

RM MFILIPPAKIS

Ανάλυση με τη χρήση SPSS

Περιεχόμενα Διάλεξης

- ✦ Έλεγχοι Υποθέσεων
- ✦ Έλεγχος Κανονικότητας
- ✦ Έλεγχος Υποθέσεων σε Ποιοτικές Μεταβλητές – χ^2 test
- ✦ T- Test
 - ✦ Ένα Δείγμα (one sample t test)

Έλεγχοι Υποθέσεων

Μια υπόθεση είναι μια αφαιρετική δήλωση της σχέσης μεταξύ δύο ή περισσότερων μεταβλητών που μπορούν να μετρηθούν.

Για να ελεγχθούν οι υποθέσεις πρέπει να διαμορφωθούν

- η Μηδενική Υπόθεση (H_0) και
- η Εναλλακτική Υπόθεση (H_1)

Η Μηδενική Υπόθεση ονομάζεται επίσης και «Υπόθεση της μη Διαφοράς»

Η Εναλλακτική Υπόθεση είναι η «θετική εκδοχή» της μηδενικής υπόθεσης

Παράδειγμα

H_0 : Δεν υπάρχει διαφορά στη βαθμολογία το μάθημα της Στατιστικής μεταξύ φοιτητών και φοιτητριών ή

$$H_0: \mu\alpha = \mu\beta$$

H_1 : Υπάρχει διαφορά στη βαθμολογία το μάθημα της Στατιστικής μεταξύ φοιτητών και φοιτητριών ή

$$H_0: \mu\alpha \neq \mu\beta$$

Η απόρριψη της H_0 και η αποδοχή της H_1 **δε θεωρείται εμπεριστατωμένη απόδειξη** ότι η εναλλακτική υπόθεση είναι πράγματι αληθινή, απλώς αποτελεί στοιχείο για την αύξηση της πεποίθησης του ερευνητή ότι η H_1 είναι σωστή.

Έλεγχοι Υποθέσεων

Είδη Τεστ

-Τεστ ενός δείγματος (one sample tests)

Συγκρίνουν το δείγμα με τον πληθυσμό και εξετάζουν αν το στατιστικό δείγμα δεν διαφέρει στατιστικά σημαντικά σε μια παράμετρο του πληθυσμού.

-Τεστ δύο δειγμάτων (two sample tests)

Ελέγχεται αν δεν υπάρχει στατιστικά σημαντική διαφορά μεταξύ δύο δειγμάτων και κατ' επέκταση αν τα δύο δείγματα αντιπροσωπεύουν διαφορετικούς πληθυσμούς.

Έλεγχοι Υποθέσεων

Βήματα Ελέγχων Υποθέσεων

Βήμα 1^ο Διατύπωση της Υπόθεσης

Βήμα 2^ο Καθορισμός του επιπέδου στατιστικής σημαντικότητας (significance level)

Βήμα 3^ο Υπολογισμός του στατιστικού τεστ (test statistic)

Βήμα 4^ο Υπολογισμός της κρίσιμης στατιστικής αξίας (critical statistical value)

Βήμα 5^ο Σύγκριση του στατιστικού με την κρίσιμη στατιστική αξία

Βήμα 6^ο Εξαγωγή συμπεράσματος

Ο γενικός κανόνας για την απόρριψη ή την αποδοχή της μηδενικής υπόθεσης όταν χρησιμοποιείται η p-value (κρίσιμη στατιστική αξία – sig.) είναι:

Εάν $p > 0.05$ η H_0 γίνεται αποδεκτή σε επίπεδο σημαντικότητας $\alpha = 0,05$

Εάν $p < 0.05$ η H_0 απορρίπτεται σε επίπεδο σημαντικότητας $\alpha = 0,05$

Το επίπεδο σημαντικότητας α μπορεί να είναι και διαφορετικό, δηλ. $\alpha = 0,01$

Έλεγχος Κανονικότητας

Για να ελέγξουμε αν η κατανομή μιας μεταβλητής ακολουθεί την **Κανονική Κατανομή**, εφαρμόζουμε το τεστ Kolmogorov Smirnov ή, για μικρά δείγματα, το Shapiro Wilk.

Η υπόθεση που ελέγχουμε είναι:

H_0 : Η μεταβλητή ακολουθεί την Κανονική Κατανομή

H_1 : Η μεταβλητή δεν ακολουθεί την Κανονική Κατανομή

Κριτήριο Απόρριψης: Απορρίπτουμε την H_0 αν $\text{sig} < 0.05$

Έλεγχος Κανονικότητας

Dataset advertisements

Τριάντα περιοδικά κατατάχθηκαν με βάση το εκπαιδευτικό επίπεδο των αναγνωστών τους. Στη συνέχεια επιλέχθηκαν τρία περιοδικά από κάθε ένα από τα τρία εκπαιδευτικά επίπεδα, και από αυτά, επιλέχθηκαν τυχαία 6 διαφημίσεις. Για τις διαφημίσεις αυτές καταγράφηκαν ο αριθμός των λέξεων στο κείμενο, ο αριθμός των προτάσεων στο κείμενο και ο αριθμός των λέξεων στο κείμενο που είχαν περισσότερες από 3 συλλαβές. Οι μεταβλητές που χρησιμοποιούνται είναι οι εξής:

- words = Αριθμός λέξεων στο κείμενο της διαφήμισης
- sentences = Αριθμός προτάσεων στο κείμενο της διαφήμισης
- syllable = Αριθμός λέξεων με περισσότερες από 3 συλλαβές στο κείμενο της διαφήμισης
- magazine = 1. Scientific American, 2. Fortune, 3. The New Yorker, 4. Sports Illustrated, 5. Newsweek, 6. People, 7. National Enquirer, 8. Grit, 9. True Confessions
- ed_level = Επίπεδο Εκπαίδευση (1: Υψηλό Μορφωτικό Επίπεδο, 2: Μέτριο Μορφωτικό Επίπεδο, 3: Χαμηλό Μορφωτικό Επίπεδο)
- age = Ηλικιακή Κατηγορία (1: Άτομα ηλικίας έως 25 ετών, 2: Άτομα ηλικίας μεταξύ 25 και 40 ετών, 3: Άτομα ηλικίας μεγαλύτερης των 40 ετών)

Έλεγχος Κανονικότητας

advertisements.sav [DataSet1] - SPSS Data Editor

	words	senten
1	205	
2	203	
3	229	
4	208	
5	146	
6	230	
7	215	
8	153	
9	205	
10	80	
11	208	
12	89	
13	49	
14	93	
15	46	
16	34	
17	39	

25 :

File Edit View Data Transform Analyze Graphs Utilities Add-ons Window Help

Reports

Descriptive Statistics

123 Frequencies...

Tables

Descriptives...

Explore...

Crosstabs...

Ratio...

P-P Plots...

Q-Q Plots...

