- 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

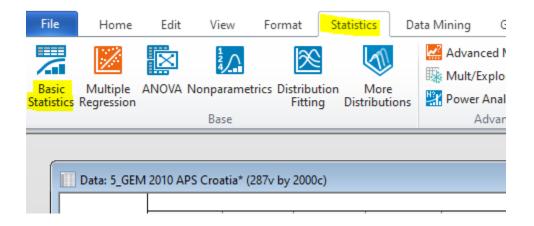
- 1. Make a frequency table of variable 155. Interpret it.
- 2. Make a descriptive statistics of variable 156. Interpret it.

## Solution:

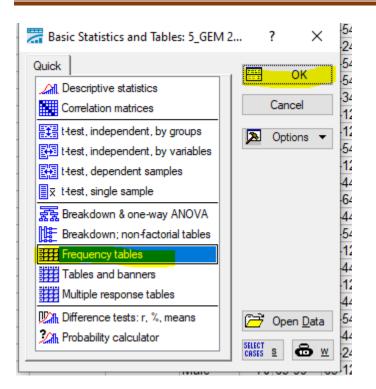
Import: File - Open - 5\_GEM 2010 APS Croatia.sta

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

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



Solutions: Selected statistical tests MoER



Frequency Tables: 5_GEM 2010 APS Croatia ?	× ar
	ar
Variables: none	iummary ar
Quick Advanced Options Descr. Normality Ca	ncel
	ar
Select the variables for the analysis ? ×	ns ▼ ar
	ar
151 - ENG_OMRE 161 - HRHHINC OK	oup ar
152 - ENG_OMOP 162 - HRREDUC 153 - ENG_BARE 163 - HRREGION	ar
154 - ENG_EXRE 164 - HRSTRATA Cancel	ar
155 - GENDER 165 - HRCITY	ar
156 - AGE 166 - SUB [Bundles]	ar
158 - AGE9C 168 - SUBO	ar
159 - OCCU 169 - SUBOA	ar
160 - HHSIZE 170 - SUBOANW	ar
< >>	🔂 👱 ar
	omnts
Select All Spread Zoom	ar
Select variables:	n ar
	e ar
155	) ar
Show appropriate variables only	u-u4 ⊤ar
	5-120   ar

🚟 Frequency Tables: 5_GEM 2010 APS Croatia	? ×
GENDER	Summary a
Quick Advanced Options Descr. Normality	Cancel
Summary: Frequency tables	Doptions 🔻
Histograms	By Group
Descriptive statistics	
<u>3</u> D histograms, bivariate distributions	
	SELECT CASES S
	Wghtd momnts
	MD deletion
	Casewise
iviaie 4	

	Frequency table	requency table: GENDER: DEMA. What is your gender? (5_GEM 2010 APS Croatia.sta)			
Category	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:

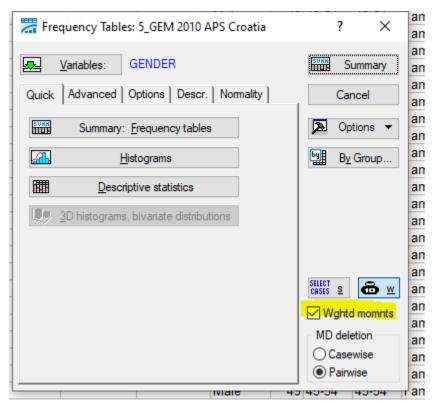
1. click on weight (bottom right corner), click 'On', weight variable: weight\_I

Spreadsheet C	Case Weights ?	×			
Weight variabl	le: WEIGHT_L				
Status On Off	Double-click on edit press F2 to choose all variables. Values selected variable wil as case multipliers, otherwise noted in th respective analysis module.	from list of of I be used unless ne or analysis			
55		F	emale S	el:OFF	Weight:OFF

Setting Spreadsheet Case Weights			?	×
You have chosen to enable Case We Spreadsheet Case Weights will be in which are created from this Spreadsh automatically saved with the Spreads explicitly turn them off.	effect for all Statistics and eet, and these Case Weig	Graphs		
Also, any Macros run on this Spreads because these conditions are part of a macro, while analyzing this data set part of the macro.	the data set. However, if	you record	-	
To set Case Weights which will be re Analysis level by clicking the WEIGH dialogs.				•
Don't display this warning again			OK	

Female Sel:OFF Weight: ON

2. in frequency table window click wghtd momnts



	Frequency table: GENDER: DEMA. What is your gender? (5_GEM 2010 APS Croatia.sta)			
Category	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

Quick	ОК
Descriptive statistics	
Correlation matrices	Cancel
ttest, independent, by groups	>> Options ▼
🕶 t-test, independent, by variables	
t-test, dependent samples	
xxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxxx	
Breakdown; non-factorial tables	
Frequency tables	
Tables and banners	
Multiple response tables	
🕅 Difference tests: r, %, means	🗁 Open Data
Ch Probability calculator	

in descriptive statistics window click wghtd momnts

Descriptive Statistics: 5_GEM 2010 APS Croatia	? ×
👥 Variables: AGE	Summary
Quick Advanced Robust Normality Prob. & Scatterplots Categ. plots Options	Cancel
Summary: Statistics       Emergency tables       Emergency tables         Imm       Frequency tables       Imm       Histograms	Deptions 👻
Box & whisker plot for all variables	
Graphical comparative summary display	
	DF =
	MD deletion
	Pairwise

	Descriptive Statistics (5_GEM 2010 APS Croatia.sta)				
Variable	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).

- 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\_I and wghtd momnts

## **Guidelines for making a solution:**

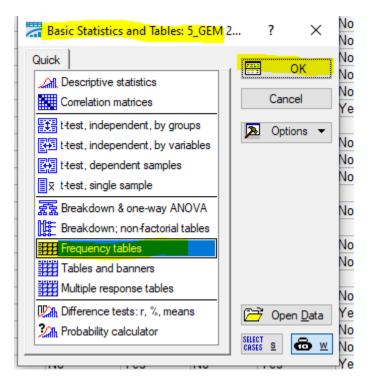
Import 5\_GEM 2010 APS Croatia.sta

- 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



Frequency Tables: 5_0	GEM 2010 AP	S Croatia	?	×
➡ Variables: FRF.	AIL10		SUNN SU	mmary
Quick Advanced Option	ns Descr.	Normality	Can	cel
Summary: Freq	uency tables		🔈 Opt	ions 🔻
Histogr	ams		By By (	Group
Descriptive	statistics			
1 3D histograms, biva	riate distribution	ns		
			SELECT CRSES S	@ ⊻
			Vghtd	momnts
			- MD dele	
			Casev	
110	110		Pairwi	se

