

ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

ΔΠΜΣ

«ΕΠΙΣΤΗΜΕΣ ΤΗΣ ΑΓΩΓΗΣ ΜΕΣΩ ΚΑΙΝΟΤΟΜΩΝ ΤΕΧΝΟΛΟΓΙΩΝ
ΚΑΙ ΒΙΟΙΑΤΡΙΚΩΝ ΠΡΟΣΕΓΓΙΣΕΩΝ»

«Μεθοδολογία Εκπαιδευτικής Έρευνας και
Στατιστική »

Σημειώσεις Εργαστηρίου SPSS

2^ο μέρος: «ΕΛΕΓΧΟΙ ΥΠΟΘΕΣΕΩΝ»

ΕΥΣΤΑΘΙΑ ΠΑΠΑΓΕΩΡΓΙΟΥ - ΠΑΝΑΓΙΩΤΑ ΛΑΛΟΥ

2.1 Έλεγχοι Κανονικότητας

Για να αποφασιστεί εάν θα χρησιμοποιηθεί παραμετρικός ή μη παραμετρικός έλεγχος, αρχικά εξετάζουμε αν οι παρατηρήσεις μας ακολουθούν κανονική κατανομή.

Οι έλεγχοι που θα χρησιμοποιήσουμε για τον έλεγχο κανονικότητας είναι:

- Έλεγχος Kolmogorov-Smirnov
- Έλεγχος Shapiro - Wilk

One-Sample Kolmogorov-Smirnov (K-S) Test

Το κριτήριο *K-S* είναι ένας μη παραμετρικός έλεγχος που χρησιμοποιείται για να εξετάσει την καλή προσαρμογή ενός τυχαίου δείγματος σε μία δεδομένη κατανομή, η οποία μπορεί να είναι Κανονική, Ομοιόμορφη, Poisson ή Εκθετική.

Για την επιλογή του κατάλληλου ελέγχου, μας ενδιαφέρει να εξετάσουμε αν το δείγμα έχει καλή προσαρμογή σε κανονική κατανομή. Η υπόθεση που εξετάζεται είναι εάν η κατανομή των δεδομένων είναι η κανονική κατανομή, ή όχι.

$H_0 : f(x) = N(\mu, \sigma^2)$ (Η μεταβλητή ακολουθεί κανονική κατανομή)

$H_1 : f(x) \neq N(\mu, \sigma^2)$ (Η μεταβλητή δεν ακολουθεί κανονική κατανομή)

Η διαδρομή που ακολουθείται είναι:

Analyze / Nonparametric tests / Legacy Dialogs / 1 sample K-S,

Έλεγχος Shapiro - Wilk

Το κριτήριο Shapiro-Wilk είναι ένας ακόμα πολύ γνωστός μη παραμετρικός έλεγχος για το αν οι παρατηρήσεις μίας μεταβλητής προέρχεται από κανονική κατανομή.

Οι υποθέσεις του ελέγχου είναι:

$$H_0 : f(x) = N(\mu, \sigma^2)$$

$$H_1 : f(x) \neq N(\mu, \sigma^2)$$

Έλεγχος Shapiro-Wilk προτείνεται όταν το δείγμα μας είναι σχετικά μικρό, $n \leq 25$

Η διαδρομή που ακολουθείται είναι:

Analyze / Descriptive Statistics / Explore / Plots

Άσκηση 2.1 Να γίνει ο έλεγχος κανονικότητας στις μεταβλητές Βάρος και Αθρωματικός δείκτης της άσκησης 1.4 με τους ελέγχους Kolmogorov-Smirnov (K-S) Test και Shapiro Wilk.

Λύση:

Θέλουμε να ελέγξουμε την υπόθεση:

H_0 : Η μεταβλητή ακολουθεί κανονική κατανομή

έναντι της

H_1 : Η μεταβλητή δεν ακολουθεί κανονική κατανομή

α) Με τον έλεγχο K-S:

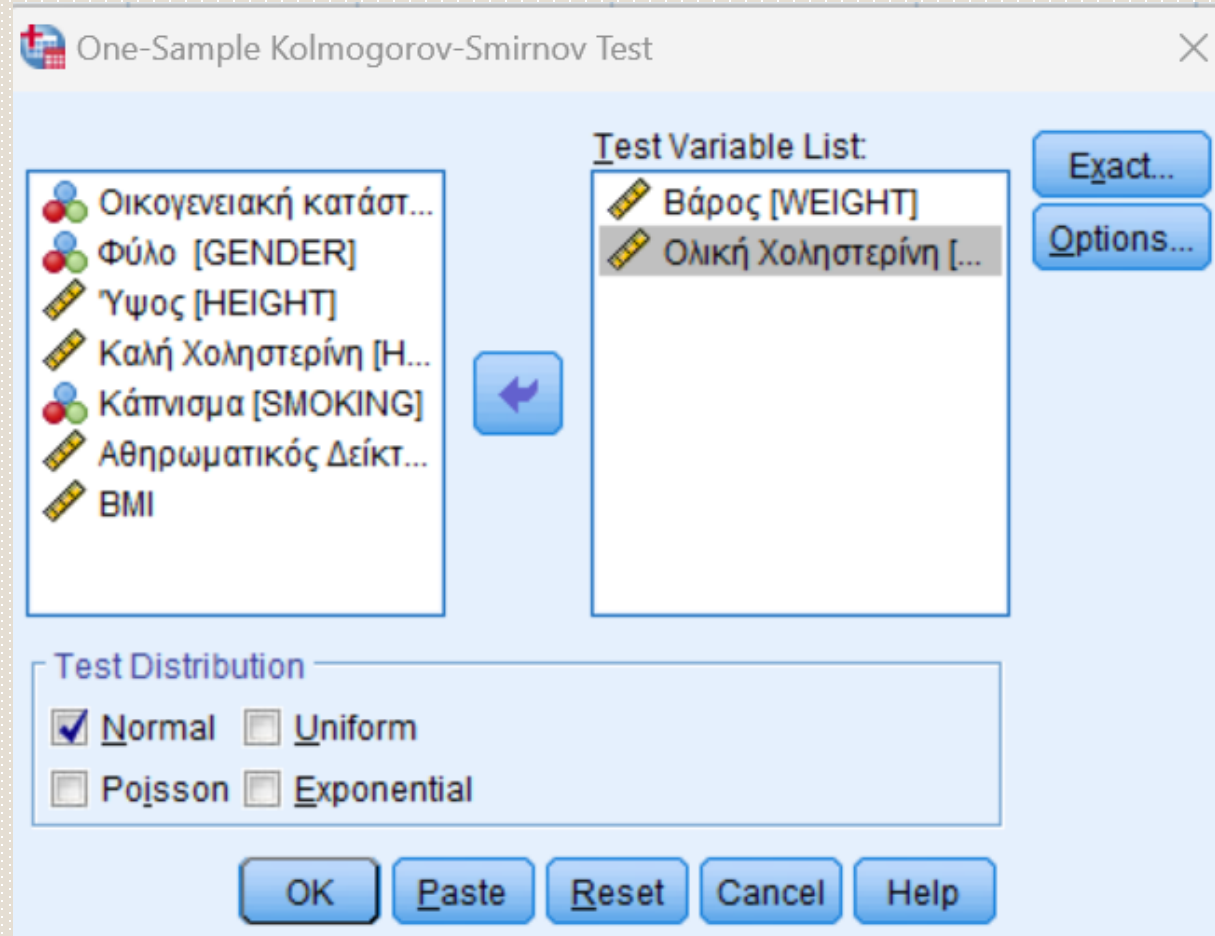
Analyze / Nonparametric tests / Legacy Dialogs / 1 sample K-S

File Edit View Data Transform Analyze Graphs Utilities Extensions Window Help

	Protein	var
1	10	
2	9	
3	7	
4	19	
5	20	
6	16	
7	8	
8	9	
9	26	
10	11	
11	11	
12	8	
13	13	
14	100	
15	70	
16	92	
17	87	
18	73	
19	82	
20	99	
21	44	
22	42	
23	62	
24	61	
25	47	
26	26	

- Reports
- Descriptive Statistics
- Bayesian Statistics
- Tables
- Compare Means
- General Linear Model
- Generalized Linear Models
- Mixed Models
- Correlate
- Regression
- Loglinear
- Neural Networks
- Classify
- Dimension Reduction
- Scale
- Nonparametric Tests**
 - One Sample...
 - Independent Samples...
 - Related Samples...
 - Legacy Dialogs
 - Chi-square...
 - Binomial...
 - Runs...
 - 1-Sample K-S...**
 - 2 Independent Samples...
 - K Independent Samples...
 - 2 Related Samples...
 - K Related Samples...
- Forecasting
- Survival
- Multiple Response
- Missing Value Analysis...
- Multiple Imputation
- Complex Samples
- Simulation...
- Quality Control
- ROC Curve...
- Spatial and Temporal Modeling...
- Direct Marketing

Μεταφέρουμε με το μπλε βέλος στο δεξί πλαίσιο τις μεταβλητές ' Βάρος' και 'Ολική Χοληστερίνη' που θέλουμε να ελέγξουμε. ενώ η επιλογή **Normal** είναι προ-επιλεγμένη στην περιοχή **Test Distributio**.



Πατώντας OK, εμφανίζεται ο πίνακας **One-Sample Kolmogorov-Smirnov Test** στον οποίον αναγράφονται τα αποτελέσματα του ελέγχου K-S:

- Το πλήθος (**N**) του δείγματος.
- Οι εκτιμήσεις των παραμέτρων της κανονικής κατανομής, **Normal Parameters**, η οποία προσεγγίζει τα δεδομένα μας δηλ. μέσο(**Mean**) και τυπική απόκλιση (**Std. Deviation**).
- Οι ακραίες διαφορές, **Most Extreme Differences**, μεταξύ της εμπειρικής συνάρτησης κατανομής και της εκτιμημένης κανονικής κατανομής
- Η τιμή της στατιστικής συνάρτησης,
- Το (δίπλευρο), **Asymp. Sig. (2-tailed)**, δηλ. το p -value του ελέγχου όπου έχει γίνει διόρθωση σημαντικότητας του Lilliefors).

One-Sample Kolmogorov-Smirnov Test

	Βάρος	Αθρωματικό ς Δείκτης
N	20	20
Normal Parameters ^{a,b}	Mean	91,05
	Std. Deviation	20,814
Most Extreme Differences	Absolute	,118
	Positive	,118
	Negative	-,115
Test Statistic	,118	,171
Asymp. Sig. (2-tailed)	,200 ^{c,d}	,129 ^c

a. Test distribution is Normal.

b. Calculated from data.

c. Lilliefors Significance Correction.

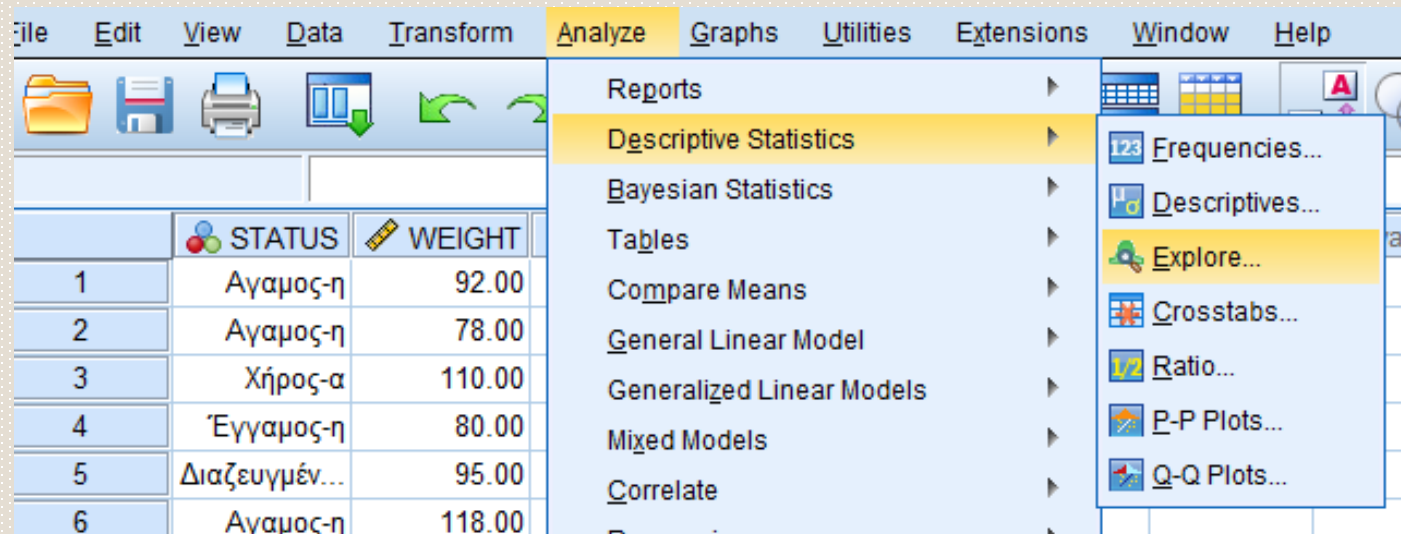
d. This is a lower bound of the true significance.