Compare Means

General Linear Model

Generalized Linear Models

Mixed Models

Correlate

Regression

Loglinear

Classify

Data Reduction

Scale

Nonparametric Tests

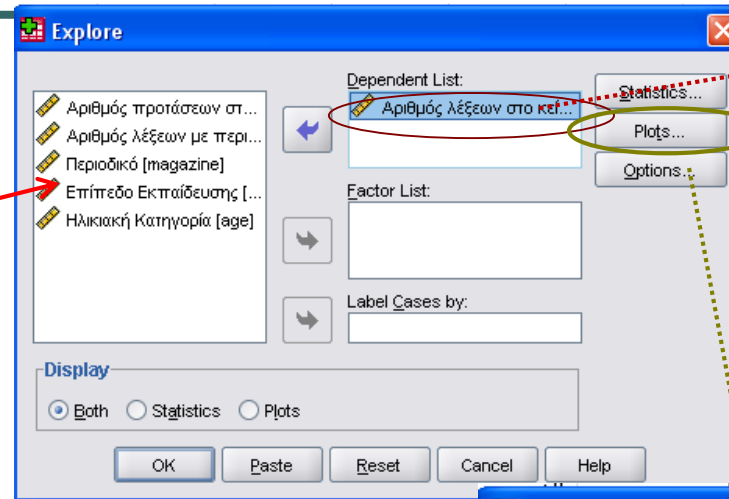
Time Series

Survival

Multiple Response

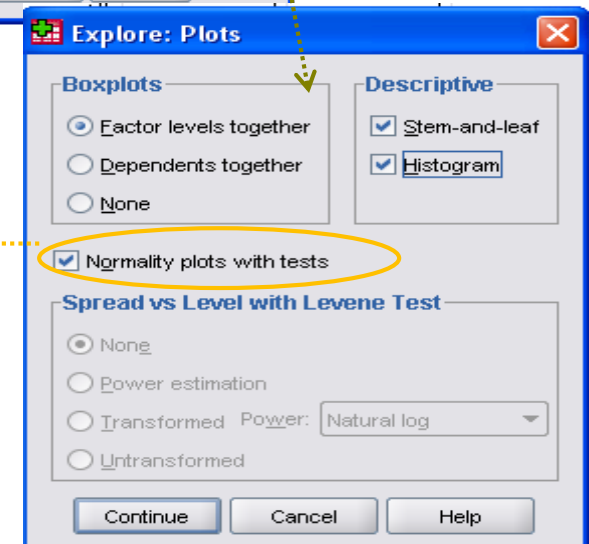
Quality Control

ROC Curve...



Η ποσοτική μεταβλητή για την οποία θα κάνουμε έλεγχο κανονικότητας

Η επιλογή για το τεστ Κανονικότητας



Έλεγχος Κανονικότητας

Tests of Normality

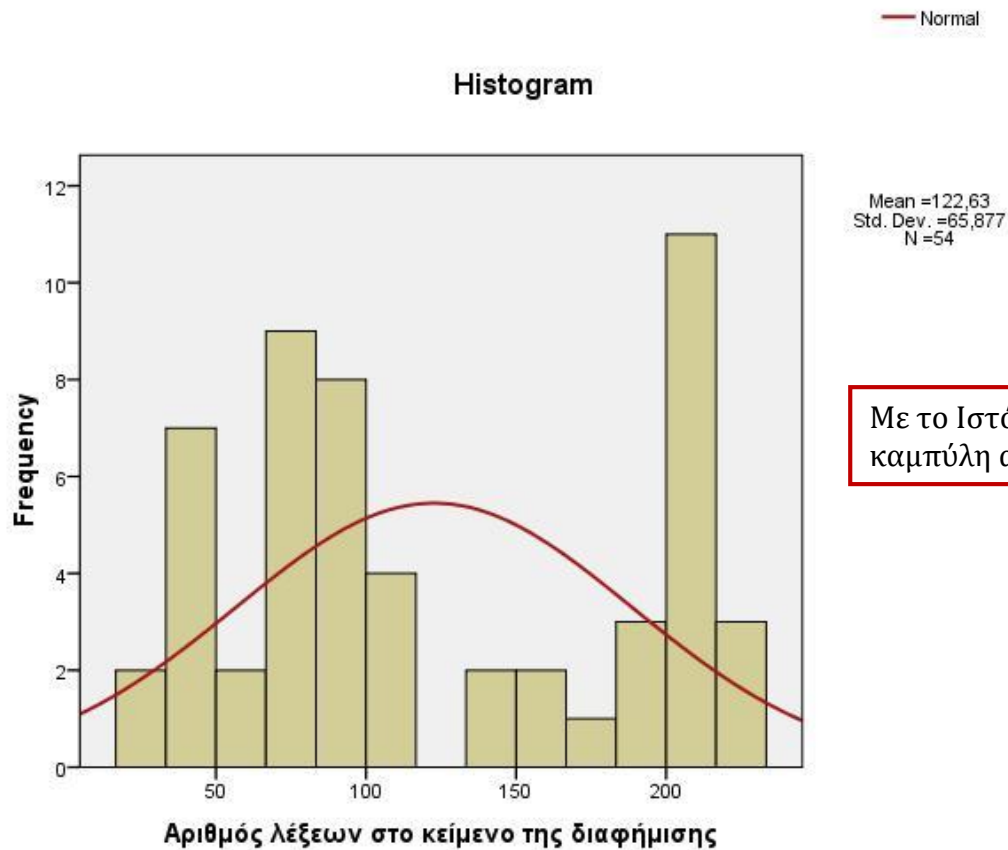
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Αριθμός λέξεων στο κείμενο της διαφήμισης	,170	54	,000	,883	54	,000

a. Lilliefors Significance Correction

Με το τεστ Kolmogorov Smirnov, το $\text{sig} = 0.000 < 0.05$. Επομένως απορρίπτουμε την H_0 . Δηλαδή η μεταβλητή δεν ακολουθεί την Κανονική Κατανομή

Με το τεστ Shapiro-Wilk, το $\text{sig} = 0.000 < 0.05$. Επομένως απορρίπτουμε την H_0 . Δηλαδή η μεταβλητή δεν ακολουθεί την Κανονική Κατανομή

