



Εργαστήριο Κυκλοφορίας,
Μεταφορών
και Διαχείρισης Εφοδιαστικής
Αλυσίδας

ΣΤΑΤΙΣΤΙΚΗ ΑΝΑΛΥΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ – ΧΡΗΣΗ SPSS

Αντικείμενα διάλεξης

- ❑ Σύντομη εισαγωγή στη στατιστική
- ❑ Λίγα λόγια για το στατιστικό πακέτο SPSS
- ❑ Επεξεργασία αποτελεσμάτων με τη χρήση του SPSS

Η επιστήμη της στατιστικής

- **Στατιστική είναι ένα σύνολο αρχών και μεθοδολογιών για** (*Ronald Fisher, 1890-1962*):
 - Το σχεδιασμό της διαδικασίας συλλογής δεδομένων
 - Τη συνοπτική και αποτελεσματική παρουσίαση των συλλεχθέντων δεδομένων
 - Την ανάλυση και εξαγωγή αντίστοιχων συμπερασμάτων.
- **2 βασικές μορφές:**
 - **Περιγραφική στατιστική**, η οποία ασχολείται με την περιγραφή των δεδομένων του δείγματος
 - **Επαγωγική στατιστική**, η οποία ασχολείται με την εξαγωγή χρήσιμων συμπερασμάτων για τον πληθυσμό.

Μεταβλητές

- Τα χαρακτηριστικά ως προς τα οποία εξετάζουμε έναν πληθυσμό καλούνται **μεταβλητές**. Οι δυνατές τιμές που μπορεί να πάρει μια μεταβλητή ονομάζονται τιμές της μεταβλητής.
- **2 κατηγορίες μεταβλητών:**
 - Ποιοτικές ή κατηγορικές
 - Ποσοτικές

Οι **ποιοτικές** μεταβλητές μπορεί να είναι είτε **ονομαστικού τύπου** στις οποίες οι τιμές αναφέρονται μόνο σε κατηγορίες, π.χ. ηλικιακή ομάδα, είτε **διατακτικού τύπου** στις οποίες οι συγκρίσεις της μορφής «μεγαλύτερη», «μικρότερη», «ίση» έχουν νόημα, π.χ. σε ερωτηματολόγιο ικανοποίησης.

Οι **ποσοτικές** μεταβλητές μπορεί να είναι είτε συνεχείς είτε διακριτές. **Συνεχείς** είναι οι μεταβλητές οι οποίες μπορούν να λάβουν οποιαδήποτε τιμή σε ένα διάστημα τιμών. **Διακριτές** είναι οι μεταβλητές που μπορούν να πάρουν διακριτές (μεμονωμένες) τιμές.

Οι **ποσοτικές** μεταβλητές μπορούν επίσης να διαχωριστούν σε άλλες 2 κατηγορίες με βάση τον τρόπο μέτρησής τους. Όταν εκτός από τη διάταξη των τιμών τους μας ενδιαφέρει και η μεταξύ τους απόσταση, π.χ. έτη ζωής, τότε μιλάμε για **ποσοτικές μεταβλητές οι οποίες μετρώνται σε κλίμα διαστήματος**.

Όταν εκτός από τη διάταξη και το μέγεθος του διαστήματος μεταξύ των τιμών έχει έννοια και ο λόγος των τιμών, τότε μιλάμε για **ποσοτικές μεταβλητές που μετρώνται σε κλίμακα λόγου**, π.χ. προϊόντα .

Δειγματοληψία

- 4 Βασικές τεχνικές δειγματοληψίας:
 - **Απλή τυχαία δειγματοληψία:** Επιλέγουμε τυχαία στοιχεία ή μονάδες (μετακινούμενους) από το σύνολο του πληθυσμού*
 - **Στρωματοποιημένη δειγματοληψία:** Χωρίζουμε τον πληθυσμό σε στρώματα και στη συνέχεια επιλέγουμε τυχαία τα στοιχεία (μετακινούμενους) από κάθε στρώμα
 - **Δειγματοληψία κατά ομάδες:** Χωρίζουμε τον πληθυσμό σε πολλές ομάδες (όχι στρώματα) με την κάθε ομάδα να περιέχει ένα πλήθος στοιχείων, και στη συνέχεια επιλέγουμε τυχαία ομάδες από το σύνολο των ομάδων και συμπεριλαμβάνουμε στο δείγμα όλες τις μονάδες των επιλεγμένων ομάδων
 - **Συστηματική δειγματοληψία:** Επιλέγουμε τυχαία ένα στοιχείο και στη συνέχεια ακολουθώντας ένα (σταθερό) «βήμα» επιλέγουμε τα υπόλοιπα στοιχεία.

***Πληθυσμός:**

- ✓ **Αντικειμενικός πληθυσμός:** το σύνολο των ατόμων ή στοιχείων των οποίων ένα ή περισσότερα χαρακτηριστικά θέλουμε να εξετάσουμε.
- ✓ **Υπό μελέτη πληθυσμός:** είναι υποσύνολο συνήθως του αντικειμενικού πληθυσμού, μπορεί και να ταυτίζεται.
- ✓ **Δειγματοληπτικό πλαίσιο:** είναι το σύνολο των ατόμων ή στοιχείων που έχουν την πραγματική δυνατότητα επιλογής στο δείγμα (η πηγή του δείγματος).

Μέση τιμή, διάμεσος, επικρατούσα τιμή

- **Μέση τιμή ή αριθμητικός μέσος (mean, average):**

$$\bar{x} = \frac{t_1 + t_2 + \dots + t_v}{v} = \frac{\sum_{i=1}^v t_i}{v} = \frac{1}{v} \sum_{i=1}^v t_i$$

$$\bar{x} = \frac{x_1V_1 + x_2V_2 + \dots + x_\kappa V_\kappa}{V_1 + V_2 + \dots + V_\kappa} = \frac{\sum_{i=1}^{\kappa} x_i V_i}{\sum_{i=1}^{\kappa} V_i} = \frac{1}{V} \sum_{i=1}^{\kappa} x_i V_i$$

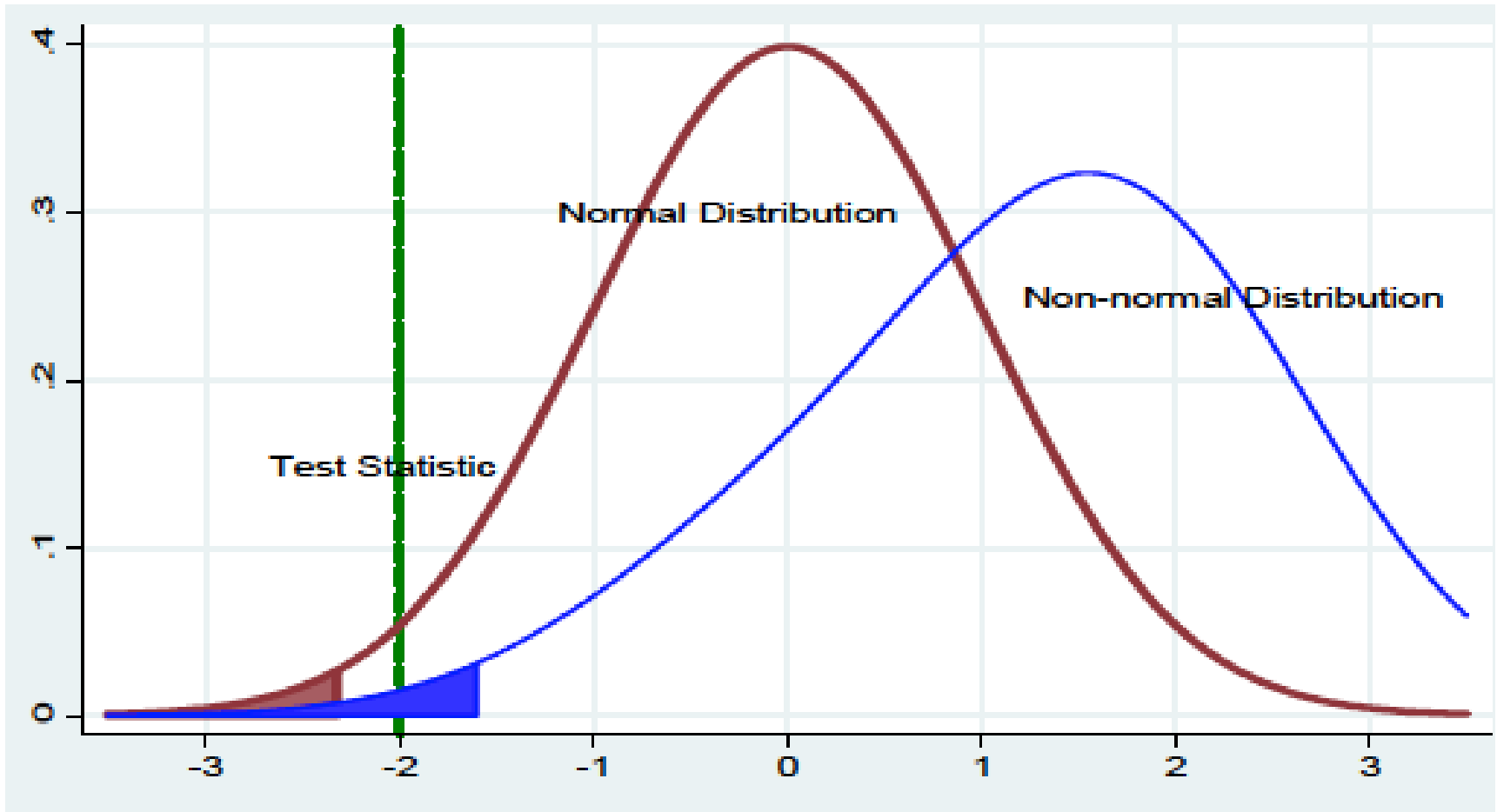
- **Διάμεσος (median):** ένα δείγμα n παρατηρήσεων το τοποθετούμε σε αύξουσα σειρά, αν το n είναι περιττός η διάμεσος είναι η μεσαία παρατήρηση, αν το n είναι άρτιος η διάμεσος είναι το ημίαθροισμα των δύο μεσαίων παρατηρήσεων.
- **Επικρατούσα τιμή (mode):** στην περίπτωση μη ομαδοποιημένων παρατηρήσεων είναι η παρατήρηση με τη μεγαλύτερη συχνότητα.

Επαγωγική στατιστική ανάλυση

Έλεγχος κανονικότητας (1/2)

- Όταν το δείγμα είναι μικρό, και η κανονικότητα δεν είναι σίγουρη δημιουργούνται διαγράμματα P-P και Q-Q και έλεγχοι Shapiro-Wilk W ($N \leq 2000$), Shapiro-Francia W ($N \leq 5000$), Kolmogorov-Smirnov D ($N > 2000$), and Jarque-Bera.
- Αν η κανονικότητα δεν ισχύει, τότε χρησιμοποιούνται μη-παραμετρικές μέθοδοι Kolmogorov-Smirnov, Kruscal-Wallis, Wilcoxon Rank-Sum.

