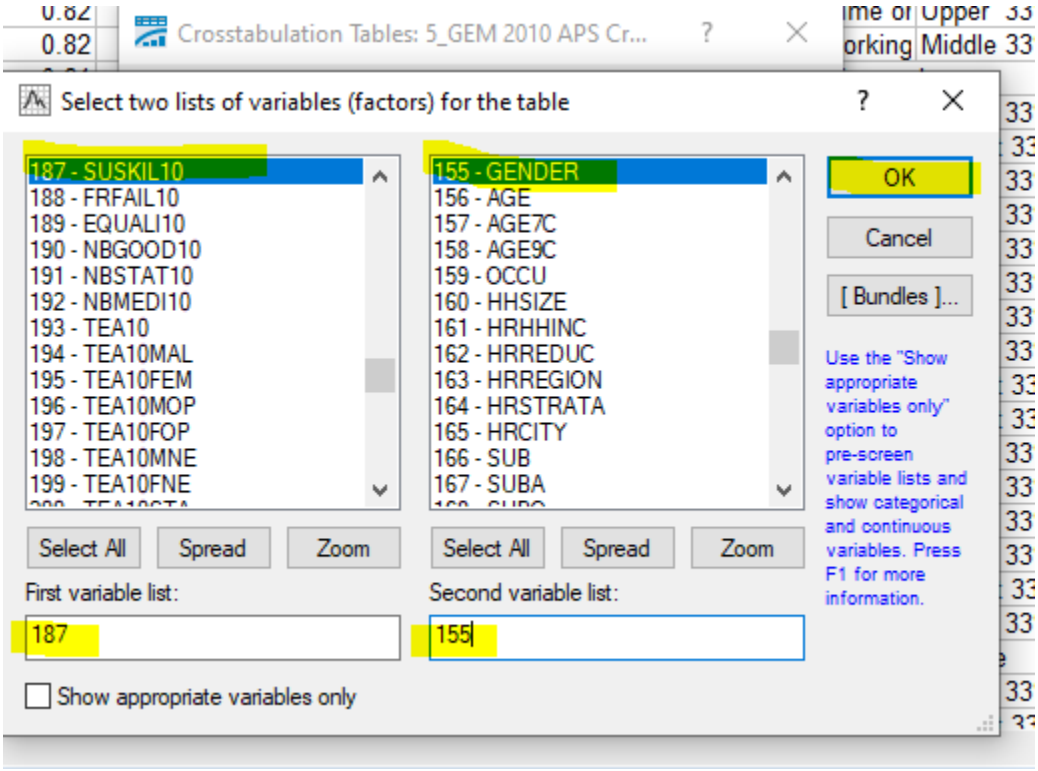
H_0: there is no dependence between gender and have knowledge and skills to start a business

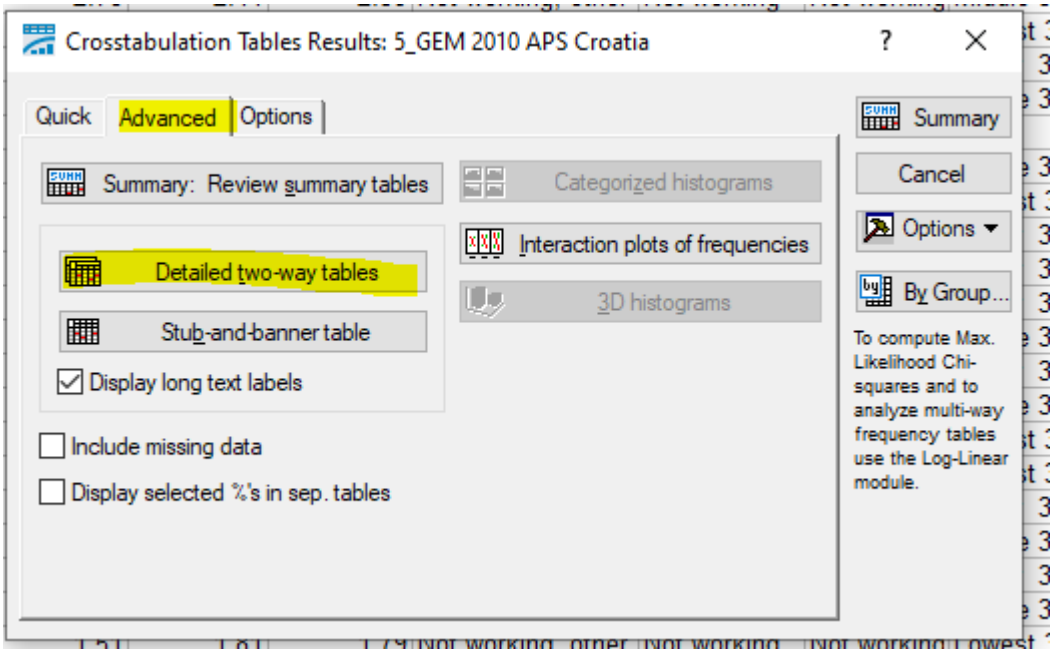
H_1: there is dependence between gender and have knowledge and skills to start a business

