

ΒΙΟΣΤΑΤΙΣΤΙΚΗ (ΕΡΓΑΣΤΗΡΙΟ)

με χρήση του Στατιστικού Πακέτου SPSS v19

ΓΕΝΙΚΟ ΤΜΗΜΑ ΜΑΘΗΜΑΤΙΚΩΝ

Τουβιάς Θωράς ΕΠΙΣΤ. ΣΥΝΕΡΓ.
Παπαγεωργίου Έφη ΕΠΙΚ. ΚΑΘ.

ΚΕΦΑΛΑΙΟ 3 Έλεγχοι Στατιστικών Υποθέσεων





Εφαρμογή 3.1. Σε τέσσερα άτομα με αυξημένες τιμές των τριγλυκεριδίων του ορού (mg/dl) χορηγήθηκε για ένα μήνα φάρμακο που πιστεύεται ότι ελαττώνει τα επίπεδα των τριγλυκεριδίων. Οι τιμές των τριγλυκεριδίων στα τέσσερα αυτά άτομα πριν και μετά τη χορήγηση του φαρμάκου ήταν:

<u>Πριν τη χρήση</u>	<u>Μετά τη χρήση</u>
180	120
200	220
240	130
230	160

- Βρείτε ένα **95%** δ.ε. για την διαφορά των μέσων στα επίπεδα των τριγλυκεριδίων πριν και μετά την χορήγηση.
- Ελαττώνει τα επίπεδα των τριγλυκεριδίων το φάρμακο αυτό;
- Υπάρχει συσχέτιση των τιμών των τριγλυκεριδίων πριν και μετά τη χορήγηση του φαρμάκου;

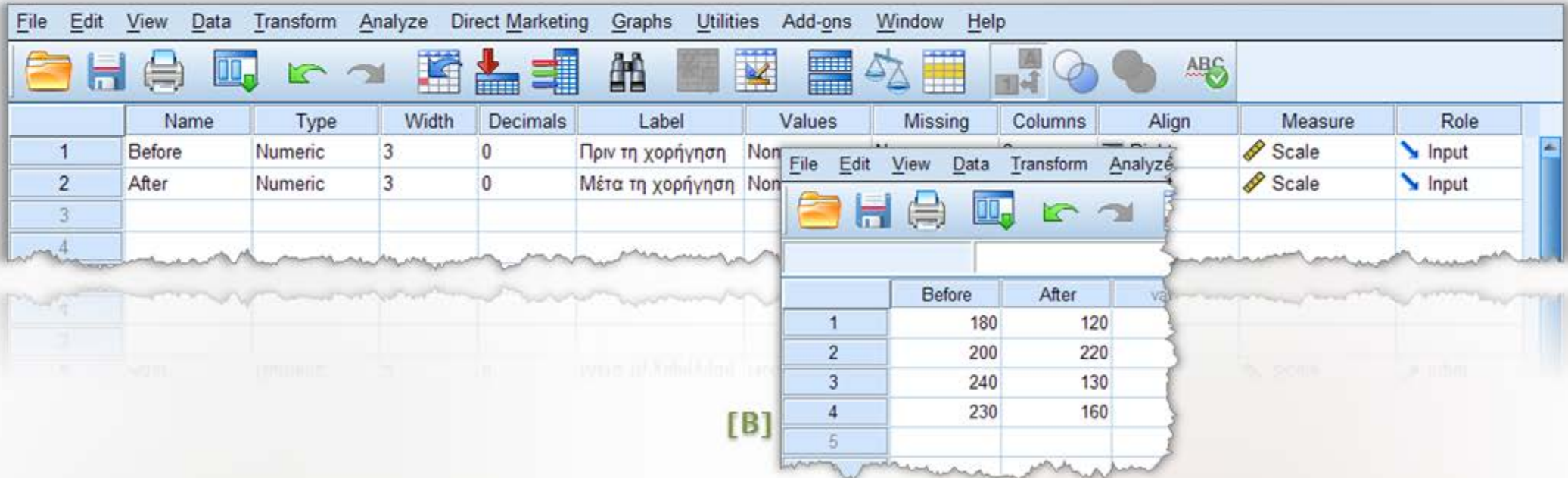
Έχουμε τον εξής έλεγχο υποθέσεων: $H_0: \mu_1 = \mu_2$ και $H_1: \mu_1 > \mu_2$ με κριτήριο

$$R = \left\{ \frac{\bar{z}}{s_z} \sqrt{n} > t_{n-1; \alpha} \right\},$$

όπου $z_i = x_i - y_i$, $i = 1, 2, 3, 4$, $\alpha = 0,05$, οπότε εκτελούμε t -έλεγχο για ζευγαρωτές παρατηρήσεις.

Μεταβαίνουμε αρχικά σε κατάσταση **Variable View** και δημιουργούμε δύο ποσοτικές (**Scale**) αριθμητικές (**Numeric**) μεταβλητές με ονόματα **Before** και **After**, και ετικέτες **Πριν τη χορήγηση** και **Μετά τη χορήγηση** αντίστοιχα [A].

Στη συνέχεια, σε κατάσταση **Data View**, εισάγουμε τις παρατηρήσεις του πίνακα της εφαρμογής [B].



[A]

Name	Type	Width	Decimals	Label	Measure	Role
1 Before	Numeric	3	0	Πριν τη χορήγηση	Scale	Input
2 After	Numeric	3	0	Μετά τη χορήγηση	Scale	Input
3						
4						

[B]

	Before	After
1	180	120
2	200	220
3	240	130
4	230	160
5		

Από το menu **Analyze** επιλέγουμε **Compare Means** > **Paired-Samples T Test**^[A]. Στο εμφανιζόμενο παράθυρο διαλόγου μεταφέρουμε από τη λίστα (αριστερά) τις δύο μεταβλητές μας^[B] (με το βέλος απόθεσης) στη λίστα **Paired Variable(s)**^[Γ]. Με χρήση του **Options** μπορούμε να επιλέξουμε άλλο διάστημα εμπιστοσύνης εκτός του προεπιλεγμένου **95%**^[Δ]. Επικυρώνουμε τις επιλογές μας με **Continue** και τέλος με **OK**.

The image shows a sequence of three screenshots from the SPSS software interface, illustrating the steps to perform a Paired-Samples T Test. Yellow arrows indicate the flow of the process.

- [A]** The first screenshot shows the **Analyze** menu open, with **Compare Means** > **Paired-Samples T Test...** selected.
- [B]** The second screenshot shows the **Paired-Samples T Test** dialog box. Two variables, "Πριν τη χορήγηση ..." and "Μετά τη χορήγηση (After)", are being moved from the left list to the **Paired Variables** list on the right.
- [Γ]** The third screenshot shows the **Paired-Samples T Test** dialog box with the two variables in the **Paired Variables** list. The **Options** button is highlighted.
- [Δ]** The fourth screenshot shows the **Paired-Samples T Test - Options** dialog box. The **Confidence Interval Percentage** is set to **95** %.

Εμφανίζεται λοιπόν στον **Viewer** ο πίνακας του ελέγχου των ζευγαρωτών παρατηρήσεων **Paired Samples Test**, όπου παρατηρούμε ότι ένα **95%** δ.ε. για τη διαφορά των πληθυσμιακών μέσων των τριγλυκεριδίων (πριν-μείον-μετά τη φαρμακευτική αγωγή) είναι το **(-31,669, 141,669)**. Π.χ. εάν θεωρηθεί ότι $\mu_1 = \bar{x} = 212,5$, τότε για τη (πληθυσμιακή) μέση τιμή μ_2 των τριγλυκεριδίων μετά τη φαρμακευτική αγωγή, θα έχουμε (στο **95%** των περιπτώσεων) $-31,669 \leq 212,5 - \mu_2 \leq 141,669$, δηλ. **$70,84 \leq \mu_2 \leq 244,169$** .

Επίσης, ο έλεγχος έχει επίπεδο σημαντικότητας **Sig. (2-tailed)** ίσο με **0,137 (> $\alpha = 0,05$)**. Άρα δεχόμαστε την υπόθεση **H₀**, δηλ. το φάρμακο δεν ελαττώνει, γενικά, τα επίπεδα των τριγλυκεριδίων.

Τέλος, ο πίνακας συσχέτισης των ζευγαρωτών παρατηρήσεων **Paired Samples Correlations**, δίνει συντελεστή συσχέτισης **Correlation** ίσο με **-0,074 (< 0,5)**, δηλ. δεν υπάρχει καμία συσχέτιση μεταξύ των μετρήσεων πριν και μετά τη χορήγηση του φαρμάκου.

Paired Samples Statistics					
		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Πριν τη χορήγηση	212,50	4	27,538	13,769
	Μετά τη χορήγηση	157,50	4	45,000	22,500

Paired Samples Correlations				
		N	Correlation	Sig.
Pair 1	Πριν τη χορήγηση & Μετά τη χορήγηση	4	-,074	,926

Paired Samples Test									
		Paired Differences				t	df	Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower				Upper
Pair 1	Πριν τη χορήγηση - Μετά τη χορήγηση	55,000	54,467	27,234	-31,669	141,669	2,020	3	,137



Εφαρμογή 3.2. Στην Εφαρμογή 3.1 εάν θεωρηθεί ότι η φυσιολογική μέση τιμή των τριγλυκεριδίων είναι **200 mg/dl**, τότε:

- Βρείτε ένα **95%** δ.ε. για την διαφορά του φυσιολογικού μέσου από τους μέσους των επίπεδων των τριγλυκεριδίων τόσο πριν όσο και μετά τη φαρμακευτική αγωγή. Που πρέπει να κυμαίνονται οι παρατηρήσεις για να θεωρούνται (κατά **95%**) φυσιολογικές;
- Παραμένουν φυσιολογικά τα επίπεδα των τριγλυκεριδίων πριν και μετά τη φαρμακευτική αγωγή;

Έχουμε τους εξής ελέγχους υποθέσεων: $H_0: \mu_i = \mu_0$ και $H_0: \mu_i \neq \mu_0$, $i = 1, 2$, όπου μ_1 και μ_2 οι πληθυσμιακοί μέσοι (των τιμών των τριγλυκεριδίων) πριν και μετά τη φαρμακευτική αγωγή αντίστοιχα, ενώ $\mu_0 = 200$ η τιμή ελέγχου, δηλ. ο θεωρούμενος ως φυσιολογικός πληθυσμιακός μέσος. Επομένως, εκτελούμε δύο μονοδειγματικούς t -ελέγχους για τις διαφορές $\mu_1 - \mu_0$ και $\mu_2 - \mu_0$.

Από το menu **Analyze** επιλέγουμε **Compare Means** > **One-Samples T Test**^[A]. Στο εμφανιζόμενο παράθυρο διαλόγου μεταφέρουμε από τη λίστα (αριστερά) τις δύο μεταβλητές μας (με το βέλος απόθεσης) στη λίστα **Test Variable(s)** και εισάγουμε την τιμή ελέγχου **200** στη θέση **Test Value**^[B]. Με χρήση του **Options** μπορούμε να επιλέξουμε άλλο διάστημα εμπιστοσύνης εκτός του προεπιλεγμένου **95%**^[Γ]. Επικυρώνουμε τις επιλογές μας με **Continue** και τέλος με **OK**.

The image shows a sequence of three screenshots from the SPSS software interface, illustrating the steps to perform a one-sample t-test:

- [A]** The **Analyze** menu is open, and the path **Compare Means** > **One-Sample T Test...** is highlighted.
- [B]** The **One-Sample T Test** dialog box is shown. The **Test Variable(s)** list contains two variables: "Πριν τη χορήγηση ..." and "Μετά τη χορήγηση ...". The **Test Value** is set to **200**.
- [Γ]** The **One-Sample T Test: O...** sub-dialog box is shown. The **Confidence Interval Percentage** is set to **95**%. Under **Missing Values**, the option **Exclude cases analysis by analysis** is selected.