Έλεγχος κανονικότητας (2/2)



t-test and Analysis of Variance (1/3)

Στη μηδενική υπόθεση H_0 :

- Ο έλεγχος one sample t-test ελέγχει αν η μέση τιμή του πληθυσμού είναι ίση με μία συγκεκριμένη τιμή.
- Ο έλεγχος two sample dependent t-test (paired t-test) ελέγχει αν η διαφορά μεταξύ τιμών ζευγών είναι ίση με το μηδέν (π.χ. πριν-μετά).
- Οι δύο αυτοί έλεγχοι είναι ΙΔΙΟΙ.

t-test and Analysis of Variance (2/3)

- Όταν τα δείγματα προέρχονται από δύο πληθυσμούς χρησιμοποιείται ο έλεγχος independent sample t-test που ελέγχει τις μέσες τιμές των δύο δειγμάτων
- Όταν τα δύο δείγματα έχουν την ίδια διακύμανση χρησιμοποιείται ο έλεγχος independent samples t-test με pooled variance
- Αν διαφέρει και η διακύμανσή τους τότε χρησιμοποιείται ο έλεγχος folded F test.

t-test and Analysis of Variance (3/3)

- Ενώ ο έλεγχος independent sample t-test συγκρίνει μέσες τιμές δύο δειγμάτων, ο έλεγχος one-way ANOVA (Analysis of Variance) συγκρίνει περισσότερες, χρησιμοποιώντας F στατιστικούς υπολογισμούς
- Ο έλεγχος t-test θεωρείται one-way ANOVA με σύγκριση δύο δειγμάτων και έναν βαθμό ελευθερίας
- Το μέγεθος t statistic είναι η τετραγωνική ρίζα του μεγέθους F statistic της ANOVA ($F=t^2$)
- Ο έλεγχος folded F test είναι διαφορετικός και χρησιμοποιείται όταν ελέγχεται η ισότητα δύο διακυμάνσεων.

ANOVA

- Χρησιμοποιείται για τη σύγκριση μέσων όρων τριών ή και περισσότερων ανεξάρτητων ομάδων (επεμβάσεων) ως προς μία μεταβλητή
- Απαραίτητες προϋποθέσεις είναι η μεταβλητή να είναι συνεχής (continuous), τα δεδομένα να έχουν κανονικότητα (όχι ακραίες τιμές) και να τηρείται η ομοιογένεια των παραλλακτικότητων
- Η ANOVA θα δείξει αν τουλάχιστον δύο μ.ο. διαφέρουν σημαντικά μεταξύ τους (σημαντικό F μέσου τετραγώνου επεμβάσεων)
- Ακολουθεί ο διαχωρισμός των μέσων όρων με ένα από τα συνήθη τεστ (π.χ. LSD, Tuckey, Duncan κλπ).

Παράδειγμα ANOVA

Μεταβλητή → χρόνος σε λεπτά για μεταφορά x τόνων εμπορεύματος

Επεμβάσεις → 4 πιθανές διαδρομές

α: μέσα από την πόλη (μ.ο. 58 λεπτά)

β: εσωτερική περιφερειακή (μ.ο. 46 λεπτά)

γ: εξωτερική περιφερειακή (μ.ο. 41 λεπτά)

δ: εθνική οδός (μ.ο. 32 λεπτά)

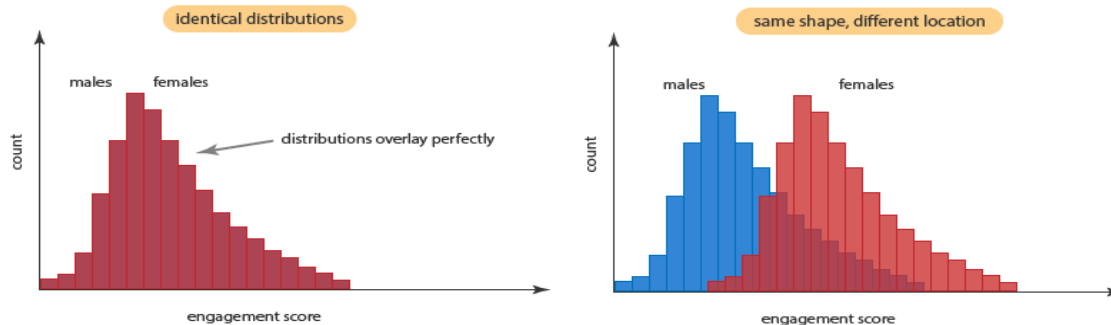
Επαναλήψεις → 20 δρομολόγια σε κάθε διαδρομή

➤ Η ανάλυση της παραλλακτικότητας θα μας δείξει αν τουλάχιστον δύο μέσοι όροι διαφέρουν μεταξύ τους (F σημαντικό).

➤ Ο διαχωρισμός των μ.ο. θα δείξει σε επίπεδο σημαντικότητας π.χ. 5% ποιοι μ.ο. διαφέρουν μεταξύ τους.

Mann-Whitney U-test

- Χρησιμοποιείται για τη σύγκριση διαφορών μεταξύ δύο ανεξάρτητων ομάδων
- Η εξαρτημένη μεταβλητή μπορεί να είναι διάταξης (ordinal) ή συνεχής (continuous)
- Τα δεδομένα της εξαρτημένης μεταβλητής δεν ακολουθούν κανονική κατανομή
- Οι κατανομές θα πρέπει ωστόσο να έχουν παρόμοια διάταξη (σχήμα)

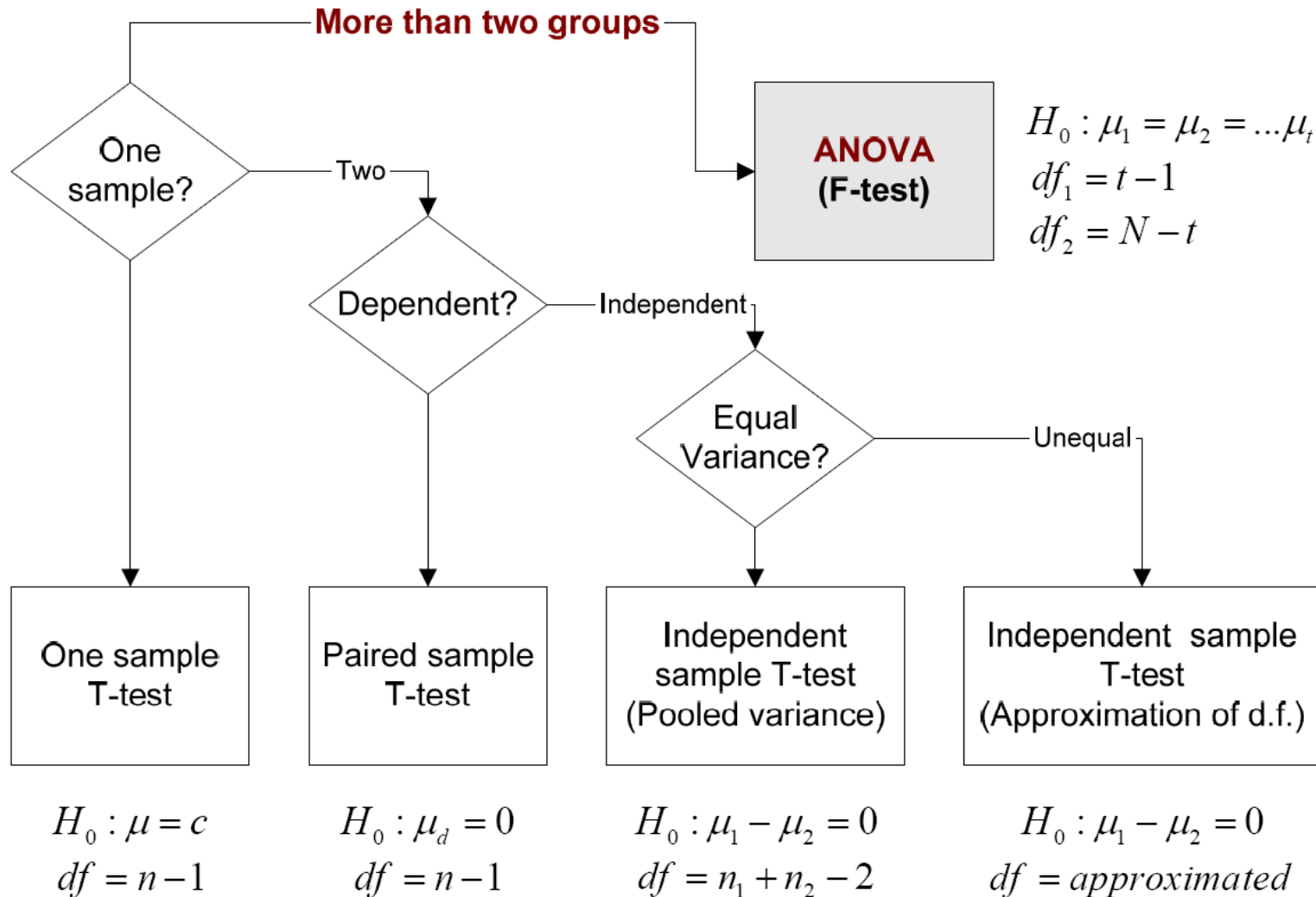


- Είναι εν μέρει η μη παραμετρική προσέγγιση του t-test
- Μπορεί να χρησιμοποιηθεί για να μελετηθούν π.χ. οι προτιμήσεις διαφόρων ανθρώπων και πώς επηρεάζονται ανά τοποθεσία.

Kruskal-Wallis H test

- Μη παραμετρικό τεστ βασισμένο σε δεδομένα που είναι σε κατάταξη (π.χ. αύξουσα)
- Χρησιμοποιείται για να καθορίσει εάν υπάρχουν στατιστικά σημαντικές διαφορές μεταξύ δύο ή περισσότερων ομάδων μίας ανεξάρτητης μεταβλητής σε μία εξαρτημένη συνεχή μεταβλητή (continuous) ή μεταβλητή διάταξης (ordinal)
- Θεωρείται η μη παραμετρική προσέγγιση της ANOVA ή μια επέκταση του U-test καθώς επιτρέπει τη σύγκριση περισσότερων των δύο ομάδων

