Ακ. έτος | 2020-2021

Σημειώσεις μαθήματος

Στατιστικές μέθοδοι στην Κοινωνική Έρευνα ΙΙ

Π.Μ.Σ.
Κοινωνική Πολιτική: Μέθοδοι και Εφαρμογές

Μαρία Συμεωνάκη

Πάντειο Πανεπιστήμιο
Τμήμα Κοινωνικής Πολιτικής

Απλή Γραμμική Συσχέτιση και Παλινδρόμηση

Διαγράμματα διασποράς (Ποσοτικά δεδομένα)

Τα διαγράμματα διασποράς (Scatterplots), αποτελούν ένα γραφικό τρόπο αναζήτησης της σχέσης μεταξύ δύο συνεχών μεταβλητών (π.χ. ύψους και βάρους). Τα πιο συνηθισμένα διαγράμματα διασποράς είναι αυτά στα οποία κάθε παρατήρηση παριστάνεται σε ένα σύστημα αξόνων, όπου ο κάθε άξονας αντιστοιχεί σε μια μεταβλητή. Από τον τρόπο με τον οποίο κατανέμονται τα σημεία αυτά στο επίπεδο μπορούμε να πάρουμε πολλές πληροφορίες για τον τρόπο συσχέτισης των δύο μεταβλητών. Είναι γενικά καλό τα διαγράμματα διασποράς να γίνονται πριν από τον έλεγχο της συσχέτισης, γιατί τα διαγράμματα αυτά μας δίνουν μια ένδειξη για το αν υπάρχει γραμμική ή καμπυλόγραμμη σχέση μεταξύ των δύο μεταβλητών. Ο έλεγχος της γραμμικής συσχέτισης έχει νόημα μόνο αν οι μεταβλητές συνδέονται γραμμικά. Επιπλέον, τα διαγράμματα διασποράς θα μας δείξουν αν οι μεταβλητές σχετίζονται θετικά (που σημαίνει ότι όταν οι τιμές της μιας μεταβλητής αυξάνονται, αυξάνονται επίσης και οι τιμές της άλλης μεταβλητής) ή αρνητικά (όταν αυξάνονται οι τιμές της μιας μεταβλητής οι τιμές της άλλης μεταβλητής μειώνονται). Επίσης, μας δίνουν μια ένδειξη της έντασης της σχέσης μεταξύ των μεταβλητών. Αν η σχέση είναι ασθενής, τότε τα σημεία είναι διασκορπισμένα σε όλο το επίπεδο, ενώ αν η σχέση είναι ισχυρή, τότε τα σημεία βρίσκονται σε μια ζώνη γύρω από μια ευθεία. Για να γίνει ένα διάγραμμα διασποράς πρέπει να δηλωθούν τουλάχιστον δύο συνεχής μεταβλητές. Με το SPSS μπορούμε να δημιουργήσουμε είτε ένα απλό διάγραμμα διασποράς (standard Scatterplot), είτε ένα διάγραμμα διασποράς με αλληλεπίδραση (Interactive Scatterplot).