Εμφανίζεται λοιπόν στον **Viewer** ο πίνακας του μονοδειγματικού *t*-ελέγχου **One-Sample Test**, όπου παρατηρούμε ότι τα επίπεδα σημαντικότητας **Sig. (2-tailed)** των *t*-ελέγχων $\mu_1 - \mu_0$ και $\mu_2 - \mu_0$ ισούνται με **0,431** ($> \alpha = 0,05$) και με **0,155** ($> \alpha = 0,05$) αντίστοιχα, οπότε και δεχόμαστε τις αρχικές υποθέσεις H_0 , δηλ. ότι τα επίπεδα των τριγλυκεριδίων πριν και μετά τη φαρμακευτική αγωγή θεωρούνται φυσιολογικά (δηλ. ίσα με την τιμή ελέγχου $\mu_0 = 200$) σε επίπεδα σημαντικότητας **0,431** και **0,155** αντίστοιχα.

Ένα **95%** δ.ε. για τη διάφορα των $\mu_1 - \mu_0$ και $\mu_2 - \mu_0$, είναι τα **(-31,32, 56,32)** και **(-114,11, 29,11)** αντίστοιχα. Άρα, $168,68 = \mu_0 - 31,32 \leq \mu_1 \leq \mu_0 + 56,32 = 256,32$ και $85,89 = \mu_0 - 114,11 \leq \mu_2 \leq \mu_0 + 29,11 = 229,11$.

Επομένως, τα φυσιολογικά επίπεδα τιμών των τριγλυκεριδίων, πριν τη χορήγηση του φαρμάκου, σε όλο των πληθυσμό (δηλ. σε ποσοστό **95%**), κυμαίνονται από **168,68** έως και **256,32**, ενώ μετά τη χορήγηση του φαρμάκου, κυμαίνονται από **85,89** έως και **229,11**. ▶

One-Sample Statistics				
	N	Mean	Std. Deviation	Std. Error Mean
Πριν τη χορήγηση	4	212,50	27,538	13,769
Μετά τη χορήγηση	4	157,50	45,000	22,500

One-Sample Test						
	Test Value = 200					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Πριν τη χορήγηση	,908	3	,431	12,500	-31,32	56,32
Μετά τη χορήγηση	-1,889	3	,155	-42,500	-114,11	29,11



► Το γεγονός ότι τα παραπάνω διαστήματα εμπιστοσύνης αποτελούν και τα διαστήματα στα οποία κυμαίνονται οι φυσιολογικές τιμές (των τριγλυκεριδίων), τόσο πριν όσο και μετά την φαρμακευτική αγωγή, οφείλεται στο ότι προηγουμένως έχουμε κάνει δεκτές τις αρχικές υποθέσεις $H_0: \mu_1 = \mu_0$ και $H_0: \mu_2 = \mu_0$, δηλ. οι (πληθυσμιακοί) μέσοι μ_1 και μ_2 θεωρούνται πια ως φυσιολογικοί. Εάν λοιπόν επιλέγαμε ως μ_0 τιμές κοντά στα άκρα των διαστημάτων αυτών, τα επίπεδα σημαντικότητας των νέων μονοδειγματικών t -ελέγχων θα ήταν οριακά μεγαλύτερα (ή ίσα) με $\alpha = 0,05$, και επομένως θα δεχόμασταν πάλι, έστω και οριακά, τις αρχικές μας υποθέσεις περί ισότητας των μέσων, δηλ. οι νέοι πληθυσμιακοί μέσοι θα ισούταν με τους νέες τιμές ελέγχου μ_0 . Άρα λοιπόν, οποιαδήποτε τιμή μέσα στα διαστήματα εμπιστοσύνης (που υπολογίσαμε πριν) χαρακτηρίζεται από την ισότητά της με την δεδομένη $\mu_0 = 200$, δηλ. χαρακτηρίζεται ως “φυσιολογική”.

Τέλος, το γεγονός ότι τα επίπεδα των τριγλυκεριδίων του πληθυσμού δεν φαίνεται να επηρεάζονται σημαντικά από την φαρμακευτική αγωγή (όπως δείχθηκε και στην **Εφαρμογή 3.1**) και να παραμένουν “φυσιολογικά”, ενισχύεται και από το γεγονός ότι το επίπεδο σημαντικότητας **Sig. (2-tailed)**, για τη διαφορά του πληθυσμιακού μέσου μ_2 από τον φυσιολογικό μέσο μ_0 , είναι ίσο με **0,155 ($> \alpha = 0,05$)**, δηλ. και πάλι δεν φαίνεται το φάρμακο να μεταβάλλει σημαντικά τα επίπεδα των τριγλυκεριδίων από τα θεωρούμενα ως φυσιολογικά.



Εφαρμογή 3.3. Πέντε άτομα θερμομετρήθηκαν το πρωί και το βράδυ και βρέθηκαν να έχουν τις παρακάτω θερμοκρασίες ($^{\circ}\text{C}$):

<u>Πρωί</u>	<u>Βράδυ</u>
36,1	36,4
36,4	36,8
36,0	36,6
36,5	36,8
36,2	36,5

- Βρείτε ένα **95%** δ.ε. για την διαφορά των μέσων θερμοκρασιών πρωί και βράδυ.
- Μπορούμε να συμπεράνουμε ότι, γενικά, διαφέρει (αυξάνεται ή μειώνεται και κατά πόσο) η θερμοκρασία των ανθρώπων πρωί και βράδυ;
- Εάν θεωρήσουμε ως φυσιολογική θερμοκρασία τους **$36,5^{\circ}\text{C}$** , να υπολογισθούν τα διαστήματα που κυμαίνονται γενικά οι θερμοκρασίες, **1** στους **2** ανθρώπους, το πρωί και το βράδυ. Ποια τα διαστήματα που αναμένουμε να κυμαίνονται οι θεωρούμενες ως φυσιολογικές τιμές των θερμοκρασιών;

Εφαρμογή 3.4. Να απαντηθούν τα ερωτήματα της Εφαρμογής 3.1, θεωρώντας ότι ο πίνακας αυτής δίνει τις τιμές των τριγλυκεριδίων τυχαίων 4 ατόμων πριν και τυχαίων 4 ατόμων μετά τη χορήγηση ενός φαρμάκου για τη μείωση των τριγλυκεριδίων.

- Για τον έλεγχο των υποθέσεων: $H_0: \mu_1 = \mu_0$ και $H_1: \mu_1 \neq \mu_0$, εκτελούμε t -έλεγχο για ανεξάρτητα δείγματα.
- Για να πραγματοποιήσουμε t -έλεγχο για ανεξάρτητα δείγματα στο SPSS πρέπει οι παρατηρήσεις μας να είναι στοιχεία μιας ενιαίας μεταβλητής, δηλ. πρέπει να μπουν σε μια στήλη. Αντιγράφοντας (ή αποκόπτοντας) λοιπόν τα στοιχεία της μεταβλητής **After** με δεξί click και **Copy** (ή **Cut**)^[A] τα επικολλούμε με δεξί click και **Paste**^[B] στη μεταβλητή **Before** και με διαγράφουμε την **After** με δεξί click στο όνομά της και **Clear**^[Γ]. ►

The figure shows three sequential screenshots of the SPSS Data Editor window, illustrating the data manipulation process:

- [A]**: The 'After' column is selected. A right-click context menu is open, and the 'Copy' option is highlighted.
- [B]**: The 'Before' column is selected. A right-click context menu is open, and the 'Paste' option is highlighted.
- [Γ]**: The 'After' column is selected. A right-click context menu is open, and the 'Clear' option is highlighted.

Yellow arrows indicate the flow of data from the 'After' column in [A] to the 'Before' column in [B], and from the 'After' column in [B] to the 'Clear' option in [Γ].

- ▶ Στη συνέχεια, μεταβαίνοντας σε κατάσταση **Data View**, μετονομάζουμε τη μοναδική μας μεταβλητή σε **TG** και αλλάζουμε την ετικέτα σε **Τριγλυκερίδια**.

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	TG	Numeric	3	0	Τριγλυκερίδια	None	None	8	Right	Scale	Input
2											

- ▶ Δημιουργούμε μια νέα ποσοτική (**Scale**) και διατάξιμη (**Ordinal**) μεταβλητή με όνομα **Index** και ετικέτα **Τμηματικός Δείκτης**^[A]. Επίσης, επιλέγοντας **Values**, αντιστοιχούμε στην τιμή **1** την ετικέτα **Πριν** και στην τιμή **2** την ετικέτα **Μετά**^[B].
- ▶ Επιστρέφοντας σε κατάσταση **Data View**, εισάγουμε στα 4 πρώτα κελιά της **Index** την αριθμητική τιμή **1** και στα επόμενα 4 την τιμή **2**, οπότε αυτόματα εμφανίζονται στα κελιά οι ετικέτες τους^[Γ]. Οι τιμές αυτές αποτελούν και τους δείκτες που ομαδοποιούν την μεταβλητή **TG**.

[A]

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	TG	Numeric	3	0	Τριγλυκερίδια	None	None	8	Right	Scale	Input
2	Index	Numeric	1	0	Τμηματικός Δείκ...	None	None	8	Right	Ordinal	Input
3											

[B]

Value Labels

Value:

Label:

1 = "Πριν"

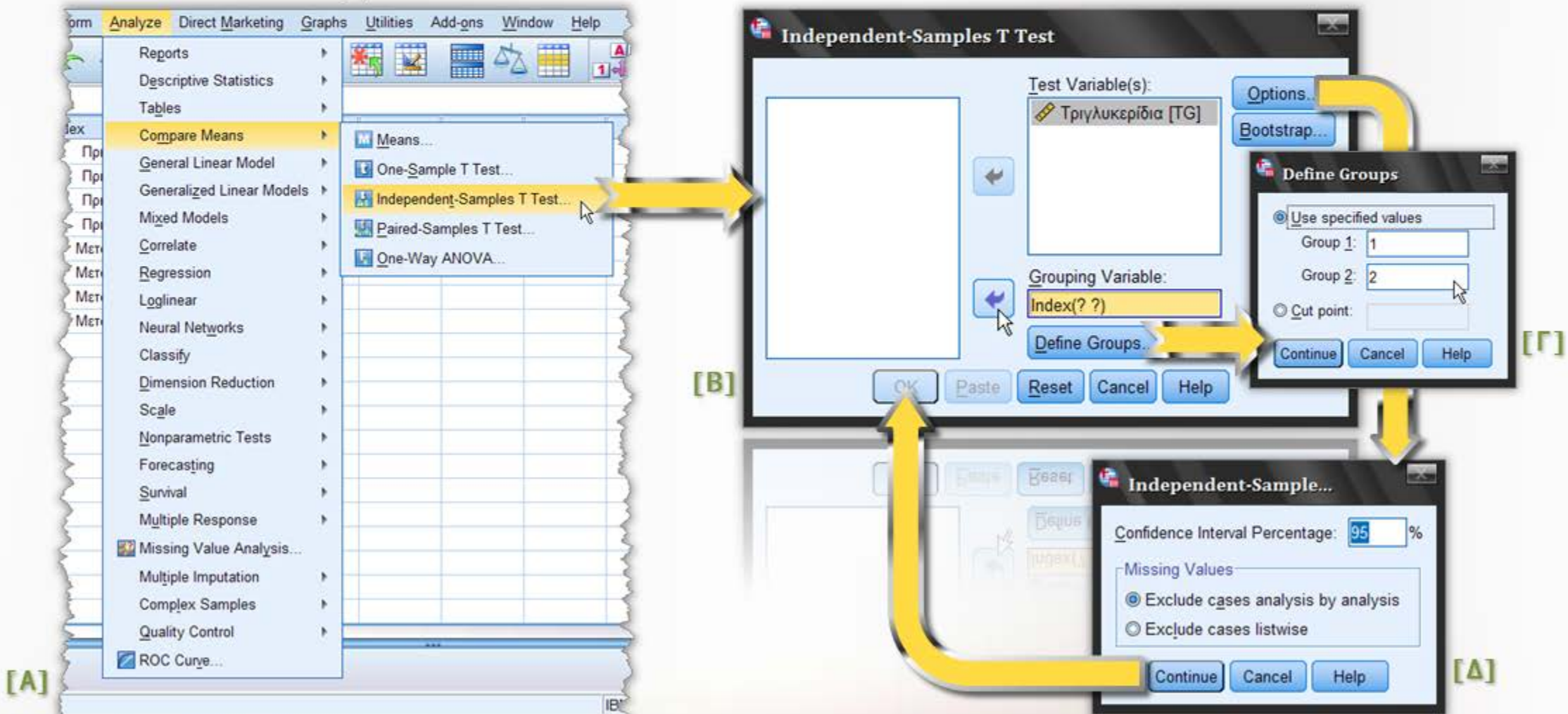
2 = "Μετά"

OK Cancel Help

[Γ]

	TG	Index
1	180	1
2	200	1
3	240	1
4	230	1
5	120	2
6	220	2
7	130	2
8	160	2

Από το menu **Analyze** επιλέγουμε **Compare Means** > **Independent-Samples T Test**^[A]. Στο εμφανιζόμενο παράθυρο διαλόγου μεταφέρουμε από τη λίστα (αριστερά) τη μεταβλητή **TG** (με το βέλος απόθεσης) στη λίστα **Test Variable(s)**, ενώ την **Index** στη λίστα **Grouping Variable**^[B]. Κατόπιν, με χρήση του **Define Groups** εισάγουμε τις τιμές **1** και **2** στις θέσεις **Group 1** και **Group 2** αντίστοιχα^[Γ]. Μπορούμε ακόμα, να επιλέξουμε άλλο διάστημα εμπιστοσύνης εκτός του προεπιλεγμένου **95%**^[Δ]. Επικυρώνουμε τις επιλογές μας με **Continue** και τέλος με **OK**.