Λογικό διάγραμμα επιλογής στατιστικού τεστ

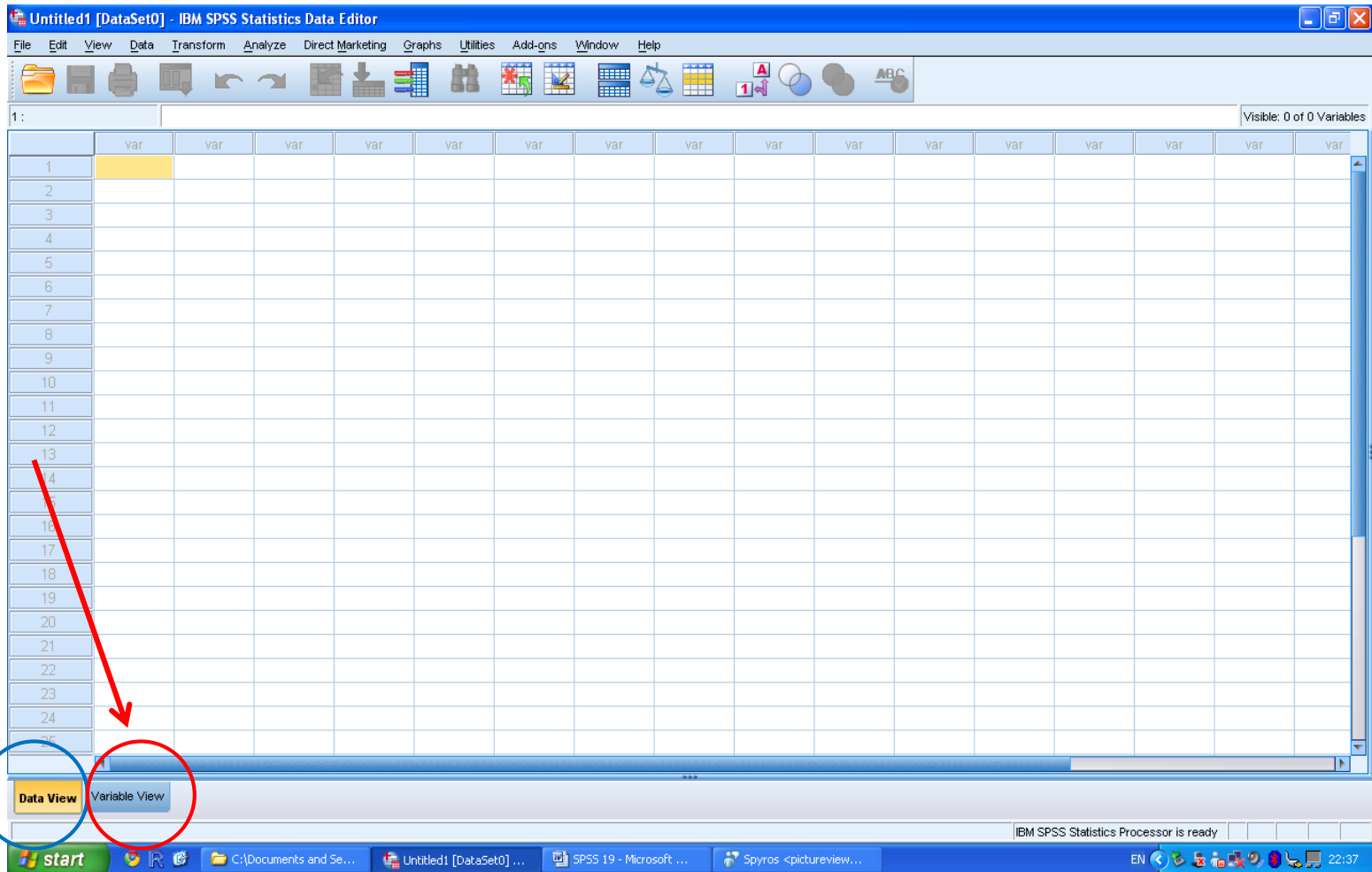


Λίγα λόγια για το στατιστικό πακέτο SPSS



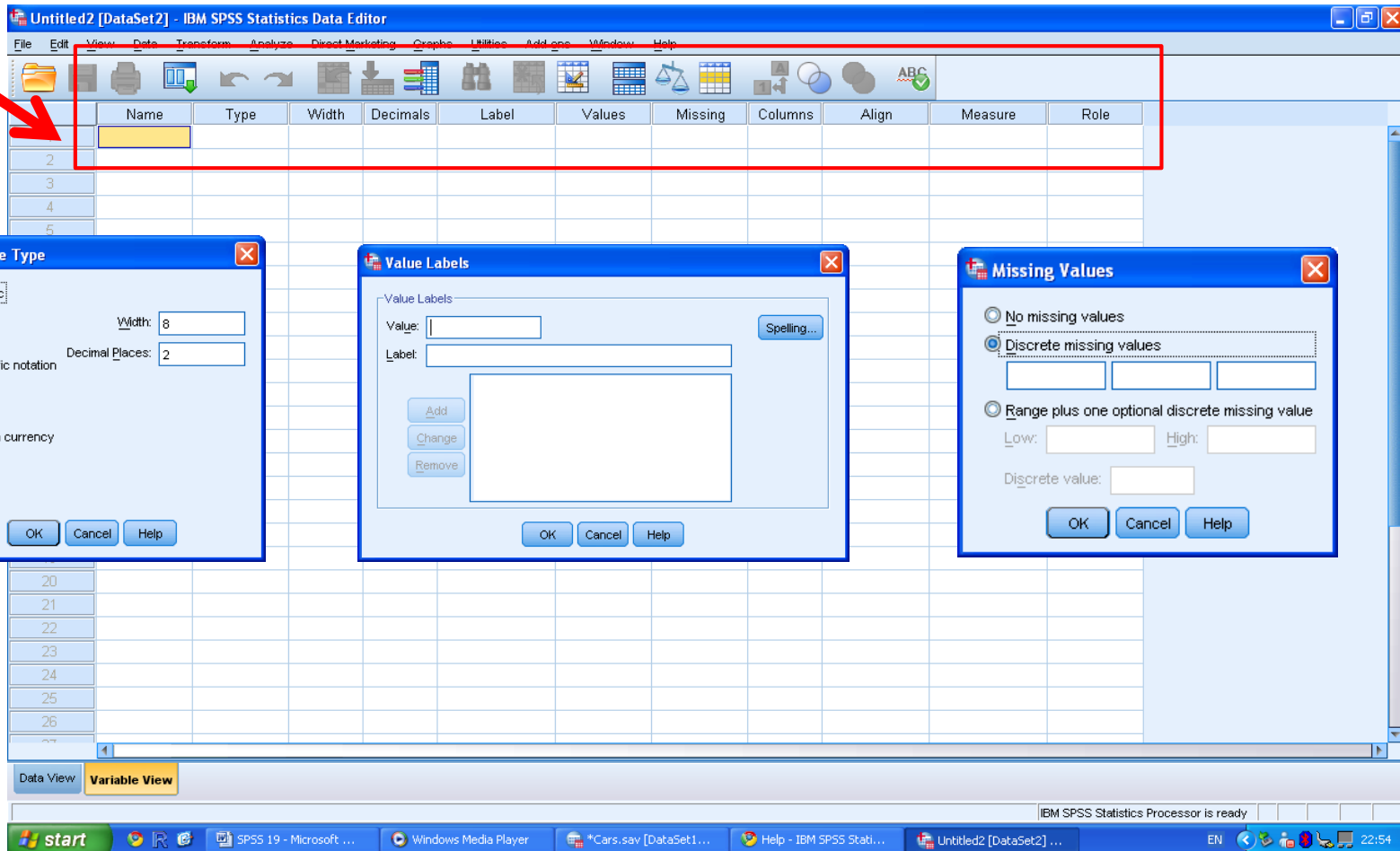
Βασικά «παράθυρα» του SPSS

- **Data editor:**



Βασικά «παράθυρα» του SPSS

Variable view:



The screenshot shows the IBM SPSS Statistics Data Editor in Variable View. A red box highlights the menu bar and toolbar, with a red arrow pointing to the 'View' menu. Three dialog boxes are open in the foreground:

- Variable Type:** Shows 'Numeric' selected with a width of 8 and 2 decimal places.
- Value Labels:** Shows fields for 'Value' and 'Label' with 'Add', 'Change', and 'Remove' buttons.
- Missing Values:** Shows 'Discrete missing values' selected with input fields for 'Low', 'High', and 'Discrete value'.

The background shows the Variable View grid with columns: Name, Type, Width, Decimals, Label, Values, Missing, Columns, Align, Measure, and Role. The status bar at the bottom indicates 'Variable View' is active.

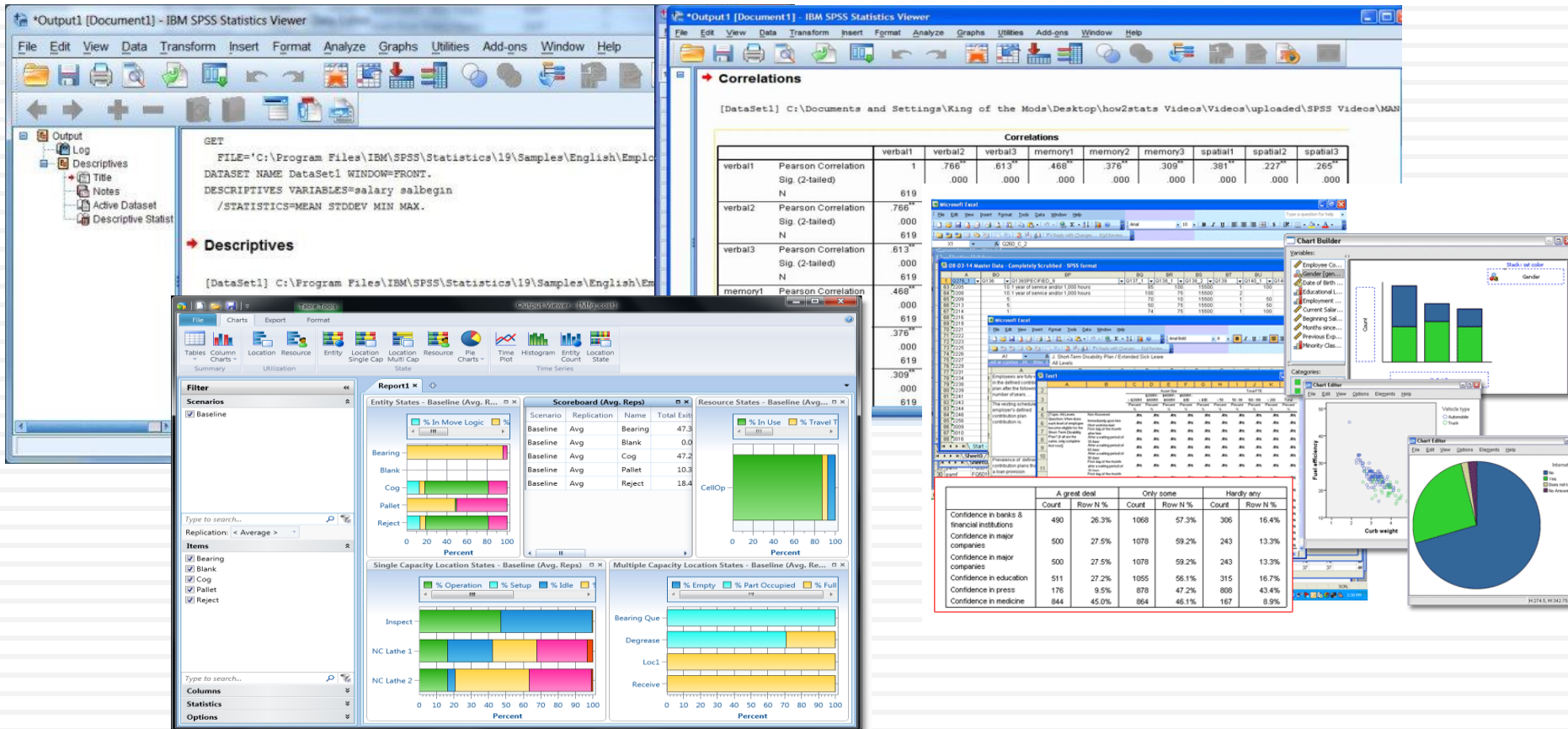
Βασικά «παράθυρα» του SPSS

Analyze:

The screenshot shows the IBM SPSS Statistics Data Editor interface. The 'Analyze' menu is open, displaying a list of statistical analysis options. The main window displays a data grid with one variable named 'VAR00001' and a value of 1.00 in the first row. The status bar at the bottom indicates 'Data View' and 'Variable View'.

	VAR00001	var
1	1,00	
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		
17		
18		
19		
20		
21		
22		
23		
24		
25		

Επεξεργασία αποτελεσμάτων με τη χρήση του SPSS



The image displays a collage of IBM SPSS Statistics Viewer windows. The main window shows the 'Correlations' output for a dataset named 'DataSet1'. Below the correlation table, there are several 'Report' windows showing various charts and data summaries.