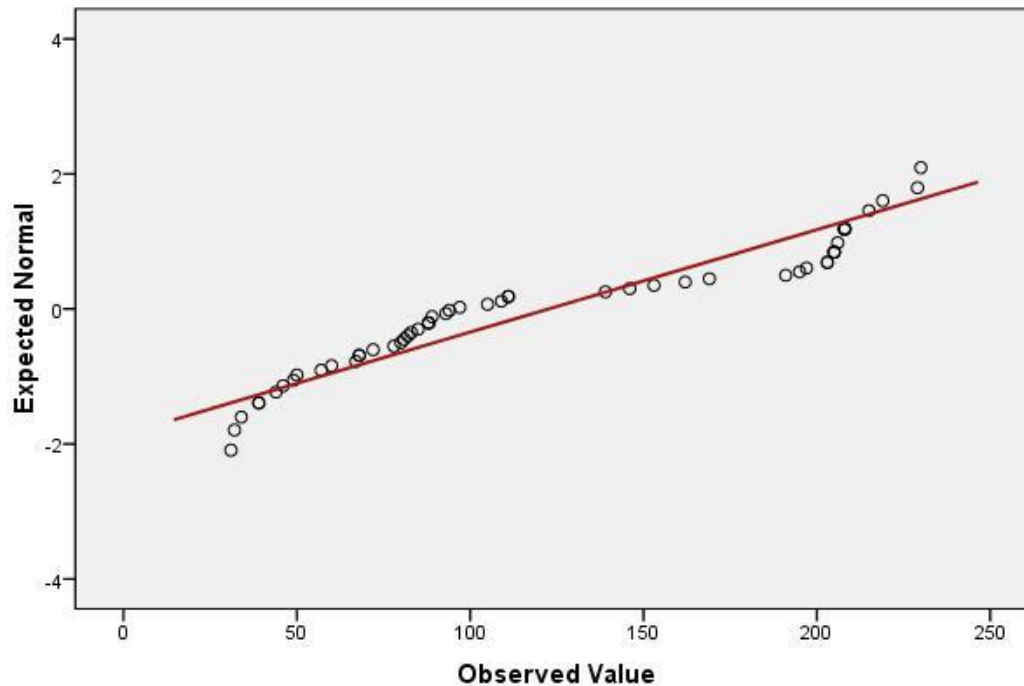
Έλεγχος Κανονικότητας



Με το Ιστόγραμμα ελέγχουμε την κανονικότητα. Η καμπύλη αφορά κανονική κατανομή.

Έλεγχος Κανονικότητας

Normal Q-Q Plot of Αριθμός λέξεων στο κείμενο της διαφήμισης



Η κανονικότητα ελέγχεται και με το Q-Q Plot. Όταν όλα τα σημεία βρίσκονται κοντά στην γραμμή η κατανομή θεωρείται κανονική.

Έλεγχος Υποθέσεων σε Ποιοτικές Μεταβλητές – χ^2

Για να προσδιορίσουμε τη σχέση μεταξύ δύο ποιοτικών μεταβλητών χρησιμοποιούμε το χ^2 test.

Οι υποθέσεις που ελέγχουμε είναι οι εξής:

- H_0 :** Δεν υπάρχει σχέση ανάμεσα στις δύο μεταβλητές
(οι παρατηρηθείσες συχνότητες είναι ίσες με τις αναμενόμενες συχνότητες)
- H_1 :** Υπάρχει σχέση ανάμεσα στις δύο μεταβλητές
(οι παρατηρηθείσες συχνότητες και οι αναμενόμενες συχνότητες διαφέρουν)

Προϋπόθεση εφαρμογής του χ^2 test

- όλες οι αναμενόμενες συχνότητες πρέπει να είναι > 5
- εναλλακτικά το 85% των αναμενόμενων συχνοτήτων πρέπει να είναι μεταξύ 2 και 5

Εάν δεν ισχύουν οι προϋποθέσεις αυτές εφαρμόζουμε:

Exact Test ή Monte Carlo Test

Έλεγχος Υποθέσεων σε Ποιοτικές Μεταβλητές – χ^2

Υπολογίζουμε τις παρατηρηθείσες και τις αναμενόμενες τιμές καθώς και τα αντίστοιχα ποσοστά.

Υπολογίζουμε το χ^2

advertisements.sav [DataSet1] - SPSS Data Editor

	words	senten		
1	205			
2	203			
3	229			
4	208			
5	146			
6	230			
7	215			
8	153			
9	205			
10	80			
11	208			
12	89			
13	49			
14	93			
15	46			
16	34			
17	39	5	6	3

Crosstabs

Row(s): Επίπεδο Εκπαίδευσης [ed...]

Column(s): Ηλικιακή Κατηγορία [age]

Layer 1 of 1

☐ Display clustered bar charts

☐ Suppress tables

OK Paste Reset

Crosstabs: Statistics

☒ Chi-square

☐ Correlations

Nominal

☐ Contingency coefficient

☐ Phi and Cramer's V

☐ Lambda

☐ Uncertainty coefficient

Ordinal

☐ Gamma

☐ Somers' d

☐ Kendall's tau-b

☐ Kendall's tau-c

Nominal by Interval

☐ Eta

☐ Kappa

☐ Risk

☐ McNemar

☐ Cochran's and Mantel-Haenszel statistics

Test common odds ratio equals: 1

Continue Cancel Help

Crosstabs: Cell Display

Counts

☒ Observed

☒ Expected

Percentages

☒ Row

☒ Column

☐ Total

Residuals

☐ Unstandardized

☐ Standardized

☒ Adjusted standardized

Noninteger Weights

☒ Round cell counts

☐ Round case weights

☐ Truncate cell counts

☐ Truncate case weights

☐ No adjustments

Continue Cancel Help

Έλεγχος Υποθέσεων σε Ποιοτικές Μεταβλητές – χ^2

Επίπεδο Εκπαίδευσης * Ηλικιακή Κατηγορία Crosstabulation

			Ηλικιακή Κατηγορία			
			Άτομα ηλικίας έως 25 ετών	Άτομα ηλικίας μεταξύ 25 και 40 ετών	Άτομα ηλικίας μεγαλύτερης των 40 ετών	Total
Επίπεδο Εκπαίδευσης	Υψηλό Μορφωτικό Επίπεδο	Count	4	7	7	18
		Expected Count	5,0	7,3	5,7	18,0
		% within Επίπεδο Εκπαίδευσης	22,2%	38,9%	38,9%	100,0%
		% within Ηλικιακή Κατηγορία	26,7%	31,8%	41,2%	33,3%
	Μέτριο Μορφωτικό Επίπεδο	Count	5	10	3	18
		Expected Count	5,0	7,3	5,7	18,0
		% within Επίπεδο Εκπαίδευσης	27,8%	55,6%	16,7%	100,0%
		% within Ηλικιακή Κατηγορία	33,3%	45,5%	17,6%	33,3%
	Χαμηλό Μορφωτικό Επίπεδο	Count	6	5	7	18
		Expected Count	5,0	7,3	5,7	18,0
		% within Επίπεδο Εκπαίδευσης	33,3%	27,8%	38,9%	100,0%
		% within Ηλικιακή Κατηγορία	40,0%	22,7%	41,2%	33,3%
	Total	Count	15	22	17	54
		Expected Count	15,0	22,0	17,0	54,0
		% within Επίπεδο Εκπαίδευσης	27,8%	40,7%	31,5%	100,0%
		% within Ηλικιακή Κατηγορία	100,0%	100,0%	100,0%	100,0%

Count: Οι πραγματικές τιμές, **Expected count:** Οι εκτιμώμενες τιμές, **Residual:** Η διαφορά ανάμεσα στις πραγματικές & τις εκτιμώμενες τιμές,

Έλεγχος Υποθέσεων σε Ποιοτικές Μεταβλητές – χ^2

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)
Pearson Chi-Square	4,010 ^a	4	,405
Likelihood Ratio	4,225	4	,376
Linear-by-Linear Association	,184	1	,668
N of Valid Cases	54		

a. 0 cells (.0%) have expected count less than 5. The minimum expected count is 5,00.