Εμφανίζεται λοιπόν στον **Viewer** ο πίνακας του ελέγχου των ανεξάρτητων παρατηρήσεων **Independent Samples Test**, όπου παρατηρούμε ότι το επίπεδο σημαντικότητας **Sig.** του ελέγχου ισότητας των διακυμάνσεων (Levene's test) ισούται με **0,478 ($> \alpha = 0,05$)**. Επομένως, δεχόμαστε την ισότητα των διακυμάνσεων των δύο πληθυσμών (των τριγλυκεριδίων), δηλ. την ομοιογένεια των πληθυσμών. Άρα, δεχόμαστε (οριακά) τελικά την υπόθεση **H₀** καθώς το επίπεδο σημαντικότητας **Sig. (2-tailed)** του t-ελέγχου της ισότητας των μέσων (στην περίπτωση ισότητας των διακυμάνσεων) ισούται με **0,082 ($> \alpha = 0,05$)**. Συμπεραίνουμε λοιπόν ότι η φαρμακευτική αγωγή δεν φαίνεται να μεταβάλλει (οριακά) τα επίπεδα των τριγλυκεριδίων.

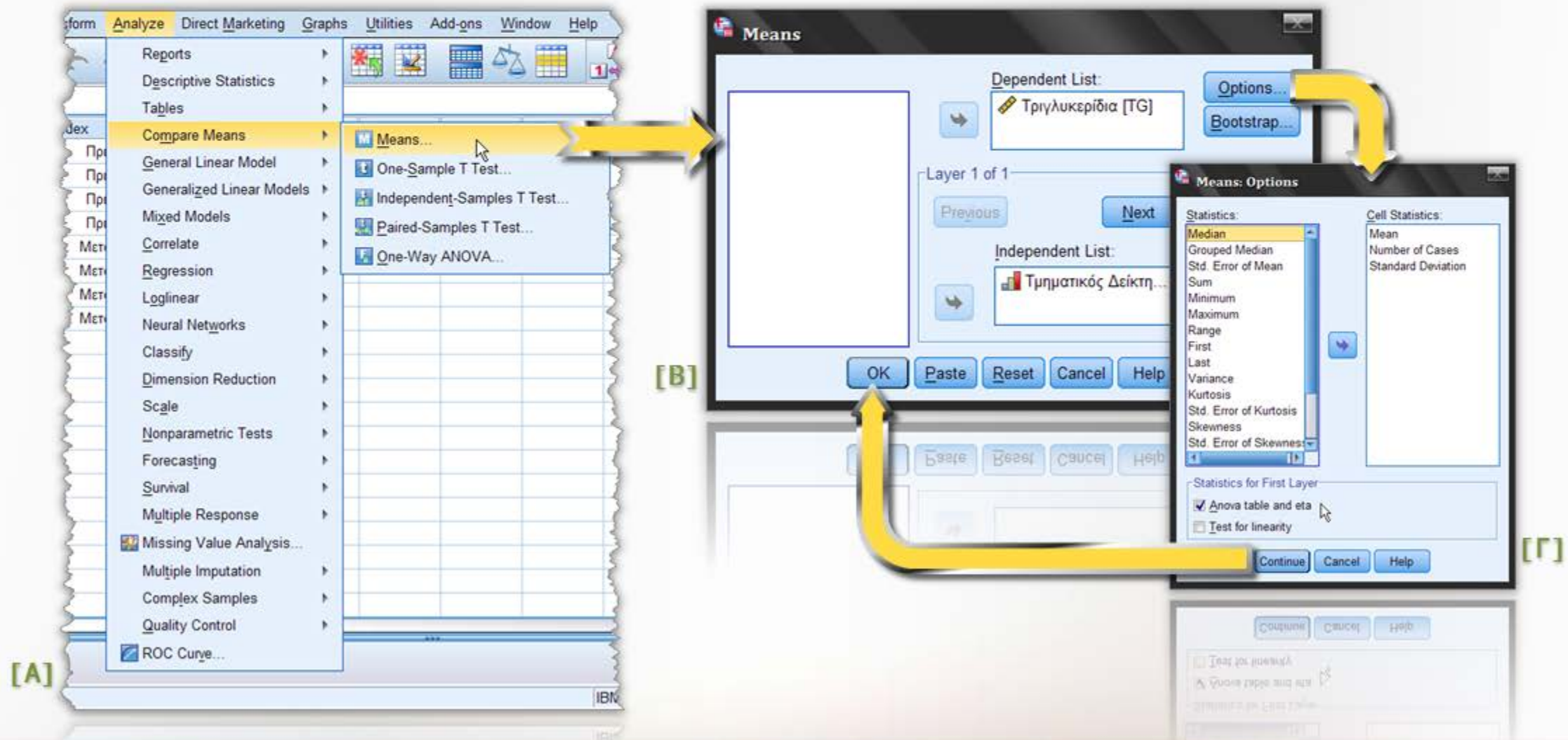
Επίσης, ένα **95%** δ.ε. για τη διάφορα των (πληθυσμιακών) μέσων των τιμών των τριγλυκεριδίων (πριν και μετά τη χορήγηση του φαρμάκου) είναι το **(-9.546, 119.546)** καθώς δεχθήκαμε προηγουμένως την ισότητα των διακυμάνσεων.

Group Statistics				
Τμηματικός Δείκτης(A)	N	Mean	Std. Deviation	Std. Error Mean
Τριγλυκερίδια Πριν	4	212,50	27,538	13,769
Μετά	4	157,50	45,000	22,500

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
Τριγλυκερίδια	Equal variances assumed	,571	,478	2,085	6	,082	55,000	26,379	-9,546	119,546
	Equal variances not assumed			2,085	4,971	,092	55,000	26,379	-12,929	122,929

IBM SPSS Statistics Processor is ready

Εναλλακτικά, από το menu **Analyze** επιλέγουμε **Compare Means** > **Means^[A]**. Στο εμφανιζόμενο παράθυρο διαλόγου μεταφέρουμε από τη λίστα (αριστερά) τη μεταβλητή **TG** (με το βέλος απόθεσης) στη λίστα **Dependent List**, ενώ την **Index** στη λίστα **Independent List^[B]**. Κατόπιν, με χρήση του **Options** μπορούμε να επιλέξουμε από τη λίστα **Statistics** διάφορα άλλα στατιστικά μέτρα ενώ τσεκάρουμε την ενέργεια **Anova table and eta^[Γ]**. Επικυρώνουμε τις επιλογές μας με **Continue** και τέλος με **OK**.

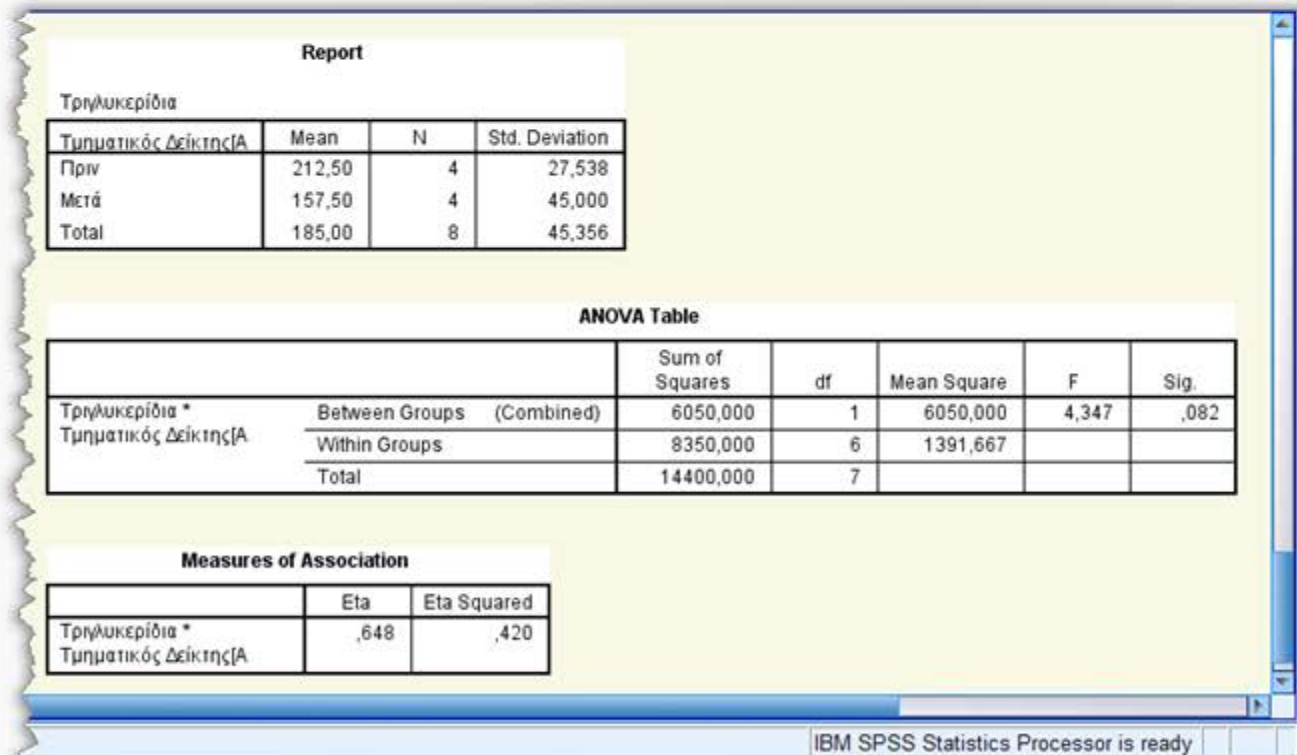


Εμφανίζεται έτσι στον **Viewer** ο πίνακας **Report**, όπου περιέχει τους μέσους των τριγλυκεριδίων πριν και μετά τη φαρμακευτική αγωγή.

Επίσης, λόγω του ότι το επίπεδο σημαντικότητας **Sig.** του πίνακα **ANOVA Table** είναι ίσο με **0,082 (> $\alpha = 0,05$)**, δεχόμαστε την υπόθεση (οριακά). Άρα το φάρμακο δεν μεταβάλλει (οριακά) τα επίπεδα των τριγλυκεριδίων.