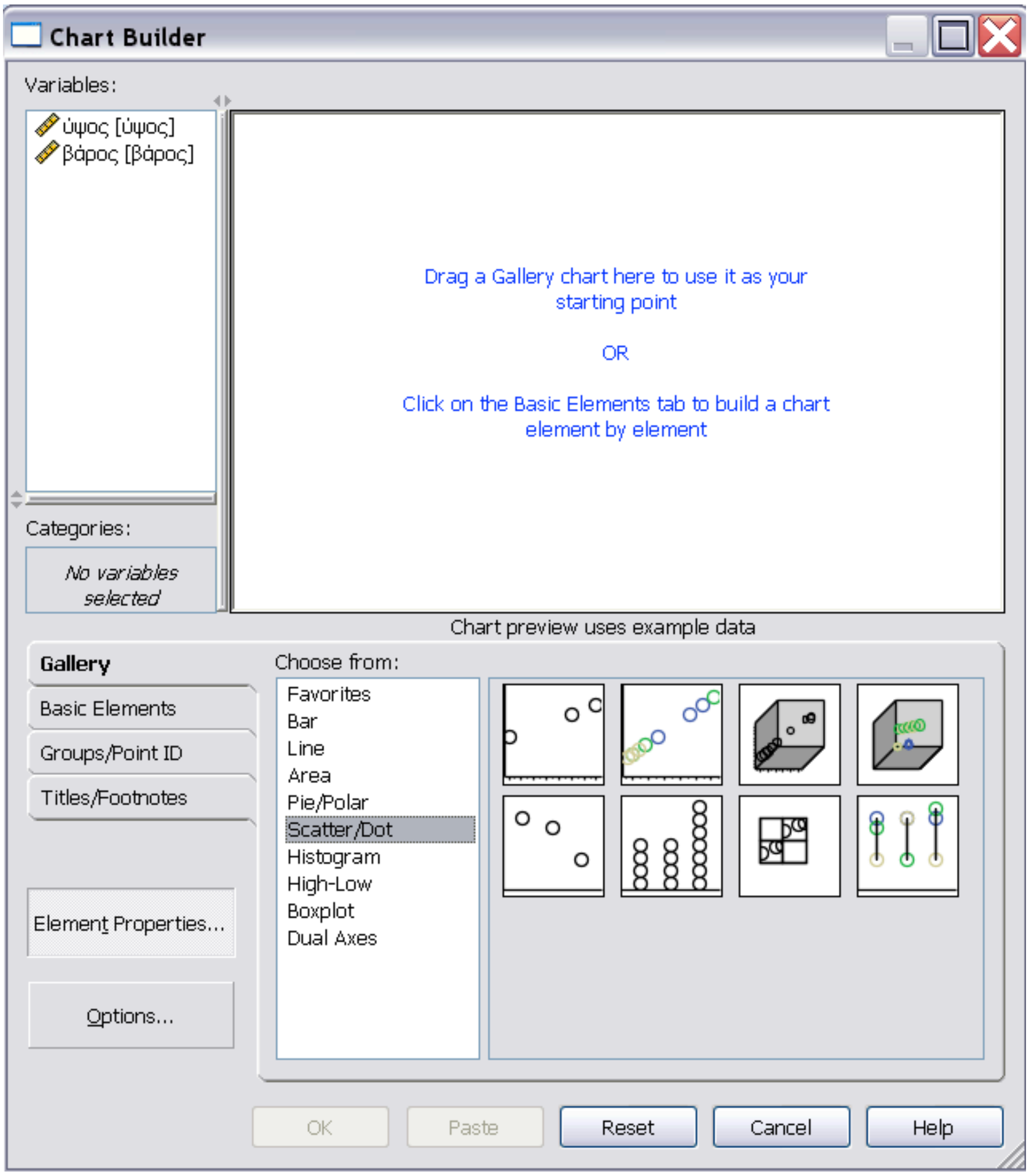
Απλό διάγραμμα διασποράς

Παράδειγμα 1:

Έστω ότι έχουμε τα παρακάτω δεδομένα και έστω ότι θέλουμε να εξετάσουμε αν οι μεταβλητές ύψος και βάρος σχετίζονται μεταξύ τους.