Correlations Table:

	verbal1	verbal2	verbal3	memory1	memory2	memory3	spatial1	spatial2	spatial3
verbal1	1	.766 ^{**}	.613 ^{**}	.468 ^{**}	.376 ^{**}	.309 ^{**}	.381 ^{**}	.227 ^{**}	.265 ^{**}
verbal2		1	.000	.000	.000	.000	.000	.000	.000
verbal3			1	.000	.000	.000	.000	.000	.000
memory1				1	.000	.000	.000	.000	.000
memory2					1	.000	.000	.000	.000
memory3						1	.000	.000	.000
spatial1							1	.000	.000
spatial2								1	.000
spatial3									1

Report Windows:

- Entity States - Baseline (Avg. Re...):** Shows a horizontal bar chart for 'Bearing', 'Blank', 'Cog', 'Pallet', and 'Reject' with categories like '% In Move Logic' and '% In Use'.
- Scoreboard (Avg. Re...):** A table showing 'Total Exit' for 'Bearing', 'Blank', 'Cog', 'Pallet', and 'Reject'.
- Resource States - Baseline (Avg...):** Shows a horizontal bar chart for 'CellOp' with categories like '% In Use' and '% Travel T'.
- Single Capacity Location States - Baseline (Avg. Re...):** Shows a horizontal bar chart for 'Inspect', 'NC Lath-1', and 'NC Lath-2' with categories like '% Operation', '% Setup', and '% Idle'.
- Multiple Capacity Location States - Baseline (Avg. Re...):** Shows a horizontal bar chart for 'Bearing Que', 'Decrease', 'Loc1', and 'Receive' with categories like '% Empty', '% Part Occupied', and '% Full'.

Other Windows:

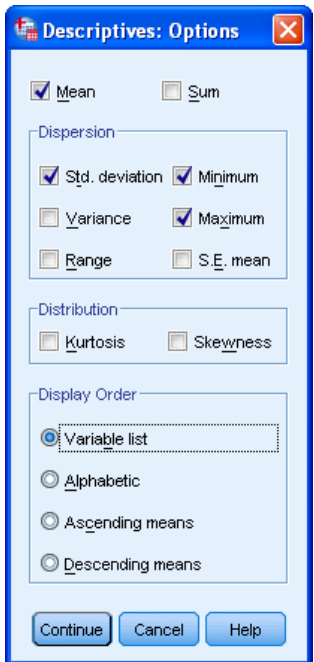
- Microsoft Excel:** Shows a data table with columns for 'A', 'B', 'C', 'D', 'E', 'F', 'G', 'H', 'I', 'J', 'K', 'L', 'M', 'N', 'O', 'P', 'Q', 'R', 'S', 'T', 'U', 'V', 'W', 'X', 'Y', 'Z'.
- Chart Builder:** Shows a bar chart with 'Count' on the y-axis and 'Gender' on the x-axis.
- Chart Editor:** Shows a scatter plot with 'Fuel efficiency' on the y-axis and 'Curb weight' on the x-axis.
- Chart Editor:** Shows a pie chart with 'Internet' on the x-axis and 'Count' on the y-axis.

Confidence in various sectors table:

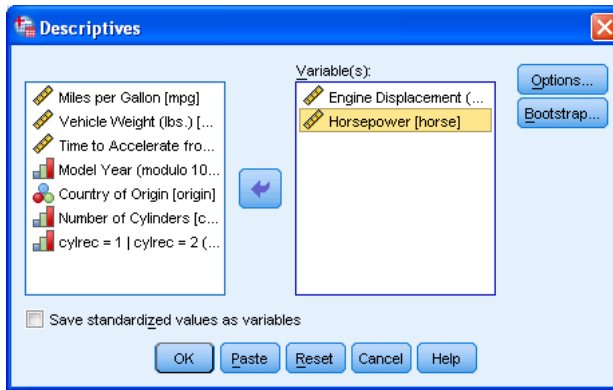
	A great deal		Only some		Hardly any	
	Count	Row N %	Count	Row N %	Count	Row N %
Confidence in banks & financial institutions	490	26.3%	1068	57.3%	306	16.4%
Confidence in major companies	500	27.5%	1078	59.2%	243	13.3%
Confidence in major companies	500	27.5%	1078	59.2%	243	13.3%
Confidence in education	511	27.2%	1055	58.1%	315	16.7%
Confidence in press	178	9.5%	870	47.2%	808	43.4%
Confidence in medicine	844	45.0%	864	46.1%	187	8.9%

Περιγραφικά μέτρα για συνεχείς μεταβλητές

- Analyze → Descriptive statistics → Descriptives



Mean Sum
 Dispersion
 Std. deviation Minimum
 Variance Maximum
 Range S.E. mean
 Distribution
 Kurtosis Skewness
 Display Order
 Variable list
 Alphabetic
 Ascending means
 Descending means
 Continue Cancel Help

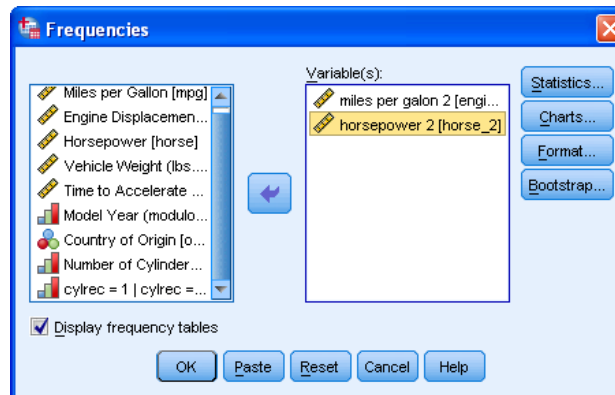


Variable(s):
 Engine Displacement (...)
 Horsepower [horse]
 Save standardized values as variables
 OK Paste Reset Cancel Help

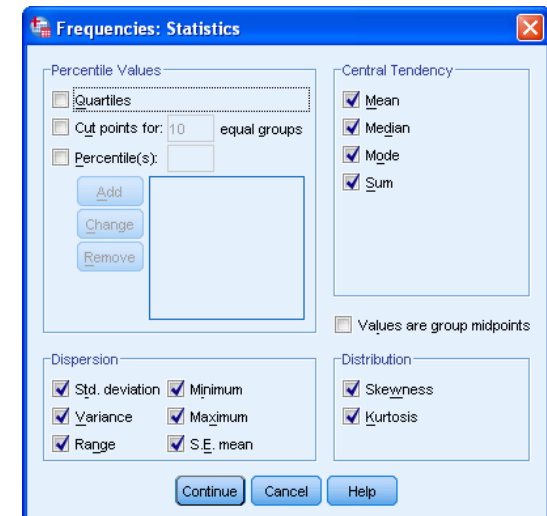
και / ή

- Analyze → Descriptive statistics → Frequencies

Δεν το
 τσεκάρω!
 



Variable(s):
 miles per galon 2 [engi...]
 horsepower 2 [horse_2]
 Display frequency tables
 Statistics... Charts... Format... Bootstrap...
 OK Paste Reset Cancel Help



Percentile Values
 Quartiles
 Cut points for: 10 equal groups
 Percentile(s):
 Add Change Remove
 Central Tendency
 Mean
 Median
 Mode
 Sum
 Values are group midpoints
 Dispersion
 Std. deviation Minimum
 Variance Maximum
 Range S.E. mean
 Distribution
 Skewness
 Kurtosis
 Continue Cancel Help

Περιγραφικά μέτρα για κατηγορικές μεταβλητές

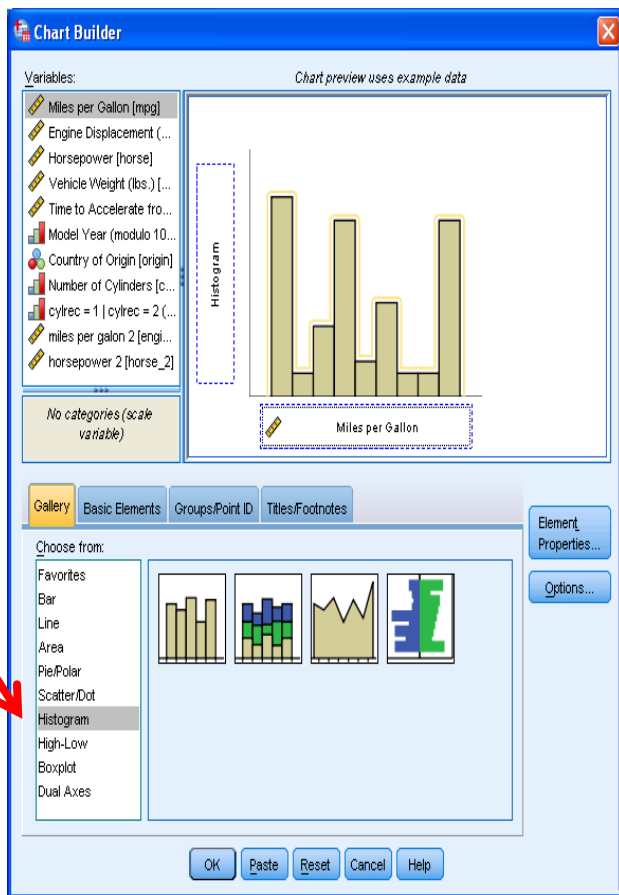
- Analyze → Descriptive statistics → Frequencies



To
τσεκάρω!

Ιστογράμματα

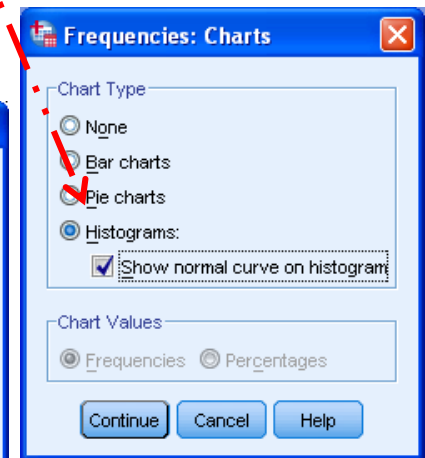
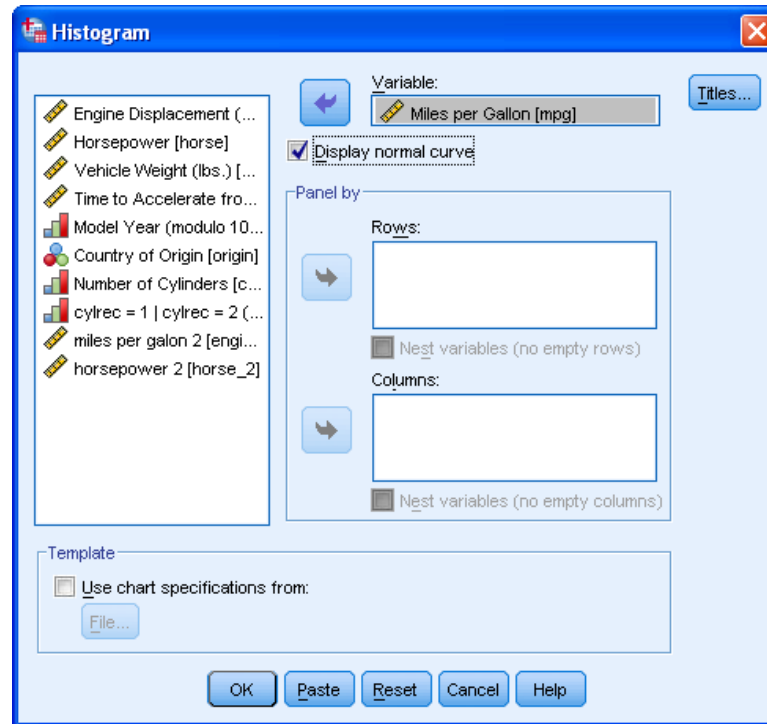
1 Graphs → Chart builder