Τέλος, επειδή η τιμή του **Eta Squared** του πίνακα **Measures of Association** είναι ίση με **0,42 (< 0,5)**, συμπεραίνουμε ότι δεν υπάρχει ιδιαίτερη σχέση των τιμών των τριγλυκεριδίων πριν και μετά τη φαρμακευτική αγωγή, καθώς αυτή επηρεάζει μόνο κατά **42%** τις τιμές των τριγλυκεριδίων του δείγματος.

ΣΗΜΕΙΩΣΗ: Ο τρόπος αυτός είναι προτιμότερος όταν υπάρχουν πάνω από τρεις ανεξάρτητες δειγματικές ομάδες, καθώς δίνει περισσότερα στοιχεία για τη γραμμικότητα των μέσων.



Report

Τριγλυκερίδια	Mean	N	Std. Deviation
Τμηματικός Δείκτης(A)			
Πριν	212,50	4	27,538
Μετά	157,50	4	45,000
Total	185,00	8	45,356

ANOVA Table

		Sum of Squares	df	Mean Square	F	Sig.
Τριγλυκερίδια * Τμηματικός Δείκτης(A)	Between Groups (Combined)	6050,000	1	6050,000	4,347	,082
	Within Groups	8350,000	6	1391,667		
	Total	14400,000	7			

Measures of Association

	Eta	Eta Squared
Τριγλυκερίδια * Τμηματικός Δείκτης(A)	,648	,420



Εφαρμογή 3.5. Για να μελετηθούν οι συνέπειες της έκθεσης σε κάδμιο, 10 σκύλοι εκτέθηκαν σε οξείδιο του καδμίου και άλλοι 10 χρησιμοποιήθηκαν ως μάρτυρες (ομάδα ελέγχου). Στο τέλος του πειράματος προσδιορίστηκαν οι τιμές της αιμοσφαιρίνης (g/dl) στους σκύλους και βρέθηκαν τα εξής αποτελέσματα:

<u>Εκτεθειμένοι</u>	<u>Μη Εκτεθειμένοι</u>
14,6	15,5
15,8	17,9
16,4	15,5
14,6	16,7
14,9	17,6
14,3	16,8
14,7	16,7
17,2	16,8
16,8	17,2
16,1	18,0

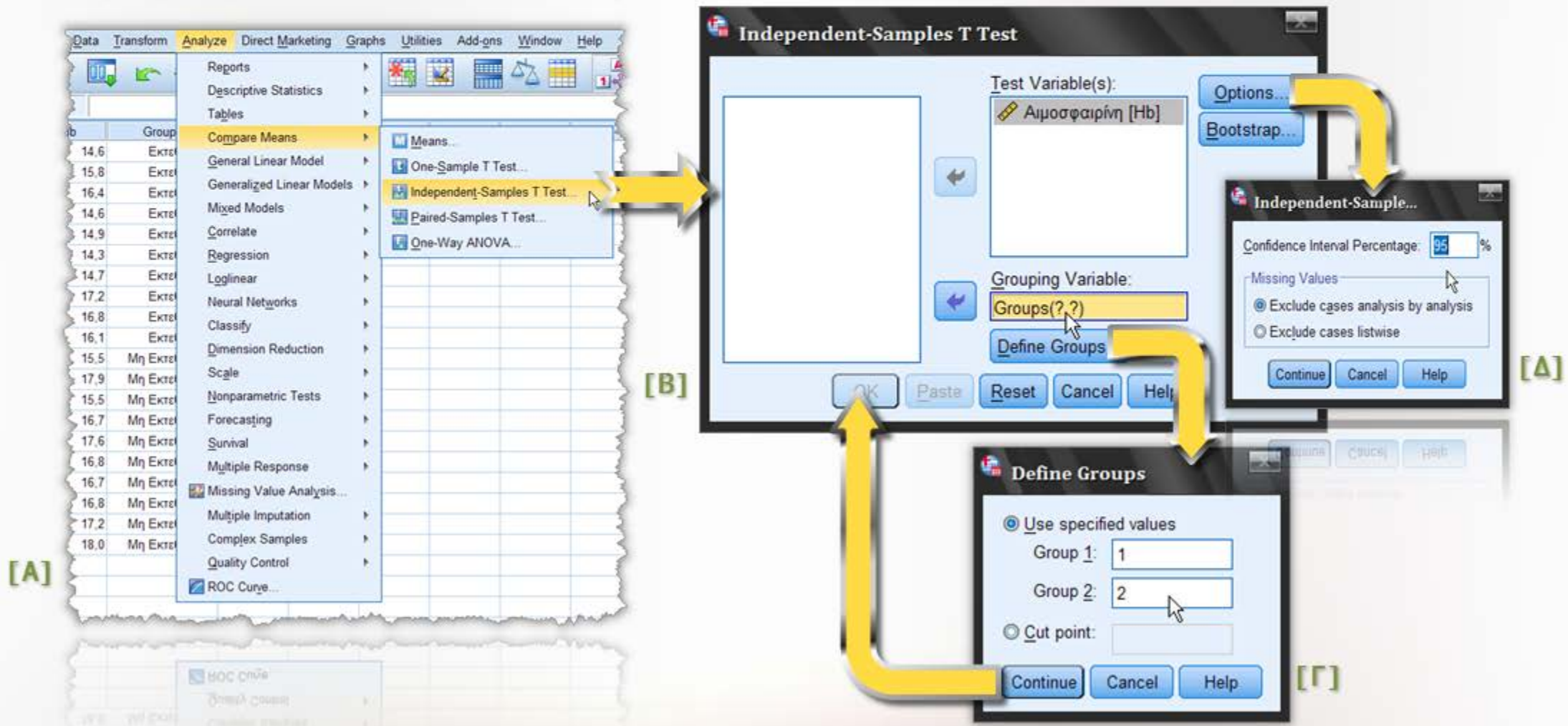
Να ελεγχθεί εάν, γενικά, επιδρά (και κατά πόσο) η συγκεκριμένη έκθεση σε κάδμιο στην αιμοσφαιρίνη των σκύλων.

- Μεταβαίνουμε αρχικά σε κατάσταση **Variable View** και δημιουργούμε μια ποσοτική (**Scale**) αριθμητική (**Numeric**) μεταβλητή με όνομα **Hb** και ετικέτα **Αιμοσφαιρίνη** καθώς και μια διατάξιμη (**Ordinal**) αριθμητική μεταβλητή με όνομα **Groups** και ετικέτα **Ομάδες**^[A]. Επιλέγοντας το κελί **Values** αυτής, αντιστοιχίζουμε στις αριθμητικές τιμές **1** και **2** τις ετικέτες **Εκτεθειμένοι** και **Μη Εκτεθειμένοι** αντίστοιχα^[B].
- Στη συνέχεια, σε κατάσταση **Data View**, εισάγουμε όλες τις παρατηρήσεις του πίνακα της εφαρμογής στα κελιά της μεταβλητής **Hb** (πρώτα τις τιμές της στήλης **Εκτεθειμένοι** και μετά τις τιμές της στήλης **Μη Εκτεθειμένοι**). Έπειτα, στα 10 πρώτα κελιά της μεταβλητής **Groups** εισάγουμε την τιμή **1**, και στα επόμενα 10 την τιμή **2**^[B].

The screenshot illustrates the SPSS workflow for creating and populating variables. It shows the Variable View window with two variables: 'Hb' (Scale, Numeric, 2 width, 1 decimal) and 'Groups' (Ordinal, Numeric, 1 width, 0 decimal). The 'Value Labels' dialog box is open for the 'Groups' variable, showing the mapping: 1 = "Εκτεθειμένοι" and 2 = "Μη Εκτεθειμένοι". A yellow arrow points from the 'Value Labels' dialog to the 'Values' cell in the Variable View window. Below, the Data View window shows a dataset with 12 rows and two columns: 'HS' and 'HS_Group'. The 'HS_Group' column contains the values 'Εκτεθειμένοι' for the first 10 rows and 'Μη Εκτεθειμένοι' for the last 2 rows.

	HS	HS_Group
1	14,6	Εκτεθειμένοι
2	15,8	Εκτεθειμένοι
3	16,4	Εκτεθειμένοι
4	14,6	Εκτεθειμένοι
5	14,9	Εκτεθειμένοι
6	14,3	Εκτεθειμένοι
7	14,7	Εκτεθειμένοι
8	17,2	Εκτεθειμένοι
9	16,8	Εκτεθειμένοι
10	16,1	Εκτεθειμένοι
11	15,5	Μη Εκτεθειμένοι
12	17,9	Μη Εκτεθειμένοι

Από το menu **Analyze** επιλέγουμε **Compare Means** > **Independent Samples T Test**^[A]. Στο εμφανιζόμενο παράθυρο διαλόγου τοποθετούμε (με το βέλος απόθεσης) στη λίστα **Test Variable(s)** τη μεταβλητή **Hb** και στη λίστα **Grouping Variable** τη μεταβλητή **Groups**^[B]. Κατόπιν, επιλέγοντας **Define Groups** εισάγουμε τις τιμές **1** και **2** στις θέσεις **Group 1** και **Group 2** αντίστοιχα^[Γ]. Με χρήση του **Options** μπορούμε να επιλέξουμε άλλο διάστημα εμπιστοσύνης εκτός του προεπιλεγμένου **95%**^[Δ]. Επικυρώνουμε τις επιλογές μας με **Continue** και τέλος με **OK**.



Εμφανίζεται λοιπόν στον **Viewer** ο πίνακας του ελέγχου των ανεξάρτητων παρατηρήσεων **Independent Samples Test**, όπου παρατηρούμε ότι το επίπεδο σημαντικότητας **Sig.** του ελέγχου ισότητας των διακυμάνσεων (Levene's test) ισούται με **0,211 ($> \alpha = 0,05$)**. Επομένως, δεχόμαστε την ισότητα των διακυμάνσεων των δύο πληθυσμών (των εκτεθειμένων και μη, σκύλων), δηλ. την ομοιογένεια των πληθυσμών. Απορρίπτουμε λοιπόν την αρχική υπόθεση (ισότητας των μέσων) καθώς το επίπεδο σημαντικότητας **Sig. (2-tailed)** του *t*-ελέγχου ισότητας των μέσων, στην περίπτωση ισότητας των διακυμάνσεων, ισούται με **0,006 ($< \alpha = 0,05$)**. Άρα, η έκθεση σε κάδμιο στους σκύλους όντως μεταβάλλει (σημαντικά) τα επίπεδα της αιμοσφαιρίνης (Hb) των σκύλων από αυτά της ομάδας ελέγχου. Επίσης, ένα **95%** δ.ε. για τη διάφορα των (πληθυσμιακών) μέσων των τιμών της Hb είναι το **(-2.2346, -0.4254)** καθώς δεχθήκαμε πριν την ισότητα των διακυμάνσεων, δηλ. το κάδμιο μειώνει την Hb των σκύλων, σε ποσοστό **95%**, από **-0.4254** έως και **-2.2346** gr/dl.

Group Statistics				
Ομάδες	N	Mean	Std. Deviation	Std. Error Mean
Αιμοσφαιρίνη				
Εκτεθειμένοι	10	15,540	1,0480	,3314
Μη Εκτεθειμένοι	10	16,870	,8693	,2749

Independent Samples Test										
		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference	
									Lower	Upper
Αιμοσφαιρίνη	Equal variances assumed	1,683	,211	-3,089	18	,006	-1,3300	,4306	-2,2346	-,4254
	Equal variances not assumed			-3,089	17,406	,007	-1,3300	,4306	-2,2368	-,4232

IBM SPSS Statistics Processor is ready



Εφαρμογή 3.6. Συνεχίζοντας την *Εφαρμογή 3.5*, εκθέτουμε άλλους 5 σκύλους στη μισή ποσότητα οξειδίου του καδμίου από ό,τι στην *Εφαρμογή 3.5*. Μετρώντας την αιμοσφαιρίνη αυτών έχουμε:

Μισο-εκτεθειμένοι

15,9

15,3

16,2

15,5

16,5

Να ελεγχθεί, γενικά, πως επιδρά το κάδμιο στην αιμοσφαιρίνη των σκύλων.