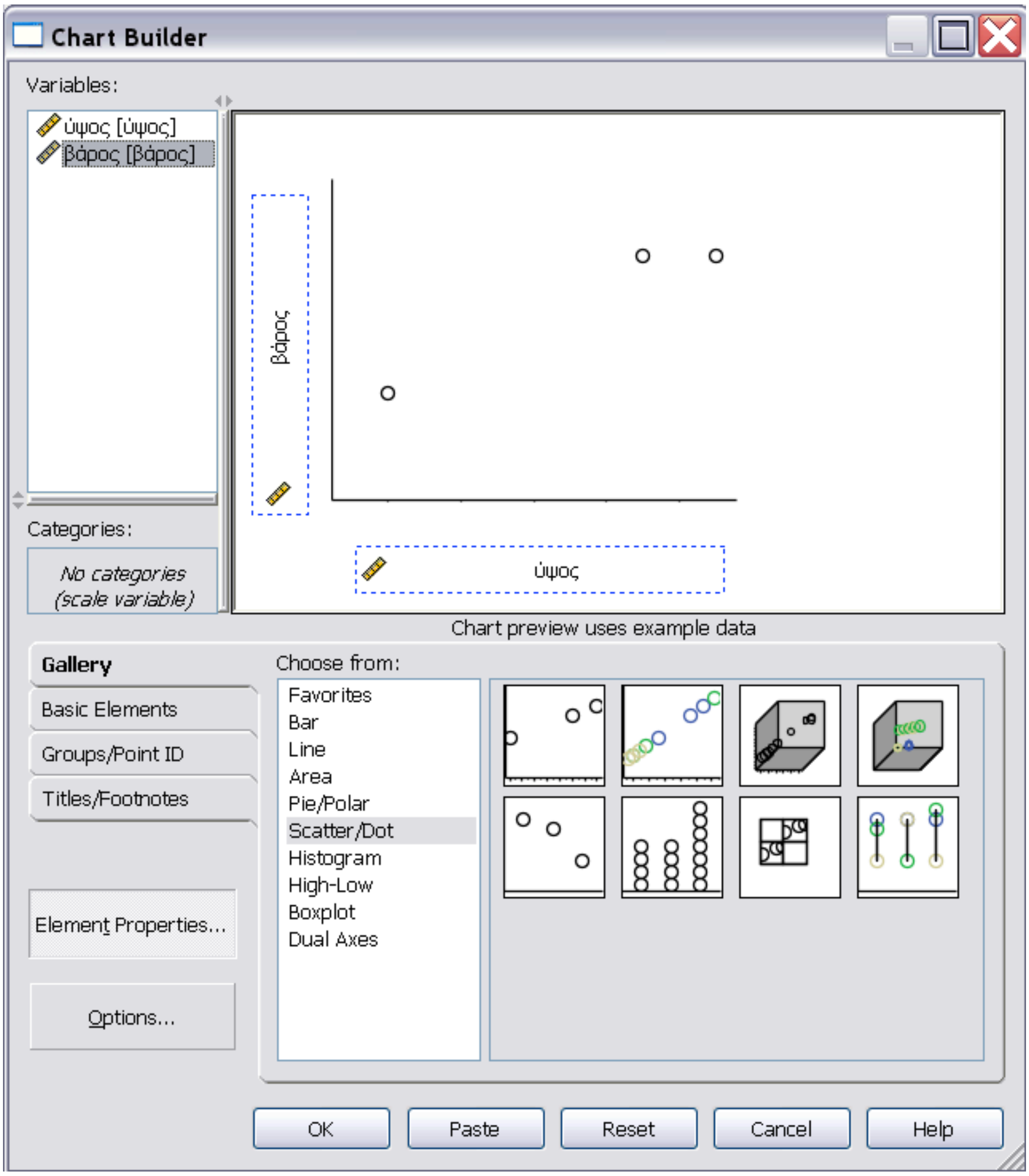
	ύψος	βάρος
1	150	45
2	145	43
3	167	60
4	145	40
5	165	61
6	132	40
7	180	70
8	156	55
9	170	65
10	154	45
11	167	54
12	156	45
13	159	50
14	182	72
15	167	65
16	171	67
17	156	45
18	178	72
19	159	56
20	164	58
21	166	62
22	178	73
23	171	65
24	159	54
25	166	53
26	167	57
27	180	75
28	175	68
29	172	59
30	164	57