	Frequency table: FRF APS Croatia.sta)	AIL10: FRFAIL adapted to	make it fit for national	level agg (5_GEM 2010
Catego ry	Count	Cumulative Count	Percent	Cumulative Percent
No	979.0269	979.027	60.65842	60.6584
Yes	632.3383	1611.365	39.17833	99.8368
Missin g	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

Basic Statistics and Tables: 5_GEN	
uick	ОК
Call Descriptive statistics	
Correlation matrices	Cancel
🗱 t-test, independent, by groups	Options
🕶 t-test, independent, by variables	
t-test, dependent samples	
⊼ t-test, single sample	
嘉쿭 Breakdown & one-way ANOVA	
Breakdown; non-factorial tables	
Frequency tables	
Tables and banners	
Multiple response tables	
Difference tests: r, %, means	(귿 Open Data
Ch Probability calculator	

Descriptive Statistics: 5_GEM 2010 APS Croatia	? ×
Jariables: TEA10	Summary
Quick Advanced Robust Normality Prob. & Scatterplots Categ. plots Options	Cancel
Summary: Statistics Graphs 1 🚟 Graphs 2	🔈 Options 🔻
Erequency tables	By Group
Box & whisker plot for all variables	
Graphical comparative summary display	SELECT S
	CRSES S W
	DF =
	● W-1 ○ N-1 MD deletion
	Pairwise

	Descriptive Statistics (5_GEM 2010 APS Croatia.sta)						
Variable	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.

- 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)

- 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'

Solutions: Selected statistical tests MoER

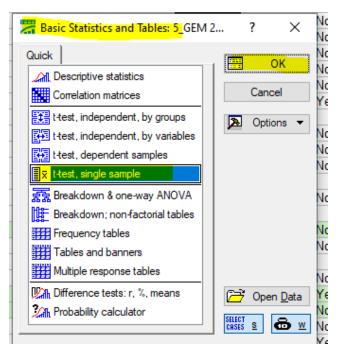
Spreadsheet Case Selection Conditions		?	×
Selections Display Subset/Random Sampling			
Enable Selection Conditions	Review <u>V</u> ariables:	Clear All	
Include cases			
Specific, selected by:			
By expression: TEA10=yes		$\sim$	
Functions		$\sim$	
or case number:			
Exclude cases (from the set of cases defined in the 'Include c	ases' section)		
By expression:		$\sim$	
Functions		~	
or case number:			
By case number: Enter case numbers and/or ranges. Example: 1, 3, 5	i-12		
By expression: Use the same operators, functions, and syntax as in Use variable names or v1, v2 v0 is the case nun Examples: (a) v1=0 OR age>18 (b) gender="MALE"	iber (v0<4 means cases 1-3)		
In case of conflict, variable names take precedence values by appending \$, as in "value"\$.		ecify text	
🗁 Open 📳 Save As	ОК	Cancel	
_GEM 2010 APS Croatia C976,V184	No	Sel:OFF	Weight: ON

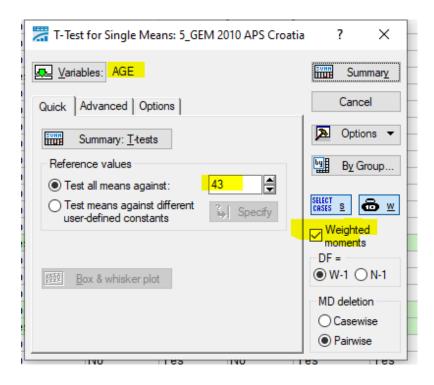


### 3. Test the hypotheses in Statistica.

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

click wghtd momnts





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

The results show that  $p < \alpha$  (0,05) and we should reject H<sub>0</sub>.

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.

- 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

	Descriptive Statistics (6_GEM 2010 APS Master.sta)							
Variable	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.

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

p<0,05 - reject H<sub>0</sub>

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.

- 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)

 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:**

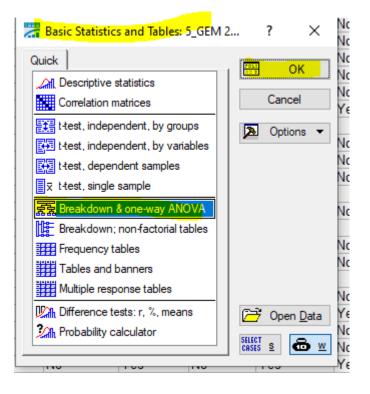
Import 5\_GEM 2010 APS Croatia.sta

- 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



## Solutions: Selected statistical tests MoER

									. ř
No	No	The Statistics	s by (	Groups (Brea	kdown): 5_(	GEM 2010 APS	Croatia	?	×s
No	No								
No	Yes	Individual ta	bles	Lists of tables	3			OK OK	S
No	No								S
No	No	va Va	ariable	es				Cancel	s
No	No	Dependen	t no	ne					
No	No	Grouping:	no					Options	-
No	No	circuping.							
No	No		5					CASES S	w S
A Selec	ct the depe	ndent variable	s and	d grouping v	ariables		? ×		s
191 - NE 192 - NE 193 - TE 194 - TE 195 - TE 196 - TE 197 - TE 198 - TE	AUALI10 3GOOD10 3STAT10 3MEDI10 A10 A10MAL A10FEM A10FEM A10FOP A10FOP A10FNE	ad Zoom	~	193 - TEA10 194 - TEA10 195 - TEA10 196 - TEA10 197 - TEA10 198 - TEA10 200 - TEA10 200 - TEA10 201 - TEA10 202 - TEA10 203 - TEA10 203 - TEA10 204 - TEA10	IMAL IFEM IMOP IFOP IFNE ISTA ISTA IOPP INEC IOTH	Zoom	OK Cancel [Bundles] Use the "Show appropriate variables only" option to pre-screen variable lists and show categorical and continuous	MD deletion Casewise Pairwise	s I-1 S S S S Yes Yes
Depende	ent variables			Grouping var	iables:		variables. Press		
188				193			F1 for more information.	3.13	30
100				133			monnation.	3.37	41
Show	annmniate	variables only						1.10	43
JIOW	appropriate	valiables of lly					.:	0.51	29
76.99	30.		<b>b</b> ð.2	0	18.09	60.87	10.05	5.31	85

	-Way Tables of Descriptive Statistics (5_GEM 2010 APS Croatia.sta) Smallest N for any ariable: 1611										
TEA10	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.