- Λόγο των τριών ανεξαρτήτων δειγμάτων και για χάριν πληρότητας, μπορούμε να κάνουμε τη λεγόμενη ανάλυση ANOVA.

- Μεταβαίνουμε αρχικά σε κατάσταση **Variable View** και δημιουργούμε μια ποσοτική (**Scale**) αριθμητική (**Numeric**) μεταβλητή με όνομα **Hb** και ετικέτα **Αιμοσφαιρίνη** καθώς και μια διατάξιμη (**Ordinal**) αριθμητική μεταβλητή με όνομα **Kd_Groups** και ετικέτα **Ομάδες Έκθεσης^[A]**. Επιλέγοντας το κελί **Values** αυτής, αντιστοιχίζουμε στις αριθμητικές τιμές **1, 2** και **3** τις ετικέτες **Εκτεθειμένος, Μισο-εκτεθειμένος** και **Μη Εκτεθειμένος** αντίστοιχα^[B].
- Στη συνέχεια, σε κατάσταση **Data View**, εισάγουμε όλες τις παρατηρήσεις του πίνακα της **Εφαρμογής 3.4** καθώς και τη νέα στήλη αυτής της εφαρμογής στα κελιά της μεταβλητής **Hb** (πρώτα τις τιμές της στήλης **Εκτεθειμένοι**, μετά της στήλης **Μισο-εκτεθειμένοι** και τέλος της στήλης **Μη Εκτεθειμένοι**). Έπειτα, στα 10 πρώτα κελιά της μεταβλητής **Kd_Groups** εισάγουμε την τιμή **1**, στα επόμενα 5 την τιμή **2**, και στα επόμενα 10 την τιμή **3^[B]**.

The screenshot shows the SPSS interface. On the left, the **Variable View** table is visible:

	Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure
1	Hb	Numeric	2	1	Αιμοσφαιρίνη	None	None	8	Right	Scale
2	Kd_Groups	Numeric	1	0	Ομάδες Έκθεσης	None	None	12	Right	Ordinal

A yellow arrow points from the **Values** cell for **Kd_Groups** to the **Value Labels** dialog box. The dialog box shows:

Value Labels

Value:

Label:

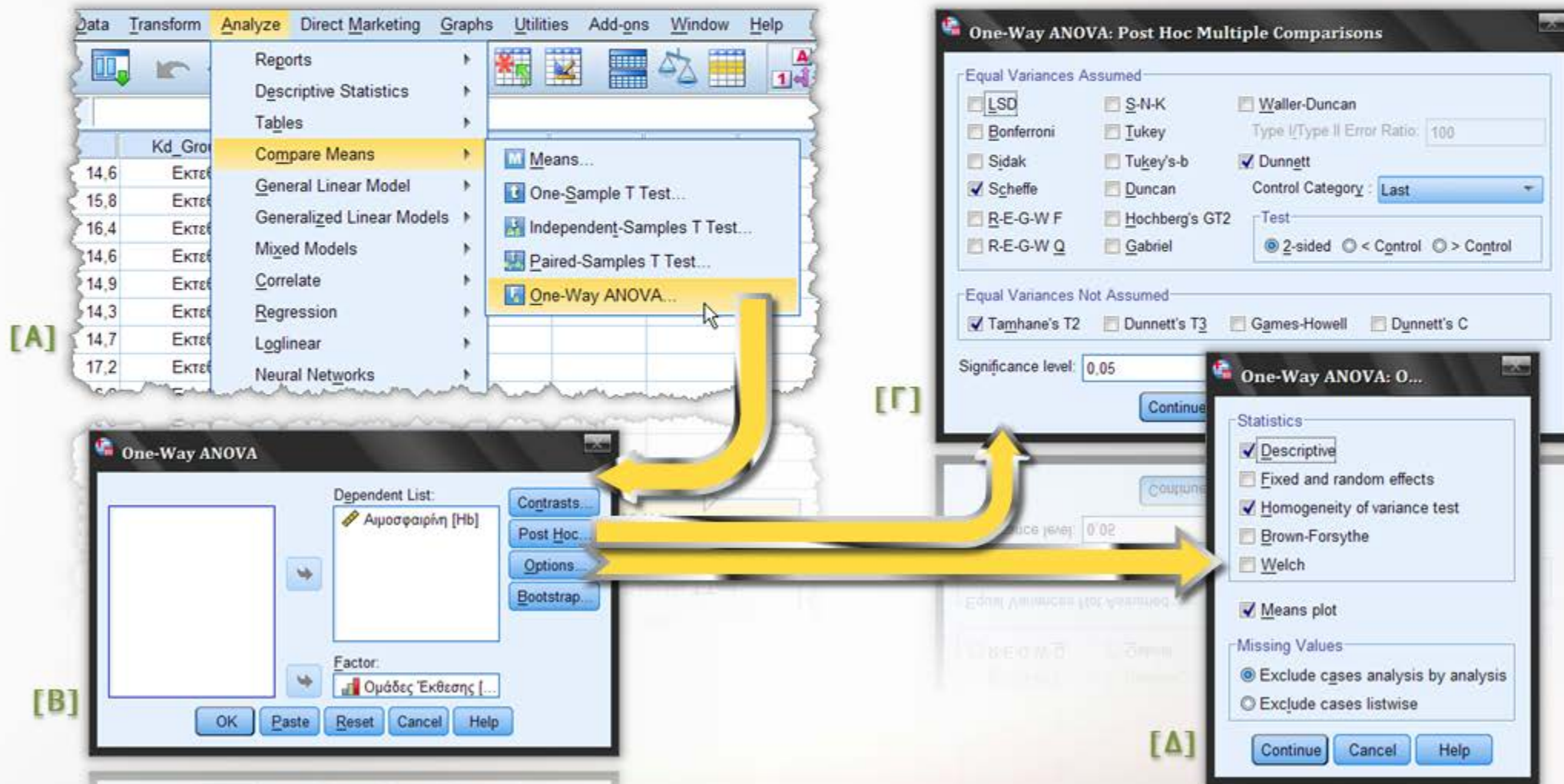
- 1 = "Εκτεθειμένος"
- 2 = "Μισο-εκτεθειμένος"
- 3 = "Μη Εκτεθειμένος"

Buttons: Add, Change, Remove, OK, Cancel, Help, Spelling.

On the right, the **Data View** table is shown:

	Hb	Kd_Groups
1	14.6	Εκτεθειμένος
2	15.8	Εκτεθειμένος
3	16.4	Εκτεθειμένος
4	14.6	Εκτεθειμένος
5	14.9	Εκτεθειμένος
6	14.3	Εκτεθειμένος
7	14.7	Εκτεθειμένος
8	17.2	Εκτεθειμένος
9	16.8	Εκτεθειμένος
10	16.1	Εκτεθειμένος
11	15.9	Μισο-εκτεθειμένος
12	15.3	Μισο-εκτεθειμένος
13	16.2	Μισο-εκτεθειμένος
14	15.5	Μισο-εκτεθειμένος
15	16.5	Μισο-εκτεθειμένος
16	15.5	Μη Εκτεθειμένος
17	17.9	Μη Εκτεθειμένος
18	15.5	Μη Εκτεθειμένος
19	16.7	Μη Εκτεθειμένος
20	17.6	Μη Εκτεθειμένος
21	16.8	Μη Εκτεθειμένος
22	16.7	Μη Εκτεθειμένος
23	16.8	Μη Εκτεθειμένος
24	17.2	Μη Εκτεθειμένος
25	18.0	Μη Εκτεθειμένος

Από το menu **Analyze** επιλέγουμε **Compare Means** > **One-Way ANOVA**^[A]. Στο εμφανιζόμενο παράθυρο διαλόγου τοποθετούμε (με το βέλος απόθεσης) στη λίστα **Test Variable(s)** τη μεταβλητή **Hb** και στη λίστα **Factor** τη μεταβλητή **Kd_Groups**^[B]. Κατόπιν, επιλέγοντας **Post-Hoc** τσεκάρουμε τους εμφανιζόμενους δείκτες^[Γ], ενώ επιλέγοντας **Options** τσεκάρουμε τις ενέργειες **Descriptive**, **Homogeneity of variance test**, καθώς και τη **Mean plot**^[Δ]. Επικυρώνουμε τις επιλογές μας με **Continue** και τέλος με **OK**.



The image shows a sequence of four screenshots illustrating the steps to perform a One-Way ANOVA in SPSS:

- [A]** The **Analyze** menu is open, and **Compare Means > One-Way ANOVA...** is selected.
- [B]** The **One-Way ANOVA** dialog box is shown. The dependent variable is **Αιμοσφαιρίνη [Hb]** and the factor is **Ομάδες Έκθεσης [...]**.
- [Γ]** The **One-Way ANOVA: Post Hoc Multiple Comparisons** dialog box is shown. Under **Equal Variances Assumed**, **Scheffe**, **Duncan**, and **Dunnett** are checked. Under **Equal Variances Not Assumed**, **Tamhane's T2** is checked. The significance level is set to 0.05.
- [Δ]** The **One-Way ANOVA: O...** dialog box is shown. Under **Statistics**, **Descriptive**, **Homogeneity of variance test**, and **Means plot** are checked. Under **Missing Values**, **Exclude cases analysis by analysis** is selected.

Εμφανίζεται λοιπόν στον **Viewer** οι πίνακες της ανάλυσης ANOVA. Στον πίνακα **Descriptives** μπορούμε να δούμε τα **95%** δ.ε. των μέσων (τιμών Hb) στις διάφορες ομάδες έκθεσης σε κάδμιο.

Στον πίνακα **Test of Homogeneity of Variances** παρατηρούμε ότι το επίπεδο σημαντικότητας **Sig.** του ελέγχου ομογένειας των διακυμάνσεων (Levene's-test) ισούται με **0,097 (> $\alpha = 0,05$)**, και επομένως, δεχόμαστε την ισότητα των διακυμάνσεων των τριών ομάδων των πληθυσμών, δηλ. την ομοιογένεια των τριών πληθυσμών.

Στον πίνακα **ANOVA** το επίπεδο σημαντικότητας **Sig.** της ανάλυσης ANOVA, ισούται με **0,01 (< $\alpha = 0,05$)**, οπότε απορρίπτουμε την αρχική υπόθεση ισότητας των (πληθυσμιακών) μέσων. Συμπεραίνουμε λοιπόν ότι η έκθεση σε κάδμιο στους σκύλους όντως μεταβάλλει τα επίπεδα της αιμοσφαιρίνης σε μια τουλάχιστον πληθυσμιακή τους ομάδα. ►

Αιμοσφαιρίνη	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Εκτεθειμένος	10	15,540	1,0480	,3314	14,790	16,290	14,3	17,2
Μισο-Εκτεθειμένος	5	15,880	,4919	,2200	15,269	16,491	15,3	16,5
Μη Εκτεθειμένος	10	16,870	,8693	,2749	16,248	17,492	15,5	18,0
Total	25	16,140	1,0591	,2118	15,703	16,577	14,3	18,0

Αιμοσφαιρίνη	Levene Statistic	df1	df2	Sig.
	2,593	2	22	,097

Αιμοσφαιρίνη	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	9,267	2	4,633	5,774	,010
Within Groups	17,653	22	,802		
Total	26,920	24			

► Στον πίνακα **Multiple Comparisons** παρατηρούμε τις αλληλεπιδράσεις των διαφόρων ομάδων έκθεσης σε κάδμιο. Π.χ. το επίπεδο σημαντικότητας **Sig. (Scheffe)** του δείκτη Scheffe (της διαφοράς των μέσων) ισούται με **0,789 ($> \alpha = 0,05$)** μεταξύ των **Εκτεθειμένων** και των **Μισο-εκτεθειμένων** σε κάδμιο, δηλ. δεχόμαστε την αρχική υπόθεση ισότητας των (πληθυσμιακών τους) μέσων, και επομένως η μισή ποσότητα καδμίου δεν επιδρά στο αίμα των σκύλων. Αντίθετα, το επίπεδο σημ. του ίδιου δείκτη μεταξύ των **Εκτεθειμένων** και των **Μη Εκτεθειμένων** σε κάδμιο, ισούται με **0,011 ($< \alpha = 0,05$)**, δηλ. απορρίπτουμε την αρχική υπόθεση ισότητας των (πληθυσμιακών) μέσων, και επομένως το καδμίου (στην πλήρη ποσότητα της **Εφαρμογής 3.4**) επιδρά στο αίμα των σκύλων.