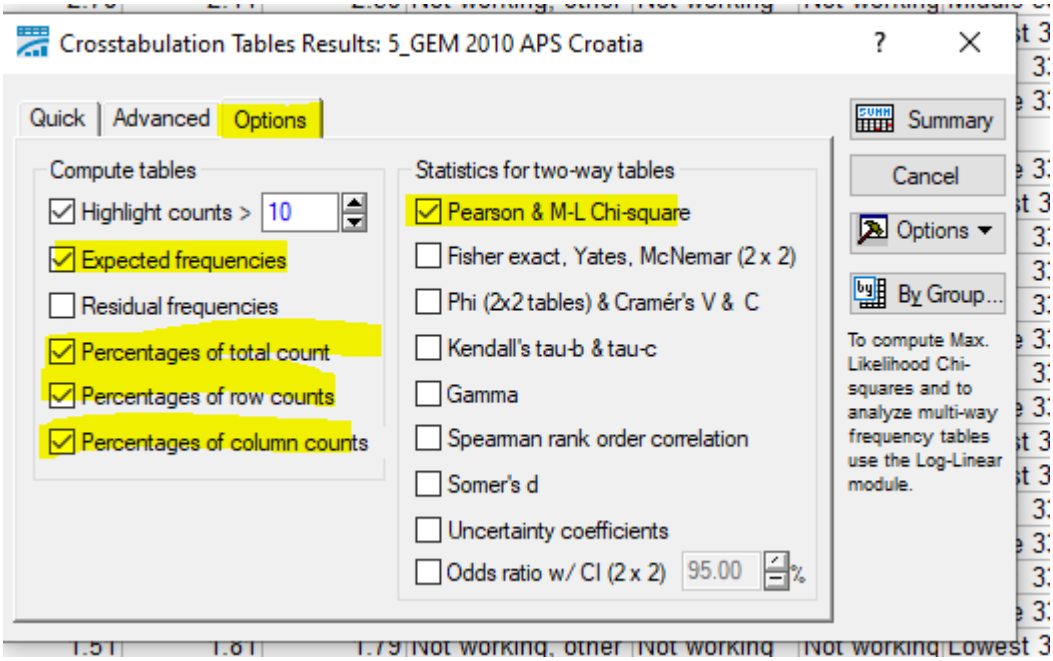
2. Calculate crosstable: Statistics - Basic statistics and tables - Tables and banners - Weighted moments - Stub and banner - First variable list: 187, Second variable list: 155 - OK - Options: Pearson and ML Chi square, Expected frequencies, Percentages of total count, Percentages of row counts, Percentages of column counts - Advanced: Detailed two-way tables

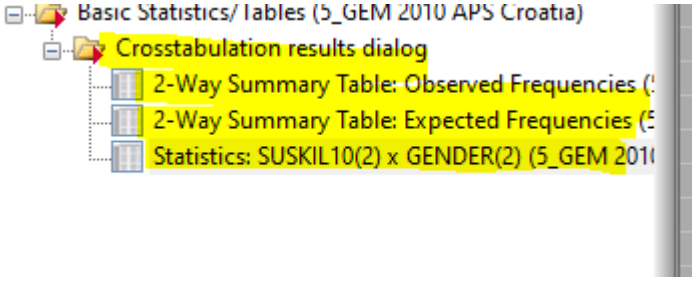

Basic Statistics and Tables: 5_GEM 2...
Quick
Descriptive statistics
Correlation matrices
t-test, independent, by groups
t-test, independent, by variables
t-test, dependent samples
t-test, single sample
Breakdown & one-way ANOVA
Breakdown; non-factorial tables
Frequency tables
Tables and banners
Multiple response tables
Difference tests: r, %, means
Probability calculator
OK
Cancel
Options
Open Data
SELECT CASES