Σύμφωνα με τη τιμή του p -value:

- Sig=0,2>0,05 για το βάρος και
- Sig=0,129>0,05 για τον αθηρωματικό δείκτη

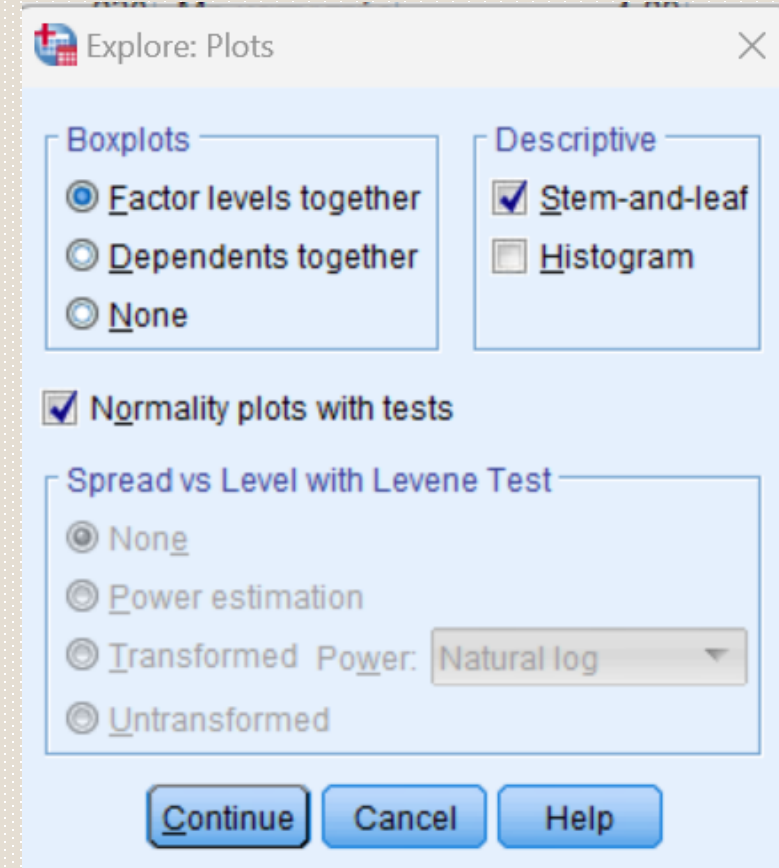
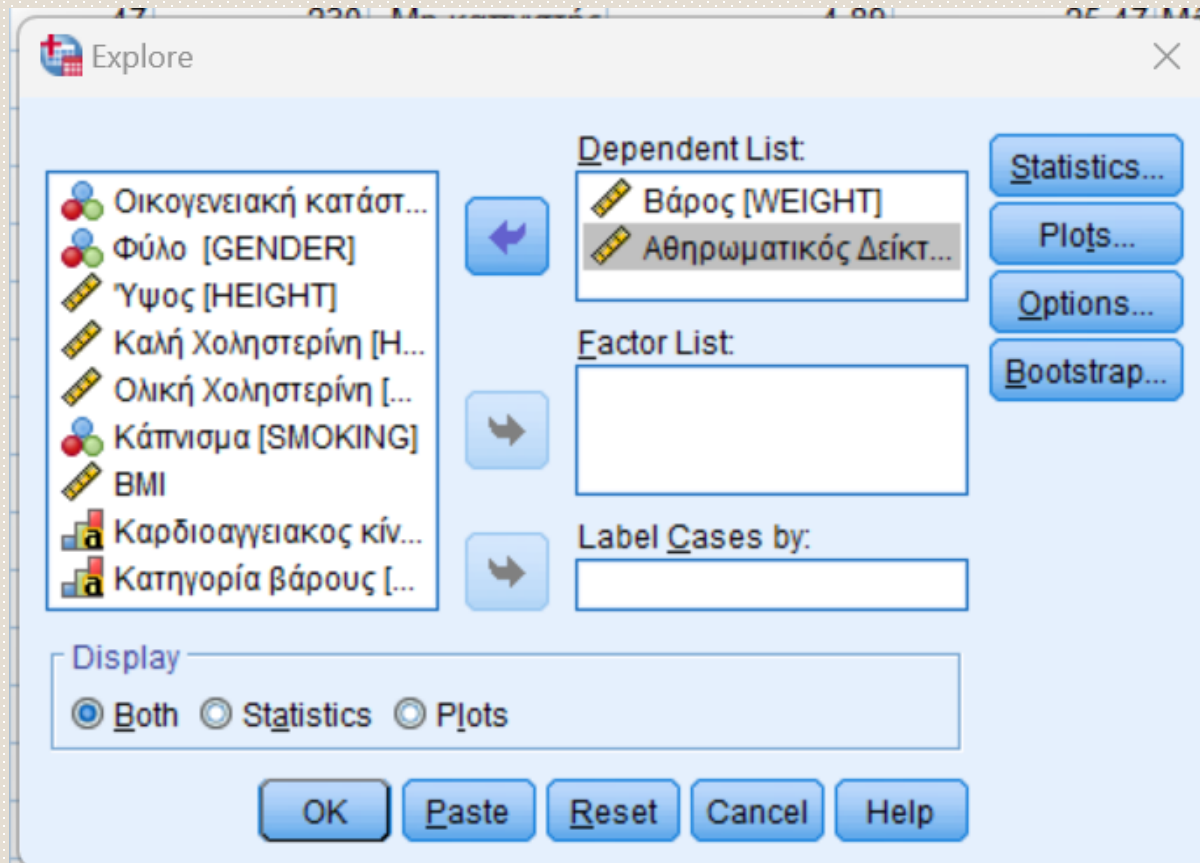
Και στις δύο περιπτώσεις Sig>0,05 ,άρα αποδεχόμαστε την υπόθεση H_0 και επομένως η κατανομή και στα δύο δείγματα είναι κανονική

β) Με τον έλεγχο Shapiro – Wilk: *Analyze / Descriptive Statistics / Explore*



Στο παράθυρο που ανοίγει, τοποθετούμε τις δύο μεταβλητές Βάρος και Αθηρωματικός Δείκτης στο πλαίσιο **Dependent List**. Στη συνέχεια πατάμε το κουμπί **Plots**. Στο παράθυρο Explore Plots που ανοίγει επιλέγουμε το **Normality plots with tests**. Πατάμε **Continue** και τέλος **OK**.

(Στο Display μπορούμε να αφήσουμε τη ήδη υπάρχουσα επιλογή, δηλαδή το Both ή να επιλέξουμε το Plots αν δεν μας ενδιαφέρει να δούμε τον πίνακα με τα περιγραφικά μέτρα. Το Statistics δεν μπορούμε να το επιλέξουμε διότι δεν μας δίνεται η δυνατότητα ελέγχου κανονικότητας)



Όπως έχουμε δει στο ά μέρος, με τη συγκεκριμένη διαδρομή προκύπτουν πίνακες με περιγραφικά μέτρα αλλά και γραφήματα όπως το θηκόγραμμα. Τα αποτελέσματα για τον έλεγχο κανονικότητας που μας ενδιαφέρουν βρίσκονται στον πίνακα Tests of Normality.

Tests of Normality						
	Kolmogorov-Smirnov ^a			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Βάρος	,118	20	,200 [*]	,956	20	,474
Αθρωματικός Δείκτης	,171	20	,129	,948	20	,332

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

a. Lilliefors Significance Correction

Στον πίνακα παρουσιάζονται τα αποτελέσματα και των δύο ελέγχων. Για το Shapiro Wilk:

$Sig=0,474 > 0,05$ για το βάρος

$Sig=0,332 > 0,05$ για το αθρωματικό δείκτη.

Επομένως και με τον έλεγχο Shapiro Wilk τα δύο δείγματα ακολουθούν κανονική κατανομή

2.2 Παραμετρικοί Έλεγχοι

2.2.1 Independent Samples T-test

Άσκηση 2.2: Έχουμε δύο ομάδες ατόμων που δύο ομάδες πάσχουν από πολλαπλή σκλήρυνση. Η πρώτη ομάδα (10 άτομα) ακολουθούν μόνο φαρμακευτική αγωγή, ενώ οι υπόλοιποι (8 άτομα) εκτός από τη φαρμακευτική θεραπεία ακολουθούν και πρόγραμμα φυσιοθεραπειών. Όλοι οι ασθενείς υποβλήθηκαν σε τεστ αξιολόγησης κινητικότητας TUG (Timed Up and Go Test), και ο χρόνος ολοκλήρωσης της δοκιμασίας του test φαίνεται στον παρακάτω πίνακα.

α) Να εξεταστεί αν υπάρχει διαφορά στο μέσο χρόνο ολοκλήρωσης του TUG μεταξύ αυτών που κάνουν φυσικοθεραπείες κι αυτών που δεν κάνουν. $\alpha=0,05$

β) Να βρεθεί 95% διάστημα εμπιστοσύνης για τη διαφορά των μέσων τιμών του χρόνου στις

Φυσικοθεραπείες	
ΟΧΙ	ΝΑΙ
18	13
20	14
12	17
15	12
23	18
17	15
30	22
20	21
21	
15	

Λύση:

Στην άσκηση αυτή διερευνούμε τη σχέση μεταξύ φυσιοθεραπείας (διχοτόμος μεταβλητή με τιμές ναι – όχι) και του χρόνου ολοκλήρωσης του τεστ (ποσοτική συνεχής μεταβλητή)

Οι υποθέσεις του ελέγχου:

H_0 : Η μέση τιμή του χρόνου **δεν** διαφέρει μεταξύ αυτών που κάνουν φυσικοθεραπείες κι αυτών που δεν κάνουν

H_1 : Η μέση τιμή του χρόνου διαφέρει μεταξύ αυτών που κάνουν φυσικοθεραπείες κι αυτών που δεν κάνουν

ή

H_0 : Δεν υπάρχει σχέση μεταξύ φυσικοθεραπείας και χρόνου ολοκλήρωσης του τεστ

H_1 : Υπάρχει σχέση μεταξύ φυσικοθεραπείας και χρόνου ολοκλήρωσης του τεστ

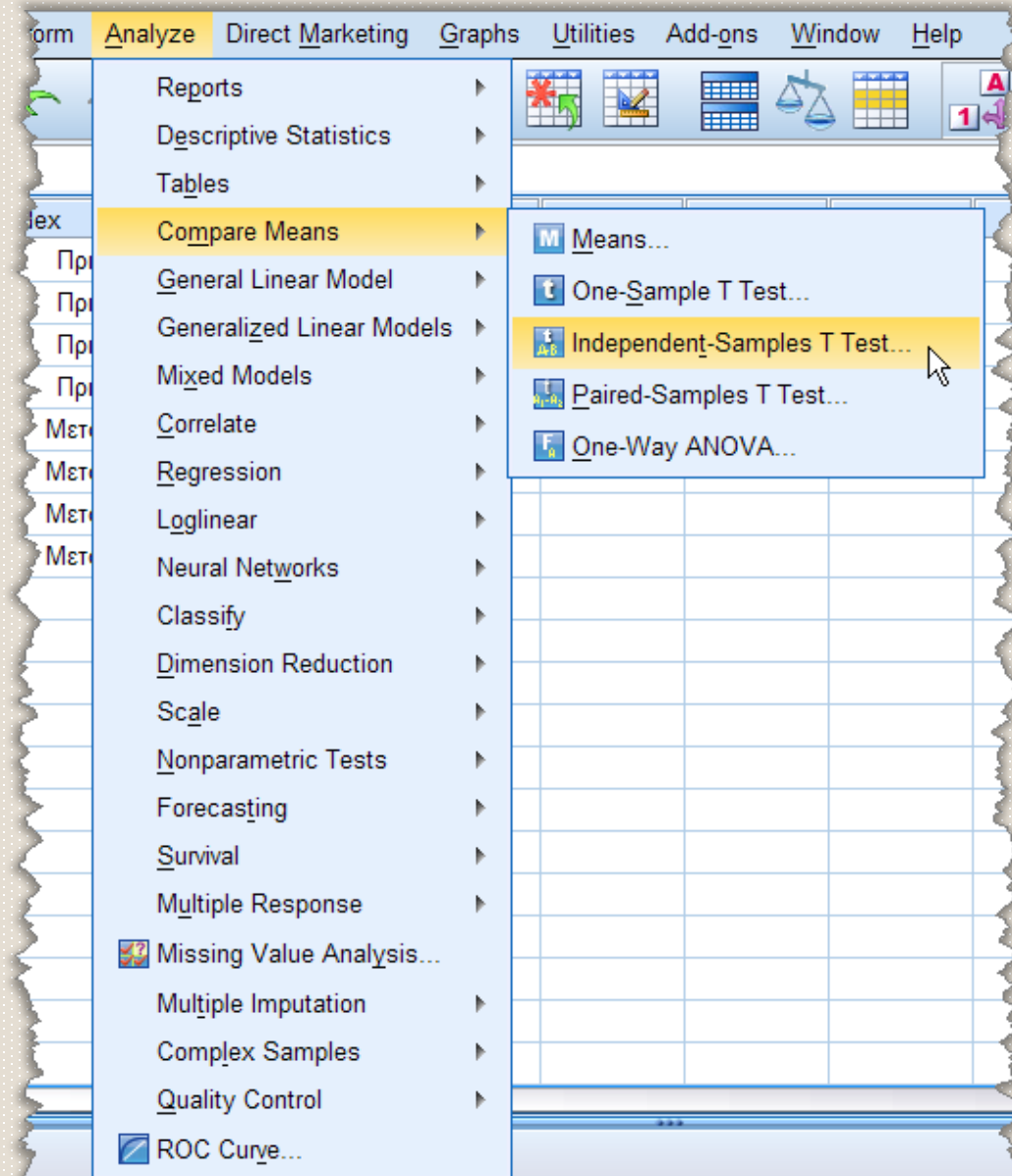
Έχουμε κάνει τον έλεγχο κανονικότητας δείγμα και τα δύο δείγματα ακολουθούν κανονική κατανομή, επομένως συνεχίζουμε με έλεγχο ισότητας μέσων τιμών για ανεξάρτητα δείγματα.