Η προϋπόθεση ισχύει, επομένως βλέπουμε το Pearson chi – square Test

Chi-Square=4.010
Significant=0.405 >0.05 άρα
δεχόμαστε την H_0 , δηλαδή δεν
υπάρχει σχέση μεταξύ των δύο
μεταβλητών.

Σημείωση:

Στους πίνακες 2x2 το Fisher's Exact Test εμφανίζεται με την αρχική εντολή

Έλεγχος Υποθέσεων σε Ποιοτικές Μεταβλητές – χ^2

advertisments.sav [DataSet1] - SPSS Data Editor

FileEditViewDataTransform

10:

	words	senten
1	205	
2	203	
3	229	
4	208	
5	146	
6	230	
7	215	
8	153	
9	205	
10	80	
11	208	
12	89	
13	49	
14	93	
15	46	
16	34	
17	39	

Analyze

GraphsUtilitiesAdd-onsWindowHelp

Reports

Descriptive Statistics

Tables

Compare Means

General Linear Model

Generalized Linear Models

Mixed Models

Correlate

Regression

Loglinear

Classify

Data Reduction

Scale

Nonparametric Tests

Time Series

Survival

Multiple Response

Quality Control

ROC Curve...

123

Frequencies...

Descriptives...

Explore...

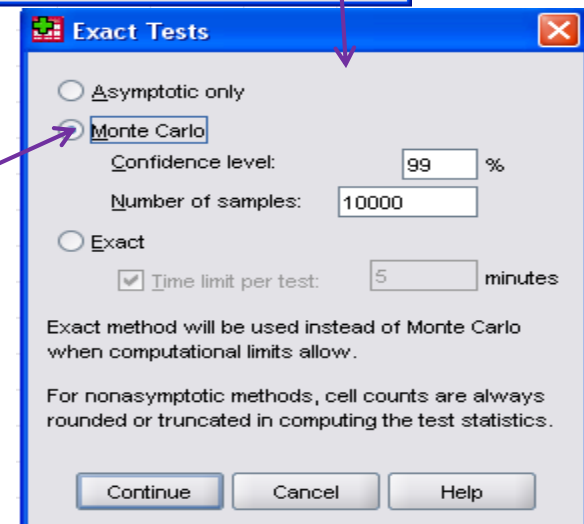
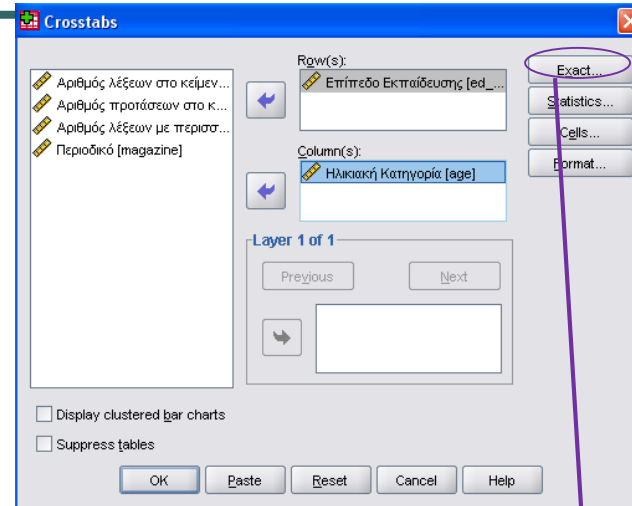
Crosstabs...

Ratio...

P-P Plots...

Q-Q Plots...

	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1
	1



Το Monte Carlo test, εφόσον δεν ισχύουν οι προϋποθέσεις εφαρμογής του χ^2 test

Έλεγχος Υποθέσεων σε Ποιοτικές Μεταβλητές – χ^2

Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Monte Carlo Sig. (2-sided)			Monte Carlo Sig. (1-sided)		
				Sig.	99% Confidence Interval		Sig.	99% Confidence Interval	
					Lower Bound	Upper Bound		Lower Bound	Upper Bound
Pearson Chi-Square	4,010 ^a	4	,405	,412 ^b	,399	,424			
Likelihood Ratio	4,225	4	,376	,397 ^b	,384	,410			
Fisher's Exact Test	4,078			,397 ^b	,384	,410			
Linear-by-Linear Association	,184 ^c	1	,668	,742 ^b	,731	,754	,377 ^b	,365	,390
N of Valid Cases	54								

a. 0 cells (,0%) have expected count less than 5. The minimum expected count is 5,00.

b. Based on 10000 sampled tables with starting seed 2000000.

c. The standardized statistic is -,429.

Sig = 0.397 > 0.05 άρα και δεχόμαστε την H_0 , δηλαδή δεν υπάρχει σχέση μεταξύ των δύο μεταβλητών.

T-test

Το T-test χρησιμοποιείται

- Για να ελέγχουμε τη διαφορά μεταξύ του μέσου μιας μεταβλητής και μιας τιμής που έχουμε ορίσει.
- Όταν υπάρχει μία ποσοτική μεταβλητή και μία κατηγορική μεταβλητή με δύο κατηγορίες (π.χ. ναι-όχι, Άνδρας-Γυναίκα, Πρωτεύουσα-Επαρχία) και θέλουμε να ελέγχουμε τη διαφορά των μέσων μεταξύ των κατηγοριών της κατηγορικής μεταβλητής.
- Για να συγκρίνουμε τους μέσους μια μεταβλητής για δύο εξαρτημένα δείγματα.

Για την εφαρμογή του t-test πρέπει να ισχύουν ορισμένες προϋποθέσεις.
Εάν δεν ισχύουν προχωρούμε στα αντίστοιχα μη παραμετρικά τεστ.

One Sample T-test

Το t-test για ένα δείγμα χρησιμοποιείται για να ελέγξουμε τη διαφορά μεταξύ του μέσου μιας μεταβλητής και μιας τιμής που έχουμε ορίσει.

H_0 : Η μέση τιμή της μεταβλητής είναι ίση με την τιμή μ_0 ή

$$H_0: \mu = \mu_0$$

H_1 : Η μέση τιμή της μεταβλητής δεν είναι ίση με την τιμή μ_0 ή

$$H_1: \mu \neq \mu_0$$

Προϋπόθεση Εφαρμογής του τεστ

Η μεταβλητή πρέπει να ακολουθεί Κανονική Κατανομή

Εάν δεν ισχύει η προϋπόθεση προχωρούμε στο αντίστοιχο μη παραμετρικό Wilcoxon Test

One Sample T-test

Επιστρέφοντας στο αρχείο ICT_exp_sav έστω ότι θέλουμε να ελέγξουμε την υπόθεση ότι τα μέσα έξοδα για Τεχνολογίες Πληροφορικής είναι ίσα με 3.