Τα ίδια συμπεράσματα έχουμε και για τους δείκτες Tamhane και Dunnett. ►

Multiple Comparisons

Dependent Variable: Αιμοσφαιρίνη

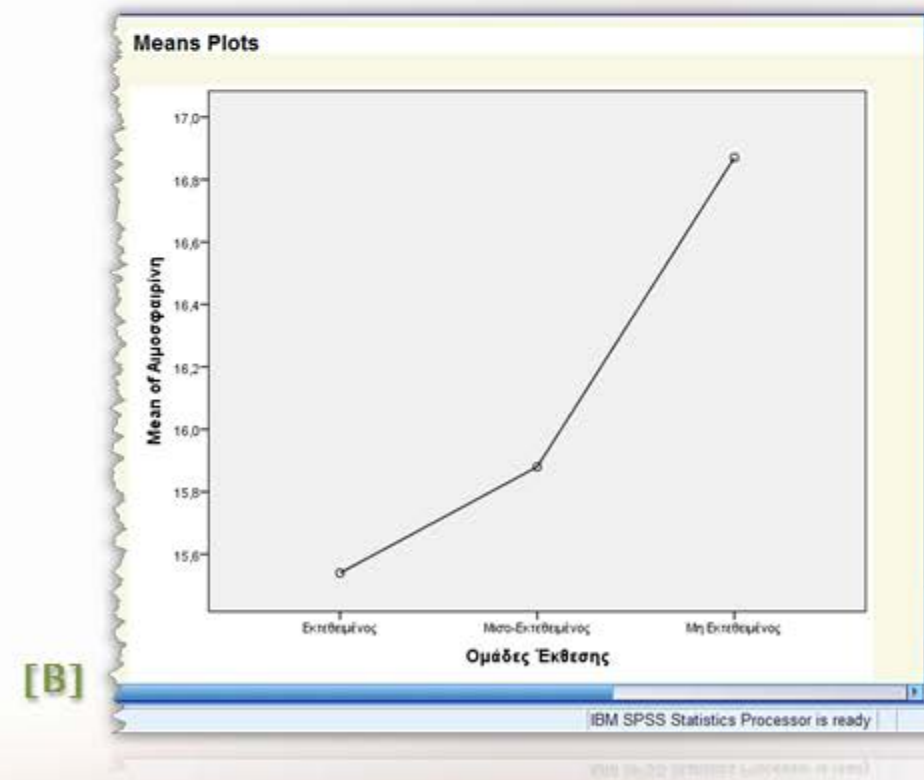
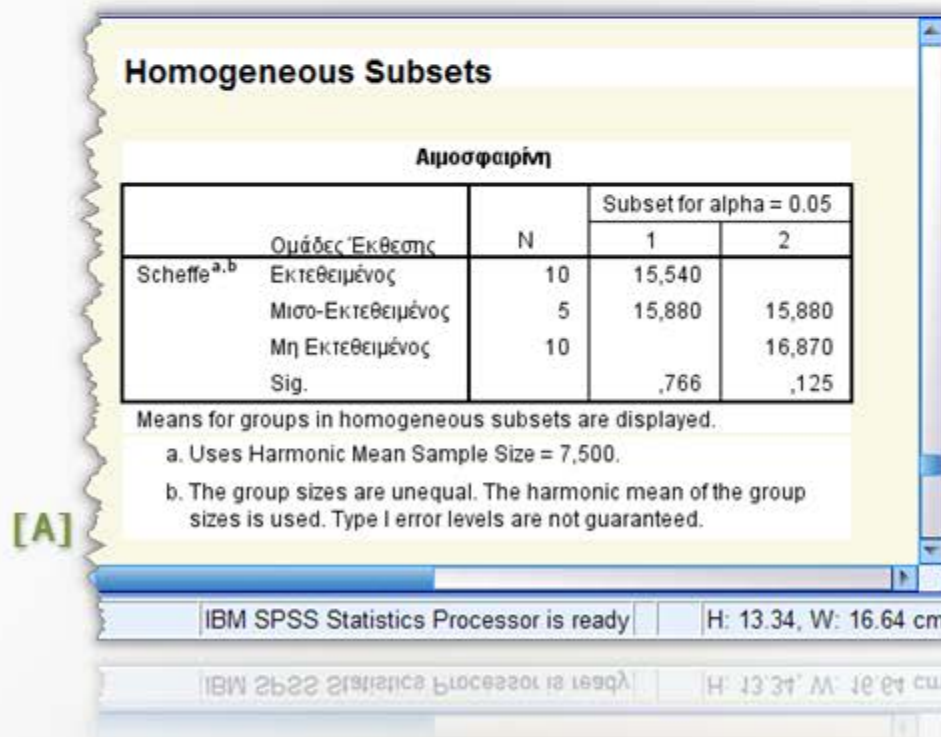
	(I) Ομάδες Έκθεσης	(J) Ομάδες Έκθεσης	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
						Lower Bound	Upper Bound
Scheffe	Εκτεθειμένος	Μισο-Εκτεθειμένος	-,3400	,4906	,789	-1,628	,948
		Μη Εκτεθειμένος	-1,3300*	,4006	,011	-2,381	-,279
	Μισο-Εκτεθειμένος	Εκτεθειμένος	,3400	,4906	,789	-,948	1,628
		Μη Εκτεθειμένος	-,9900	,4906	,154	-2,278	,298
Μη Εκτεθειμένος	Εκτεθειμένος	1,3300*	,4006	,011	,279	2,381	
	Μισο-Εκτεθειμένος	,9900	,4906	,154	-,298	2,278	
Tamhane	Εκτεθειμένος	Μισο-Εκτεθειμένος	-,3400	,3978	,793	-1,429	,749
		Μη Εκτεθειμένος	-1,3300*	,4306	,019	-2,467	-,193
	Μισο-Εκτεθειμένος	Εκτεθειμένος	,3400	,3978	,793	-,749	1,429
		Μη Εκτεθειμένος	-,9900*	,3521	,045	-1,958	-,022
Μη Εκτεθειμένος	Εκτεθειμένος	1,3300	,4306	,019	,193	2,467	
	Μισο-Εκτεθειμένος	,9900*	,3521	,045	,022	1,958	
Dunnnett t (2-sided) ^b	Εκτεθειμένος	Μη Εκτεθειμένος	-1,3300*	,4006	,006	-2,281	-,379
	Μισο-Εκτεθειμένος	Μη Εκτεθειμένος	-,9900	,4906	,102	-2,155	,175

*. The mean difference is significant at the 0.05 level.
b. Dunnnett t-tests treat one group as a control, and compare all other groups against it.

IBM SPSS Statistics Processor is ready

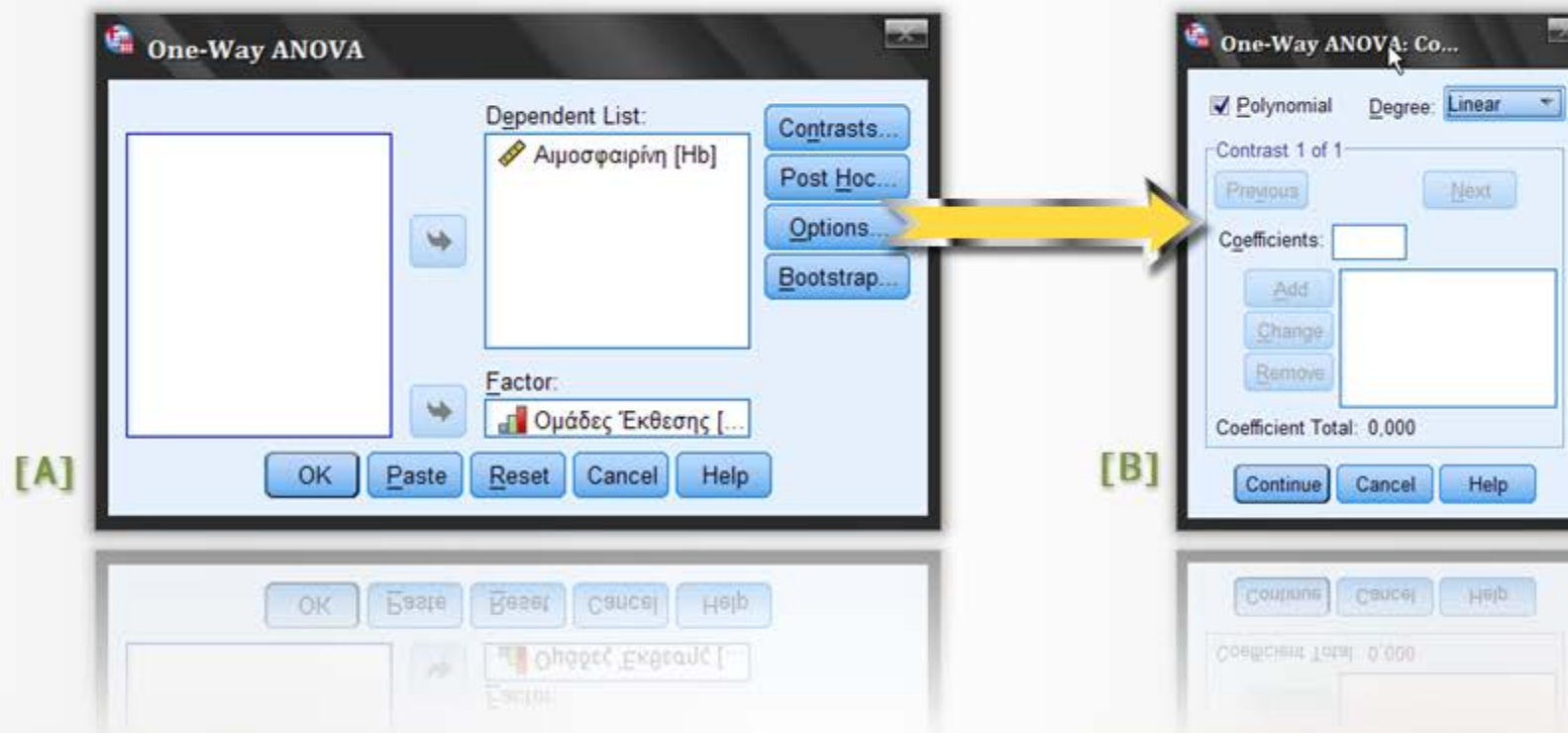
► Ο πίνακα **Homogeneous Subsets**^[A] μας δίνει τις ομογενείς πληθυσμιακές υποομάδες (έκθεσης σε κάδμιο), δηλ. τις ομάδες όπου γίνεται δεκτή η αρχική υπόθεση της ισότητας των (πληθυσμιακών) μέσων. Παρατηρούμε λοιπόν ότι οι δύο υποομάδες των **Εκτεθειμένων** και **Μισο-εκτεθειμένων** όπως και οι δύο υποομάδες των **Μισο-εκτεθειμένων** και **Μη Εκτεθειμένων** έχουν ίσους (πληθυσμιακούς) μέσους καθώς τα επίπεδα σημαντικότητας **Sig. (Scheffe)** του δείκτη Scheffe είναι **0,766 ($> \alpha = 0,05$)** και **0,125 ($> \alpha = 0,05$)** αντίστοιχα. Μεταξύ λοιπόν αυτών των ομάδων δεν υπάρχει σημαντική επίδραση ως αναφορά την έκθεση σε κάδμιο.