Για να εφαρμοστεί ο έλεγχος αυτός πρέπει οι τιμές και των δύο δειγμάτων να μπουν στην ίδια στήλη. Επομένως δημιουργούμε μια ποσοτική μεταβλητή (scale) στην οποία δίνουμε όνομα TIME και μια δεύτερη ποιοτική μεταβλητή (nominal) την οποία ονομάζουμε INDEX. Η μεταβλητή INDEX παίρνει τιμές 1 και 2, Βάζουμε 1 όταν η αντίστοιχη τιμή της 1^{ης} στήλης ανήκει στο 1^ο δείγμα και 2 όταν ανήκει στο 2^ο

	TIME	INDEX	
1	18	1	
2	20	1	
3	12	1	
4	15	1	
5	23	1	
6	17	1	
7	30	1	
8	20	1	
9	21	1	
10	15	1	
11	13	2	
12	14	2	
13	17	2	
14	12	2	
15	18	2	
16	15	2	
17	22	2	
18	21	2	
19			

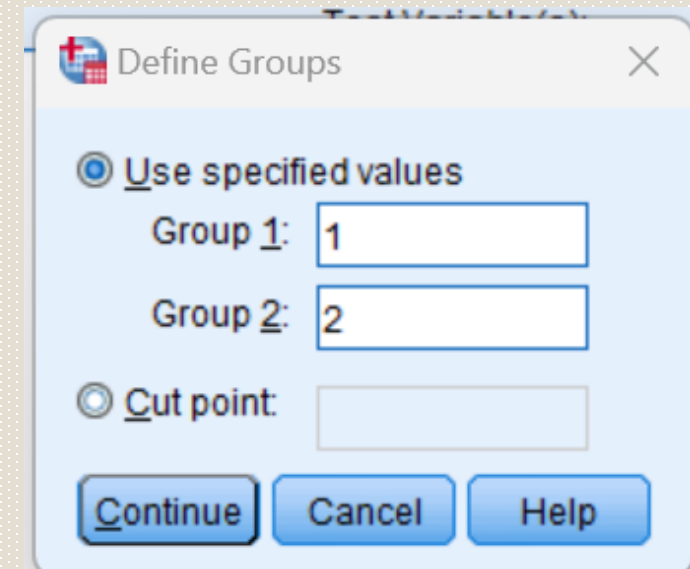
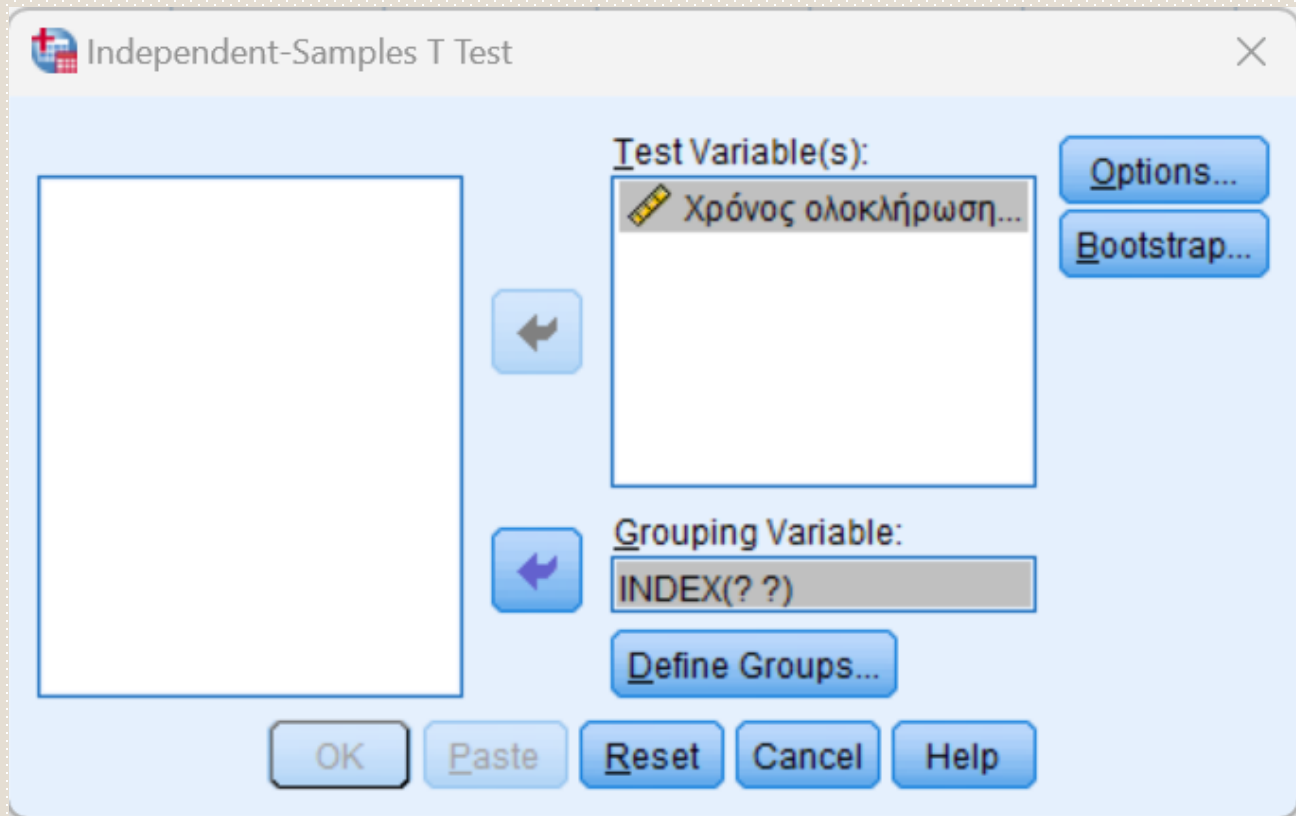
Η διαδρομή στο SPSS είναι:

***Analyze / Compare Means /
Independent Samples T-Test***



Στο **Test Variable(s)** τοποθετήστε την ποσοτική μεταβλητή (TIME) και στο **Grouping Variable** τοποθετήστε τη διχοτόμο μεταβλητή (INDEX)

Επιλέγουμε **Define Groups**, οπότε προκύπτει το πλαίσιο διαλόγου Define Groups. Βάζουμε τις τιμές 1 και 2 στα κελιά, πατάμε **Continue** και τέλος **OK**.



Στον πίνακα **Group Statistics** παρουσιάζονται:

- Ο αριθμός των παρατηρήσεων (N) σε κάθε κατηγορία της διχοτόμου μεταβλητής.
- Ο μέσος (Mean) της ποσοτικής μεταβλητής σε κάθε κατηγορία της διχοτόμου μεταβλητής.
- Η τυπική απόκλιση (Std. Deviation) της ποσοτικής μεταβλητής σε κάθε κατηγορία της διχοτόμου μεταβλητής.
- Το τυπικό σφάλμα του μέσου (Std. Error Mean) της ποσοτικής μεταβλητής σε κάθε κατηγορία της διχοτόμου μεταβλητής

	Δείκτης	N	Mean	Std. Deviation	Std. Error Mean
Χρόνος ολοκλήρωσης τεστ	1	10	19,10	5,043	1,595
	2	8	16,50	3,665	1,296

Τα αποτελέσματα του ελέγχου παρουσιάζονται στον 2^ο πίνακα: ***Independent Samples Test***

Στον πίνακα αυτό παρουσιάζονται τα αποτελέσματα δύο ελέγχων:
Ελέγχου Levene και ελέγχου ισότητας μέσων τιμών (T-Test).

A) Έλεγχος του Levene

Με τον έλεγχο αυτό, εξετάζεται η ισότητα ή όχι των διασπορών στους πληθυσμούς. Μας ενδιαφέρει το αποτέλεσμα του ελέγχου για να δούμε με ποια από τις δύο γραμμές του ελέγχου μέσων τιμών θα συνεχίσουμε.

Οι υποθέσεις του ελέγχου είναι:

$$H_0: \sigma_1^2 = \sigma_2^2$$

$$H_1: \sigma_1^2 \neq \sigma_2^2$$

Εαν ισχύει η μηδενική υπόθεση (ισότητα διασπορών), τότε στον έλεγχο T-Test, ισχύουν τα αποτελέσματα της 1^{ης} γραμμής. Εάν ισχύει η εναλλακτική, τότε στο T-Test ισχύει η δεύτερη γραμμή.

Το *p-value* του ελέγχου Levene είναι $\text{Sig}=0,588>0,05$. Επομένως δεχόμαστε την H_0 (ισότητα διασπορών) και άρα στο έλεγχο t-test κοιτάμε το sig της 1^{ης} γραμμής:

$\text{Sig}=0,24>0,05$

Επομένως αποδεχόμαστε την H_0 δηλαδή την ισότητα μέσων τιμών. Άρα ΔΕΝ υπάρχει διαφορά στο μέσο χρόνο ολοκλήρωσης του TUG μεταξύ αυτών που κάνουν φυσικοθεραπείες κι αυτών που δεν κάνουν. Στην ίδια γραμμή βλέπουμε την τιμή της στατιστικής συνάρτησης: $t=1,22$ και το διάστημα εμπιστοσύνης για τη διαφορά των μέσων τιμών στους δυο πληθυσμούς:

$[-1.917, 7,117]$

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
χρόνος ολοκλήρωσης τεστ	Equal variances assumed	,306	,588	1,220	16	,240	2,600	2,131	-1,917	7,117
	Equal variances not assumed			1,265	15,897	,224	2,600	2,055	-1,758	6,958

Άσκηση 2.3: Να εξεταστεί αν υπάρχει διαφορά στη μέση τιμή Δείκτη Μάζας Σώματος (BMI) μεταξύ ανδρών και γυναικών, χρησιμοποιώντας τα δεδομένα της άσκησης 1.4. Επιπλέον να βρεθεί 95% διάστημα εμπιστοσύνης για τη διαφορά των μέσων.

Άσκηση 2.4: Μετρήσαμε την αρτηριακή πίεση σε δυο ανεξάρτητες ομάδες παιδιών, αποτελούμενες από 8 και 10 παιδιά αντίστοιχα. Στην 1^η ομάδα τα παιδιά έχουν υπερτασικούς γονείς, ενώ οι γονείς των παιδιών της 2^{ης} ομάδας δεν παρουσιάζουν υπέρταση. Υπάρχει διαφορά στη μέση τιμή πίεσης παιδιών υπερτασικών και παιδιών μη υπερτασικών γονιών

ΟΜΑΔΑ Α									
100	102	96	106	110	120	112	90		
ΟΜΑΔΑ Β									
104	88	100	98	102	92	96	100	96	97

2.2.2 Paired Samples T-Test

Άσκηση 2.5: Η αρτηριακή πίεση 10 ασθενών πριν (X) και μετά (Y) τη χορήγηση φαρμάκου κατά της πίεσης είναι:

X	13	15	18	14	12	13	15	16	18	19
Y	12	13	15	15	14	13	13	14	14	13

Να ελεγχθεί σε επίπεδο σημαντικότητας 5%, εαν το συγκεκριμένο φάρμακο είναι αποτελεσματικό κατά της πίεσης.

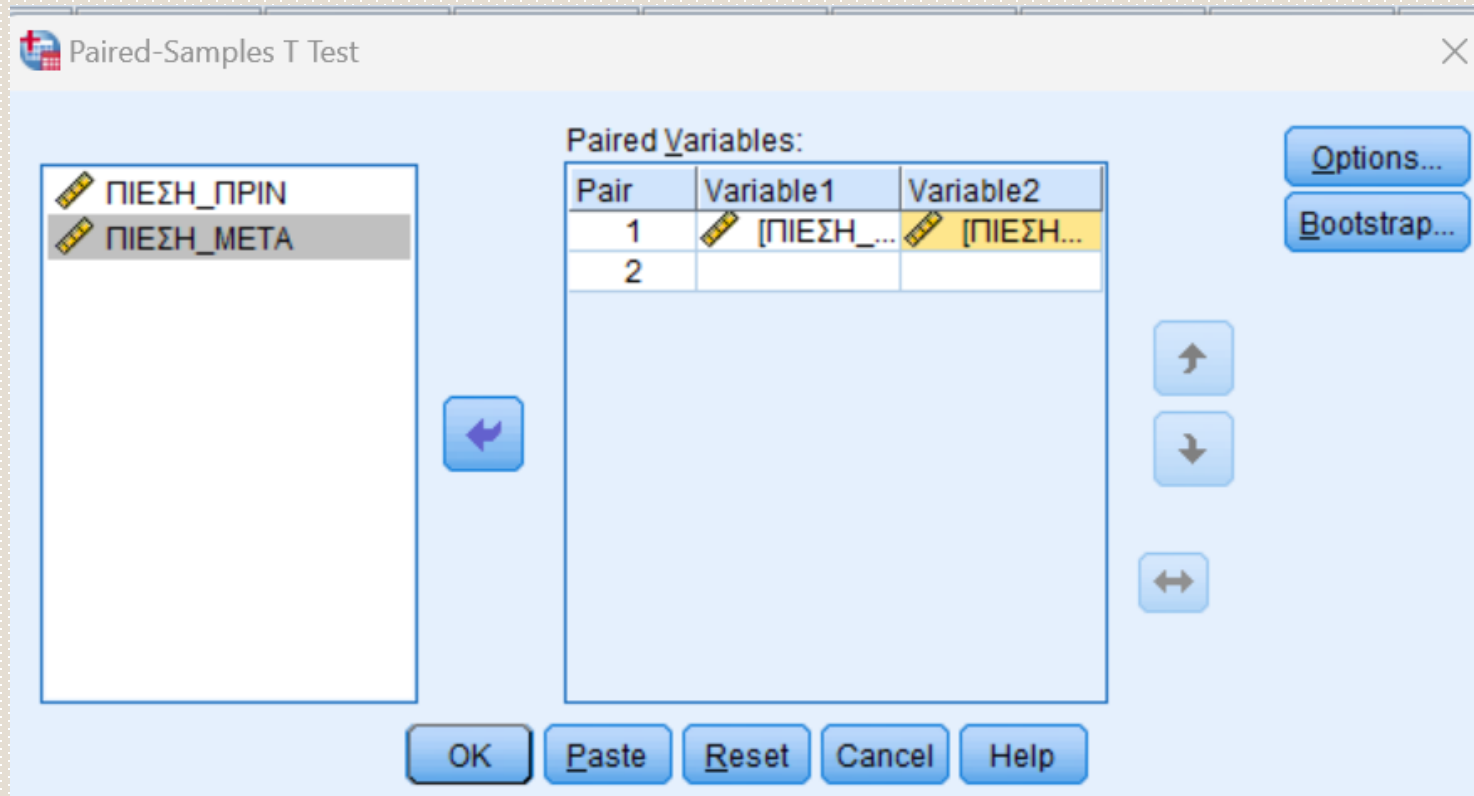
Λύση

Εφαρμόζοντας έλεγχο κανονικότητας διαπιστώνουμε ότι και τα δύο δείγματα ακολουθούν κανονική κατανομή. Επομένως θα εφαρμόσουμε έλεγχο ισότητας μέσων τιμών για δείγματα κατά ζεύγη, (Paired Samples T-Test)

- Για την εφαρμογή του ελέγχου, πρέπει να καταχωρούμε τα δείγματα σε δύο διαφορετικές στήλες. Επομένως ορίζουμε δυο ποσοτικές μεταβλητές (scale) με name: ΠΙΕΣΗ_ΠΡΙΝ και ΠΙΕΣΗ_ΜΕΤΑ και καταχωρούμε τις τιμές στο Data View.
- Η διαδρομή στη συνέχεια είναι:
- **Analyze / Compare Means / Paired Samples T-Test**

	ΠΙΕΣΗ_ΠΡΙΝ	ΠΙΕΣΗ_ΜΕ...	var
	13	12	
	15	13	
	18	15	
	14	15	
	12	14	
	13	13	
	15	13	
	16	14	
	18	14	
	19	13	

- Στο πλαίσιο διαλόγου Paired Samples T Test,
- Στο Paired Variables, στη στήλη Variable 1 τοποθετούμε την ποσοτική μεταβλητή που αφορά στο πρώτο «δείγμα»(ΠΙΕΣΗ_ΠΡΙΝ) και στη στήλη Variable 2 τοποθετούμε την ποσοτική μεταβλητή που αφορά στο δεύτερο «δείγμα» (ΠΙΕΣΗ ΜΕΤΑ). Πατάμε *OK*.