$$H_0: \mu=3$$

$$H_1: \mu \neq 3$$

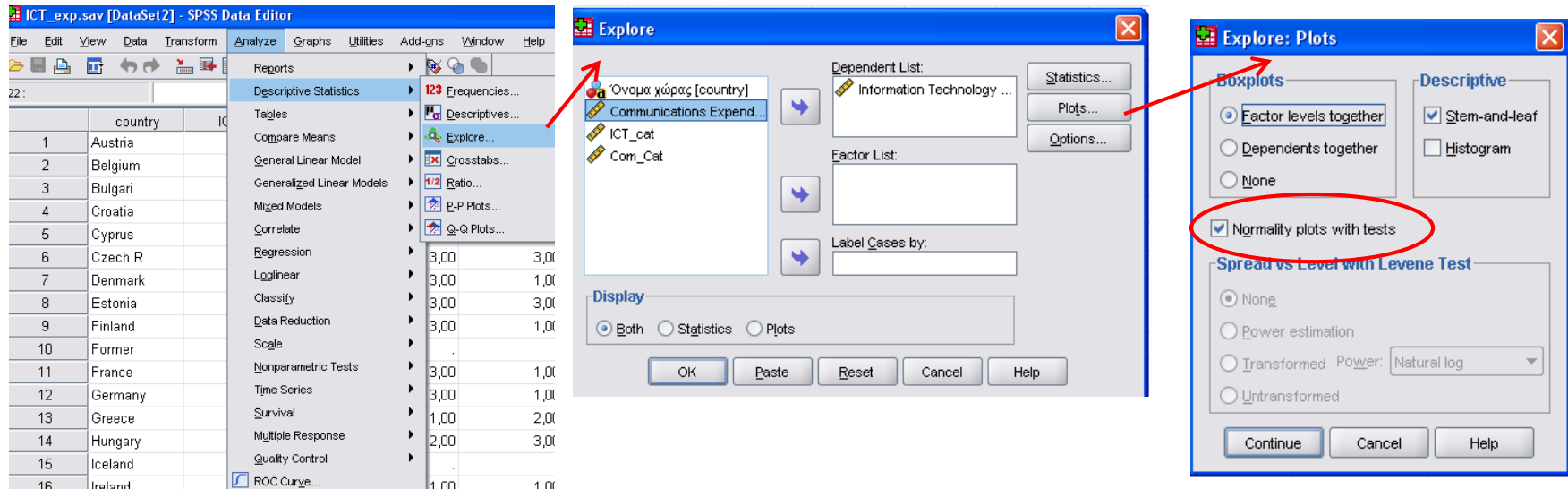
Αρχικά θα ελέγξουμε εάν ισχύει η **προϋπόθεση**, δηλαδή εάν η μεταβλητή Έξοδα για Τεχνολογίες Πληροφορικής ακολουθεί Κανονική Κατανομή.

Η υπόθεση που ελέγχουμε είναι:

H_0 : Η μεταβλητή ακολουθεί την Κανονική Κατανομή

H_1 : Η μεταβλητή δεν ακολουθεί την Κανονική Κατανομή

One Sample T-test



Tests of Normality

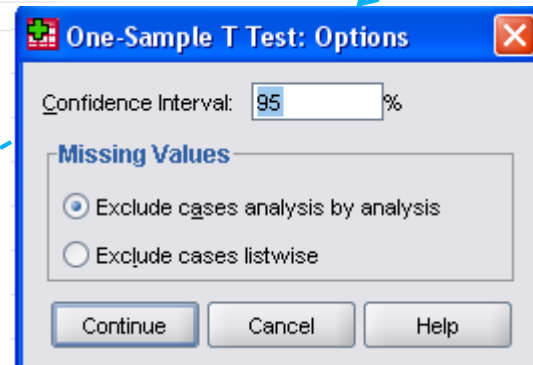
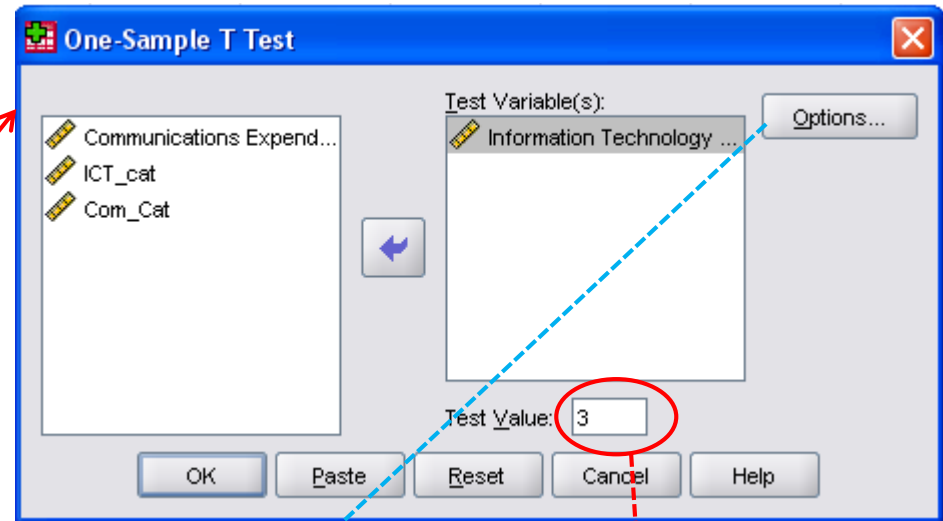
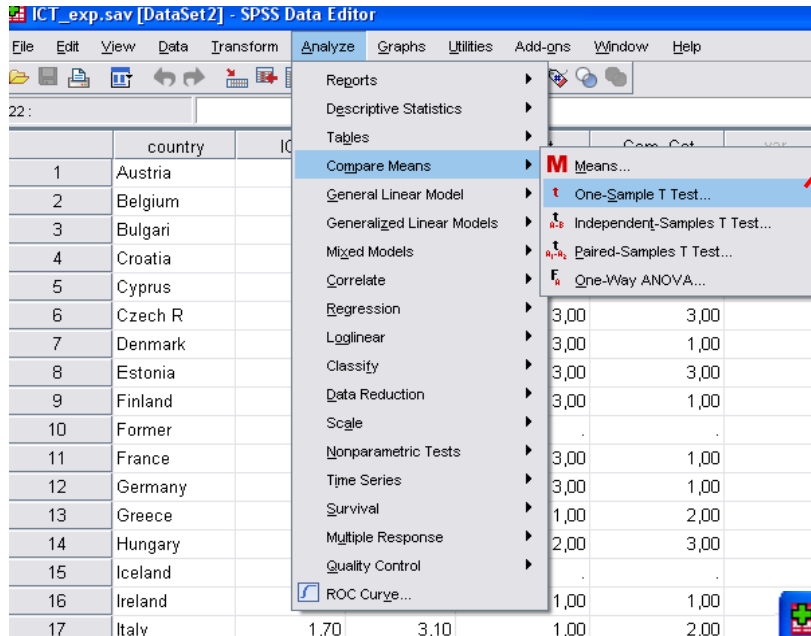
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Information Technology Expenditure	,112	28	,200 [*]	,965	28	,452

a. Lilliefors Significance Correction

*. This is a lower bound of the true significance.

Το δείγμα μας είναι >30, επομένως από το test Kolmogorov-Smirnov: Sig=0,200>0,05
Άρα **ισχύει** η προϋπόθεση της κανονικότητας και έτσι προχωράμε στον έλεγχο t-test.