3 Analyze → Descriptive statistics

→ Frequencies

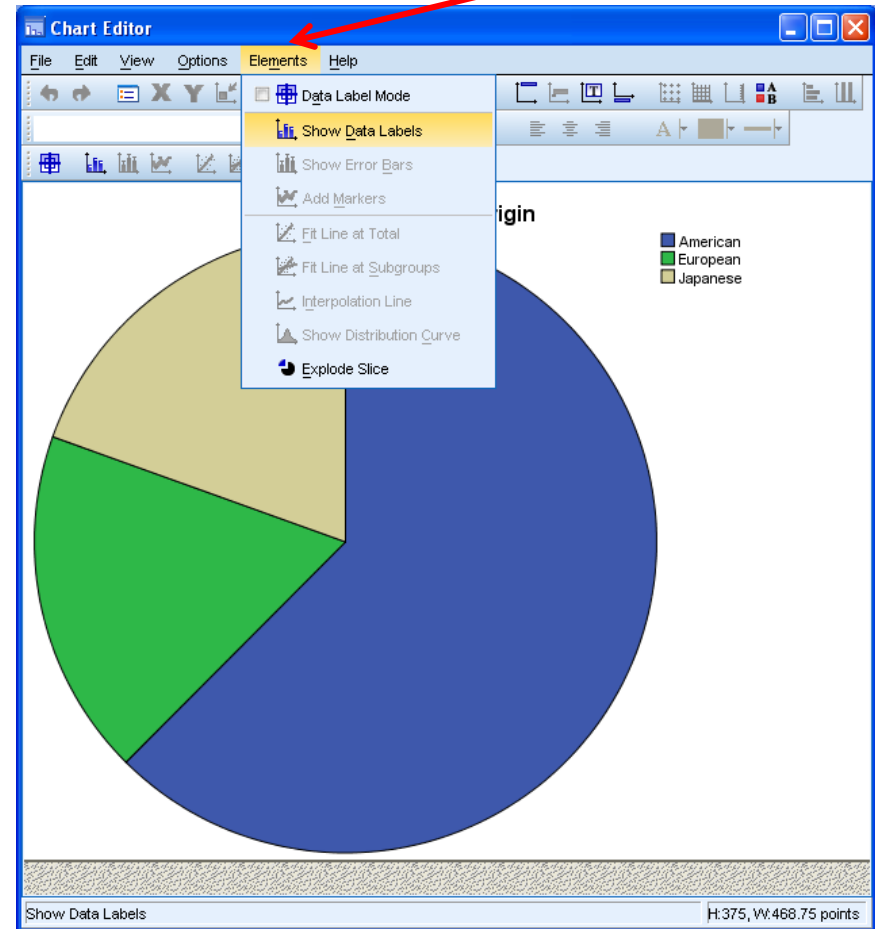
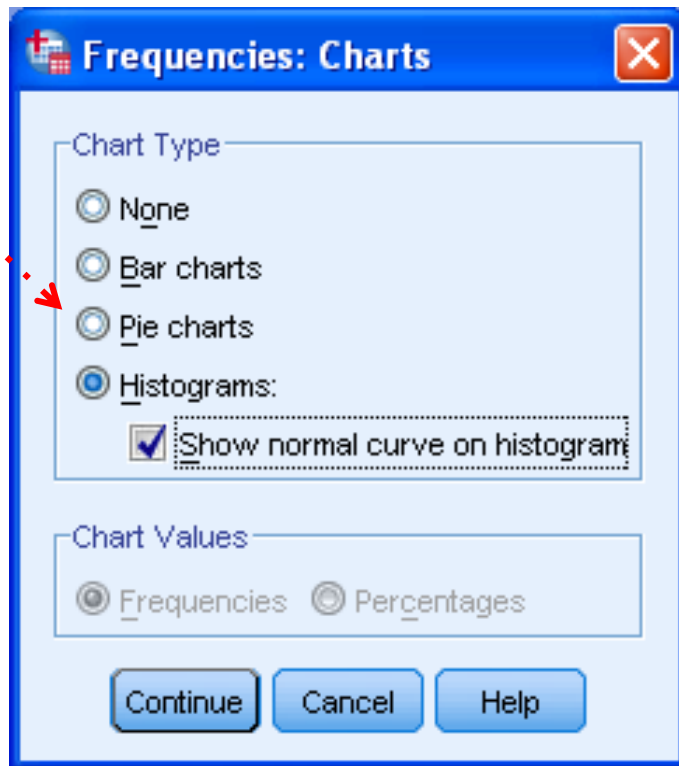
2 Legacy dialogs → Histogram



Κυκλικά διαγράμματα

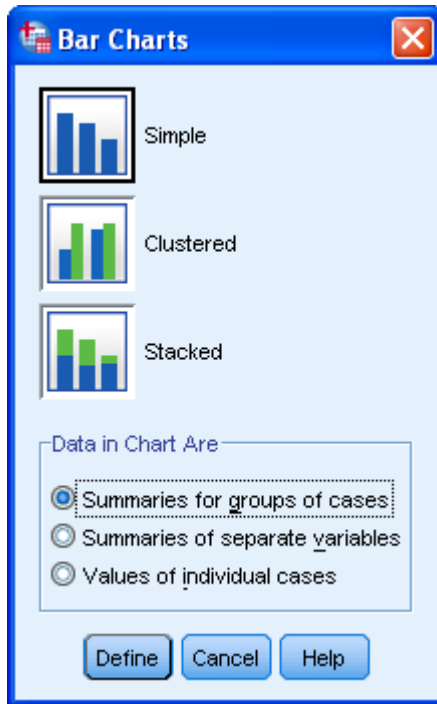
- Analyze → Descriptive statistics → Frequencies

Labels

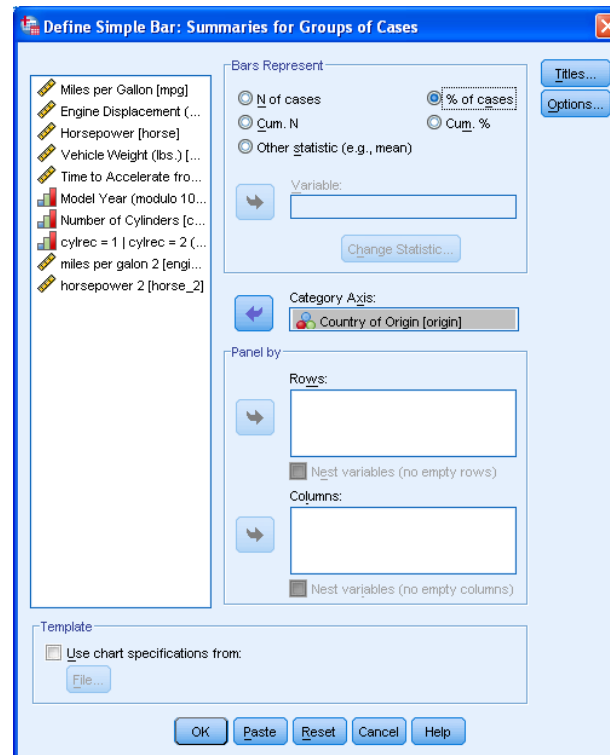


Ραβδογράμματα

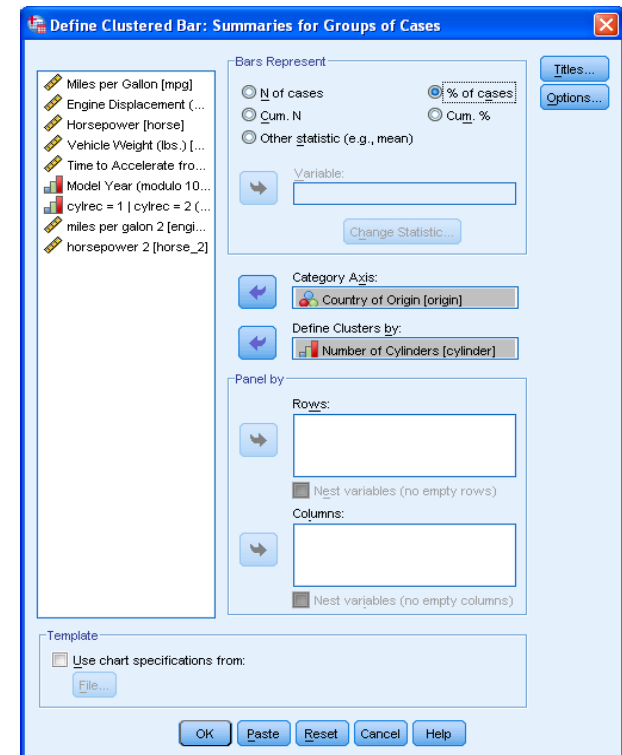
- **Graphs** → **Legacy** → **Dialogs** → **Bar**



Simple



Clustered



Έλεγχος κανονικότητας (1/3)

- 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**.

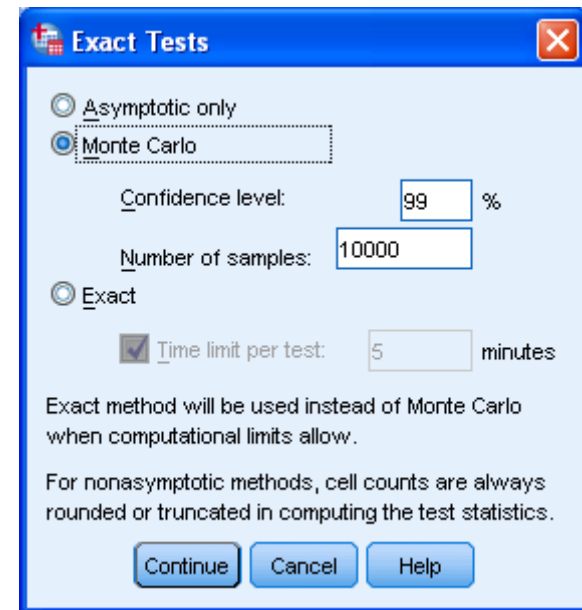
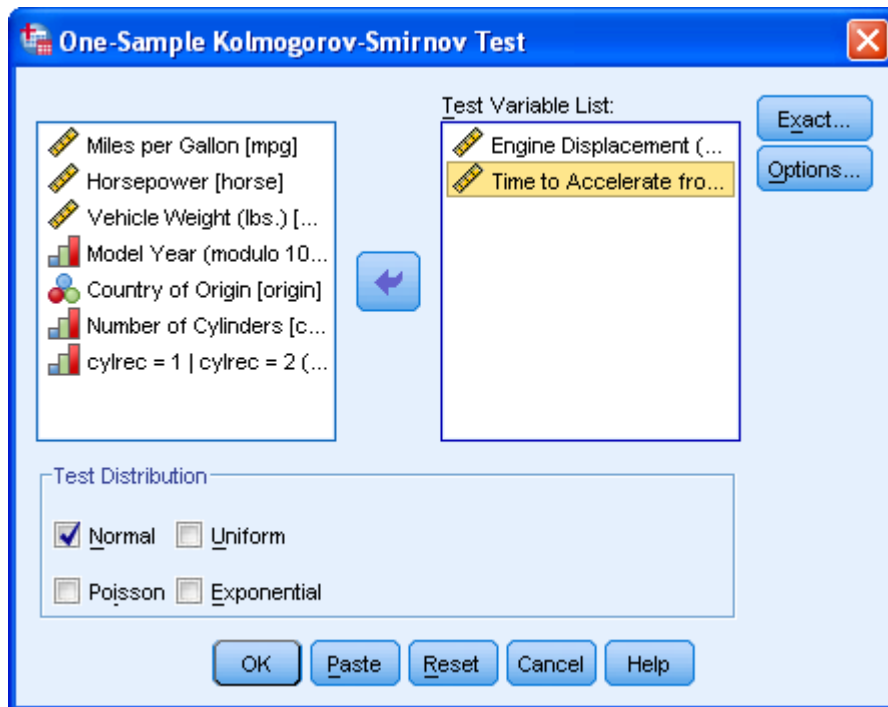


Έλεγχος κανονικότητας (2/3)

- Έλεγχος κανονικότητας \longrightarrow 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.**