Τα αποτελέσματα του ελέγχου φαίνονται στον 3^ο πίνακα: ***Paired Samples Test***

Sig=0,049<0,05 οπότε απορρίπτεται η μηδενική υπόθεση κι επομένως συμπεραίνεται πως υπάρχει στατιστικά σημαντική σχέση μεταξύ φαρμακευτικής αγωγής και συστολικής αρτηριακής πίεσης. Δηλαδή το φάρμακο είναι αποτελεσματικό.

Στον ίδιο πίνακα μπορούμε να δούμε τη τιμή της στατιστικής συνάρτησης $t=2,279.$, τους βαθμούς ελευθερίας $df=9$ καθώς και το 95% διάστημα εμπιστοσύνης.

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	ΠΙΕΣΗ_ΠΡΙΝ - ΠΙΕΣΗ_ΜΕΤΑ	1,700	2,359	,746	,012	3,388	2,279	9	,049

Άσκηση 2.6: Θέλουμε να εξετάσουμε κατά πόσο ένα νέο πρόγραμμα ασκήσεων είναι αποτελεσματικό στην αντιμετώπιση χρόνιου πόνου στη μέση. Για το λόγο αυτό μοιράστηκε σε 10 ασθενείς που υποφέρουν από πόνο στη μέση ερωτηματολόγιο από το οποίο προέκυψε ο βαθμός στη κλίμακα πόνου, (0 έως 10). Το ίδιο ερωτηματολόγιο απαντήθηκε μετά το τέλος των συνεδριών. Τα αποτελέσματα φαίνονται στον πίνακα:

ΠΡΙΝ ΤΟ ΠΡΟΓΡΑΜΜΑ	ΜΕΤΑ ΤΟ ΠΡΟΓΡΑΜΜΑ
9	7
10	5
5	3
8	5
8	4
6	3
9	4
7	5
9	6
9	4

2.2.4 One Way ANOVA

Άσκηση 2.7: Μια πολυεθνική εταιρεία για να προσλάβει τους υπαλλήλους της χρησιμοποιεί ψυχομετρικό τεστ. Στον παρακάτω πίνακα βλέπετε την επίδοση στο ίδιο τεστ, των υποψηφίων από 3 διαφορετικές χώρες: Ελβετία, Ελλάδα, Γερμανία. Υπάρχει διαφορά στην επίδοση των υποψηφίων μεταξύ των 3 χωρών;

ΕΛΒΕΤΙΑ	ΕΛΛΑΔΑ	ΓΕΡΜΑΝΙΑ
82	83	38
83	78	59
97	68	55
93	61	66
55	77	45
67	54	52
53	69	52
	51	61
	63	

Λύση:

Θέλουμε να κάνουμε τον έλεγχο:

H_0 : Η μέσες τιμές επίδοσης είναι ίσες και στις 3 χώρες

H_1 : Η μέσες τιμές επίδοσης **δεν** είναι ίσες και στις 3 χώρες

ή

H_0 : Δεν υπάρχει σχέση μεταξύ επίδοσης και χώρας

H_1 : Υπάρχει σχέση μεταξύ επίδοσης και χώρας

Τα τρία δείγματα ακολουθούν κανονική κατανομή (μετά από έλεγχο κανονικότητας)

Επομένως θα γίνει Ανάλυση διασποράς μιας κατεύθυνσης (**One Way ANOVA**)

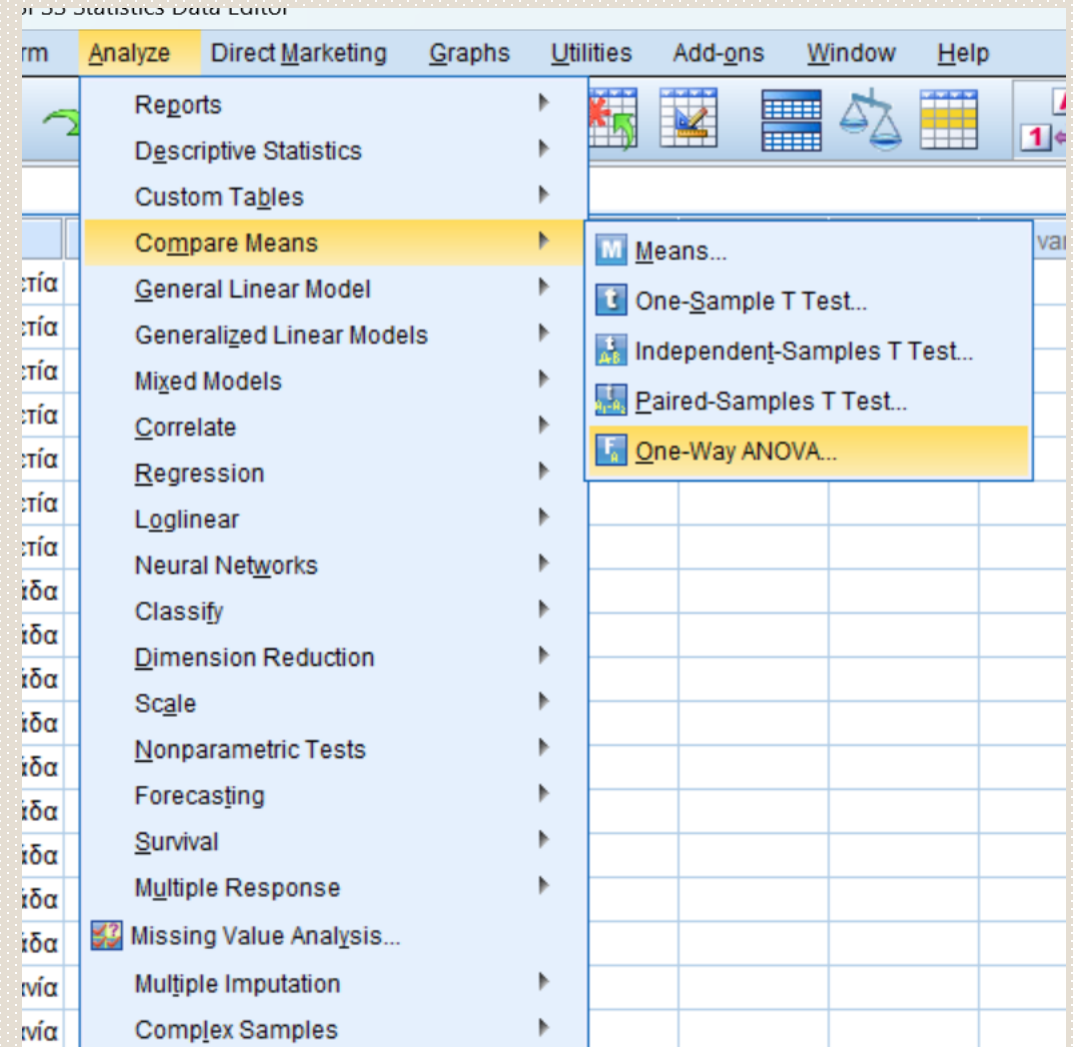
Για την εφαρμογή του ελέγχου πρέπει να καταχωρηθούν οι τιμές και των τριών δειγμάτων στην ίδια στήλη. Επομένως δημιουργούμε μια ποσοτική μεταβλητή (scale) στην οποία δίνουμε όνομα «επίδοση» και μια δεύτερη ποιοτική μεταβλητή (nominal) την οποία ονομάζουμε “index”. Η μεταβλητή index παίρνει τιμές 1, 2 και 3. Προαιρετικά μπορούμε να βάλουμε ετικέτες στις τιμές με τη ρύθμιση values. Βάζουμε τις παρακατω ετικέτες:

- 1-Ελβετία
- 2-Ελλάδα
- 3-Γερμανία

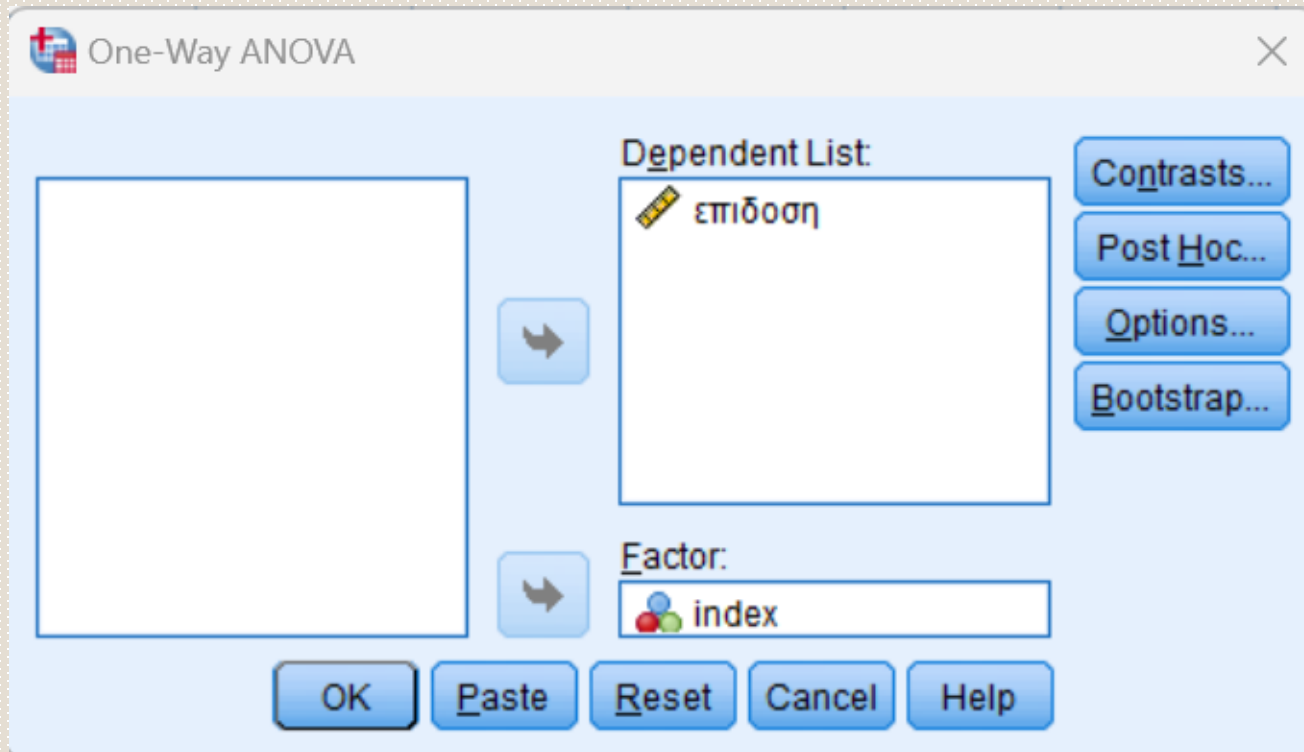
επίδοση	index	var
82	Ελβετία	
83	Ελβετία	
97	Ελβετία	
93	Ελβετία	
55	Ελβετία	
67	Ελβετία	
53	Ελβετία	
83	Ελλάδα	
78	Ελλάδα	
68	Ελλάδα	
61	Ελλάδα	
77	Ελλάδα	
54	Ελλάδα	
69	Ελλάδα	
51	Ελλάδα	
63	Ελλάδα	
38	Γερμανία	
59	Γερμανία	
55	Γερμανία	
66	Γερμανία	
45	Γερμανία	

Η διαδρομή για τον έλεγχο είναι:

Analyze / Compare Means / One-Way ANOVA



Στο **Dependent List** τοποθετήστε την ποσοτική μεταβλητή (επίδοση) και στο **Factor** τοποθετήστε την ονομαστική μεταβλητή index, και πατάμε **OK**.



Στο output λαμβάνουμε το αποτέλεσμα του ελέγχου στο πίνακα ANOVA., όπου βλέπουμε την τιμή του *p-value*:

Sig=0,00<0,05

από το οποίο διαπιστώνουμε ότι η μηδενική υπόθεση απορρίπτεται ($p\text{-value} < 0,05$), συνεπώς υπάρχει τουλάχιστον ένας μέσος που διαφέρει σημαντικά από τους υπόλοιπους.

Επίδοση	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	1901,516	2	950,758	5,896	,009
Within Groups	3386,317	21	161,253		
Total	5287,833	23			

Ωστόσο, δεν μας δίνει την πληροφορία ποιος μέσος διαφέρει από τους υπόλοιπους.

Για να προσδιοριστούν οι μέσοι που διαφέρουν μεταξύ τους, θα πρέπει να γίνουν πολλαπλές συγκρίσεις μεταξύ των ομάδων (ανά δύο).

Αφού επαναλάβουμε τη διαδικασία:

Analyze / Compare Means / One-Way ANOVA

Επιλέγουμε **Post Hoc**, οπότε προκύπτει το πλαίσιο διαλόγου One-Way ANOVA: **Post Hoc Multiple Comparisons**:

One-Way ANOVA: Post Hoc Multiple Comparisons

Equal Variances Assumed

LSD S-N-K Waller-Duncan

Bonferroni Tukey Type I/Type II Error Ratio: 100

Sidak Tukey's-b Dunnett

Scheffe Duncan Control Category: Last

R-E-G-W F Hochberg's GT2

R-E-G-W Q Gabriel

Test

2-sided < Control > Control

Equal Variances Not Assumed

Tamhane's T2 Dunnett's T3 Games-Howell Dunnett's C

Significance level: 0,05

Continue Cancel Help

Στο πλαίσιο διαλόγου One-Way ANOVA: Post Hoc Multiple Comparisons, στο Equal Variance Assumed επιλέγουμε Bonferroni (για <10 συγκρίσεις) ή Tukey ή Scheffe (για ≥συγκρίσεις). Επιλέγουμε στη συγκεκριμένη περίπτωση **Bonferroni**, στη συνέχεια **Continue** και τέλος **OK**.

Στη στήλη (I) παρουσιάζεται η μια κατηγορία της ανεξάρτητης μεταβλητής που συμπεριλαμβάνεται στον επιμέρους έλεγχο της υπόθεσης.

Στη στήλη (J) παρουσιάζεται η δεύτερη κατηγορία της ανεξάρτητης μεταβλητής που συμπεριλαμβάνεται στον επιμέρους έλεγχο της υπόθεσης.

Στη στήλη Mean Difference (I-J) παρουσιάζεται η διαφορά των μέσων τιμών της ποσοτικής μεταβλητής στις 2 κατηγορίες της ανεξάρτητης μεταβλητής που συμπεριλήφθηκαν στον επιμέρους έλεγχο της υπόθεσης.