Τέλος, το γράφημα **Mean Plots**^[B] μας παρέχει μια εικόνα της μεταβολής των δειγματικών μέσων (των τιμών Hb) για τις τρεις ομάδες έκθεσης σε κάδμιο.





Επειδή το πλήθος των δειγμάτων των τριών ομάδων μας είναι διαφορετικό, είναι προτιμότερο, στο προηγούμενο παράθυρο διαλόγου **One-Way ANOVA**^[A], να επιλέξουμε **Contrasts** και να τσεκάρουμε την ενέργεια **Polynomial**^[B]. Με τον τρόπο αυτό παίρνουμε μια διεξοδικότερη ανάλυση στον πίνακα **ANOVA** του παραθύρου **Viewer**.



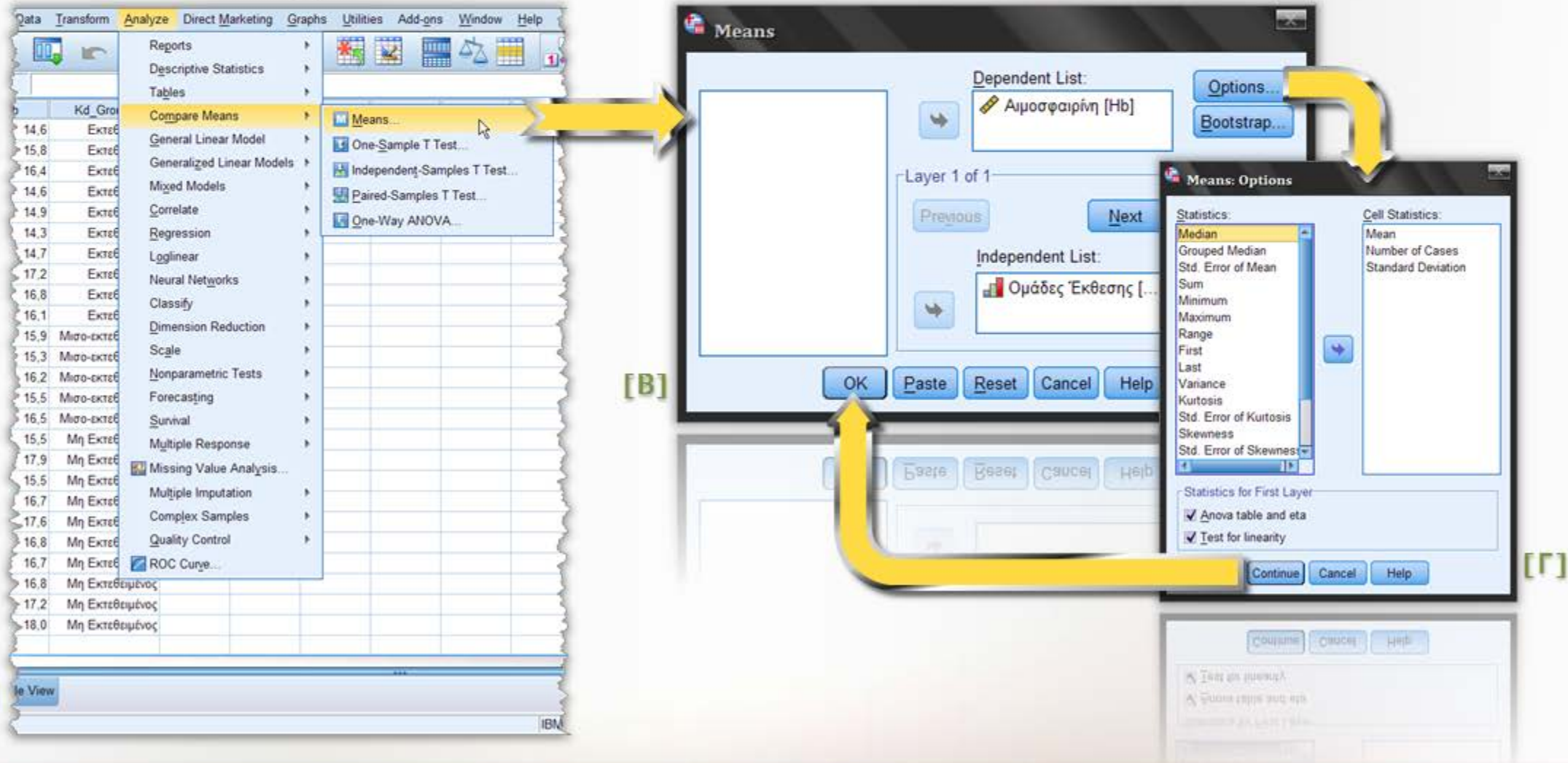
Στον πίνακα **ANOVA** λοιπόν παρατηρούμε ότι το επίπεδο σημαντικότητας **Sig. (Unweighted)** του ελέγχου των ισοσταθμισμένων μέσων (θεωρώντας ότι όλοι οι μέσοι των ομάδων έχουν το ίδιο βάρος) ισούται με **0,01** ($< \alpha = 0,05$), οπότε και απορρίπτουμε η αρχική υπόθεση περί της μη-γραμμικής σχέσης των (πληθυσμιακών) μέσων, και επομένως θεωρούμε ότι υπάρχει σημαντική γραμμική σχέση των μέσων. Στο ίδιο συμπέρασμα καταλήγουμε και στην περίπτωση των σταθμισμένων μέσων (**Weighted**).

Επίσης, το επίπεδο σημαντικότητας **Sig. (Deviation)** για τον έλεγχο της απόκλισης των μέσων από τη γραμμικότητα, ισούται με **0,476** ($> \alpha = 0,05$), οπότε δεχόμαστε την αρχική υπόθεση περί της μη-απόκλισης από τη γραμμικότητα των (πληθυσμιακών) μέσων, και επομένως, μέσω και του γραφήματος **Mean Plots**, θεωρούμε ότι υπάρχει γραμμική μείωση των τιμών Hb με την αυξανόμενη έκθεση των σκύλων σε κάδμιο.

Αιμοσφαιρίνη			Sum of Squares	df	Mean Square	F	Sig.
Between Groups	(Combined)		9,267	2	4,633	5,774	,010
	Linear Term	Unweighted	8,844	1	8,844	11,022	,003
		Weighted	8,844	1	8,844	11,022	,003
		Deviation	,422	1	,422	,527	,476
Within Groups			17,653	22	,802		
Total			26,920	24			

IBM SPSS Statistics Processor is ready | H: 13.34, W: 16.64 cm

Εναλλακτικά (και λιγότερο διεξοδικά), από το menu **Analyze** επιλέγουμε **Compare Means** > **Means**^[A]. Στο εμφανιζόμενο παράθυρο διαλόγου μεταφέρουμε από τη λίστα (αριστερά) τη μεταβλητή **Hb** (με το βέλος απόθεσης) στη λίστα **Dependent List**, ενώ την **Kd_Groups** στη λίστα **Independent List**^[B]. Κατόπιν, με χρήση του **Options** μπορούμε να επιλέξουμε από τη λίστα **Statistics** διάφορα άλλα δειγματικά μέτρα ενώ τσεκάρουμε τις ενέργειες **Anova table and eta** και **Test for linearity**^[Γ]. Επικυρώνουμε τις επιλογές μας με **Continue** και τέλος με **OK**.



Εμφανίζεται έτσι στον **Viewer** ο πίνακας **ANOVA Table**, όπου το επίπεδο σημαντικότητας **Sig.** του συνδυαστικού ελέγχου, μεταξύ των ομάδων (**Combined**), εμφανίζεται να ισούται με **0,01** ($< \alpha = 0,05$), οπότε απορρίπτουμε την αρχική υπόθεση ισότητας των (πληθυσμιακών) μέσων. Επομένως, το κάδμιο μεταβάλλει τα επίπεδα Hb μιας τουλάχιστον πληθυσμιακής ομάδας έκθεσης (εκ των τριών).

Ως αναφορά τα **Sig. (Linearity)** και **Sig. (Deviation from Linearity)** έχουμε τα ίδια συμπεράσματα όπως και στην προηγούμενη ανάλυση **One-Way ANOVA**.

Τέλος, επειδή η τιμή του **Eta Squared** του πίνακα **Measures of Association** είναι ίση με **0,344** ($< 0,5$), συμπεραίνουμε ότι δεν υπάρχει ιδιαίτερη σχέση των τιμών Hb και της έκθεσης σε κάδμιο, καθώς αυτή επιδρά μόνο κατά **34,4%** στις τιμές της Hb.

	Sum of Squares	df	Mean Square	F	Sig.
Αιμοσφαιρίνη * Ομάδες Έκθεσης	Between Groups (Combined)	2	4,633	5,774	,010
	Linearity	1	8,844	11,022	,003
	Deviation from Linearity	1	,422	,527	,476
Within Groups	17,653	22	,802		
Total	26,920	24			

	R	R Squared	Eta	Eta Squared
Αιμοσφαιρίνη * Ομάδες Έκθεσης	,573	,329	,587	,344

IBM SPSS Statistics Processor is ready



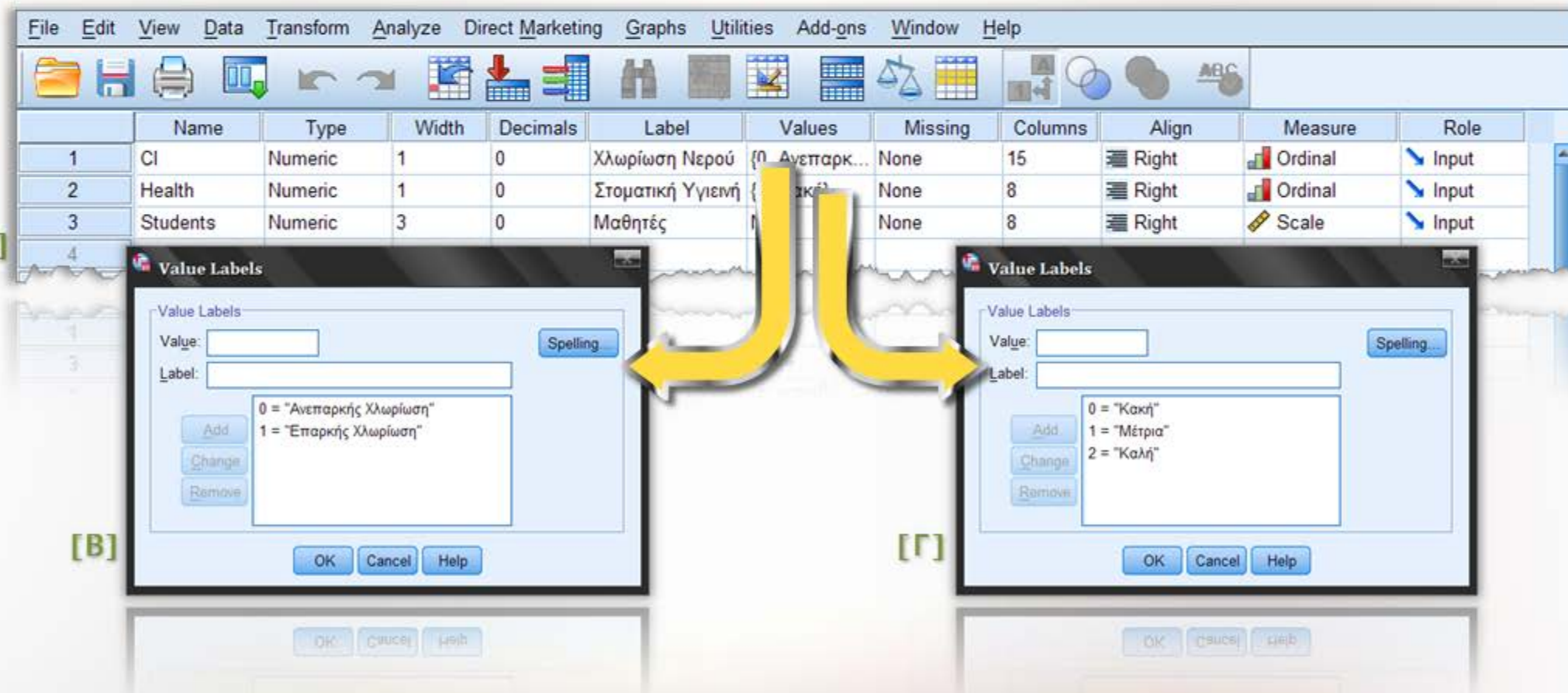
Εφαρμογή 3.6. Σε 500 μαθητές ενός δημοτικού σχολείου μελετήθηκε η σχέση της υγείας του στόματος με την χλωρίωση του νερού στην περιοχή διαμονής τους, οπότε εξάχθηκαν τα παρακάτω στοιχεία:

Χλωρίωση Νερού	Στοματική Υγιεινή		
	Κακή	Μέτρια	Κακή
Επαρκής	80	120	75
Ανεπαρκής	40	80	105

○ Σχετίζεται η στοματική υγιεινή των μαθητών με τη χλωρίωση του νερού;

▮ Εκτελούμε τον έλεγχο χ^2 για τις δεδομένες συχνότητες των μαθητών, όπως αυτές εμφανίζονται στον πίνακα της Εφαρμογής, όπου η αρχική υπόθεση αφορά την ανεξαρτησία (μη-συσχέτιση) της χλωρίωσης του νερού και της στοματικής υγιεινής των μαθητών.