Crosstabulation Tables: 5_GEM 2010 APS Cr...
Crosstabulation
Stub-and-banner
Specify tables (select variables)
No. of 2-way tables: none
Identification of levels in table factors
Use all integer codes in the selected vars
Use selected grouping codes only
Codes:
Not selected
OK
Cancel
Options
SELECT CASES
Weighted moments


Crosstabulation Tables: 5_GEM 2010 APS Cr...
Select two lists of variables (factors) for the table
187 - SUSKIL10
188 - FRFAIL10
189 - EQUALI10
190 - NBGOOD10
191 - NBSTAT10
192 - NBMEDI10
193 - TEA10
194 - TEA10MAL
195 - TEA10FEM
196 - TEA10MOP
197 - TEA10FOP
198 - TEA10MNE
199 - TEA10FNE
155 - GENDER
156 - AGE
157 - AGE7C
158 - AGE9C
159 - OCCU
160 - HHSIZE
161 - HRHHINC
162 - HRREDUC
163 - HRREGION
164 - HRSTRATA
165 - HRCITY
166 - SUB
167 - SUBA
OK
Cancel
[Bundles]...
Use the "Show appropriate variables only" option to pre-screen variable lists and show categorical and continuous variables. Press F1 for more information.
Select All
Spread
Zoom
Select All
Spread
Zoom
First variable list:
187
Second variable list:
155
Show appropriate variables only


Crosstabulation Tables Results: 5_GEM 2010 APS Croatia
Quick
Advanced
Options
Summary: Review summary tables
Categorized histograms
Detailed two-way tables
Interaction plots of frequencies
Stub-and-banner table
3D histograms
Display long text labels
Include missing data
Display selected %'s in sep. tables
Summary
Cancel
Options
By Group...
To compute Max. Likelihood Chi-squares and to analyze multi-way frequency tables use the Log-Linear module.





SUSKIL10: SUSKIL adapted to make it fit for national l	2-Way Summary Table: Expected Frequencies (5_GEM 2010 APS Croatia.sta) Marked cells have counts > 10		
	GENDER Male	GENDER Female	Row Totals
No	354.4343	360.8290	715.263
Yes	402.7072	409.9728	812.680
Totals	757.1415	770.8018	1527.943

- Check the assumption: no more than 20% of expected values may be less than 5

SUSKIL10: SUSKIL adapted to make it fit for national l	2-Way Summary Table: Observed Frequencies (5_GEM 2010 APS Croatia.sta) Marked cells have counts > 10		
	GENDER Male	GENDER Female	Row Totals
No	282.98	432.28	715.26
Column Percent	37.38%	56.08%	
Row Percent	39.56%	60.44%	
Total Percent	18.52%	28.29%	46.81%
Yes	474.16	338.52	812.68
Column Percent	62.62%	43.92%	
Row Percent	58.35%	41.65%	
Total Percent	31.03%	22.16%	53.19%
Totals	757.14	770.80	1527.94
Total Percent	49.55%	50.45%	100.00%

Among those who think they don't have knowledge and skill to start a business, there are 39,56% males and 60,44% females.
Among those who think they have knowledge and skill to start a business, there are 58,35% males and 41,65% females.

Among all males, 37,38% think they don't have knowledge and skill to start a business and 62,62% think they have knowledge and skill to start a business.
Among all females, 56,08% think they don't have knowledge and skill to start a business and 43,92% think they have knowledge and skill to start a business.

3. Calculate χ^2

Statistic	Statistics: SUSKIL10(2) x GENDER(2) (5_GEM 2010 APS Croatia.sta)		
	Chi-square	df	p
Pearson Chi-square	53.68351	df=1	p=.00000
M-L Chi-square	54.01583	df=1	p=.00000

$p<0,05$ - reject H_0

4. Make a conclusion

There is a statistically significant dependence between gender and have knowledge and skills to start a business. It can be concluded that males are more confident in their knowledge and skills to start a business compared to females.