Για να δημιουργήσουμε ένα απλό Scatterplot επιλέγουμε από τον Chart Builder το Scatter/Dot:



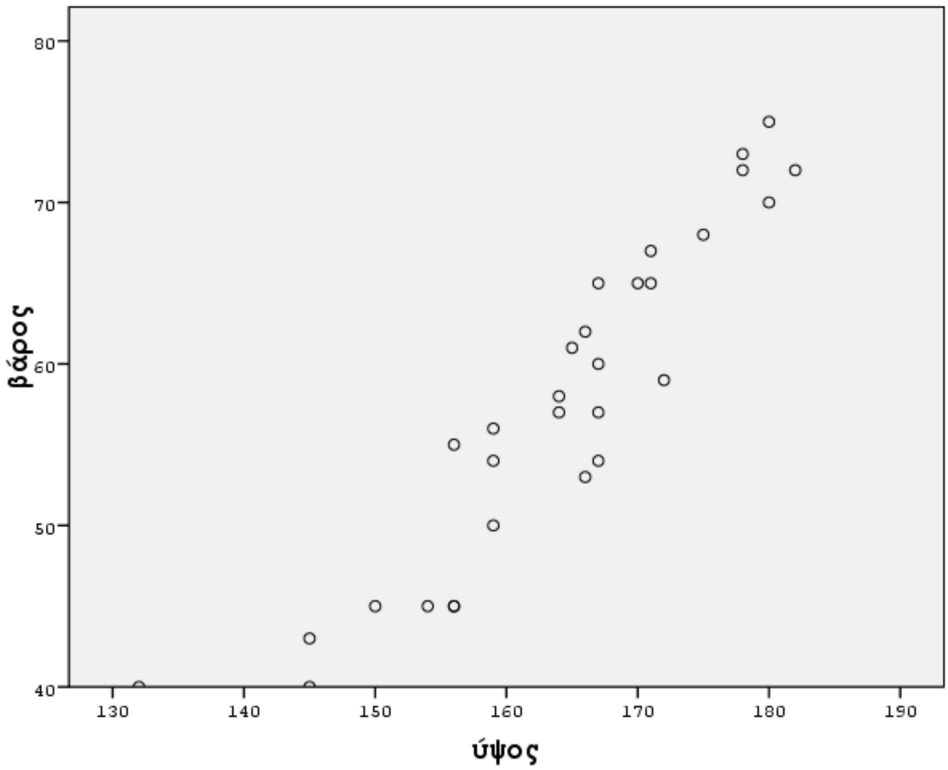
Εικόνα 1: Δημιουργία Scatterplot

Επιλέγουμε τη μορφή Simple και στη συνέχεια τη μεταφέρουμε στο πλαίσιο που βρίσκεται ακριβώς από πάνω στον Chart Builder (Εικόνα 1). Στον άξονα X τοποθετείται η **ανεξάρτητη μεταβλητή** και στον άξονα Y η **εξαρτημένη μεταβλητή**. Αν θεωρήσουμε ότι οι τιμές της μιας μεταβλητής (σε αυτό το παράδειγμα οι τιμές της μεταβλητής βάρος) εξαρτώνται από τις τιμές της άλλης (της μεταβλητής ύψος), τότε η πρώτη αναφέρεται ως εξαρτημένη και η δεύτερη ως ανεξάρτητη (Εικόνα 2).