Μεταβαίνουμε αρχικά σε κατάσταση **Variable View** και δημιουργούμε διατάξιμες (**Ordinal**) αριθμητικές (**Numeric**) μεταβλητές με ονόματα **CI** και **Health** και ετικέτες **Χλωρίωση Νερού** και **Στοματική Υγιεινή** αντίστοιχα, καθώς και μια ποσοτική (**Scale**) αριθμητική μεταβλητή με όνομα **Students** και ετικέτα **Μαθητές**^[A]. Επιλέγοντας το κελί **Values** της **CI** αντιστοιχίζουμε στις αριθμητικές τιμές **0** και **1** τις ετικέτες **Ανεπαρκής Χλωρίωση** και **Επαρκής Χλωρίωση**^[B], ενώ στο κελί **Values** της **Health** αντιστοιχίζουμε στις αριθμητικές τιμές **0**, **1** και **2** τις ετικέτες **Κακή**, **Μέτρια** και **Καλή** αντίστοιχα^[Γ].



The screenshot shows the SPSS Variable View window with three variables: CI, Health, and Students. The CI variable is set to Ordinal with values 0 and 1. The Health variable is set to Ordinal with values 0, 1, and 2. The Students variable is set to Scale. Two Value Labels dialog boxes are shown, one for CI and one for Health, with yellow arrows indicating the mapping of values to labels.

Name	Type	Width	Decimals	Label	Values	Missing	Columns	Align	Measure	Role
1	CI	Numeric	1	0	Χλωρίωση Νερού	None	15	Right	Ordinal	Input
2	Health	Numeric	1	0	Στοματική Υγιεινή	None	8	Right	Ordinal	Input
3	Students	Numeric	3	0	Μαθητές	None	8	Right	Scale	Input

[A] Value Labels dialog for CI:

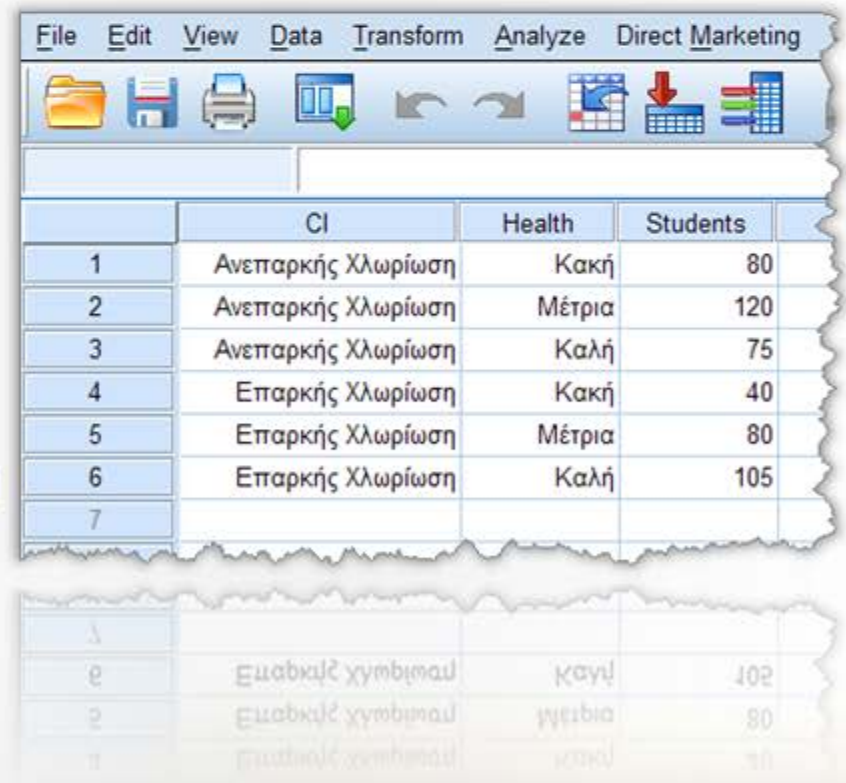
- Value: 0, Label: "Ανεπαρκής Χλωρίωση"
- Value: 1, Label: "Επαρκής Χλωρίωση"

[B] Value Labels dialog for Health:

- Value: 0, Label: "Κακή"
- Value: 1, Label: "Μέτρια"
- Value: 2, Label: "Καλή"

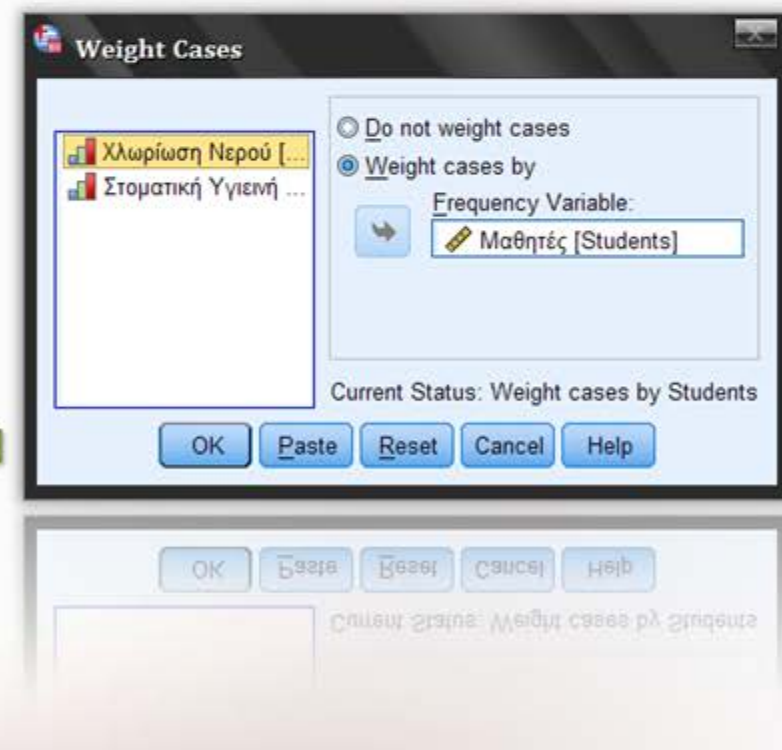
- Στη συνέχεια, σε κατάσταση **Data View**, εισάγουμε όλες τις παρατηρήσεις του πίνακα της στη μεταβλητή **Students**, εισάγοντας στα αντίστοιχα κελιά της μεταβλητής **CI** τις τιμές **0** (για επαρκή χλωρίωση) και **1** (για ανεπαρκή χλωρίωση) και στις **Health** τις τιμές **0** (για την κακή στομ. υγεία), **1** (για τη μέτρια στομ. υγεία) και **2** (για την καλή στομ. υγεία)^[A]. Αυτόματα εμφανίζονται οι ετικέτες των τιμών (όπως ορίστηκαν πριν).
- Κατόπιν, από το menu **Data** επιλέγουμε **Weight Cases**^[A], και στο εμφανιζόμενο παράθυρο διαλόγου τσεκάρουμε την ενέργεια **Weight cases by**, οπότε στη θέση **Frequency Variable** τοποθετούμε (με το βέλος απόθεσης) την μεταβλητή **Students**^[B]. Επικυρώνουμε τις επιλογές μας με **OK**.

[A]



	CI	Health	Students
1	Ανεπαρκής Χλωρίωση	Κακή	80
2	Ανεπαρκής Χλωρίωση	Μέτρια	120
3	Ανεπαρκής Χλωρίωση	Καλή	75
4	Επαρκής Χλωρίωση	Κακή	40
5	Επαρκής Χλωρίωση	Μέτρια	80
6	Επαρκής Χλωρίωση	Καλή	105
7			

[B]



Weight Cases

Do not weight cases
 Weight cases by
 Frequency Variable:

Current Status: Weight cases by Students

Από το menu **Analyze** επιλέγουμε **Descriptive Statistics** > **Crosstabs^[A]**. Στο εμφανιζόμενο παράθυρο διαλόγου μεταφέρουμε από τη λίστα των μεταβλητών (αριστερά) τις **CI** και **Health** (με το βέλος απόθεσης) στις λίστες **Row(s)** και **Column(s)** αντίστοιχα, και τσεκάρουμε την ενέργεια **Display clustered bar charts^[B]**. Επιλέγοντας **Statistics** τσεκάρουμε την ενέργεια **Chi-square^[Γ]**, ενώ επιλέγοντας **Cells** μπορούμε να τσεκάρουμε τις ενέργειες **Row**, **Column** ή/και **Total^[Δ]**. Επικυρώνουμε τις επιλογές μας με **Continue** και τέλος με **OK**.

The image shows a sequence of three screenshots from the SPSS software interface, illustrating the steps to configure Crosstabs analysis:

- [A]** The **Analyze** menu is open, and **Descriptive Statistics** > **Crosstabs...** is selected. A yellow arrow points from this menu item to the Crosstabs dialog box.
- [B]** The **Crosstabs** dialog box is shown. The **Row(s)** list contains "Χλωρίωση Νερού [CI]" and the **Column(s)** list contains "Στοματική Υγιεινή [Health]". The **Display clustered bar charts** checkbox is checked. A yellow arrow points from the **Statistics** button to the Crosstabs: Statistics dialog box.
- [C]** The **Crosstabs: Statistics** dialog box is shown. The **Chi-square** checkbox is checked. A yellow arrow points from the **Cells** button in the previous dialog to the Crosstabs: Cell Display dialog box.
- [D]** The **Crosstabs: Cell Display** dialog box is shown. The **Observed** checkbox is checked, and the **Row** checkbox is selected under the **Percentages** section. A yellow arrow points from the **Cells** button in the previous dialog to this dialog box.

Εμφανίζεται έτσι στον **Viewer** ο πίνακας συχνοτήτων των διασταυρωμένων παρατηρήσεων **Crosstabulation** καθώς και ο πίνακας **Chi-Square Tests** όπου λαμβάνουμε τα αποτελέσματα του ελέγχου χ^2 . Παρατηρούμε ότι το επίπεδο σημαντικότητας **Asymp. Sig. (2-sided)** του ελέγχου **Pearson Chi-Square** ισούται με **0** ($< \alpha = 0.05$), οπότε και απορρίπτουμε την αρχική υπόθεση της ανεξαρτησίας, δηλ. υπάρχει σημαντική (ευθεία) εξάρτηση της χλωρίωσης του νερού με τη στοματική υγεία των μαθητών και μάλιστα σε επίπ. σημαντικότητας **0**, κάτι που επιβεβαιώνεται και από τα ομαδοποιημένα ιστογράμματα **Bar Chart**, όπου φαίνεται η αισθητή βελτίωση της στοματικής υγείας με την επαρκή χλωρίωση του νερού.