Multiple Comparisons

Dependent Variable: επίδοση
Bonferroni

(I) index	(J) index	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Ελβετία	Ελλάδα	8,603	6,399	,580	-8,04	25,25
	Γερμανία	22,214*	6,572	,008	5,12	39,31
Ελλάδα	Ελβετία	-8,603	6,399	,580	-25,25	8,04
	Γερμανία	13,611	6,170	,116	-2,44	29,66
Γερμανία	Ελβετία	-22,214*	6,572	,008	-39,31	-5,12
	Ελλάδα	-13,611	6,170	,116	-29,66	2,44

*. The mean difference is significant at the 0.05 level.

Στη στήλη Std. Error παρουσιάζεται το τυπικό σφάλμα της διαφοράς των μέσων τιμών της ποσοτικής μεταβλητής στις 2 κατηγορίες της ανεξάρτητης μεταβλητής που συμπεριλήφθηκαν στον επιμέρους έλεγχο της υπόθεσης.

Στη στήλη Sig. παρουσιάζεται η τιμή *p-value* του επιμέρους ελέγχου της υπόθεσης και στις δυο τελευταίες στήλες έχουμε το διάστημα εμπιστοσύνης της διαφοράς των μέσων τιμών της ποσοτικής μεταβλητής στις 2 κατηγορίες .

Με βάση τα αποτελέσματα μας:

- Δεν υπάρχει διαφορά στην μέση επίδοση μεταξύ Ελλάδας –Ελβετίας (Sig=0,58)
- Δεν υπάρχει διαφορά στην μέση επίδοση μεταξύ Ελλάδας –Γερμανίας (Sig=0,116)
- Διαφέρει στατιστικά σημαντικά η μέση επίδοση στην Ελβετία σε σχέση με τη Γερμανία (Sig=0,008) με τη επίδοση στην Ελβετία να είναι μεγαλύτερη . (Η διαφορά των μέσων είναι 22,14)

2.3 Μη Παραμετρικοί Έλεγχοι

2.3.1 Έλεγχος Mann-Whitney

Άσκηση 2.8: Στο διπλανό πίνακα φαίνεται η κατανάλωση ενέργειας μέσα σε 24 ώρες μιας ομάδας γυναικών φυσιολογικού βάρους και μιας ομάδας παχύσαρκων γυναικών, Υπάρχει διαφορά στη κατανάλωση μεταξύ των δύο ομάδων;

Λύση

Θέλουμε να ελέγξουμε τις υποθέσεις:

H_0 Η μέση τιμή κατανάλωσης ενέργειας **δεν** διαφέρει με ταξύ λεπτών κ παχύσαρκων

H_1 Η μέση τιμή κατανάλωσης ενέργειας διαφέρει μεταξύ λεπτών και παχύσαρκων

ή

H_0 Η κατανάλωση ενέργειας δεν σχετίζεται με τη παχυσαρκία

H_1 Η κατανάλωση ενέργειας σχετίζεται με παχυσαρκία

Λεπτές ($n = 13$)	Παχύσαρκες ($n = 9$)
6.13	8.79
7.05	9.19
7.48	9.21
7.48	9.68
7.53	9.69
7.58	9.97
7.90	11.51
8.08	11.85
8.09	12.79
8.11	
8.40	
10.15	
10.88	

Μετά από έλεγχο κανονικότητας συμπεράναμε ότι τα δείγματα δεν ακολουθούν κανονική κατανομή. Επομένως θα εφαρμόσουμε μη παραμετρικό έλεγχο. Τα δείγματα έχουν διαφορετικό μέγεθος, που σημαίνει ότι είναι σίγουρα ‘ανεξάρτητα’. Επομένως ο κατάλληλος έλεγχος είναι Mann – Whitney.

Για να εφαρμοστεί ο έλεγχος αυτός πρέπει οι τιμές και των δύο δειγμάτων να μπουν στην ίδια στήλη. Επομένως δημιουργούμε μια ποσοτική μεταβλητή (scale) στην οποία δίνουμε όνομα Energy και μια δεύτερη ποιοτική μεταβλητή (nominal) την οποία ονομάζουμε INDEX. Η μεταβλητή INDEX παίρνει τιμές 1 και 2, Βάζουμε 1 όταν η αντίστοιχη τιμή της 1^{ης} στήλης ανήκει στο 1^ο δείγμα και 2 όταν ανήκει στο 2^ο

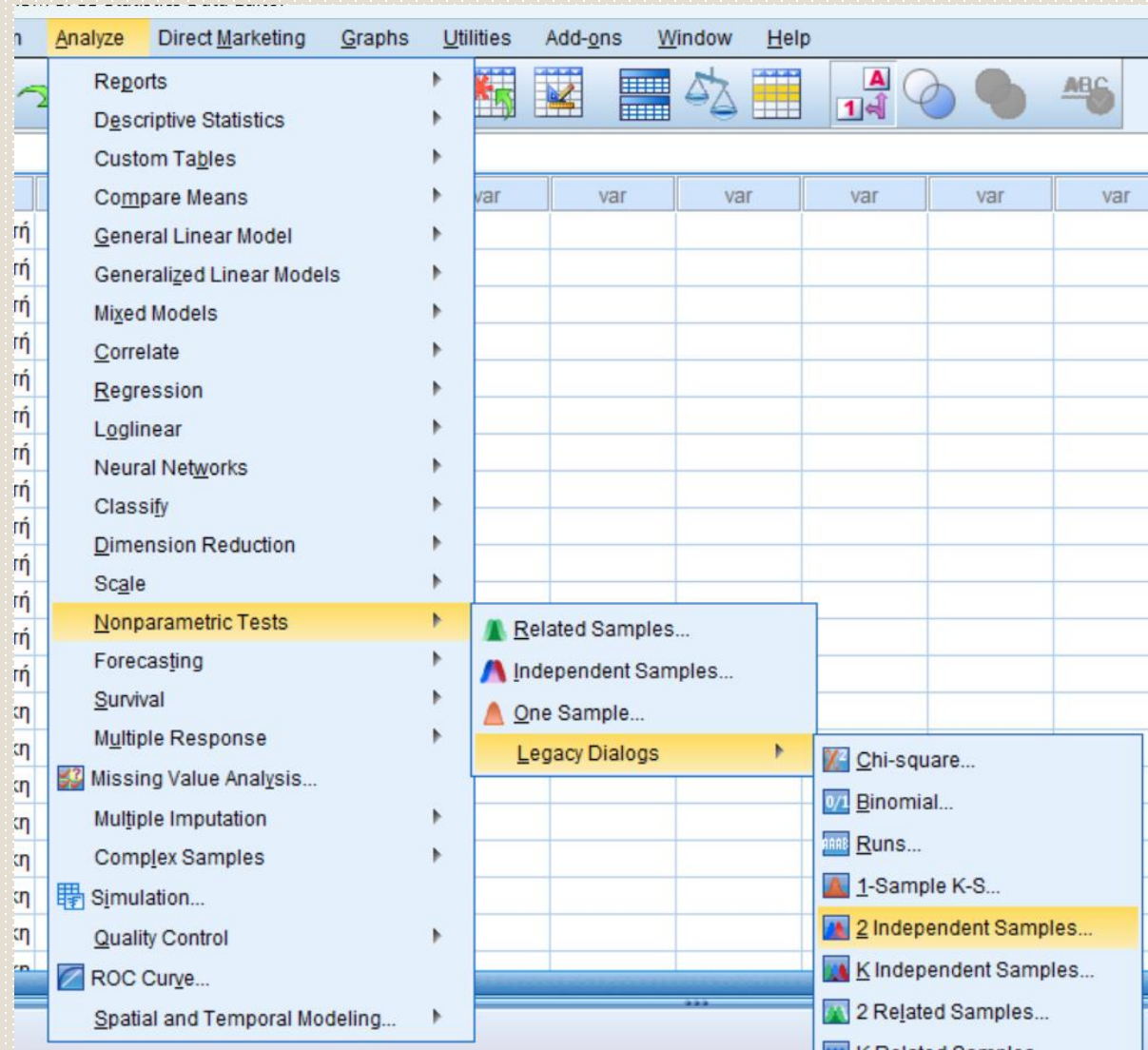
Προαιρετικά μπορούμε να βάλουμε ετικέτες με τη ρύθμιση values στις τιμές: 1: Λεπτή 2: Παχύσαρκη



	Energy	INDEX	val
1	6,13	Λεπτή	
2	7,05	Λεπτή	
3	7,48	Λεπτή	
4	7,48	Λεπτή	
5	7,53	Λεπτή	
6	7,58	Λεπτή	
7	7,90	Λεπτή	
8	8,08	Λεπτή	
9	8,09	Λεπτή	
10	8,11	Λεπτή	
11	8,40	Λεπτή	
12	10,15	Λεπτή	
13	10,88	Λεπτή	
14	8,79	Παχύσαρκη	
15	9,19	Παχύσαρκη	
16	9,21	Παχύσαρκη	
17	9,68	Παχύσαρκη	
18	9,69	Παχύσαρκη	
19	9,97	Παχύσαρκη	
20	11,51	Παχύσαρκη	
21	11,85	Παχύσαρκη	

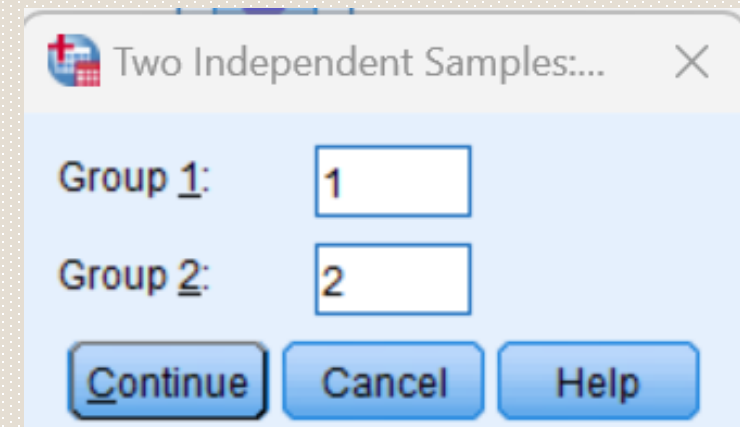
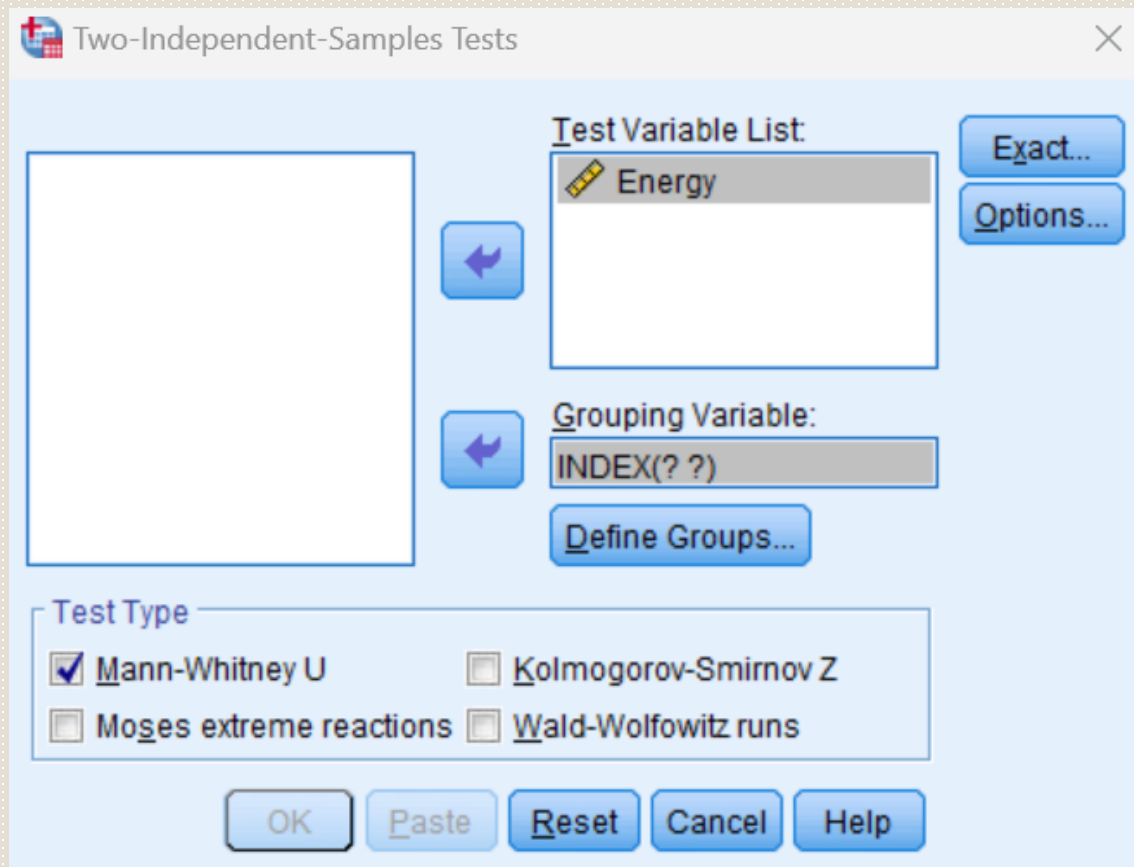
Η διαδρομή είναι:

Analyze / Nonparametric Tests / Legacy Dialogs / 2 Independent Samples



Στο **Test Variable List** τοποθετήστε την ποσοτική μεταβλητή (Energy) και στο **Grouping Variable** τοποθετήστε τη διχοτόμο μεταβλητή (INDEX)

Επιλέγουμε **Define Groups**, οπότε προκύπτει το πλαίσιο διαλόγου Define Groups. Βάζουμε τις τιμές 1 και 2 στα κελιά, πατάμε **Continue** και τέλος **OK**.



Τα αποτελέσματα του ελέγχου φαίνονται στον πίνακα Test Statistics

Στη γραμμή Z αναγράφεται η τιμή της στατιστικής συνάρτησης του ελέγχου Mann-Whitney.

Στη γραμμή Exact. Sig. [2*(1-tailed Sig.)] αναγράφεται η ακριβής τιμή του p-value αυτού (0,001) για το οποίο, όπως αναγράφεται στην υποσημείωση του πίνακα, δεν έχει γίνει διόρθωση για τις ισοπαλίες (ties).