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

### 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

Basic Statistics and Tables: 5_GEN	12 ? ×					
uick	ок	N				
Call Descriptive statistics						
Correlation matrices	Cancel					
😫 t-test, independent, by groups	Doptions	- É				
t-test, independent, by variables		No				
t-test, dependent samples		Ne				
⊼ t-test, single sample		No				
हेडू Breakdown & one-way ANOVA		No				
Breakdown; non-factorial tables						
Frequency tables		No				
Tables and banners		No				
Hultiple response tables		No				
h Difference tests: r, %, means	🗁 Open <u>D</u> at					
The Probability calculator		No				
	CRSES S	w No				
100 100 100	100	Y€				
	No No	Yes	No	10	No	Y
No No	T-Test for Indepen	dent Samples by Gro	ups: 5_GEM 2010	APS Cro	. ? ×	Y
No No				li l	-	V Y
No No	Variables: Depend	lent: none			Summary	- Y
No No	Groupin	ig: none			Cancel	1
No No Coo	le for Group 1:	Code for	Group 2:		🔈 Options 🔻	ā.
Select the dependent variables	and one grouping	variable	? ×		· ·	N
					By Group	. N
188 - FRFAIL10 189 - EQUALI10	193 - TEA10 194 - TEA10MA	^	ОК		SELECT S 🔂 W	
190 - NBGOOD10	195 - TEA10FE	N	Cancel		Weighted	1
191 - NBSTAT10 192 - NBMEDI10	196 - TEA10MO 197 - TEA10FOI		Canoci		moments	Y
193 - TEA10	198 - TEA10MN	E	[Bundles]		DF =	٦Y
194 - TEA10MAL 195 - TEA10FEM	199 - TEA10FN		Use the "Show		● W-1 ○ N-1	Y
196 - TEA10MOP	201 - TEA10OP	P	appropriate		MD deletion	V
197 - TEA10FOP 198 - TEA10MNE	202 - TEA10NE 203 - TEA10OT		variables only" option to		Casewise	Ý
199 - TEA10FNE			pre-screen		Pairwise	Y
Select All Spread Zoom	Select All	Spread Zoom	variable lists and show categorical		NU	Y
Dependent variables:	Grouping variable		and continuous variables. Press	S	Yes	N Y
188			F1 for more information.			Y
					1	-
Show appropriate variables only						

T-Test for Indep	endent Sample	s by (	Groups: 5_GEM 2010 A	APS Cro	o ?	×		lo
		1						lo
-		_			SUNH	Summarv	Y	/es
	endent: FRFAIL1	0				Samuel	Y	/es
Grou	ping: TEA10				C	ancel	Y	/es
Code for Group 1: No		Code t	for Group 2: Yes			Dellana -	Y	/es
		2	Options 💌	Ν	٥V			
Quick Advanced	Options				by E	By Group	Y	/es
_			I have a second second second			or or output	Y	/es
Display long varia	able names		Homogeneity of variand	ces	SELECT CASES S	<b>@</b> ⊻	Y	/es
Test w/ separate	variance estimat	tee	Levene's test					
			Brown & Forsythe te	est	⊠ <sup>Wei</sup>	ghted	N	lo
Multivariate test (	Hotelling's T <sup>2</sup> )					ients		
Equivalence test	(TOST) 🔝				DF =		-	lo
					• W •	1 🔿 N-1 👘	Y	/es
p-value for highlighti	na: .05 兽					1.0	Y	/es
					- MD d	eletion	N	١o
CI for estimates	95.00	%			⊖ Ca	asewise	Y	/es
					Pa	airwise	Y	/es
		110					N	٥
No	No	No		No				/es
NI-	NI-	NI-		V		NI-		1-

	T-tests; Grouping: TEA10: Involved in Total early-stage Entrepreneur (5_GEM 2010 APS Croatia.sta) Group 1: No Group 2: Yes													
Variabl e	Mean No	Mean Yes	t- value	df	p	t separ var.es t.	df	p 2- sided	Sum of w No	Sum of w Yes	Std.D ev. No		F-ratio Varian ces	
FRFAI	0.402			<mark>1609.</mark>	0.000	4.063		0.000						0.0432
L10	687	025	021	<mark>365</mark>	<mark>480</mark>	<mark>615</mark>	602	095	292	369	600	554	33	<mark>52</mark>

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 <mark>t-test</mark> 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<sub>0</sub>

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.

• 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

	2-Way Tables of Descriptive Statistics (5_GEM 2010 APS Croatia.sta) Smallest N for any variable: 60 Include condition: TEA10='yes'									
GEND ER	SUMONTUS Means	SUMONTUS Sum of w	SUMONTUS Std.Dev.							
Male	144794.2	40.63063	288519.5							
Femal e	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

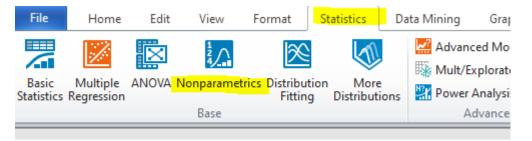
### 3. Set up hypotheses

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

H1: 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



	Mann-Whiti GENDER N					GEM 201	0 APS Cr	oatia.sta	a) By var	iable
variable	Rank Sum Female	Rank Sum Male	U	Z	p-value	Z adjusted	p-value	Valid N Female		2*1sided exact p
SUMON TUS	787.0000	1841.000	462.000 0		0.13927 6	۔ 1.48248	0.13821 3	25	47	<mark>0.14006</mark> 0

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.

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)

- 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:**

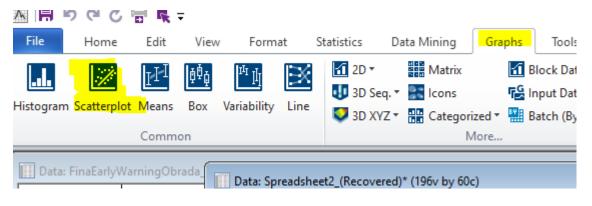
Import 6\_GEM 2010 APS Master.sta

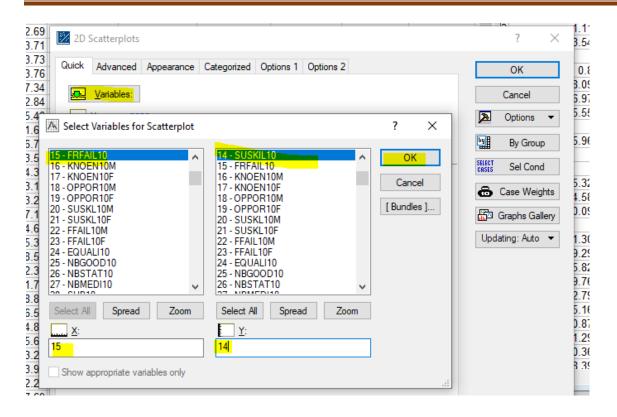
- 1. Draw scatterplots.
- 2. Test if our variables are normally distributed.
- 3. Set up hypotheses.
- 4. Test the hypotheses in Statistica. Make a conclusion.