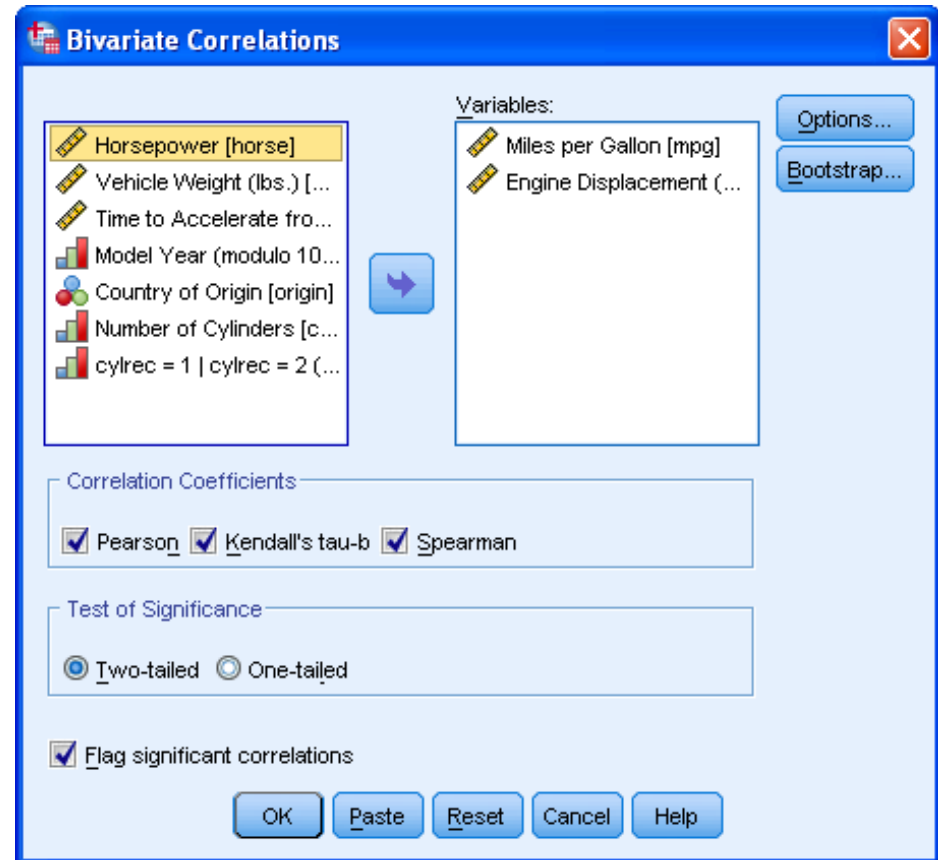
Έλεγχος κανονικότητας (3/3)

- Analyze → Nonparametric test → Legacy Dialogs → 1-sample K-S

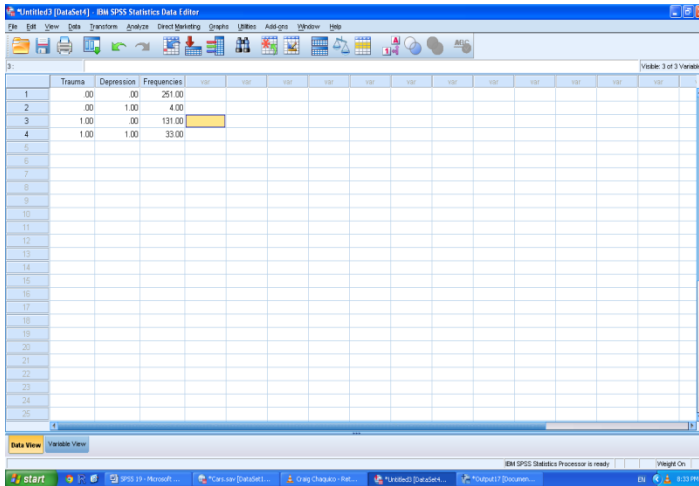


Συντελεστές γραμμικής συσχέτισης

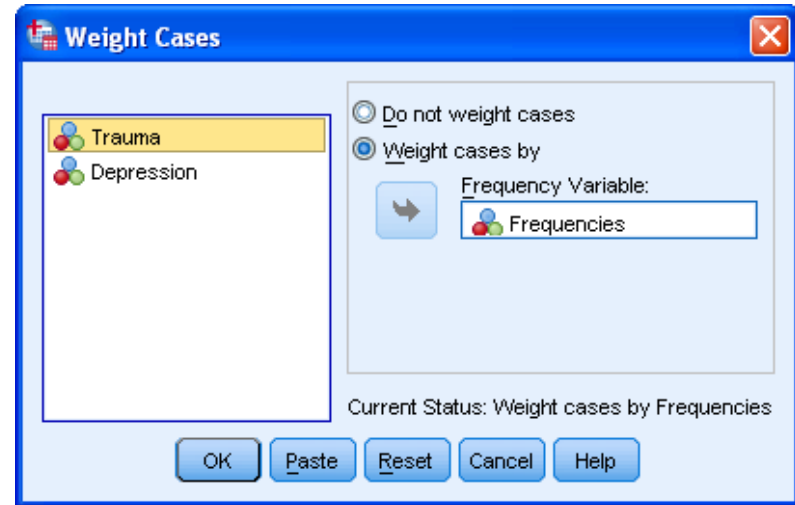
- **Analyze** → **Correlate** → **Bivariate**
- **Μηδενική υπόθεση (null hypothesis) $H_0: \rho=0$** : δεν υπάρχει γραμμική συσχέτιση μεταξύ των 2 μεταβλητών
- **Εναλλακτική υπόθεση (alternative hypothesis) $H_1: \rho \neq 0$** υπάρχει γραμμική συσχέτιση μεταξύ των 2 μεταβλητών
- Όταν ο συντελεστής συσχέτισης είναι μεγάλος, σημαίνει ότι η συσχέτιση των μεταξύ των τιμών των μεταβλητών είναι ισχυρή.



Χ² Έλεγχος ανεξαρτησίας για **κατηγορικές** μεταβλητές



	Trauma	Depression	Frequencies
1	.00	1.00	251.00
2	.00	1.00	4.00
3	1.00	.00	131.00
4	1.00	1.00	33.00



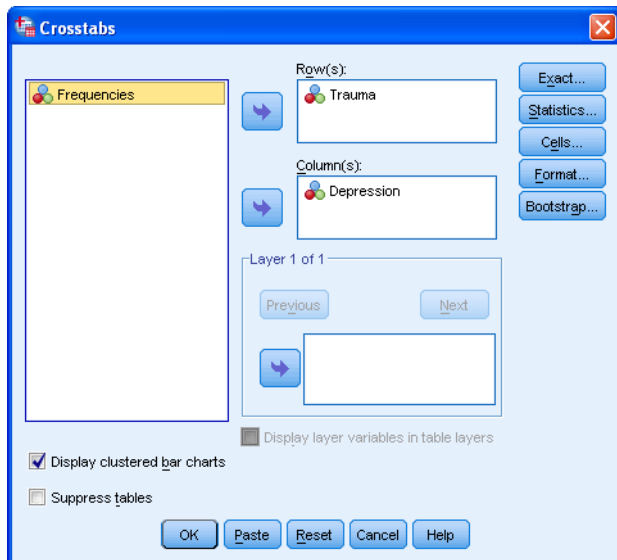
Weight Cases

Do not weight cases
 Weight cases by

Frequency Variable:

Frequencies

Current Status: Weight cases by Frequencies



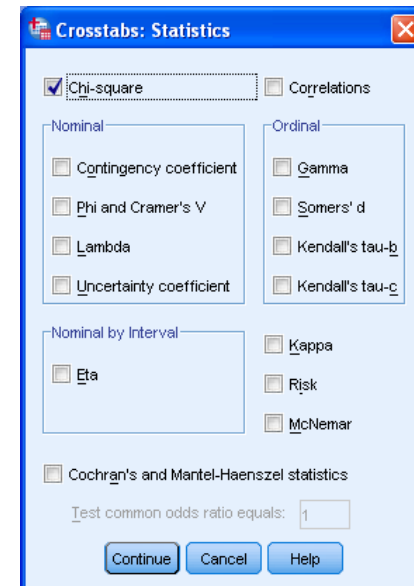
Crosstabs

Row(s): Trauma

Column(s): Depression

Layer 1 of 1

Display clustered bar charts
 Suppress tables



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

Έλεγχοι υποθέσεων για το μέσο και τη διάμεσο ενός δείγματος (t test και Wilcoxon test)

- Ο έλεγχος κανονικότητας θα καθορίσει εάν θα χρησιμοποιήσουμε το t-test [κανονική κατανομή] ή το Wilcoxon test [μη - κανονική κατανομή].

- **t-test στο SPSS:**

Analyze → Compare Means → One sample t-test

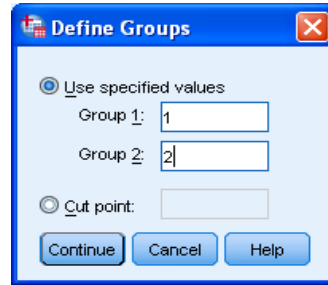
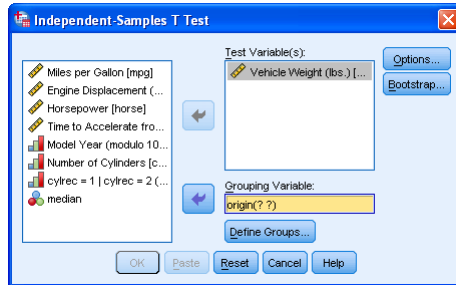
- **Wilcoxon στο SPSS:**

Analyze → Nonparametric tests → Legacy Dialogs → 2 Related Samples

Έλεγχοι υποθέσεων για τη διαφορά των μέσων 2 ανεξάρτητων δειγμάτων (t test και Mann-Whitney-Wilcoxon test)

- Ο έλεγχος κανονικότητας θα καθορίσει εάν θα χρησιμοποιήσουμε το t-test [κανονική κατανομή] ή το Mann-Whitney-Wilcoxon test [μη - κανονική κατανομή].
- **t-test στο SPSS:**

Analyze → Compare Means → Independent-Samples t-test



- **Mann-Whitney-Wilcoxon test στο SPSS:**

Analyze → Nonparametric tests → Legacy Dialogs → 2 Independent Samples

Έλεγχοι υποθέσεων για τη διαφορά των μέσων 2 εξαρτημένων δειγμάτων (t test και Wilcoxon test για δείγμα παρατηρήσεων)

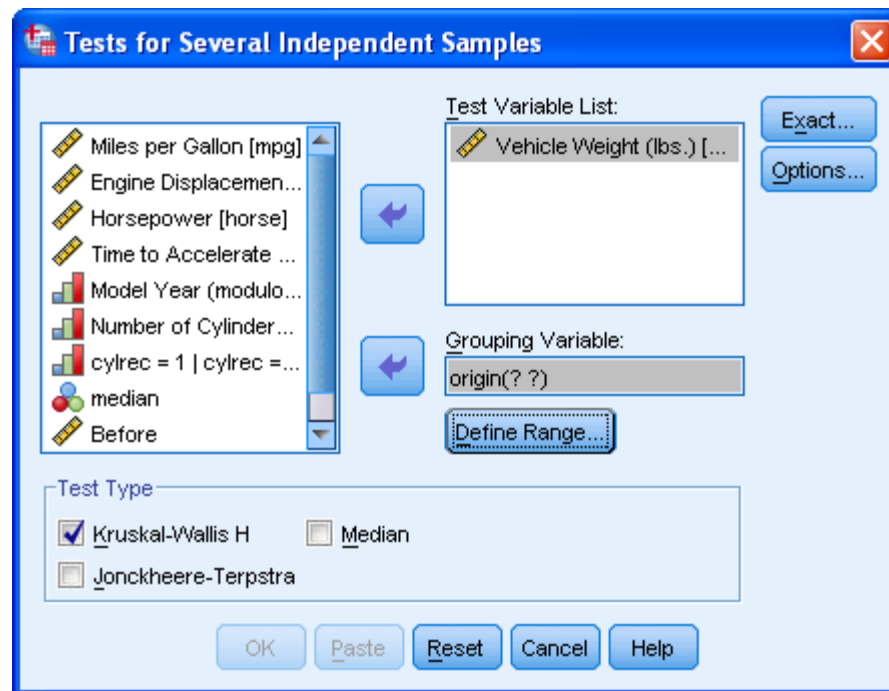
- Ο έλεγχος κανονικότητας θα καθορίσει εάν θα χρησιμοποιήσουμε το t-test [κανονική κατανομή] ή το Wilcoxon test [μη - κανονική κατανομή].
- **t-test στο SPSS:**
Analyze → Compare Means → Paired-Samples t-test
- **Wilcoxon test στο SPSS:**
Analyze → Nonparametric tests → Legacy Dialogs → 2 Related Samples