Στη γραμμή Asymp. Sig. (2-tailed Sig.) αναγράφεται η ασυμπτωτική στατιστική σημαντικότητα του αμφίπλευρου ελέγχου (0,002), η οποία είναι πολύ κοντά στην (προηγούμενη) ακριβή τιμή της σημαντικότητας.

Παρατηρούμε ότι $p\text{-value} = 0,001 < 0,05$ οπότε

απορρίπτουμε τη μηδενική υπόθεση

(της ισότητας των πληθυσμιακών μέσων)

και συμπεραίνουμε ότι διαφέρει η κατανάλωση ενέργειας

Μεταξύ λεπτών και παχύσαρκων γυναικών

	Energy
Mann-Whitney U	12,000
Wilcoxon W	103,000
Z	-3,106
Asymp. Sig. (2-tailed)	,002
Exact Sig. [2*(1-tailed Sig.)]	,001 ^b

a. Grouping Variable: INDEX

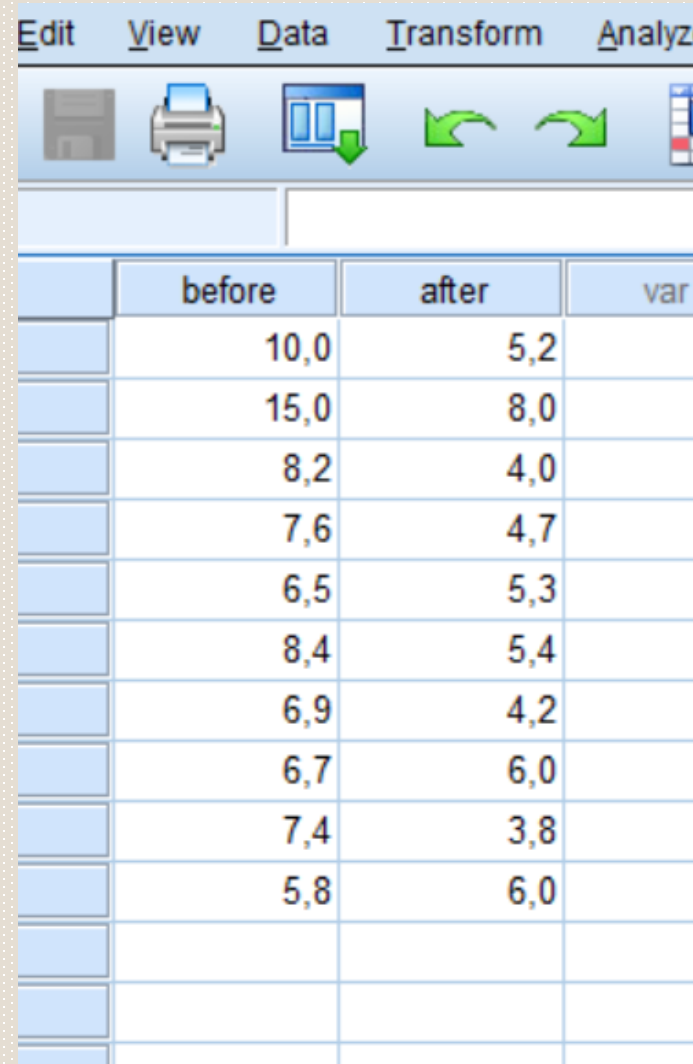
b. Not corrected for ties.

2.3.2 Έλεγχος Wilcoxon

Άσκηση 2.9: Σε 10 ασθενείς με κινητικά προβλήματα λόγω εγκεφαλικού επεισοδίου, εφαρμόστηκε ένα νέο πρόγραμμα φυσικοθεραπειών. Μετρήθηκε ο χρόνος (σε δευτερόλεπτα) που χρειάστηκε ο κάθε ασθενής να μεταφέρει ένα αντικείμενο από μία θέση σε μία άλλη. Εφαρμόστηκε το νέο πρόγραμμα και μετά την ολοκλήρωση του ξαναέγινε η ίδια μέτρηση. Στον πίνακα που ακολουθεί βλέπετε το χρόνο του κάθε ασθενή πριν και μετά τις φυσικοθεραπείες. Είναι αποτελεσματικό το πρόγραμμα;

Χρόνος πριν	Χρόνος μετά
10	5,2
15	8
8,2	4
7,6	4,7
6,5	5,3
8,4	5,4
6,9	4,2
6,7	6
7,4	3,8
5,8	6

- Εφαρμόζοντας έλεγχο κανονικότητας διαπιστώνουμε ότι και τα δύο δείγματα δεν ακολουθούν κανονική κατανομή. Επιπλέον δείγματα είναι εξαρτημένα, αφού αφορούν μετρήσεις από τον ίδιο ασθενή. Επομένως θα εφαρμόσουμε έλεγχο Wilcoxon
- Για την εφαρμογή του ελέγχου, πρέπει να καταχωρούμε τα δείγματα σε δύο διαφορετικές στήλες. Επομένως ορίζουμε δυο ποσοτικές μεταβλητές (scale) με name: before και after και καταχωρούμε τις τιμές στο Data View.

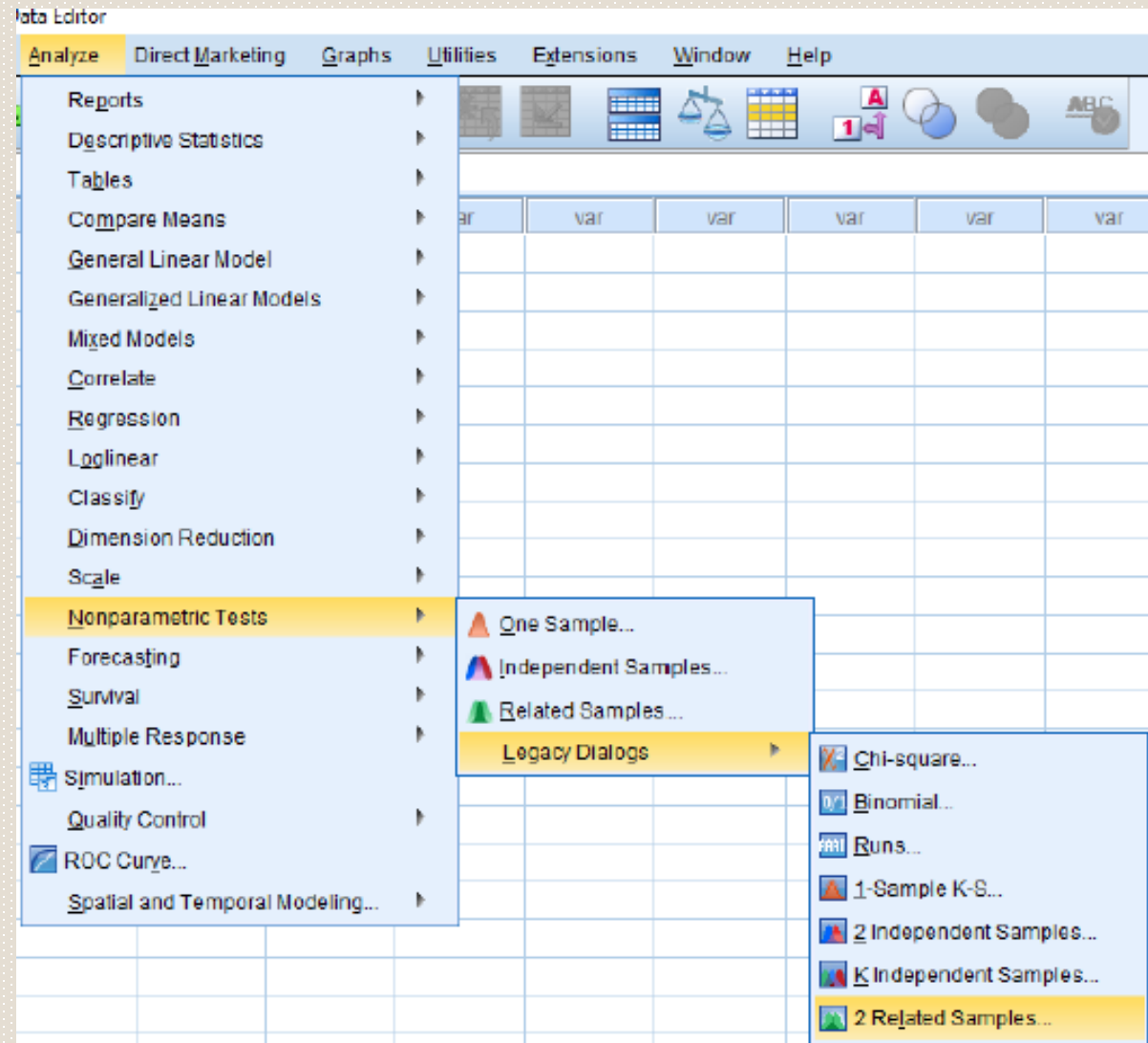


The screenshot shows the SPSS Data View window. The menu bar includes Edit, View, Data, Transform, and Analyze. The toolbar contains icons for saving, printing, opening a file, and undo/redo. The data table has three columns: 'before', 'after', and 'var'. The 'before' column contains values: 10,0, 15,0, 8,2, 7,6, 6,5, 8,4, 6,9, 6,7, 7,4, 5,8. The 'after' column contains values: 5,2, 8,0, 4,0, 4,7, 5,3, 5,4, 4,2, 6,0, 3,8, 6,0. The 'var' column is empty.

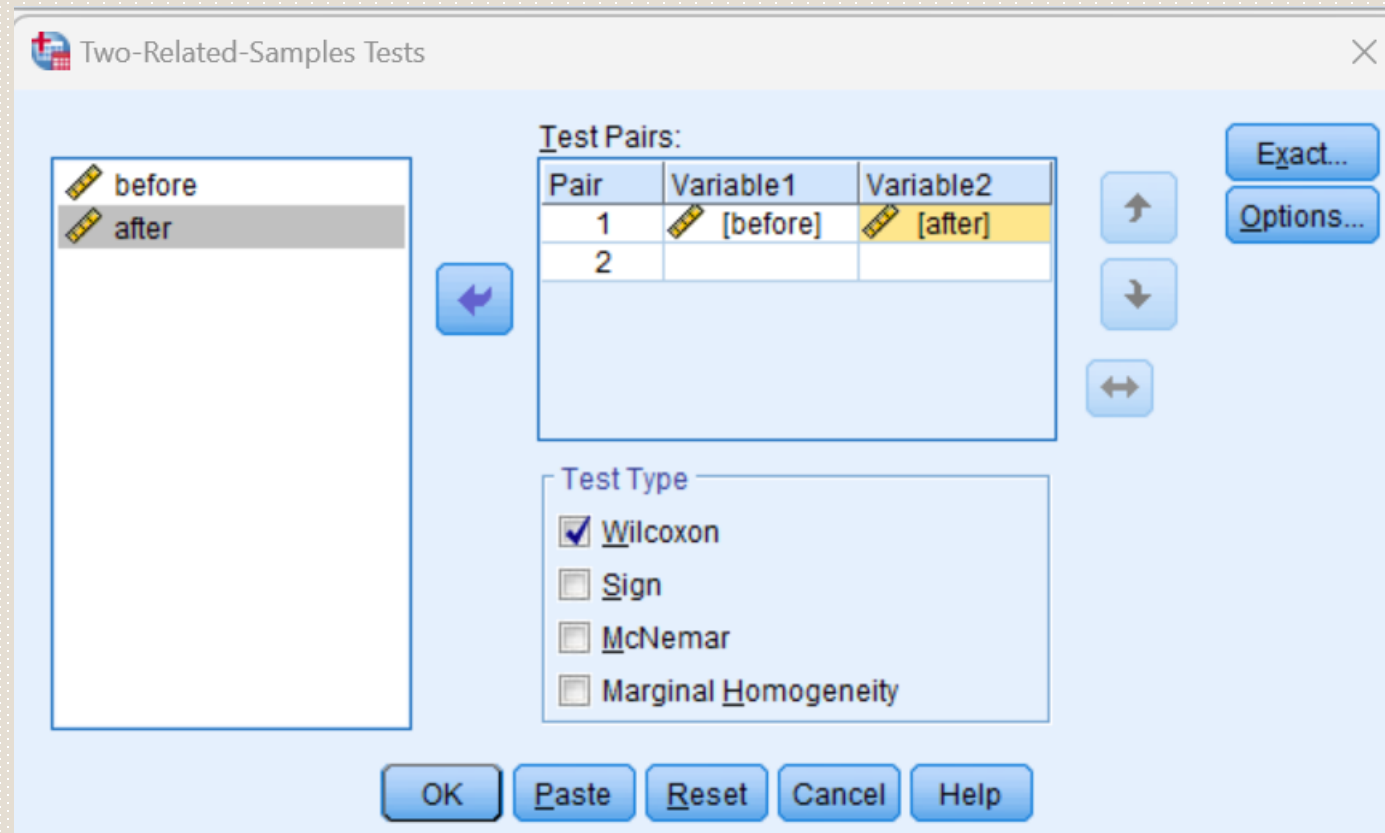
	before	after	var
	10,0	5,2	
	15,0	8,0	
	8,2	4,0	
	7,6	4,7	
	6,5	5,3	
	8,4	5,4	
	6,9	4,2	
	6,7	6,0	
	7,4	3,8	
	5,8	6,0	

Η διαδρομή στη συνέχεια είναι:

Analyze / Nonparametric Tests / Legacy Dialogs / 2 Related Samples



- Στο πλαίσιο διαλόγου Paired Samples T Test,
- Στο **Test Pairs**, στη στήλη Variable 1 τοποθετούμε την ποσοτική μεταβλητή που αφορά στο πρώτο «δείγμα»(before) και στη στήλη Variable 2 τοποθετούμε την ποσοτική μεταβλητή που αφορά στο δεύτερο «δείγμα» (after). Πατάμε **OK**.



Το αποτέλεσμα του ελέγχου φαίνεται στον πίνακα Test Statistics, στον οποίο βλέπουμε τη τιμή της στατιστικής συνάρτησης $Z=-2,7$ και την τιμή του p -value:

$$\text{Sig}=0,007 < 0,05$$

Επομένως, απορρίπτουμε τη Μηδενική υπόθεση κι επομένως η μέση τιμή χρόνου διαφέρει πριν και μετά τις φυσικοθεραπείες. Δηλαδή το πρόγραμμα φυσικοθεραπειών είναι αποτελεσματικό

	after - before
Z	-2,701 ^b
Asymp. Sig. (2-tailed)	,007

a. Wilcoxon Signed Ranks Test
b. Based on positive ranks.