Example 13

- Examine whether there is dependence between TEA10 (193) and OPPORT10 (186). (5_GEM 2010 APS Croatia.sta)

Guidelines for making a solution:
Import 5_GEM 2010 APS Croatia.sta
weight variable: weight_l
1. Set up hypotheses
2. Calculate crosstable
3. Calculate χ^2
4. Make a conclusion

Solution:

1. Set up hypotheses

Since both variables are categorical - TEA: yes, no; Opport: yes, no, we can use χ^2 test.

H_0: there is no dependence between TEA and OPPORT

H_1: there is dependence between TEA and OPPORT

2. Calculate crosstable: Statistics - Basic statistics and tables - Tables and banners - Weighted moments - Stub and banner - First variable list: 193, Second variable list: 186 - OK - Options: Pearson and ML Chi square, Expected frequencies, Percentages of total count, Percentages of row counts, Percentages of column counts - Advanced: Detailed two-way tables

- Check the assumption: no more than 20% of expected values may be less than 5

TEA10 : Involved in Total early-stage Entrepreneuria	2-Way Summary Table: Expected Frequencies (5_GEM 2010 APS Croatia.sta) Marked cells have counts > 10		
	OPPORT10 No	OPPORT10 Yes	Row Totals
No	979.596	297.8613	1277.458
Yes	55.914	17.0014	72.915
Totals	1035.510	314.8626	1350.373

TEA10 : Involved in Total early-stage Entrepreneuria	2-Way Summary Table: Observed Frequencies (5_GEM 2010 APS Croatia.sta) Marked cells have counts > 10		
	OPPORT10 No	OPPORT10 Yes	Row Totals
No	1003.37	274.09	1277.46
Column Percent	96.90%	87.05%	
Row Percent	78.54%	21.46%	
Total Percent	74.30%	20.30%	94.60%
Yes	32.14	40.77	72.91
Column Percent	3.10%	12.95%	
Row Percent	44.08%	55.92%	
Total Percent	2.38%	3.02%	5.40%
Totals	1035.51	314.86	1350.37
Total Percent	76.68%	23.32%	100.00%

Among those who are not entrepreneurially active, 78,54% think there will be no business opportunities in the next six months and 21,46% thinke there will be opportunities.
Among those who are entrepreneurially active, 44,08% think there will be no business opportunities in the next six months and 55,92% thinke there will be opportunities.

Among those who think there will be no business opportunities in the next six months, 96,9% are not entrepreneurially active while 3,1% are entrepreneurially active.
Among those who think there will be business opportunities in the next six months, 87,05% are not entrepreneurially active while 12,95% are entrepreneurially active.

3. Calculate χ^2

Statistic	Statistics: TEA10(2) x OPPORT10(2) (5_GEM 2010 APS Croatia.sta)		
	Chi-square	df	p
Pearson Chi-square	45.82403	df=1	p=.00000
M-L Chi-square	38.26639	df=1	p=.00000

$p<0,05$ - reject H_0

4. Make a conclusion

There is a statistically significant dependence between TEA and OPPORT. It can be concluded that entrepreneurially active people see more opportunities compared to TEA non active.

Example 14

- Examine whether oppinion about equal standard of living depends on gender.

- GENDER (155)
- EQUALI10 (189): most people would prefer that everyone had a similar standard of living
  (5_GEM 2010 APS Croatia.sta)