One Sample T-test



Δηλώνουμε το αντίστοιχο
διάστημα εμπιστοσύνης

Εισάγουμε την τιμή με
την οποία θέλουμε να
συγκριθεί ο μέσος της
εξαρτημένης μεταβλητής

One Sample T-test

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Information Technology Expenditure	28	2,6107	,72742	,13747

Περιγραφικά στατιστικά για τη μεταβλητή που εξετάζουμε

One-Sample Test

	Test Value = 3					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Information Technology Expenditure	-2,832	27	,009	-,38929	-,6713	-,1072

Η τιμή που έχουμε εισάγει

Η διαφορά ανάμεσα στη μέση τιμή του υπό εξέταση δείγματος (2,6107) και της μέσης τιμής (3) που θέλουμε να ελέγξουμε

Sig = 0.009 < 0.05 άρα απορρίπτουμε την H_0 , δηλαδή οι μέσες δαπάνες για Τεχνολογίες Πληροφορικής διαφέρουν στατιστικά σημαντικά από την τιμή του 3%.

Το 95% διάστημα εμπιστοσύνης για τη διαφορά ανάμεσα στη μέση τιμή του υπό εξέταση δείγματος (2,6107) και της μέσης τιμής (3) που θέλουμε να ελέγξουμε κυμαίνεται μεταξύ -0,6716 και -0,1072.

One Sample T-test

Στην περίπτωση που η εναλλακτική υπόθεση H_1 δεν ήταν η $H_1: \mu \neq 3$ αλλά μία εκ των $H_1: \mu > 3$ ή $H_1: \mu < 3$ τότε θα ίσχυαν τα εξής:

$H_0: \mu = \mu_0$ vs $H_1: \mu > \mu_0$

$p\text{-value} = 0.009/2 = 0.0045 < 0.05$

Άρα απορρίπτουμε τη μηδενική υπόθεση

$H_0: \mu = \mu_0$ vs $H_1: \mu < \mu_0$

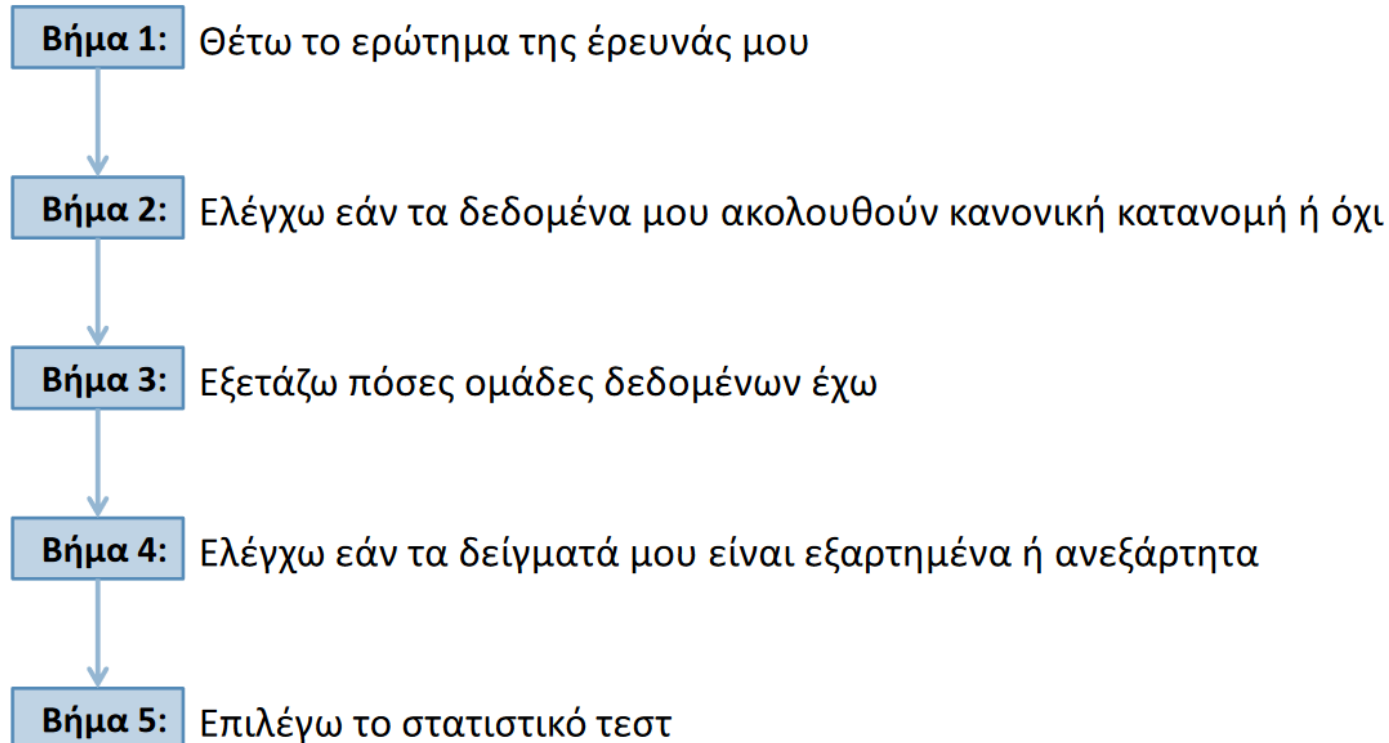
$1 - (p\text{-value}) = 1 - (0.009/2) = 0.9955 > 0.05$