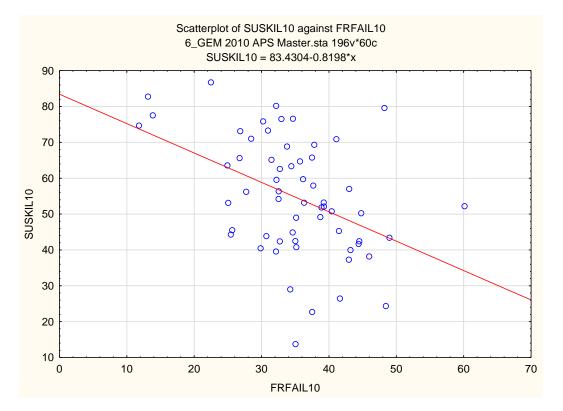
# Solution

### 1. Draw a scatterplots.

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







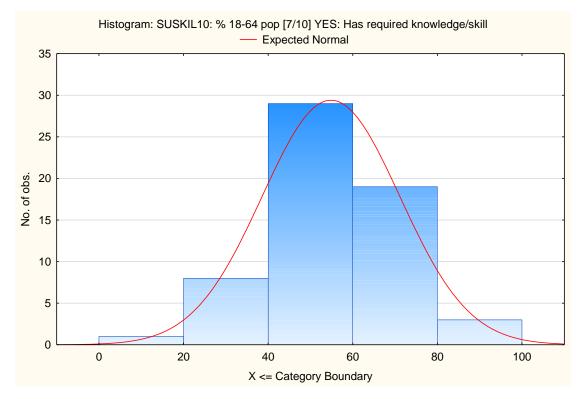
Negative correlation can be noticed - one variable increases with the decrease of the other variable.

## 2. Test if our variables are normally distributed.

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

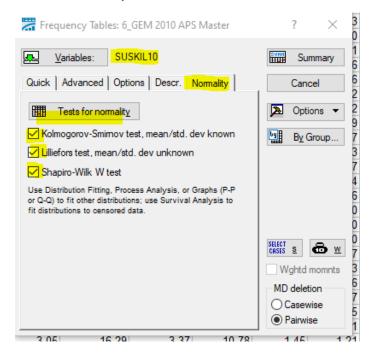
Advanced: 'Neat intervals', approximate no.: 6

Basic Statistics and Tables: Spreadsh ?	×
Quick	
Contractive statistics	
Correlation matrices Cancel	
t-test, independent, by groups	
t-test, independent, by variables	
t-test, dependent samples	
≣⊼ ttest, single sample	
嘉嘉 Breakdown & one-way ANOVA	
Breakdown; non-factorial tables	-
Frequency tables	
Tables and banners	
Multiple response tables	
Difference tests: r, %, means	<u>D</u> ata
Calculator	3 w
	h
Frequency Tables: 6_GEM 2010 APS Master	? × 3
Uariables: SUSKIL10	
Quick Advanced Options Descr. Normality	Cancel 6 Cancel 2 P Options ▼ 2 By Group 7 3 7 4 6
Summary: Frequency tables	Doptions - 2
Categorization methods for tables & graphs:	By Group 7
○ All distinct values  with text labels	3
◯ No. of exact intervals: 10	7
Intervals; approximate no.: 6	6
◯ Step size: 1	
starting at: 0 🚽 or 🗹 at minimum	SELECT CRSES S D W 7
O Integer categories vith text labels	Wghtd momnts 3
O Specific grouping codes (values)	Wghtd momnts 3 MD deletion 7 Casewise 5 Pairwise 1
O User-specified categories	Casewise 5
	Pairwise
3 05 16 29 3 37 10 78	145 121



#### It looks normal.

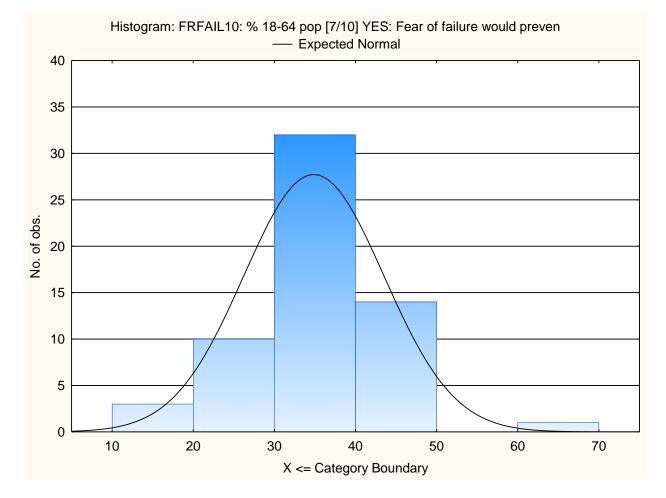
#### Test for normality:



	Tests of Normality (6_GEM 2010 APS Master.sta					laster.sta)
Variable	N	max D	K-S p	Lilliefors p	W	р
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:



	Tests of Normality (6_GEM 2010 APS Master.sta)					
	N	max D	K-S	Lilliefors	W	p
Variable		max b	р	р	••	٢
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