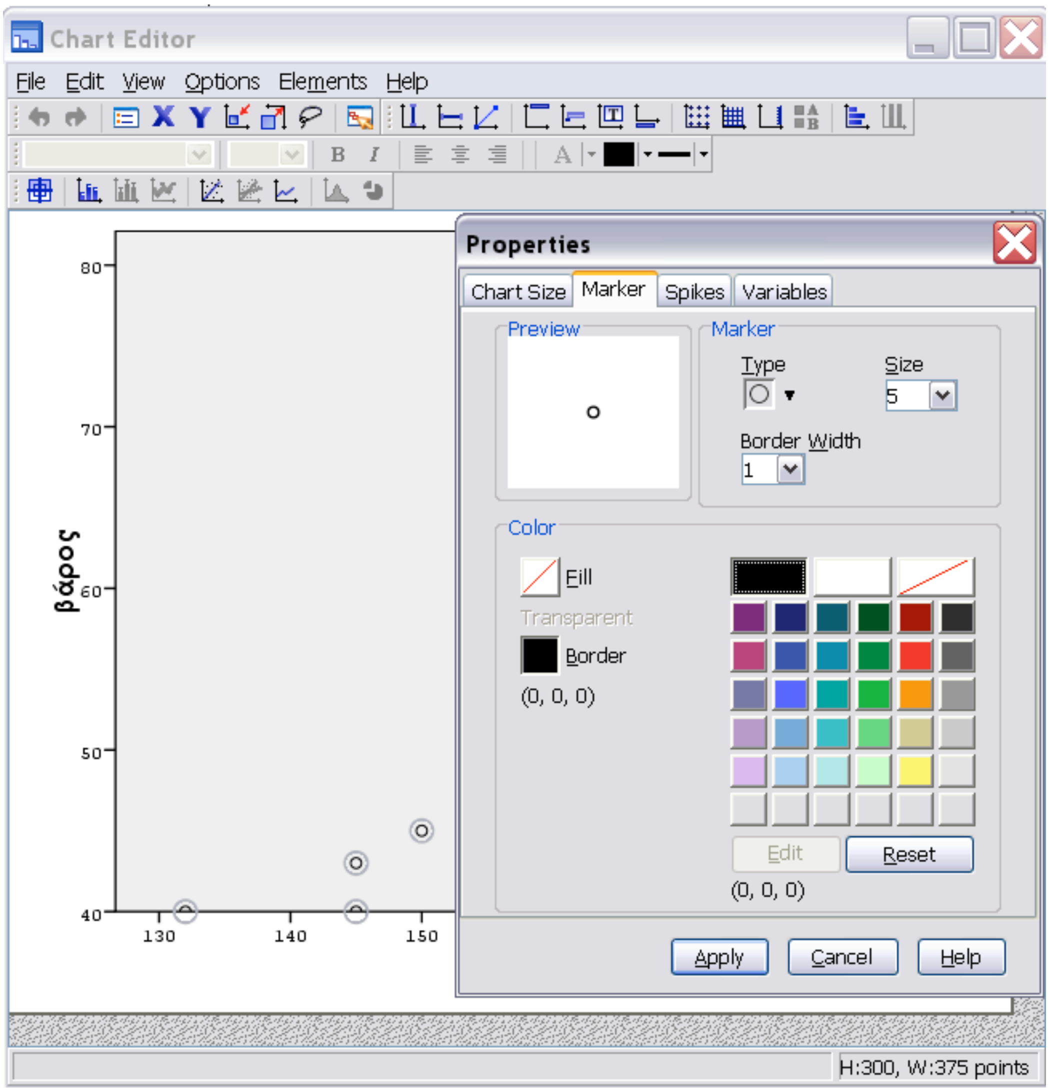
Εικόνα 2: Επιλογή είδους Scatterplot και εισαγωγή μεταβλητών

Στο παράθυρο διαλόγου Chart Builder πατάμε το OK και εμφανίζεται στον Output Viewer το παρακάτω διάγραμμα διασποράς (Εικόνα 3):