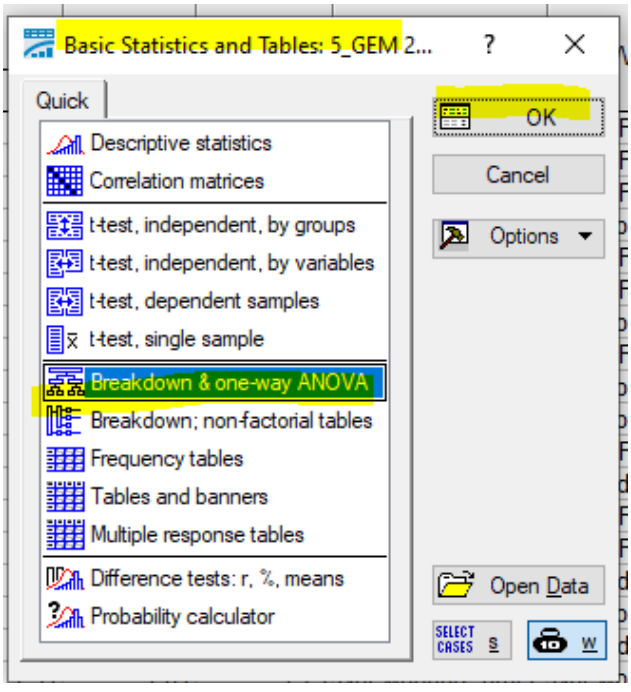
Example 15

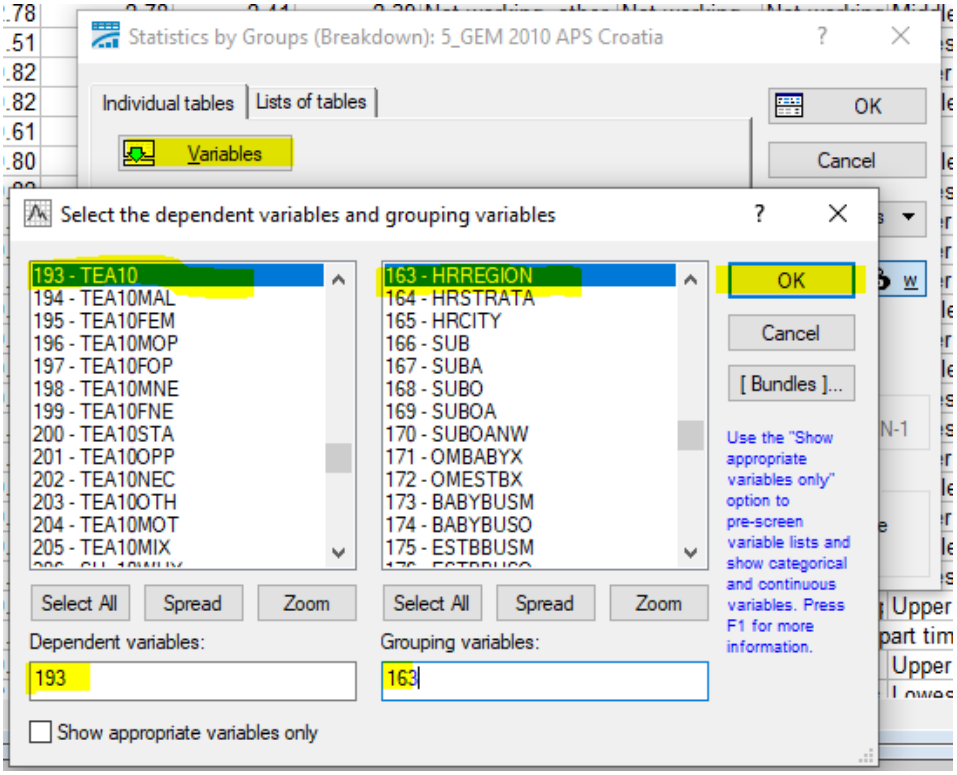
- We would like to examine whether there is a difference in TEA (193) between regions in Croatia (hrregion: 163).
  (5_GEM 2010 APS Croatia.sta)

Guidelines for making a solution:
Import 5_GEM 2010 APS Croatia.sta
weight variable: weight_l
1. Make a breakdown table
2. Set up hypotheses
3. Calculate ANOVA
4. Make a conclusion

Solution:

1. Make a breakdown table: Statistics - Basic statistics and tables - Breakdown & one-way ANOVA - Variables - Dependent: 193; Grouping: 163 - OK - Detailed two way tables





HRREGION	2-Way Tables of Descriptive Statistics (5_GEM 2010 APS Croatia.sta) N=1614 (No missing data in dep. var. list)		
	TEA10 Means	TEA10 Sum of w	TEA10 Std.Dev.
Zagreb region	0.070592	422.425	0.256446
Northern Croatia	0.046121	304.521	0.210091
Slavonia	0.039457	274.411	0.195035
Lika and Banovina	0.060421	124.091	0.239230
Istria and Kvarner	0.060086	188.550	0.238280
Dalmatia	0.051849	300.002	0.222092
All Grps	0.055188	1614.000	0.228418

Notice that samples are > 30.