Quick		[====]	OK					
Descriptive statistic	s		ОК					
Correlation matrices	S	Ca	ancel	_				
ttest, independent	, by groups		ptions 🔻	-				
t-test, independent	, by variables		puons +					
t-test, dependent s	amples							
	e			_				
蒙蒙 Breakdown & one-	way ANOVA			-				
Breakdown; non-fa	actorial tables							
Frequency tables								
Tables and banner	s							
Multiple response t	ables							
Difference tests: r,	%, means	🖻 o	pen <u>D</u> ata	-				
Children Probability calculate	or							
		CASES S	<b>⊕</b> <u>₩</u>					
	0.00	2.02	10.04	2 70		1 03	38.6	4
				-				
71 🛛 🛣 Product-Mo	ment and Part	tial Correlation	ons: 6_GE	?	)	× 30	39.9	-
71 Roduct-Mo	ment and Part			? •••• •••		51	38.0	6
71 Product-Mo 73 76 Que One variable	ment and Part	tial Correlatio		? Sur	) mmary	51 2 36	38.0 24.0	)6 )1
71 Product-Mo 73 76 One <u>variable</u> 34 End Edd	e list			? Bur		51 2 36 16	38.0 24.0 47.8	)6 )1 )3
71 Product-Mo 73 76 One <u>variable</u> 34 End Edd	e list			? I Sur		51 2 36	38.0 24.0 47.8	)6 )1
71 Product-Mo 73 76 One <u>v</u> ariable 34 Select one or two vari	ment and Part	wo lists (rect.	matrix)	? I Sur		51 2 36 16	38.0 24.0 47.8 × 4	)6 )1 )3 )2
71 Product-Mo 73 76 One <u>variable</u> 34 End Edd	e list able lists	wo lists (rect.	matrix)	? Sur		51 2 36 16	38.0 24.0 47.8 × 4 7	16 11 13 12 78
71 Product-Mo 73 76 One variable 34 Select one or two vari 14 - SUSKIL10 15 - FRFAIL10 16 - KNOEN10M	e list able lists	wo lists (rect. 15 - FRFAIL1 16 - KNOEN1 17 - KNOEN1	matrix)	? Sur		2 51 2 36 16 ? 0K	38.0 24.0 47.8 × 4 7 3 0 5	06 01 03 03 02 05 05 05 04
71         Product-Mo           73         76           76         One variable           34         End list           76         Select one or two variable           77         Select one or two variable           78         Select one or two variable           79         FFFAIL10           15 - FRFAIL10           16 - KNOEN10M           17 - KNOEN10F	e list <u>E</u> <u>T</u> able lists	wo lists (rect. 15 - FRFAIL1 16 - KNOEN1 17 - KNOEN1 18 - OPPOR1	matrix)	? Sur		51 2 36 16 ?	38.0 24.0 47.8 × 4 7 el 2	06 01 03 03 02 78 06 05 05 04 24
71         Product-Mo           73         One variable           76         One variable           34         Ent lat           Select one or two variable         Select one or two variable           15 - FRFAIL10         FRFAIL10           16 - KNOEN10M         17 - KNOEN10F           18 - OPPOR10M         19 - OPPOR10F	able lists	wo lists (rect. 15 - FRFAIL1 16 - KNOEN1 17 - KNOEN1 18 - OPPOR1 19 - OPPOR1 20 - SUSKL10	matrix)	? Sur		2 51 2 36 16 ? 0K	38.0 24.0 47.8 × 4 7 el 3 5 2 s ] 6	16 11 33 42 78 66 95 54 44 44 57
71         Product-Mo           73         One variable           76         One variable           34         End Ed.           Select one or two variable         Select one or two variable           15 - FRFAIL10         FRFAIL10           16 - KNOEN10M         FRFAIL10           17 - KNOEN10F         18 - OPPOR10F           19 - OPPOR10F         20 - SUSKL10M	able lists	wo lists (rect. <b>15-FRFAIL1</b> 16 - KNOEN1 17 - KNOEN1 18 - OPPOR1 19 - OPPOR1 20 - SUSKL1( 21 - SUSKL1)	matrix)	? Sur		2 36 16 ? Cance [Bundles	38.0 24.0 47.8 × 7 el 5 el 5 s ] 6 9	16 11 33 .2 78 36 55 54 4 4 4 57 72
71         Product-Mo           73         One variable           76         One variable           34         Ent Ett.           Select one or two variable         Select one or two variable           15 - FRFAIL10         FRFAIL10           16 - KNOEN10M         17 - KNOEN10F           18 - OPPOR10M         19 - OPPOR10F	able lists	wo lists (rect. 15 - FRFAIL1 16 - KNOEN1 17 - KNOEN1 18 - OPPOR1 19 - OPPOR1 20 - SUSKL10	matrix)	? Sur		2 36 16 ? OK Cance	38.0 24.0 47.8 × 4 7 5 5 5 5 5 8 1 9 8	06 01 03 02 05 05 05 05 04 02 02 02
71         Product-Mo           73         76           76         One variable           34         End late           75         FRFAIL10           15         FRFAIL10           16         KNOEN10M           17         KNOEN10F           18         OPPOR10F           20         SUSKL10M           21         SUSKL10F           22         FFAIL10M           23         FFAIL10M           23         FFAIL10M	able lists	15 - FRFAIL1 16 - KNOEN1 17 - KNOEN1 18 - OPPOR1 19 - OPPOR1 20 - SUSKL1( 21 - SUSKL1( 22 - FFAIL10) 23 - FFAIL101 24 - EQUAL1	matrix) 0 0 0 0 0 0 0 0 0 0 0 0 0	? Sur Cano		2 36 16 ? Cance [ Bundles Use the "St appropriate variables or	38.0 24.0 47.8 × 4 7 8 8 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9 9	06 01 03 03 02 05 05 05 05 05 05 05 05 05 05 05 05 05
71         Product-Mo           73         76           76         One variable           34         End Ed.           54         End Ed.           55         FRFAIL10           16         KNOEN10M           17         KNOEN10F           18<	able lists	15-FRFAIL1 16-KNOEN1 17-KNOEN1 18-OPPOR1 20-SUSKL10 21-SUSKL10 22-FFAIL100 23-FFAIL100 23-FFAIL101 23-FFAIL102	matrix)	? Sur		2 36 16 ? OK Cance [Bundles Use the "St appropriate	38.0 24.0 47.8 × 4 7 el 2 2 s j 6 9 s j 6 9 s j 9 9 now 8 8 nly 9	06 01 03 03 02 05 05 05 05 05 04 05 05 05 00 00
71         Product-Mo           73         76           76         One variable           34         End late           75         FRFAIL10           15         FRFAIL10           16         KNOEN10M           17         KNOEN10F           18         OPPOR10F           20         SUSKL10M           21         SUSKL10F           22         FFAIL10M           23         FFAIL10M           23         FFAIL10M	able lists	wo lists (rect. 15 - FRFAIL1 16 - KNOEN1 17 - KNOEN1 18 - OPPOR1 19 - OPPOR1 20 - SUSKL1( 21 - SUSKL1( 22 - FFAIL10) 23 - FFAIL10 23 - FFAIL10 24 - EQUAL1 25 - NBGOOL 26 - NBSTAT 27 - NBMEDI	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	? Sur		2 36 16 ? Cance [ Bundles Use the "Sh appropriate variables or option to pre-screen variable list	38.0 24.0 47.8 × 4 s ] 9 s ] 9 s ] 9 s and 7 5 5 2 2 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5	06 01 03 03 02 05 05 05 05 05 05 05 05 05 05 05 05 05
Product-Mo           73           76           76           76           76           76           76           76           76           76           76           76           76           76           76           76           76           77           76           76           76           76           76           76           76           76           76           76           76           77           76	able lists	wo lists (rect. <b>15-FRFAIL1</b> 16 - KNOEN1 17 - KNOEN1 17 - KNOEN1 19 - OPPOR1 20 - SUSKL1( 22 - FFAIL101 23 - FFAIL101 24 - EQUAL11 25 - NBGOOD 26 - NBSTAT 27 - NBMEDI 20 - SUBMEDI	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	Cano		2 36 51 36 16 ? Cance [Bundle: Use the "Sh appropriate variables or option to pre-screen variable side show catego and continu	38.0 24.0 47.8 × 4 7 el 5 el 5 2 5 2 5 2 5 2 5 2 5 2 5 2 5 2 5 2 5	06 11 33 42 58 66 95 64 44 67 72 99 90 10 11 12 12 12 12 12 12 12 12 12
Product-Mo           73         76           76         One variable           34         End Ed.           58         Select one or two variable           15 - FRFAIL10         Frequencies           16 - KNOEN10M         17 - KNOEN10F           18 - OPPOR10F         20 - SUSKL10M           20 - SUSKL10M         21 - SUSKL10F           22 - FFAIL10F         23 - FFAIL10F           24 - EQUALI10F         24 - EQUALI10           25 - NBGOOD10         10	able lists	wo lists (rect. 15 - FRFAIL1 16 - KNOEN1 17 - KNOEN1 18 - OPPOR1 19 - OPPOR1 20 - SUSKL1( 21 - SUSKL1( 22 - FFAIL10) 23 - FFAIL10 23 - FFAIL10 24 - EQUAL1 25 - NBGOOL 26 - NBSTAT 27 - NBMEDI	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	? Sur Corr		2 36 51 36 16 ? Cance [Bundles Use the "St appropriate variables or option to pre-screen variable list show categy and continu variables. F	38.0 24.0 47.8 × 47 8 × 7 6 9 8  9 9 1 9 9 1 9 9 1 9 9 1 9 9 1 9 9 1 9 9 1 9 9 1 9 9 1 9 9 1 9 9 1 9 9 1 9 9 1 9 9 1 9 9 1 9 9 1 9 9 9 1 9 9 9 1 9 9 9 1 9 9 1 9 9 9 1 9 9 9 9	11       13       14       15       16       15       14       15       14       17       12       13       14       15       16       17       10       11       11
Image: Constraint of the second sec	able lists	wo lists (rect. <b>15-FRFAIL1</b> 16 - KNOEN1 17 - KNOEN1 17 - KNOEN1 19 - OPPOR1 20 - SUSKL1( 21 - SUSKL1( 22 - FFAIL101 23 - FFAIL101 24 - EQUAL11 25 - NBGOOD 26 - NBSTAT 27 - NBMEDI 20 - SUBMEDI	matrix) 0 0M 0F 0M 0F 0M 0F 0M 0F 00 010 10 10 10 Spread	Zoor		2 36 51 36 16 ? Cance [Bundle: Use the "Sh appropriate variables or option to pre-screen variable side show catego and continu	38.0 24.0 47.8 × 4 7 2 2 2 2 2 2 2 2 2 2 2 2 2	06       11       13       .2       78       366       95       34       37       92       99       90       11       11       11
Product-Mo           76         One variable           34         Event late           35         Select one or two variable           15         FRFAIL10           16         KNOEN10M           17         KNOEN10F           18<	able lists	wo lists (rect. 15- FRFAIL1 16 - KNOEN1 17 - KNOEN1 18 - OPPOR1 20 - SUSKL1( 21 - SUSKL1( 21 - SUSKL1( 22 - FFAIL10F 23 - FFAIL10F 24 - EQUAL1 25 - NBGOOD 26 - NBSTAT 27 - NBMEDI 26 - SUB12 Select All	matrix) 0 0M 0F 0M 0F 0M 0F 0M 0F 00 010 10 10 10 Spread	Zoor		2 36 51 36 16 ? Cance [ Bundles use the "Sh appropriate variable list show catege and continu variables. F F1 for more	38.0 24.0 47.8 × 4 7 el 2 s ] 9 s and 6 orical 0 ous 4 2 ress 4 9 7 7	6 6 11 13 22 8 8 8 6 5 5 4 4 4 4 4 7 7 22 9 9 00 11 11 13 8 8 8 8 8 8 8 8 8 8 8 8 8