Άρα δεχόμαστε τη μηδενική υπόθεση

— Πώς να επιλέξω το κατάλληλο στατιστικό τεστ; —



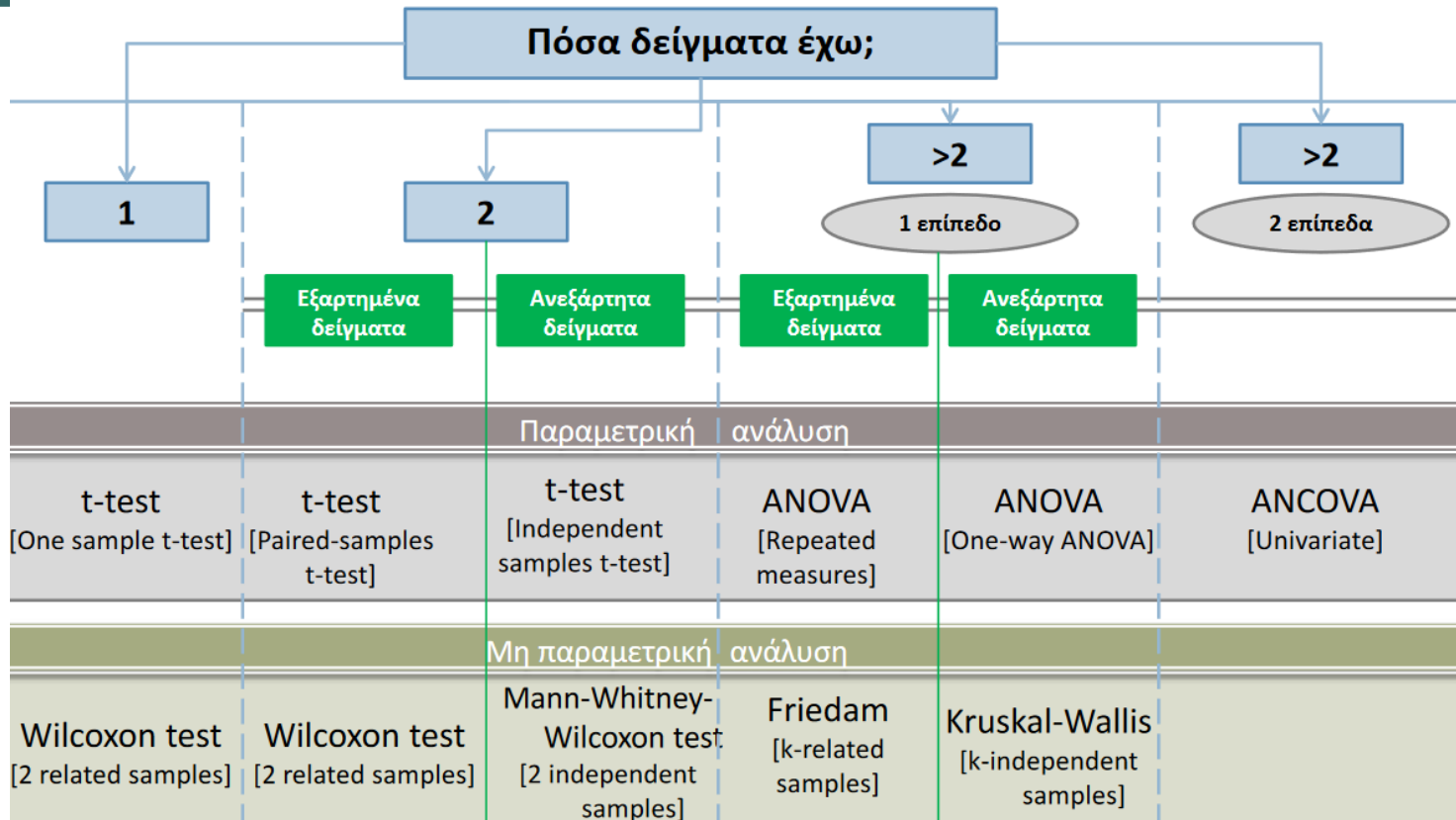
/



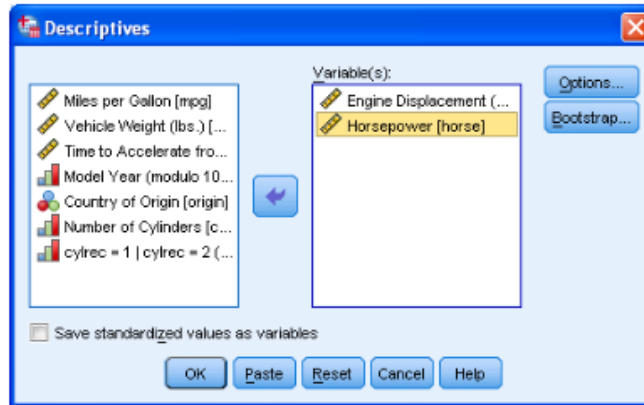
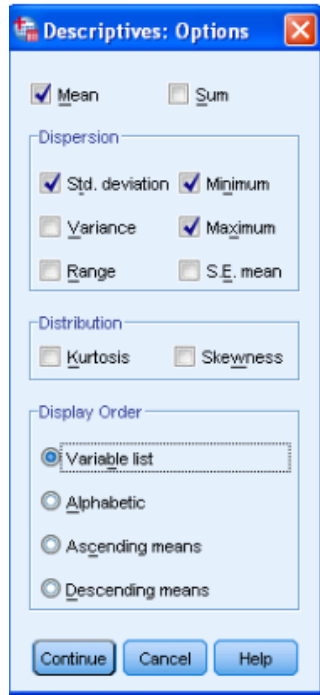
Έλεγχος κανονικότητας και επιλογή ανάλυσης



Επιλογή στατιστικού τεστ



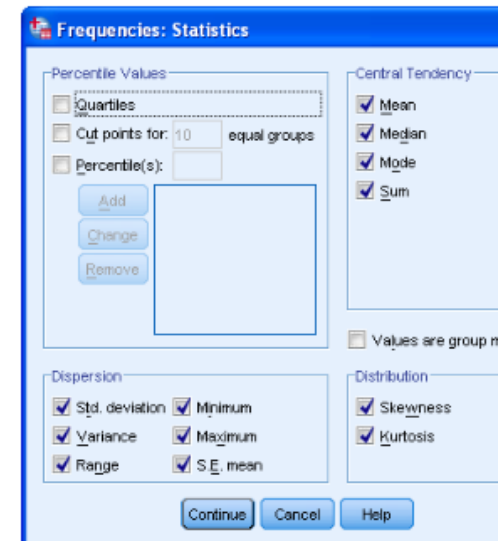
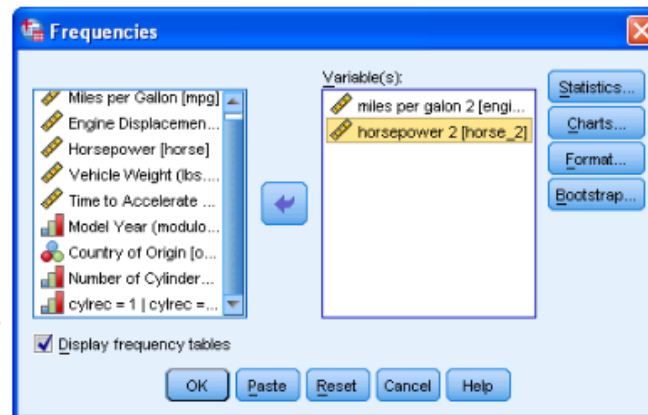
▪ **Analyze → Descriptive statistics → Descriptives**