In Zagreb region there are 7,05% of entrepreneurially active people (among working age population), in Slavonia 3,94%, in Dalmatia 5,18% etc.

We would like to test if these differences are statistically significant.

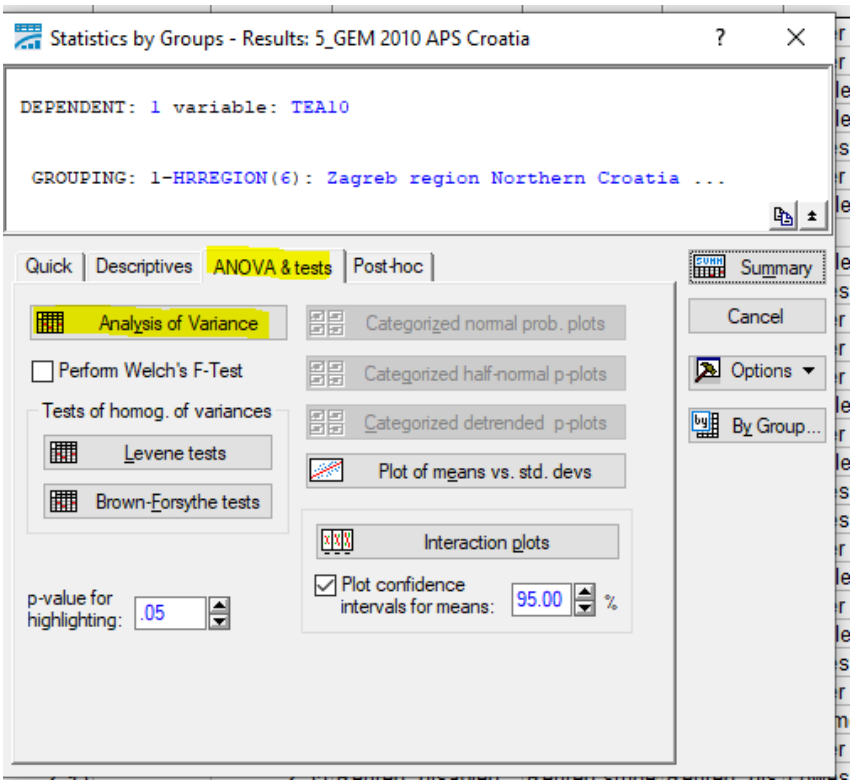
2. Set up hypotheses

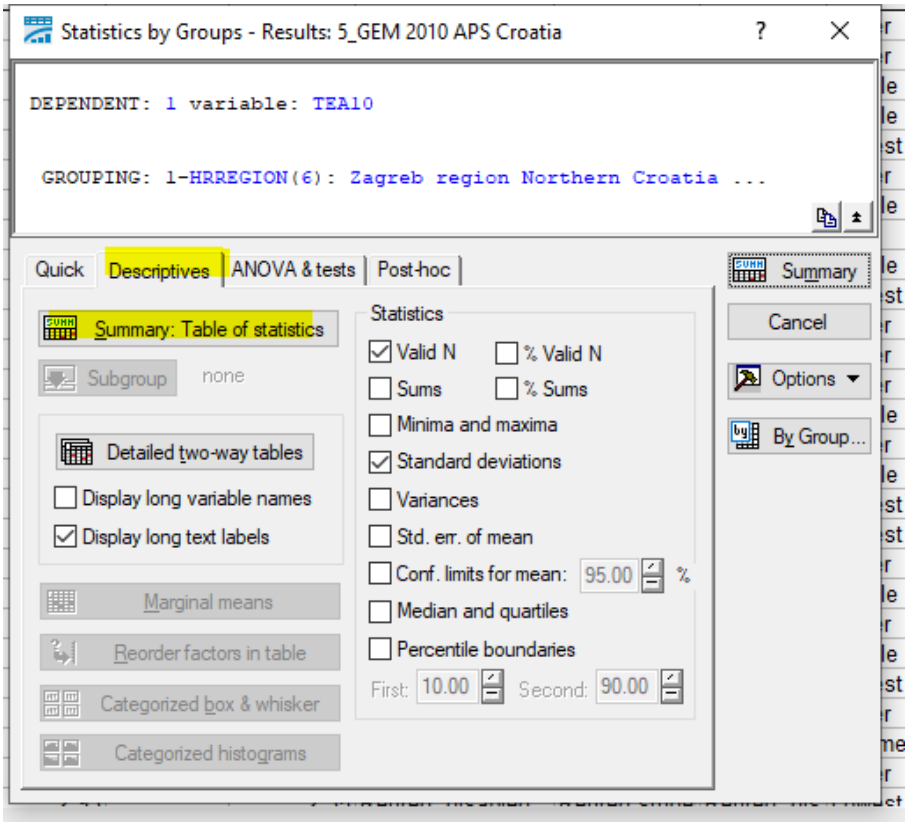
$H_0 : \mu_{zg} = \mu_{slav} = \mu_{dal} = ...$