Product-Moment and Partial Correlations: 6_G	ie ? ×
One variable list Two lists (rect. matrix)	Summary
First list: SUSKIL10	Cancel
Second list: FRFAIL10	🔈 Options 🔻
Quick Advanced Options Color maps	
Display format for correlation matrices	(Only single-list
O Display simple matrix (highlight p's)	square matrices can be saved)
Display r, p-values, and N's	SHEET
O Display detailed table of results	CRSES S
Display long variable names	Weighted moments
Extended precision calculations	DF =
p-value for highlighting: 05	● W-1 ○ N-1
$\checkmark$ Include means and std. devs. in square matrices	MD deletion
	Casewise
	O Pairwise

	Correlations (6_GEM 2010 APS Master.sta) Marked correlations are significant at p < .05000 N=60 (Casewise deletion of missing data)
е	FRFAIL10
SUSKI L10	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.

 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 REF				R REPORT
	Count	Cumulative	Percent	Cumulative	
Category		Count		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)'

Solutions: Selected statistical tests MoER

preadshee	t Case Se	ection Conditions			?	×
Selections	Display	Subset/Random Samplin	ng			
🗹 Enabl	e Selectio	n Conditions	Review	w <u>V</u> ariables:	Clear All	
Include	cases					
Spe	ecific, sele	ected by:				
By	expressior	n: CAT_GCR2='Stage 1:1	factor driven (includes tran	sition countries to	$\sim$	
	Functions				$\sim$	
or ca	ise numbe	r:				
Exclude	e cases (fr	om the set of cases define	d in the 'Include cases' se	ction)		
By	expressior	1:			$\sim$	
	Functions	3			~	
or ca	ise numbe	r:				
			- Evample: 1 2 5 12			
	umber: Ent	er case numbers and/or range	es. Example: 1, 3, 5-12 ns, and syntax as in the sprea	adsheet formulas:		
By case n	umber: Ent sion: Use Use	er case numbers and/or range e the same operators, functio e variable names or v1, v2	ns, and syntax as in the sprea v0 is the case number (v0	4 means cases 1-3	)	
By case n	umber: Ent sion: Use Use Exa In c	er case numbers and/or range e the same operators, functio e variable names or v1, v2 amples: (a) v1=0 OR age>18	ns, and syntax as in the sprea v0 is the case number (v0 b) gender='MALE' AND v4 es take precedence over varia	≪4 means cases 1-3 ⇔(v5+v6)		

### 3. Set up hypotheses

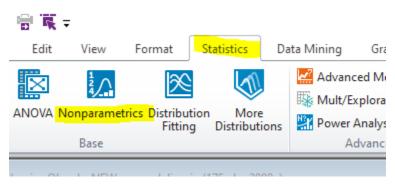
Since we have a small sample, it is recommended that we use Spearman correlation coefficient.