2.4 Έλεγχος χ^2 - Έλεγχος Ανεξαρτησίας δυο κατηγορικών μεταβλητών

Ο έλεγχος ανεξαρτησίας χ^2 εφαρμόζεται για να εξετάσουμε τη συνάφεια μεταξύ δύο ποιοτικών μεταβλητών, αν δηλαδή οι δύο μεταβλητές ενός πίνακα διασταύρωσης είναι ανεξάρτητες μεταξύ τους (αν οι μεταβολές στις τιμές της μιας δεν προκαλούν μεταβολές στις τιμές της άλλης).

Οι υποθέσεις που ελέγχουμε είναι οι εξής:

- H_0 : Δεν υπάρχει σχέση ανάμεσα στις δύο μεταβλητές και ως εκ τούτου είναι ανεξάρτητες
- H_1 : Υπάρχει σχέση ανάμεσα στις δύο μεταβλητές και άρα ΔΕΝ είναι ανεξάρτητες

Η πραγματοποίηση του ελέγχου στο SPSS, γίνεται με ταυτόχρονη κατασκευή του πίνακα διασταύρωσης (Crosstab).

Analyze / Descriptive Statistics / Crosstabs / Statistics / Chi-square

Άσκηση 2.10: Ο παρακάτω πίνακας παρουσιάζει την κατανομή 80 καρκινοπαθών (με καρκίνο του μαστού) και 160 “υγιών” κατά οικογενειακή κατάσταση. Σχετίζεται ο καρκίνος του μαστού με την οικογενειακή κατάσταση;

		Οικογενειακή κατάσταση		
		Ανύπαντρες	Παντρεμένες	Άλλα
Καρκίνος του μαστού	Ναι	30	44	6
	Όχι	30	106	24

Λύση

Θέλουμε να εξετάσουμε αν σχετίζονται δυο ποιοτικές μεταβλητές επομένως θα εφαρμόσουμε έλεγχο χ^2 . Οι υποθέσεις:

- H_0 : Ο καρκίνος μαστού και η οικογ. Κατάσταση **δεν** σχετίζονται
- H_1 : Ο καρκίνος μαστού και η οικογ. Κατάσταση **σχετίζονται**

Στο Variable View δημιουργούμε 2 ordinal μεταβλητές με ονόματα *Status* και *Canser* και μία ποσοτική με το όνομα *Women* η οποία περιέχει τη συχνότητα του κάθε κελιού. Με το values δίνουμε ετικέτες στις τιμές των ποιοτικών : Στο Status:

1-Ανύπαντρη , 2- Παντρεμένη 3-Άλλο

ενώ στο Canser:

1- Ναι, 2-Όχι

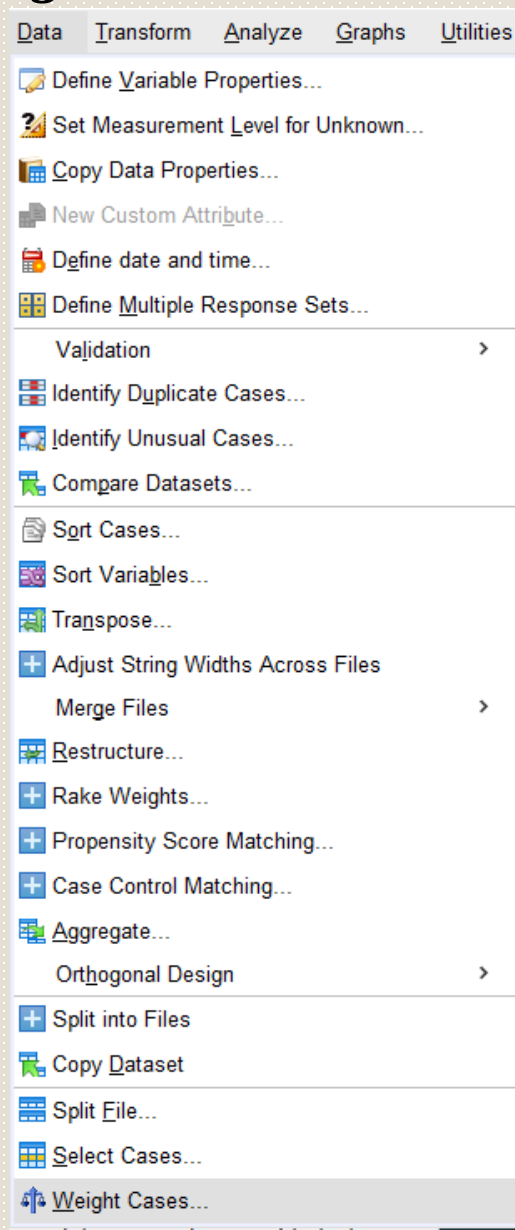
Στη συνέχεια στο Data View καταχωρούμε τις τιμές . Κάθε οριζόντια γραμμή αναφέρεται σε ένα κελί του πίνακα. Π.χ η πρώτη γραμμή αναφέρεται στο κελί της 1^{ης} γραμμής και 1^{ης} στήλης του πίνακα. Οπότε καταχωρούμε τις τιμές :

Status: 1, Canser 1, Women: 30

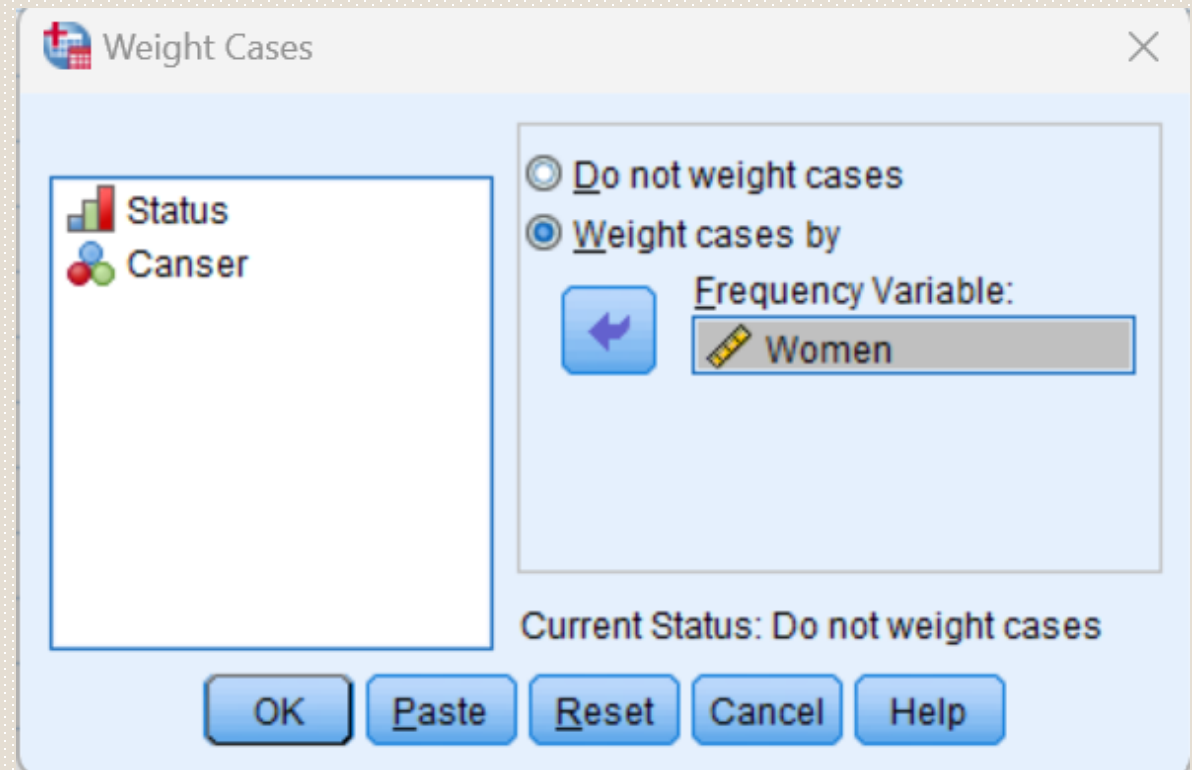
Συνεχίζουμε για όλα τα κελιά του πίνακα, οπότε στυο Data View συμπληρώνονται 6 γραμμές (αφου ο πίνακας έχει 6 κελιά).

	Status	Canser	Women
	Ανύπαντρη	Yes	30
	Παντρεμένη	Yes	44
	Άλλο	Yes	6
	Ανύπαντρη	No	30
	Παντρεμένη	No	106
	Άλλο	No	24

Στη συνέχεια επιλέγουμε:
Data / Weight Cases

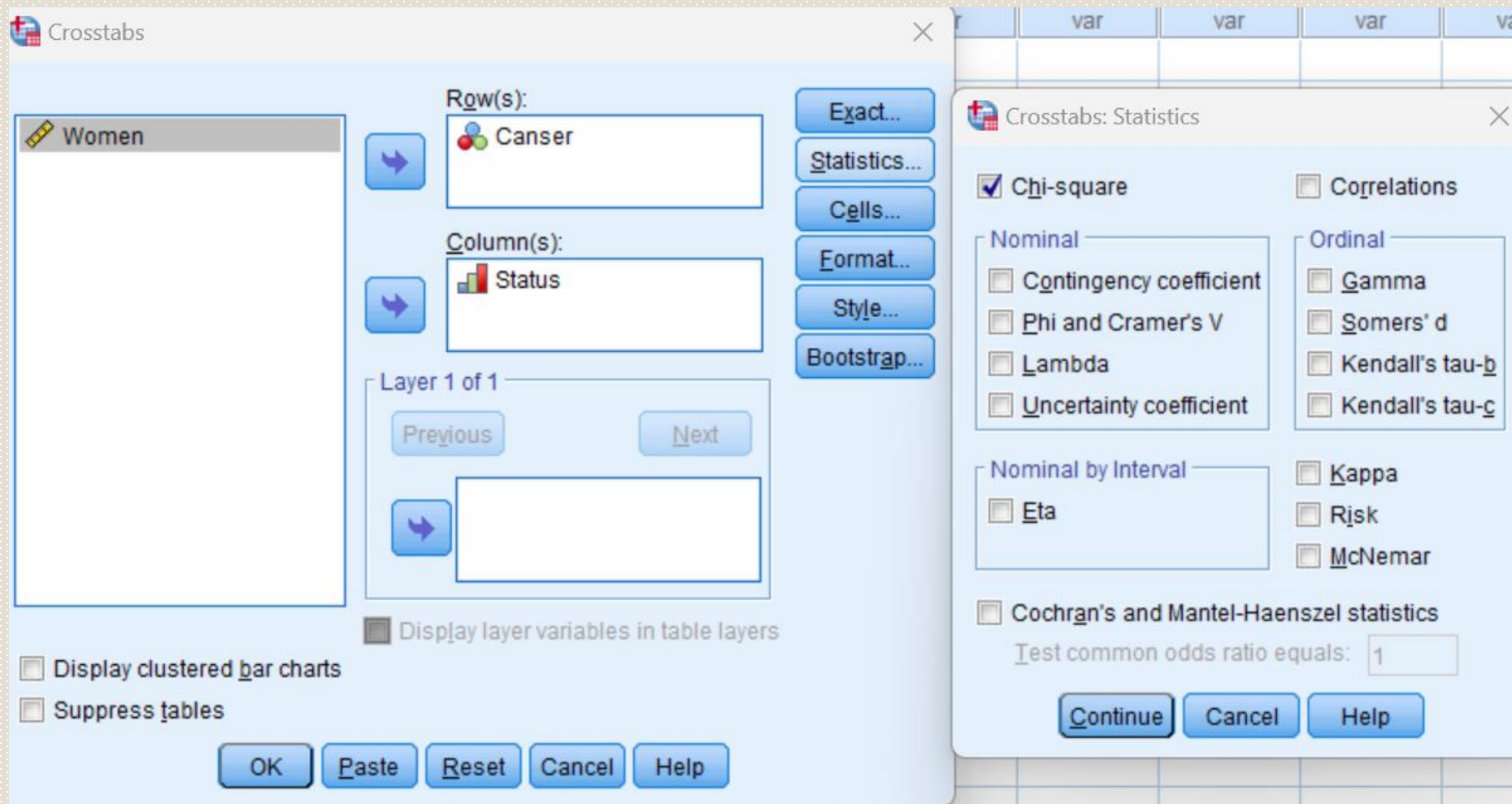


Στο εμφανιζόμενο παράθυρο διαλέγουμε την ενέργεια **Weight cases by** και στη θέση **Frequency Variable** τοποθετούμε τη ποσοτική μεταβλητή **Women**. Επικυρώνουμε τις επιλογές μας με **OK**.



• Στη συνέχεια, **Analyze** → **Descriptive Statistics** → **Crosstabs**

Μεταφέρουμε στο Rows τη μία ποιοτική μεταβλητή (π.χ Canser) και στο Columns την άλλη (Status) και από το **Statistics** επιλέγουμε **Chi-square**. Πατάμε **Continue** και τέλος **OK**.



Στα αποτελέσματα, ο 2^{ος} πίνακας είναι ο διασταυρωμένος πίνακας (Crosstab)

Status * Cancer Crosstabulation

Count

		Cancer		Total
		Yes	No	
Status	Ανύπαντρη	30	30	60
	Παντρεμένη	44	106	150
	Άλλο	6	24	30
Total		80	160	240

Τα αποτελέσματα του ελέγχου φαίνονται στον 3^ο πίνακα: **Chi-Square Tests**

Αυτό που μας ενδιαφέρει είναι η τιμή του **Pearson Chi-Square**, ενώ στο κάτω μέρος του πίνακα υπάρχει μία υποσημείωση που αφορά στις προϋποθέσεις εφαρμογής του τεστ. Επίσης στο Value φαίνεται η τιμή της στατιστικής συνάρτησης: $X^2 = 10,980$ ενώ το $df = 2$ είναι οι βαθμοί ελευθερίας.

Η τιμή του *p-value* φαίνεται στη

3^η στήλη. Εδώ έχουμε :

Asymptotic Significance = 0,004 < 0,05

Επομένως απορρίπτουμε την H_0 και

Άρα οι μεταβλητές

δεν είναι ανεξάρτητες.

Δηλαδή ο καρκίνος μαστού σχετίζεται με την οικογενειακή κατάσταση.

Chi-Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	10,980 ^a	2	,004
Likelihood Ratio	10,792	2	,005
Linear-by-Linear Association	10,183	1	,001
N of Valid Cases	240		

a. 0 cells (0,0%) have expected count less than 5. The minimum expected count is 10,00.