Ανάλυση διακύμανσης κατά ένα παράγοντα (One way ANOVA)

- Όταν έχω δείγματα που προέρχονται από περισσότερους από 2 πληθυσμούς
 - Οι υποθέσεις εφαρμογής της μεθόδου είναι πιο αυστηρές σε σχέση με την περίπτωση των 2 δειγμάτων.
 - Είναι ίδιες με την περίπτωση της απλής γραμμικής παλινδρόμησης: κανονικότητα, ανεξαρτησία και ομοσκεδαστικότητα των καταλοίπων.
 - Όταν λέμε ομοσκεδαστικότητα των καταλοίπων, εννοούμε ότι τα κατάλοιπα που δημιουργούνται να έχουν ίσες διαφορές για κάθε επίπεδο του παράγοντα.
-
- **Analyze → General linear model → Univariate**
- ή**
- **Analyze → Compare means → One way ANOVA**

Μη παραμετρική ανάλυση διακύμανσης (τεστ Kruskal-Wallis)

- Analyze → Non parametric tests → Legacy Dialogs → K-independent Samples



Ανάλυση διακύμανσης για εξαρτημένα δείγματα

- Analyze → General linear model → Repeated measures

Repeated Measures Defin...

Within-Subject Factor Name: week

Number of Levels: 4

Add
Change
Remove

Measure Name: sales

Add
Change
Remove

Define Reset Cancel Help

Repeated Measures

Within-Subjects Variables (week):
Week1(1,sales)
Week2(2,sales)
?_(3,sales)
?_(4,sales)

Model...
Contrasts...
Plots...
Post Hoc...
Save...
Options...

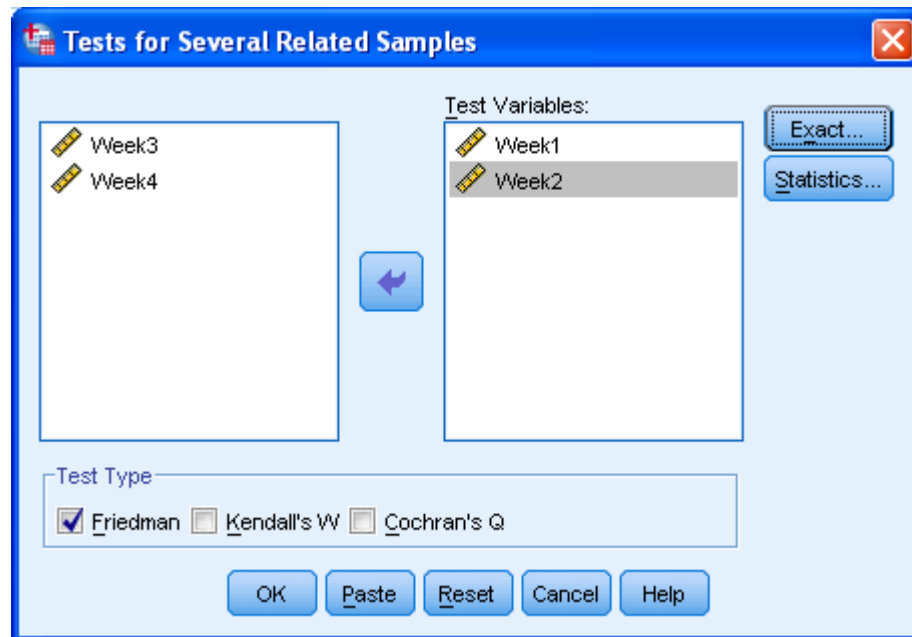
Between-Subjects Factor(s):

Covariates:

OK Paste Reset Cancel Help

Μη παραμετρική ανάλυση διακύμανσης για εξαρτημένα δείγματα

- Analyze → Non parametric tests → Legacy dialogs → K Related samples



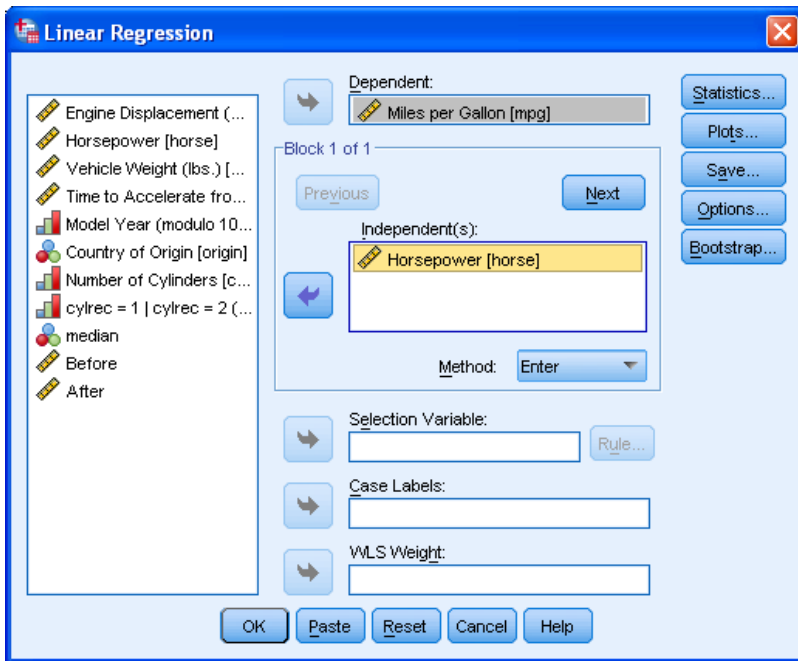
Γραμμική παλινδρόμηση

- Υποθέσεις της γραμμικής παλινδρόμησης:
 - Κανονικότητα των καταλοίπων*
 - Ανεξαρτησία των καταλοίπων
 - Ομοσκεδαστικότητα των καταλοίπων

**Με την απλή γραμμική παλινδρόμηση προσπαθούμε να εκτιμήσουμε τις τιμές της ανεξάρτητης μεταβλητής χρησιμοποιώντας τις τιμές της εξαρτημένης. Οι εκτιμώμενες τιμές θα είναι προφανώς διαφορετικές από τις πραγματικές τιμές της ανεξάρτητης μεταβλητής. Οι αποκλίσεις των τιμών των ανεξάρτητων από τις αντίστοιχες εκτιμώμενες τιμές τους ονομάζονται κατάλοιπα (ή σφάλματα).*

Απλή γραμμική παλινδρόμηση

- Analyze → Regression → Linear



Linear Regression

Dependent: Miles per Gallon [mpg]

Independent(s): Horsepower [horse]

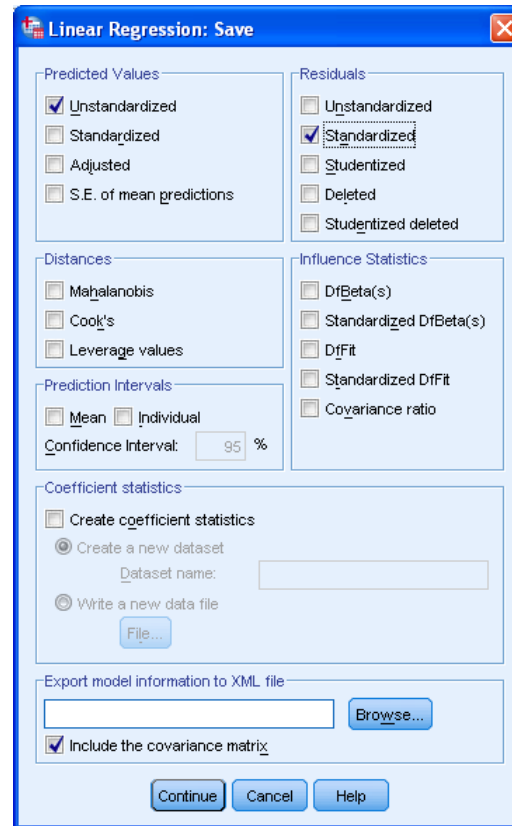
Method: Enter

Selection Variable:

Case Labels:

WLS Weight:

Buttons: OK, Paste, Reset, Cancel, Help, Statistics..., Plots..., Save..., Options..., Bootstrap...



Linear Regression: Save

Predicted Values:

- Unstandardized
- Standardized
- Adjusted
- S.E. of mean predictions

Residuals:

- Unstandardized
- Standardized
- Studentized
- Deleted
- Studentized deleted

Distances:

- Mahalanobis
- Cook's
- Leverage values

Influence Statistics:

- DfBeta(s)
- Standardized DfBeta(s)
- DfFit
- Standardized DfFit
- Covariance ratio

Prediction Intervals:

- Mean
- Individual

 Confidence Interval: 95 %

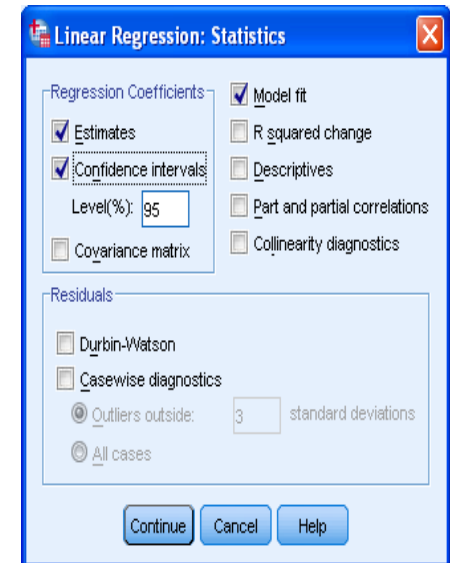
Coefficient statistics:

- Create coefficient statistics
- Create a new dataset
 - Dataset name:
- Write a new data file
 - File...