H<sub>0</sub>: no correlation

H<sub>1</sub>: correlation exists

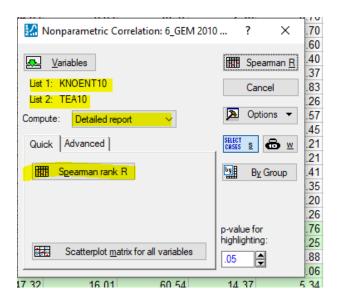
#### 4. Calculate Spearman correlation coefficient

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



## Solutions: Selected statistical tests MoER

Nonparametric Statistics: 6_GEM 2010 APS Mast	er?×
Quick	ОК
Den 2 x 2 Tables (X²/V²/Phi², McNemar, Fisher exact)	Cancel
X Observed versus expected X <sup>2</sup>	▶ Options ▼
Correlations (Spearman, Kendall tau, gamma)	
Comparing two independent samples (groups)	
Comparing multiple indep. samples (groups)	
Comparing two dependent samples (variables)	
Comparing multiple dep. samples (variables)	
Cochran Q test	(구국 Open Data
Critical descriptive statistics (median, mode,)	
64.99 20.56 63.10 1	7.58 8.06



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

p<0,05 - reject H<sub>0</sub>

### 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

		REPORT 2009-2010 - cat_gcr2='Stage 2: )'		
Category	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<sub>0</sub>: no correlation

H<sub>1</sub>: correlation exists

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

It works for the efficiency driven countries as well.

 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\_I

- 1. Set up hypotheses
- 2. Calculate crosstable
- 3. Calculate  $\chi^2$
- 4. Make a conclusion

# Solution:

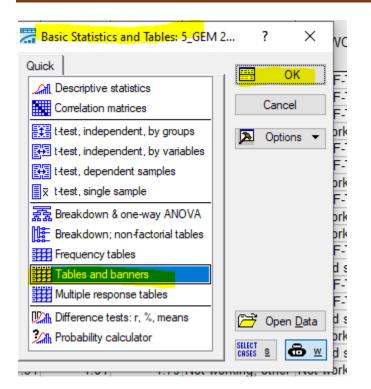
### 1. Set up hypotheses

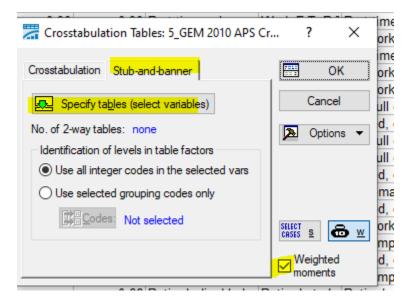
Since both variables are categorical - Gender: male, female; Suskil: yes, no, we can use  $\chi^2$  test.

H<sub>0</sub>: there is no dependence between gender and have knowledge and skills to start a business

H<sub>1</sub>: 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





0.82 Crosstabulation Tab	es: 5_GEM 2010 APS Cr ? >	orking Middle 33
K Select two lists of variables (fac	tors) for the table	? × 33
197 - SUSKIL10           188 - FRFAIL10           189 - EQUALI10           190 - NBGOOD10           191 - NBSTAT10           192 - NBMEDI10           193 - TEA10           194 - TEA10MAL           195 - TEA10FEM           196 - TEA10FOP           198 - TEA10FNE           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         Select All         Second variable list:	OK       33         Cancel       33         Image: Stress stre
Show appropriate variables only		33

# Solutions: Selected statistical tests MoER

Crosstabulation Tables Results: 5_GEN	/ 2010 APS Croatia	? ×	st
Quick Advanced Options		Summary	2
Summary: Review summary tables	Categorized histograms	Cancel	e st
Detailed two-way tables         Image: Stub-and-banner table         Display long text labels         Include missing data         Display selected %'s in sep. tables	Interaction plots of frequencies	▶ Options ▼ ▶ By Group To compute Max. Likelihood Chi-squares and to analyze multi-way frequency tables use the Log-Linear module.	e e st st

	option nonling, other prochonding pro-	. noning maa	ų
🚟 Crosstabulation Tables Results:	5_GEM 2010 APS Croatia	? ×	st 3 3:
Quick Advanced Options		Summary	e 3:
Compute tables	Statistics for two-way tables	Cancel	3
Highlight counts > 10	Pearson & M-L Chi-square	▶ Options ▼	st 3 3:
Expected frequencies	Fisher exact, Yates, McNemar (2 x 2)	par options .	3.
Residual frequencies	Phi (2x2 tables) & Cramér's V & C	By Group	3
Percentages of total count	Kendall's tau-b & tau-c	To compute Max. Likelihood Chi-	e 3: 3:
Percentages of row counts	Gamma	squares and to analyze multi-way	3
Percentages of column counts	Spearman rank order correlation	frequency tables	t 3
	Somer's d	use the Log-Linear module.	st 3
	Uncertainty coefficients		3
	Odds ratio w/ CI (2 x 2) 95.00		e 3: 3:
			3
1.51 1.61 1.7	9 NOT WORKING, OTHER NOT WORKING NO	t working Lowe	st 3

Basic Statistics/Tables (5\_GEM 2010 APS Croatia)

🗄 🖓 📴 Cro	osstabulation results dialog
	2-Way Summary Table: Observed Frequencies (!
	2-Way Summary Table: Expected Frequencies (5
	Statistics: SUSKIL10(2) x GENDER(2) (5_GEM 2010

	2-Way Summary Table: Expected Frequencies (5_GEM 2010 APS Croatia.sta) Marked cells have counts > 10			
SUSKIL10: SUSKIL adapted to make it fit for national I	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

	2-Way Summary Table: Observed Frequencies (5_GEM 2010 APS Croatia.sta) Marked cells have counts > 10			
SUSKIL10: SUSKIL adapted to make it fit for national I	GENDER Male	GENDER Female	Row Totals	
No	282.98	432.28	715.26	
Column Percent	<mark>37.38%</mark>	<mark>56.08%</mark>		
Row Percent	<mark>39.56%</mark>	<mark>60.44%</mark>		
Total Percent	18.52%	28.29%	46.81%	
Yes	474.16	338.52	812.68	
Column Percent	<mark>62.62%</mark>	<mark>43.92%</mark>		
Row Percent	<mark>58.35%</mark>	<mark>41.65%</mark>		
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 $\chi^2$

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

p<0,05 - reject H<sub>0</sub>

#### 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.

 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\_I 1. Set up hypotheses

- 2. Calculate crosstable
- 3. Calculate  $\chi^2$
- 4. Make a conclusion

# Solution:

#### 1. Set up hypotheses

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

H<sub>0</sub>: there is no dependence between TEA and OPPORT

H<sub>1</sub>: 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

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

	2-Way Summary Table: Observed Frequencies (5_GEM 2010 A Croatia.sta) Marked cells have counts > 10				
TEA10 : Involved in Total early-	OPPORT10	OPPORT10	Row		
stage Entrepreneuria	No	Yes	Totals		
No	1003.37	274.09	1277.46		
Column Percent	<mark>96.90%</mark>	<mark>87.05%</mark>			
Row Percent	<mark>78.54%</mark>	<mark>21.46%</mark>			
Total Percent	74.30%	20.30%	94.60%		
Yes	32.14	40.77	72.91		
Column Percent	<mark>3.10%</mark>	<mark>12.95%</mark>			
Row Percent	<mark>44.08%</mark>	<mark>55.92%</mark>			
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 $\chi^2$