$H_1 :$ at least one μ is different

3. Calculate ANOVA

Statistics - Basic statistics and tables - Breakdown & one-way ANOVA - Variables - Dependent: 193; Grouping: 163 - OK - ANOVA & tests - Analysis of variance





Variable	Analysis of Variance (5_GEM 2010 APS Croatia.sta) Marked effects are significant at p < .05000							
	SS Effect	df Effect	MS Effect	SS Error	df Error	MS Error	F	p
TEA10	0.204449	5.000000	0.040890	83.95343	1608.000	0.052210	0.783183	0.561747

p=0,5617

p>0,05 - don't reject H_0

4. Make a conclusion

There are no statistically significant differences in mean values of TEA between regions in Croatia. This means that entrepreneurial activity in all regions in Croatia is very similar.

Note: In this example we could have used chi-squared test as well. We would get to the same conclusion.

Example 16

- Explore the relationship between Kontinent - Continent (1) and TEA10 (37) for all GEM countires.
  (6_GEM 2010 APS Master.sta)

Guidelines for making a solution:
Import 6_GEM 2010 APS Master.sta
1. Make a breakdown table
2. Set up hypotheses
3. Calculate Kruskal Wallis ANOVA
4. Make a conclusion

Solution:

1. Make a breakdown table: Statistics - Basic statistics and tables - Breakdown & one-way ANOVA - Variables - Dependent: 37; Grouping: 1 - OK - Detailed two way tables

Kontinent	2-Way Tables of Descriptive Statistics (6_GEM 2010 APS Master.sta) N=60 (No missing data in dep. var. list)		
	TEA10 Means	TEA10 N	TEA10 Std.Dev.
Sjeverna Amerika	12.21088	6	3.27173
Azija	8.02267	12	3.39957
Afrika	21.68786	7	13.47555
Europa	6.14673	25	2.68719
Južna Amerika	20.98020	8	8.55519
Australija	29.95226	2	31.32903
All Grps	11.71278	60	10.02176

Notice that we have small samples.

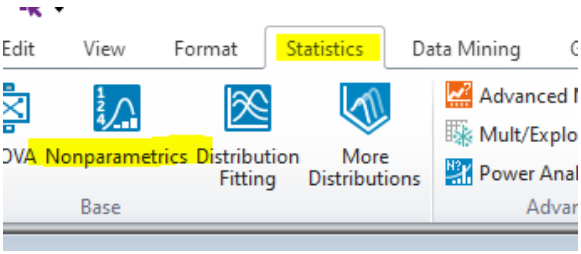
2. Set up hypotheses

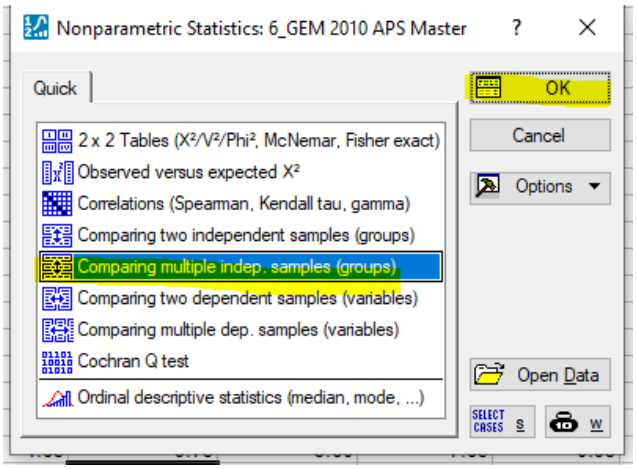
H_0: distributions of TEA are equal in all continents