και / ή

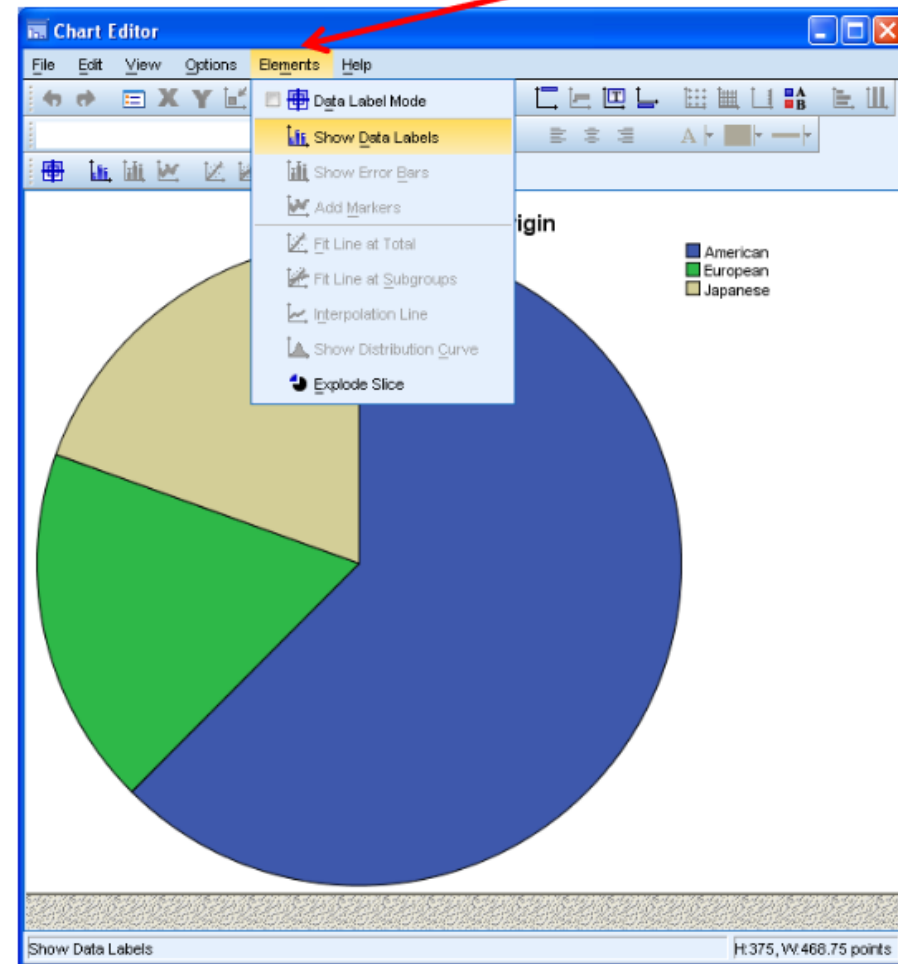
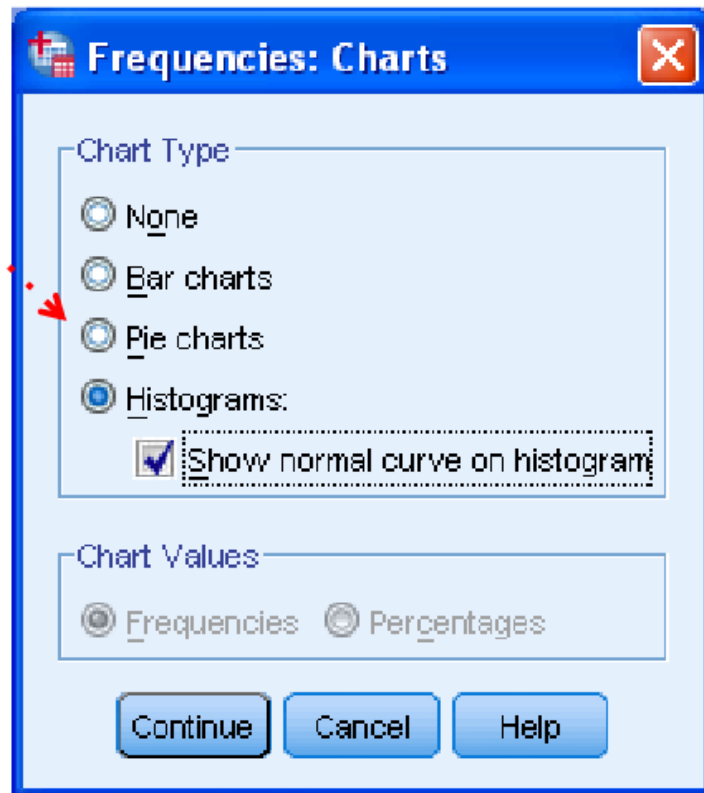
▪ **Analyze → Descriptive statistics → Frequencies**

Δεν το
τσεκάρω!



- Analyze → Descriptive statistics → Frequencies

Labels



Όταν το ιστόγραμμα συχνοτήτων των ποσοτικών μεταβλητών έχει το σχήμα «καμπάνας», τότε λέμε ότι τα δεδομένα ακολουθούν την κανονική κατανομή ή κατανέμονται κανονικά.

Ωστόσο, το ιστόγραμμα δεν είναι «ικανό» να μας δώσει την απάντηση στο αν είναι κανονικά τα δεδομένα ή εάν προέρχονται από μια κανονική κατανομή με ένα μέσο όρο ή μια διακύμανση.


Με τη βοήθεια του SPSS μπορούμε να κατασκευάσουμε 2 διαγράμματα:

Analyze → Descriptive statistics → P-P plots ή Q-Q plots

Με τα παραπάνω διαγράμματα μπορούμε να **ελέγξουμε οπτικά** την ύπαρξη κανονικότητας στα δεδομένα μας. Όσο πιο κοντά στην ευθεία είναι τα σημεία του σχήματος, τόσο πιο πολλές είναι οι ενδείξεις ότι τα δεδομένα ακολουθούν την κανονική κατανομή.

Ωστόσο, και πάλι μπορεί να κάνουμε λάθος (οπτικός έλεγχος).

Γι' αυτόν τον λόγο, κάνουμε **ΤΕΣΤ ΚΑΝΟΝΙΚΟΤΗΤΑΣ**.

- Έλεγχος κανονικότητας  Έλεγχος υποθέσεων
- **Μηδενική υπόθεση (null hypothesis) H_0 :** Η κατανομή των δεδομένων δεν διαφέρει από την κανονική κατανομή.
- **Εναλλακτική υπόθεση (alternative hypothesis) H_1 :** Η κατανομή των δεδομένων διαφέρει από την κανονική κατανομή.
- Επίπεδο στατιστικής σημαντικότητας: Ορίζεται συνήθως στο 5%.
- Το **παρατηρηθέν επίπεδο στατιστικής σημαντικότητας (p-value)** είναι η πιθανότητα η τιμή του ελέγχου να πάρει μια τιμή τόσο ακραία ή περισσότερο ακραία από αυτή που πήρε στο συγκεκριμένο δείγμα κάτω από τη μηδενική υπόθεση.
- **Αν το p-value είναι μικρότερο του 5%, τότε η μηδενική υπόθεση απορρίπτεται.**
- **Αν το p-value είναι μεγαλύτερο του 5%, τότε η μηδενική υπόθεση δεν απορρίπτεται.**

Συντελεστές γραμμικής συσχέτισης

Analyze → Correlate → Bivariate

Μηδενική υπόθεση (null hypothesis) $H_0: \rho=0$: δεν υπάρχει γραμμική συσχέτιση μεταξύ των 2 μεταβλητών

Εναλλακτική υπόθεση (alternative hypothesis) $H_1: \rho \neq 0$: υπάρχει γραμμική συσχέτιση μεταξύ των 2 μεταβλητών

Όταν ο συντελεστής συσχέτισης είναι μεγάλος, σημαίνει ότι η συσχέτιση των μεταξύ των τιμών των μεταβλητών είναι ισχυρή.