	Statistics: TEA10(2) x OPPORT10(2) (5_GEM 2010 APS Croatia.sta)				
Statistic	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<sub>0</sub>

#### 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.

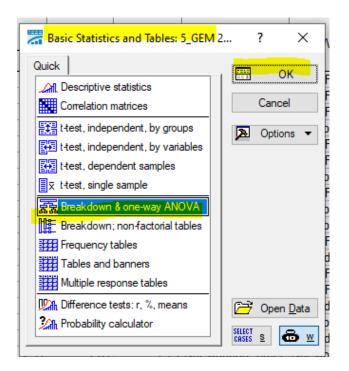
- Examine whether oppinion about equal standard of living depends on gender. •
- GENDER (155) •
- EQUALI10 (189): most people would prefer that everyone had a similar standard of ٠ living

(5\_GEM 2010 APS Croatia.sta)

- 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\_I 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



.78 .51	Statistics by G	roups (Break	down): 5_GEM 2010 APS (	Croatia	?	×s
.82 .82 .61	Individual tables	Lists of tables	]		<b></b> C	K le
.80	<mark></mark> ⊻ariable	s			Cance	
🕅 Se	lect the dependent	variables an	d grouping variables		? ×	s 🔻 ir
	TEA10	^	163 - HRREGIÓN	^	ОК	b w r
195 -	TEA10MAL TEA10FEM TEA10MOP		164 - HRSTRATA 165 - HRCITY 166 - SUB		Cancel	le
198 -	TEA10FOP TEA10MNE TEA10FNE		167 - SUBA 168 - SUBO 169 - SUBOA		[Bundles]	le
200 -	TEA10STA TEA10OPP		170 - SUBOANW 171 - OMBABYX		Use the "Show appropriate	N-1 s
203 -	TEA10NEC TEA10OTH TEA10MOT		172 - OMESTBX 173 - BABYBUSM 174 - BABYBUSO		variables only" option to pre-screen	e le
205 -	TEA10MIX	~	175 - ESTBBUSM	~	variable lists and show categorical and continuous	e le
Sele	ct All Spread	Zoom	Select All Spread	Zoom	variables. Press F1 for more	Upper
Deper	ndent variables:		Grouping variables:		information.	part tim Upper
193			163			
Sh	ow appropriate variab	les only				

### Solutions: Selected statistical tests MoER

	2-Way Tables of Descriptive Statistics (5_GEM 2010 APS Croatia.sta) N=1614 (No missing data in dep. var. list)				
HRREGION	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

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

 $H_1$ : at least one  $\mu$  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

🚟 Statistics by Groups - Result	?	×	
DEPENDENT: 1 variable:	TEA10		
GROUPING: 1-HRREGION(6	): Zagreb region Northern Croati	.a	Es ±
Quick Descriptives ANOVA &	tests Post-hoc	SONN S	u <u>m</u> mary
Analysis of Variance	Categorized normal prob. plots	Ca	ncel
Perform Welch's F-Test	er Categorized half-normal p-plots	🔈 Op	tions 🔻 👔
Tests of homog. of variances	₽₽ <u>C</u> ategorized detrended p-plots	By By	
	Plot of means vs. std. devs		       
Brown- <u>F</u> orsythe tests	Interaction plots		
p-value for	✓ Plot confidence		1
highlighting: .05	intervals for means: 95.00		ir Ie
			ן ני יר ד
			r
			1

🚟 Statistics by Groups - Results: 5_GEM 2010 APS Croatia	?	×r					
DEPENDENT: 1 variable: TEA10							
GROUPING: 1-HRREGION(6): Zagreb region Northern Croatia							
Quick Descriptives ANOVA & tests   Post-hoc		Su <u>m</u> mary le					
Summary: Table of statistics       Statistics         Subgroup       none         Sums       % Sums	N	ancel r Pptions ▼ r					
☐ Minima and maxima ↓ Detailed two-way tables ✓ Standard deviations	E E	By Group					
Display long variable names		st					
Display long text labels Std. err. of mean		st					
Conf. limits for mean:     Marginal means     Median and quartiles	95.00 🚽 %	r le					
Reorder factors in table		le					
First: 10.00 🖆 Second	90.00	st					
Categorized histograms		ne Ir					

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

p=0,5617

p>0,05 - don't reject H<sub>0</sub>

#### 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.

 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

	2-Way Tables of Descriptive Statistics (6_GEM 2010 APS Master.sta) N=60 (No missing data in dep. var. list)						
Kontinent	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.

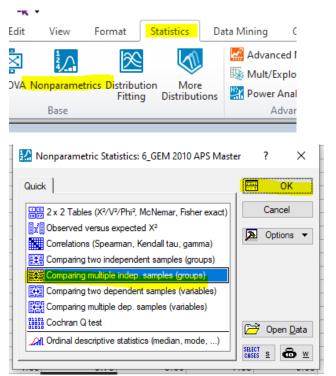
#### 2. Set up hypotheses

H<sub>0</sub>: distributions of TEA are equal in all continents

H1: 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



24 🔼 Select dep. variables a	nd an indep	o. (grouping) variable	3 ? ×	45.96 38.64 9.96 8.06
30         37 - TEA10           38 - TEA10MAL         39 - TEA10FEM           39 - TEA10FEM         40 - EB_10MAL           40 - EB_10FEM         42 - TEA10OPP           43 - TEA10NEC         44 - TEA10MOP           45 - TEA10FOP         45 - TEA10FOP           46 - TEA10MNE         47 - TEA10FNE           47 - TEA10FNE         48 - TEA10IDO           49 - TEA10DO         49 - TEA10MT1	~	I - Kontinent           2 - COUNTRY           3 - CAT_GCR1           4 - CAT_GCR2           5 - BSTÄRT10           6 - BJOBST10           7 - OWNMGE10           8 - BUSANG10           9 - FUTSUP10           10 - DISENT10           11 - EXITCT10           12 - KNOENT10           13 - OPPORT10           14 - CUEVILLE	OK Cancel [Bundles] Use the "Show appropriate variables only" option to pre-screen variable lists and show categorical and continuous	7.82
Select All Spread     Dependent variable list:     37     Show appropriate variable     Solution     Show appropriate variable     32     30.62	Zoom	Select All Sprea Indep. (grouping) varia	variables. Press F1 for more information.	9.99 3.90 4.61 1.71 3.41 9.48 40.90

# Solutions: Selected statistical tests MoER

	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					
Depend.: TEA10	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		

	Median Test, Overall Median = 7.84062; TEA10 (6_GEM 2010 APS Master.sta) Independent (grouping) variable: Kontinent Chi-Square = 23.84571 df = 5 <mark>p = .0002</mark>							
TEA10	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		
obsexp.	-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		
obsexp.	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<sub>0</sub>

### 4. Make a conclusion

Distributions of TEA are not equal in all continents.