Export model information to XML file:

- Include the covariance matrix

Buttons: Continue, Cancel, Help



Linear Regression: Statistics

Regression Coefficients:

- Estimates
- Confidence intervals
 - Level(%): 95
- Covariance matrix

Model fit:

- Model fit
- R squared change
- Descriptives
- Part and partial correlations
- Collinearity diagnostics

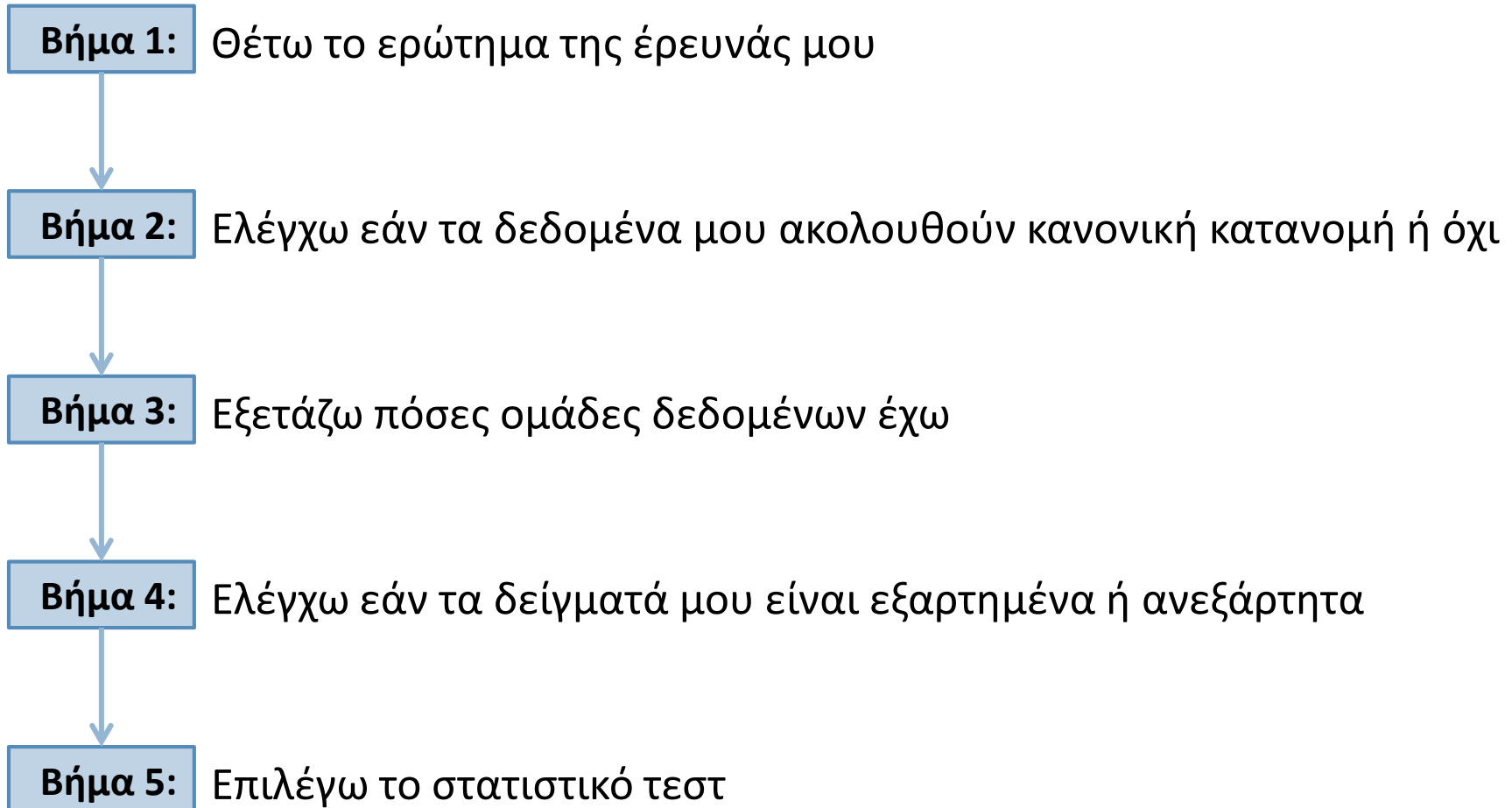
Residuals:

- Durbin-Watson
- Casewise diagnostics
 - Outliers outside: 3 standard deviations
 - All cases

Buttons: Continue, Cancel, Help

Μεθοδολογία επιλογής κατάλληλου στατιστικού ΤΕΣΤ

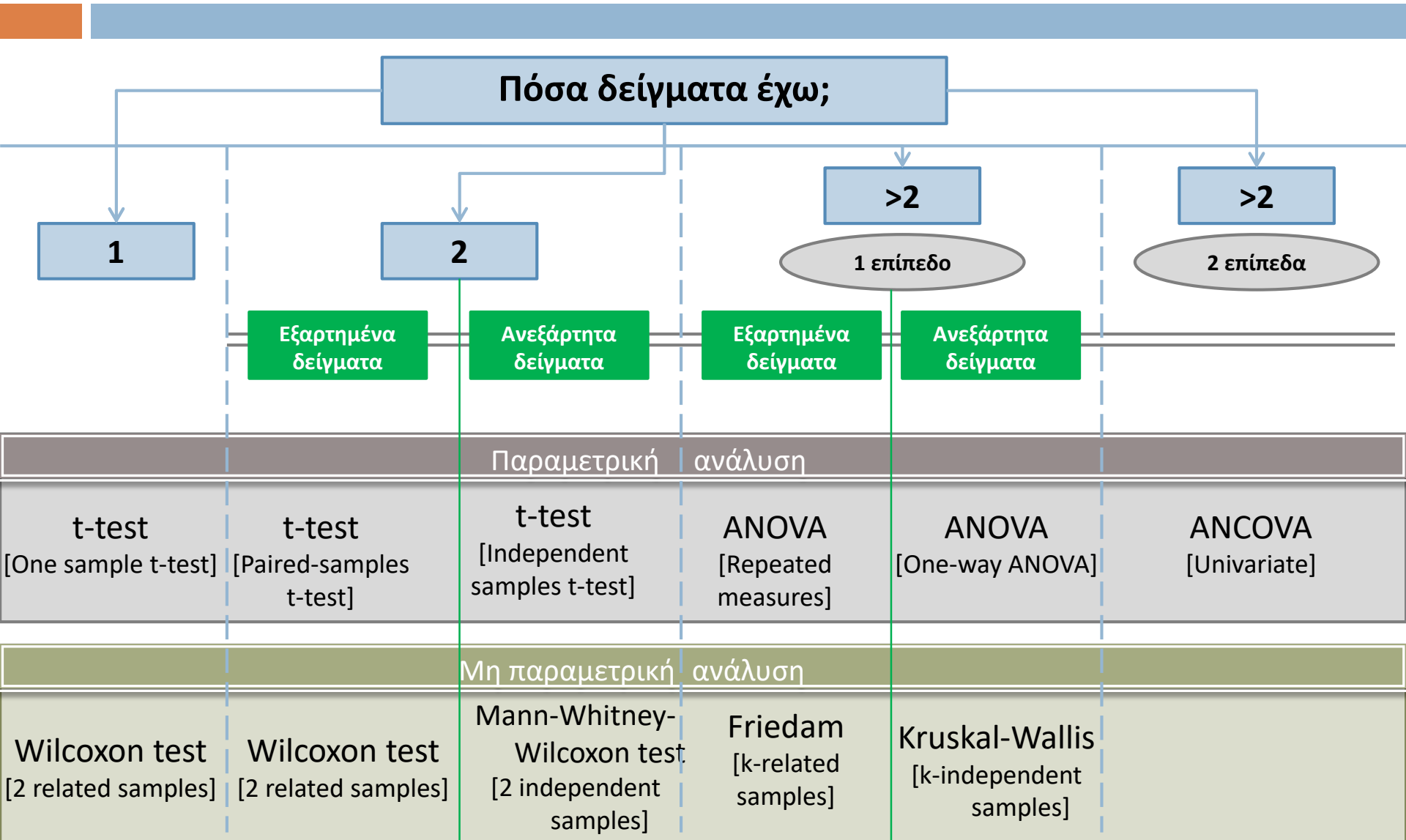
Πώς να επιλέξω το κατάλληλο στατιστικό τεστ;



Έλεγχος κανονικότητας και επιλογή ανάλυσης



Επιλογή στατιστικού τεστ



Τι γίνεται όταν έχω κατηγορικές μεταβλητές;

Ποιο τεστ πρέπει να κάνω;

Έλεγχος ανεξαρτησίας χ^2



Chi-square test

Τι ελέγγω με αυτό το τεστ;



Εάν 2 κατηγορικές μεταβλητές είναι ανεξάρτητες μεταξύ τους

Τι γίνεται εάν έχω 3 μεταβλητές;



3-variable Chi-square test



ΤΕΣΤ ΣΤΟ SPSS



CROSSTABS

Παράρτημα

One sample T-Test

Suppose we obtain n measurements y_1 through y_n that were randomly selected from a normally distributed population with unknown parameters μ and σ^2 . One example is the SAT scores of 100 undergraduate students who were randomly chosen.

The one sample t-test examines whether the unknown population mean μ differs from a hypothesized value c . This is the null hypothesis of the one sample t-test, $H_0 : \mu = c$.⁶ The t statistic is computed as follows.

$$t = \frac{\bar{y} - c}{s_{\bar{y}}} \sim t(n-1)$$

where y is a variable to be tested, $\bar{y} = \frac{\sum y_i}{n}$ is the mean of y , $s^2 = \frac{\sum (y_i - \bar{y})^2}{n-1}$ is the variance

of y , $s_{\bar{y}} = \frac{s}{\sqrt{n}}$ is the standard error of \bar{y} , and n is the number of observations. The t statistic

follows Student's t probability distribution with $n-1$ degrees of freedom (Gosset 1908).

Here we are testing if the population mean of the death rate from lung cancer is 20 per 100,000 people at the .01 significance level. The null hypothesis of this two-tailed test is $H_0 : \mu = 20$.

Paired T-test: Dependent Samples

T-tests compare the means of two samples. Two variables may or may not be independent. When each element of a sample is matched to its corresponding element of the other sample, two samples are paired. This paired t-test examines the mean of individual differences of paired measurements and thus is appropriate for pre-post situations. Suppose we want to investigate the effectiveness of a new medicine on lung cancer by checking patients before and after they took the medicine.

The paired t-test is based on the pairwise differences in values of matched observations of two samples, $d_i = y_{1i} - y_{2i}$. The difference of matched pairs is treated as a variable; the logic of the paired t-test and one sample t-test is identical.

$$t_{\bar{d}} = \frac{\bar{d} - D_0}{s_{\bar{d}}} \sim t(n-1), \text{ where } \bar{d} = \frac{\sum d_i}{n}, s_d^2 = \frac{\sum (d_i - \bar{d})^2}{n-1}, \text{ and } s_{\bar{d}} = \frac{s_d}{\sqrt{n}}.$$

The null hypothesis is that the population mean of individual differences of paired observations is D_0 (zero unless explicitly specified), $H_0 : \mu_d = D_0$. If the null hypothesis is rejected, there must be a significant difference (effect) between two samples (pre and post outcomes).

F test for Equal Variances

T-tests assume that samples are randomly drawn from normally distributed populations with unknown parameters. In addition to these random sampling and normality assumptions, you should check the equal variance assumption when examining the mean difference of two independent samples. The population variances of the two groups σ_1^2 and σ_2^2 need to be equal in order to use the pooled variance.¹² Otherwise, the t-test is not reliable due to the incorrect variance and degrees of freedom used.

In practice, unequal variances of two independent samples are less problematic when two samples have the same number of observations (balanced data) (Hildebrand et al. 2005: 362). The problem will be critical if one sample has a larger variance and a much smaller sample size compared to the other (362).

The folded form F-test is commonly used to examine whether two populations have the same variance, $H_0 : \sigma_1^2 = \sigma_2^2$. The F statistic is

$$\frac{s_L^2}{s_S^2} \sim F(n_L - 1, n_S - 1)$$

where L and S respectively indicate groups with larger and smaller sample variances.

Βιβλιογραφία

- Τσαγρής Μιχαήλ, 2011. “Στατιστική με τη χρήση του πακέτου IBM SPSS 19”. Αθήνα και Nottingham 2011.
- Park, Hun Myoung, 2009. *Comparing Groups Means: T-tests and One-way ANOVA Using STATA, SAS, R, and SPSS. Working Paper*. The University Information Technology Services (UITs) Center for Statistical and Mathematical Computing, Indiana University.”
<http://www.indiana.edu/~statmath/stat/all/ttest>