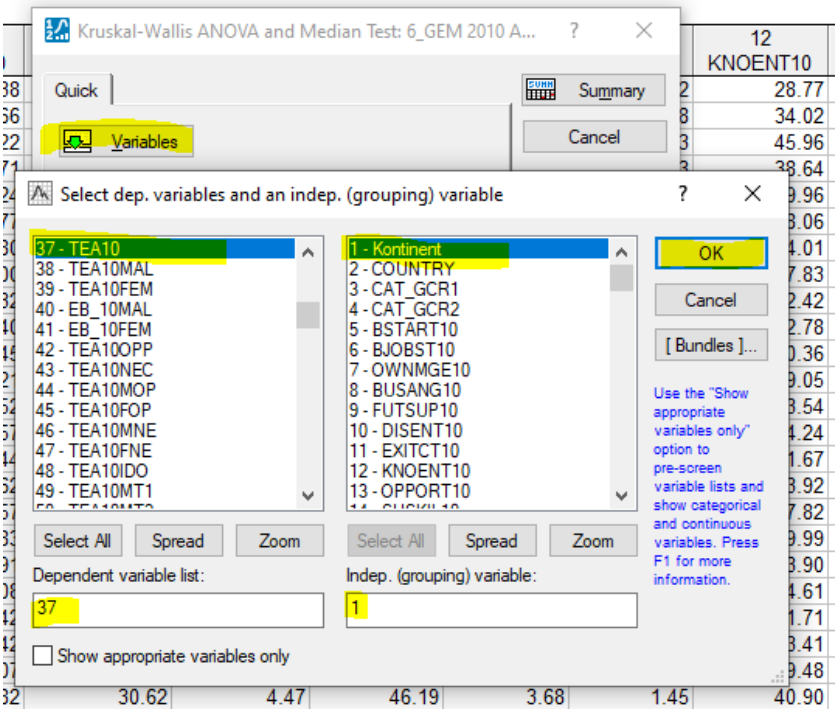
H_1: distributions of TEA are not equal in all continents

3. Calculate Kruskal Wallis ANOVA

Statistics - Nonparametrics - Comparing multiple indep. samples (groups) - Dependent variable list: 37, Indep. (grouping) variable: 1 - OK - Summary







Depend.: TEA10	Kruskal-Wallis ANOVA by Ranks; TEA10 (6_GEM 2010 APS Master.sta) Independent (grouping) variable: Kontinent Kruskal-Wallis test: H (5, N= 60) =30.55097 p =.0000			
	Code	Valid N	Sum of Ranks	Mean Rank
Sjeverna Amerika	61	6	247.0000	41.16667
Azija	62	12	320.0000	26.66667
Afrika	63	7	304.0000	43.42857
Europa	64	25	463.0000	18.52000
Južna Amerika	65	8	406.0000	50.75000
Australija	66	2	90.0000	45.00000

Dependent: TEA10	Median Test, Overall Median = 7.84062; TEA10 (6_GEM 2010 APS Master.sta) Independent (grouping) variable: Kontinent Chi-Square = 23.84571 df = 5 p = .0002						
	Sjeverna Amerika	Azija	Afrika	Europa	Južna Amerika	Australija	Total
<= Median: observed	1.00000	5.00000	2.00000	21.00000	0.00000	1.000000	30.00000
expected	3.00000	6.00000	3.50000	12.50000	4.00000	1.000000	
obs.-exp.	-2.00000	-1.00000	-1.50000	8.50000	-4.00000	0.000000	
> Median: observed	5.00000	7.00000	5.00000	4.00000	8.00000	1.000000	30.00000
expected	3.00000	6.00000	3.50000	12.50000	4.00000	1.000000	
obs.-exp.	2.00000	1.00000	1.50000	-8.50000	4.00000	0.000000	
Total: observed	6.00000	12.00000	7.00000	25.00000	8.00000	2.000000	60.00000

$p<0,05$ - reject H_0

4. Make a conclusion

Distributions of TEA are not equal in all continents.