Άσκηση 2.11: Στον διπλανό πίνακα βλέπετε την επιλογή θεραπείας 20 ασθενών με πόνο στο γόνατο (1=φαρμακευτική αγωγή, 2=εγχείρηση, 3=φυσικοθεραπείες) και το μορφωτικό τους επίπεδο (1=Λύκειο, 2=Τριτοβάθμια, 3=Μεταπτυχιακό/Διδακτορικό). Σχετίζεται η επιλογή θεραπείας με το μορφωτικό επίπεδο;

ΜΟΡΦΩΤΙΚΟ ΕΠΙΠΕΔΟ	ΘΕΡΑΠΕΙΑ
1	1
3	3
2	2
2	2
2	1
1	3
1	3
2	3
2	1
2	2
3	3
1	3
2	2
1	3
2	2
1	1
3	3
2	3
2	2
2	2

• Λύση

Στο Variable View δημιουργώ τις μεταβλητές EDUCATION και THERAPY (και οι δύο scale)

Με τη ρύθμιση Values βάζω ετικέτες στις Μεταβλητές:

Για την EDUCATION:

1=Λύκειο,

2=Τριτοβάθμια,


3= Μεταπτυχιακό/Διδακτορικό

Και για τη THERAPY:

1=φαρμακευτική αγωγή

2=εγχείρηση

3=φυσικοθεραπείες



The screenshot shows the SPSS Variable View window. The 'EDUCATION' variable is defined with 12 values, and the 'THERAPY' variable is defined with 12 values. The values are as follows:

EDUCATION	THERAPY
ΜΕΤΑΠΤΥΧΙΑΚΟ/ΔΙΔΑΚΤΟΡΙ...	ΦΥΣΙΚΟΘΕΡΑΠΕΙΕΣ
ΤΡΙΤΟΒΑΘΜΙΑ	ΕΓΧΕΙΡΗΣΗ
ΤΡΙΤΟΒΑΘΜΙΑ	ΕΓΧΕΙΡΗΣΗ
ΤΡΙΤΟΒΑΘΜΙΑ	ΦΑΡΜΑΚΟ
ΛΥΚΕΙΟ	ΦΥΣΙΚΟΘΕΡΑΠΕΙΕΣ
ΛΥΚΕΙΟ	ΦΥΣΙΚΟΘΕΡΑΠΕΙΕΣ
ΤΡΙΤΟΒΑΘΜΙΑ	ΦΥΣΙΚΟΘΕΡΑΠΕΙΕΣ
ΤΡΙΤΟΒΑΘΜΙΑ	ΦΑΡΜΑΚΟ
ΤΡΙΤΟΒΑΘΜΙΑ	ΕΓΧΕΙΡΗΣΗ
ΜΕΤΑΠΤΥΧΙΑΚΟ/ΔΙΔΑΚΤΟΡΙ...	ΦΥΣΙΚΟΘΕΡΑΠΕΙΕΣ
ΛΥΚΕΙΟ	ΦΥΣΙΚΟΘΕΡΑΠΕΙΕΣ
ΤΡΙΤΟΒΑΘΜΙΑ	ΕΓΧΕΙΡΗΣΗ
ΛΥΚΕΙΟ	ΦΥΣΙΚΟΘΕΡΑΠΕΙΕΣ
ΤΡΙΤΟΒΑΘΜΙΑ	ΕΓΧΕΙΡΗΣΗ
ΛΥΚΕΙΟ	ΦΑΡΜΑΚΟ
ΜΕΤΑΠΤΥΧΙΑΚΟ/ΔΙΔΑΚΤΟΡΙ...	ΦΥΣΙΚΟΘΕΡΑΠΕΙΕΣ
ΤΡΙΤΟΒΑΘΜΙΑ	ΦΥΣΙΚΟΘΕΡΑΠΕΙΕΣ
ΤΡΙΤΟΒΑΘΜΙΑ	ΕΓΧΕΙΡΗΣΗ
ΤΡΙΤΟΒΑΘΜΙΑ	ΕΓΧΕΙΡΗΣΗ

Όπως βλέπουμε, έχουμε διαφορετικό τρόπο καταχώρησης των δεδομένων σε σχέση με τη προηγούμενη άσκηση. Με αυτό τον τρόπο εισαγωγής δεδομένων παραλείπουμε το βήμα στάθμισης: *Data / Weight Cases*

Πηγαίνουμε κατευθείαν: **Analyze → Descriptive Statistics → Crosstabs**

Μεταφέρουμε στο Rows τη μία ποιοτική μεταβλητή (π.χ EDUCATION) και στο Columns την άλλη (THERAPY) και από το **Statistics** επιλέγουμε **Chi-square**. Πατάμε **Continue** και τέλος **OK**.

Το αποτέλεσμα του ελέγχου φαίνεται στον 3^ο πίνακα : Chi – Square Tests

	Value	df	Asymptotic Significance (2-sided)
Pearson Chi-Square	11,279 ^a	4	,024
Likelihood Ratio	14,342	4	,006
Linear-by-Linear Association	,579	1	,447
N of Valid Cases	20		

a. 9 cells (100,0%) have expected count less than 5. The minimum expected count is ,60.

Η τιμή του *p-value* είναι: $0,024 < 0,05$

Επομένως απορρίπτεται η Μηδενική υπόθεση κι άρα η θεραπεία σχετίζεται με το μορφωτικό επίπεδο.

Παρατήρηση: Η υποσημείωση κάτω απ τον πίνακα, μας λέει ότι δεν ληρούνται οι προϋποθέσεις εφαρμογής του ελέγχου.

2.5 Γραμμική συσχέτιση δύο ποσοτικών μεταβλητών

Η ύπαρξη γραμμικής σχέσης μεταξύ δύο ποσοτικών μεταβλητών, εξακριβώνεται με τους ελέγχους υπόθεσης :

- του **συντελεστή συσχέτισης Pearson** (για κανονικά δεδομένα)
- του **συντελεστή συσχέτισης Spearman** (για ποσοτικές που τουλάχιστον μία δεν ακολουθεί κανονική κατανομή ή για την περίπτωση που μια ή και οι δυο μεταβλητές είναι σε διατεταγμένη κλίμακα)
- Οι υποθέσεις του ελέγχου είναι:

H_0 : Οι μεταβλητές μας δεν συσχετίζονται γραμμικά

H_1 : Οι μεταβλητές μας συσχετίζονται γραμμικά

Για να υπολογίσετε τους συντελεστές συσχέτισης του Pearson, του Spearman από την μπάρα με τις βασικές επιλογές του SPSS επιλέξτε:

Analyze / Correlate / Bivariate

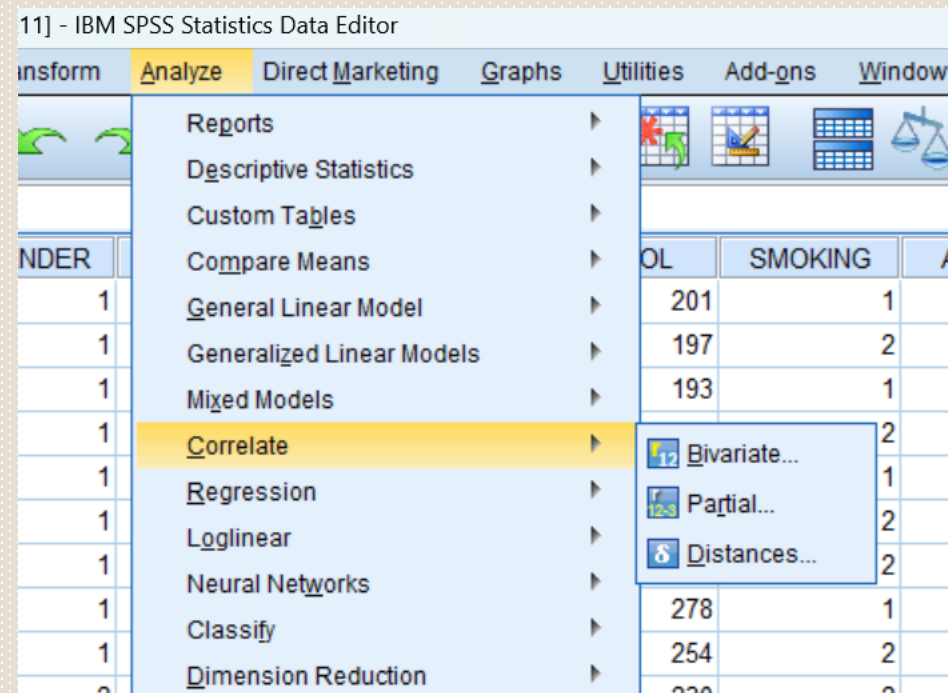
Άσκηση 2.12: Να εξετάσετε αν συσχετίζονται γραμμικά το 'Βάρος' και ο 'Αθηρωματικός Δείκτης' χρησιμοποιώντας τα δείγματα της άσκησης 1.4.

Λύση:

Εξετάζουμε τα δύο δείγματα ως προς τη κανονικότητα και ο έλεγχος K-S δείχνει ότι και οι δύο ακολουθούν κανονική κατανομή. Επομένως ο κατάλληλος έλεγχος είναι ο Pearson.

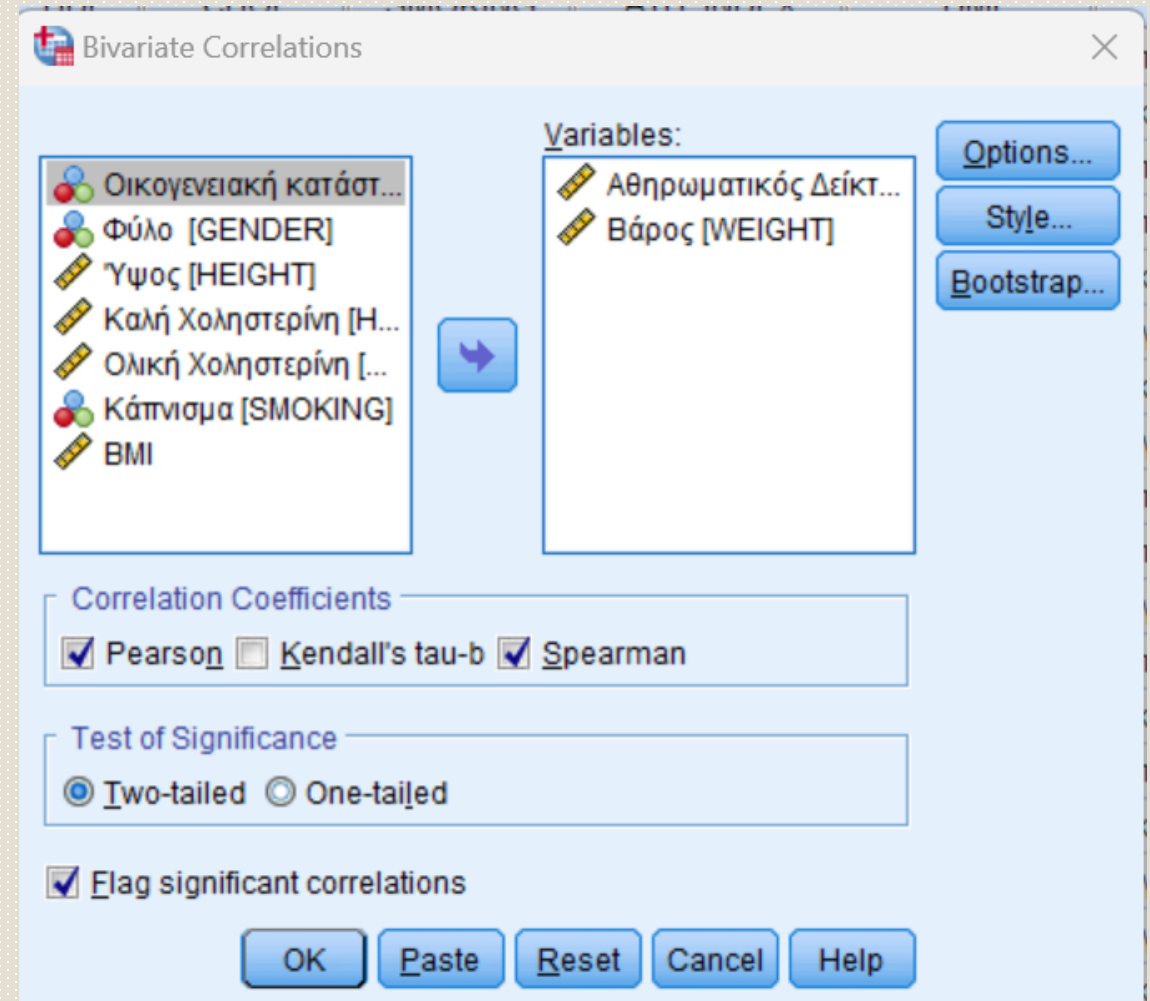
Ακολουθούμε τη διαδρομή:

Analyze / Correlate / Bivariate



Στο *Variables* τοποθετούμε τις μεταβλητές για τις οποίες θέλουμε να υπολογίσουμε τους συντελεστές συσχέτισης, ενώ στο *Correlation Coefficients* επιλέγουμε τους συντελεστές συσχέτισης που θέλουμε να υπολογίσουμε. Επομένως επιλέγουμε *Pearson*

Έπειτα, επιλέγουμε **OK**, οπότε προκύπτει ο πίνακας Correlations:



Ο πίνακας των αποτελεσμάτων είναι ο ίδιος και στην περίπτωση του συντελεστή συσχέτισης του Spearman.

Στη σειρά Pearson Correlation παρουσιάζεται ο συντελεστής συσχέτισης του Pearson και στη σειρά Sig. (2 tailed) παρουσιάζεται η τιμή *p-value*

Correlations			
		Αθρωματικός Δείκτης	Βάρος
Αθρωματικός Δείκτης	Pearson Correlation	1	,692**
	Sig. (2-tailed)		,001
	N	20	20
Βάρος	Pearson Correlation	,692**	1
	Sig. (2-tailed)	,001	
	N	20	20

** . Correlation is significant at the 0.01 level (2-tailed).

Σύμφωνα με τ' αποτελέσματα, η τιμή του *p-value* είναι:

$$\text{Sig}=0,001 < 0,05$$

Επομένως απορρίπτεται η μηδενική υπόθεση , δηλαδή το βάρος και ο αθηρωματικός δείκτης συσχετίζονται.

Η τιμή του συντελεστή συσχέτισης Pearson είναι 0,692, γεγονός που σημαίνει ότι έχουν θετική γραμμική συσχέτιση.