Εικόνα 3: Απλό διάγραμμα διασποράς

Από αυτό το διάγραμμα φαίνεται ότι τα σημεία δεν είναι τυχαία διεσπαρμένα στο επίπεδο. Πιο συγκεκριμένα, καθώς μεγαλώνει το βάρος, μεγαλώνει και το ύψος και η σχέση τους φαίνεται λίγο πολύ να είναι γραμμική. Επίσης, μας ενδιαφέρει να παρατηρήσουμε αν υπάρχουν κάποια ακραία ζεύγη τιμών στο διάγραμμά μας (εδώ π.χ. το ζεύγος $(132,40)$) και αν υπάρχουν να τα επανεξετάσουμε. Στη συνέχεια μπορούμε να μορφοποιήσουμε το διάγραμμά μας (να αλλάξουμε τον τύπο των σημείων, το χρώμα τους, το μέγεθός τους, κλπ, κάνοντας διπλό κλικ πάνω σε ένα σημείο), να προσθέσουμε τίτλο, κοκ (Εικόνα 4).

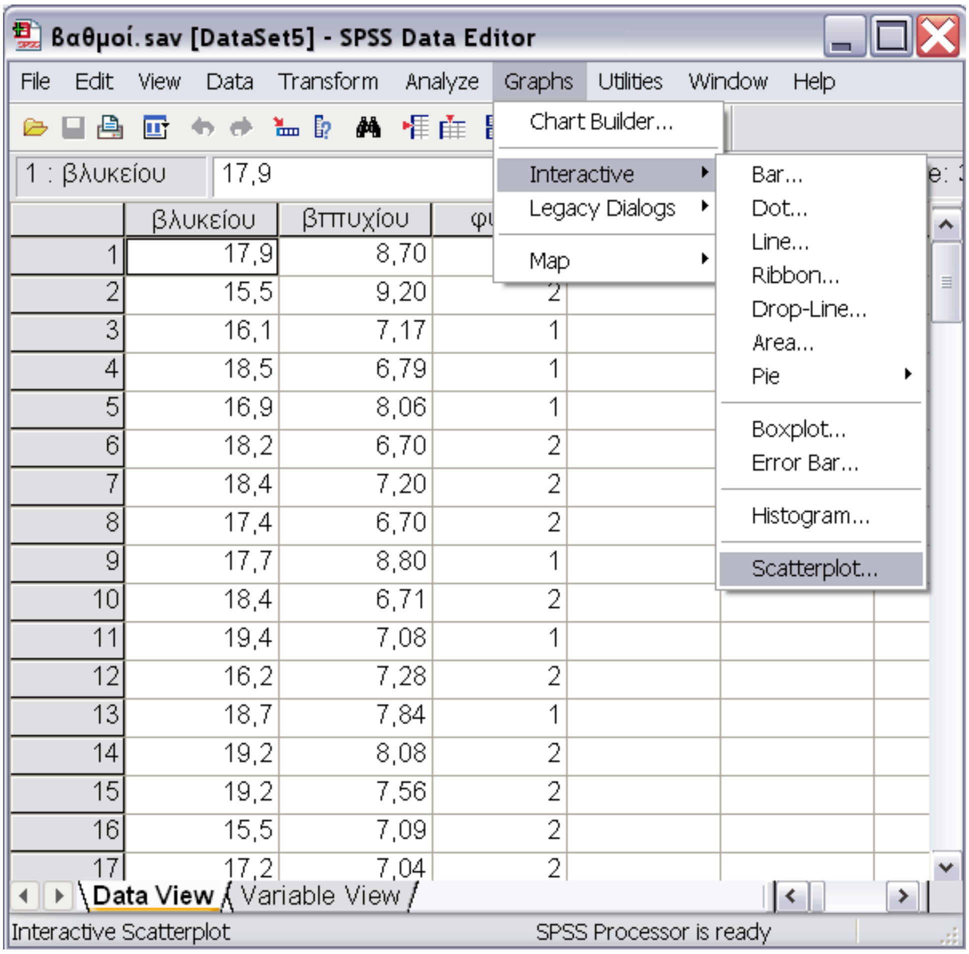


Εικόνα 4: Μορφοποίηση Scatterplot

Κατασκευή διαγράμματος διασποράς με αλληλεπίδραση

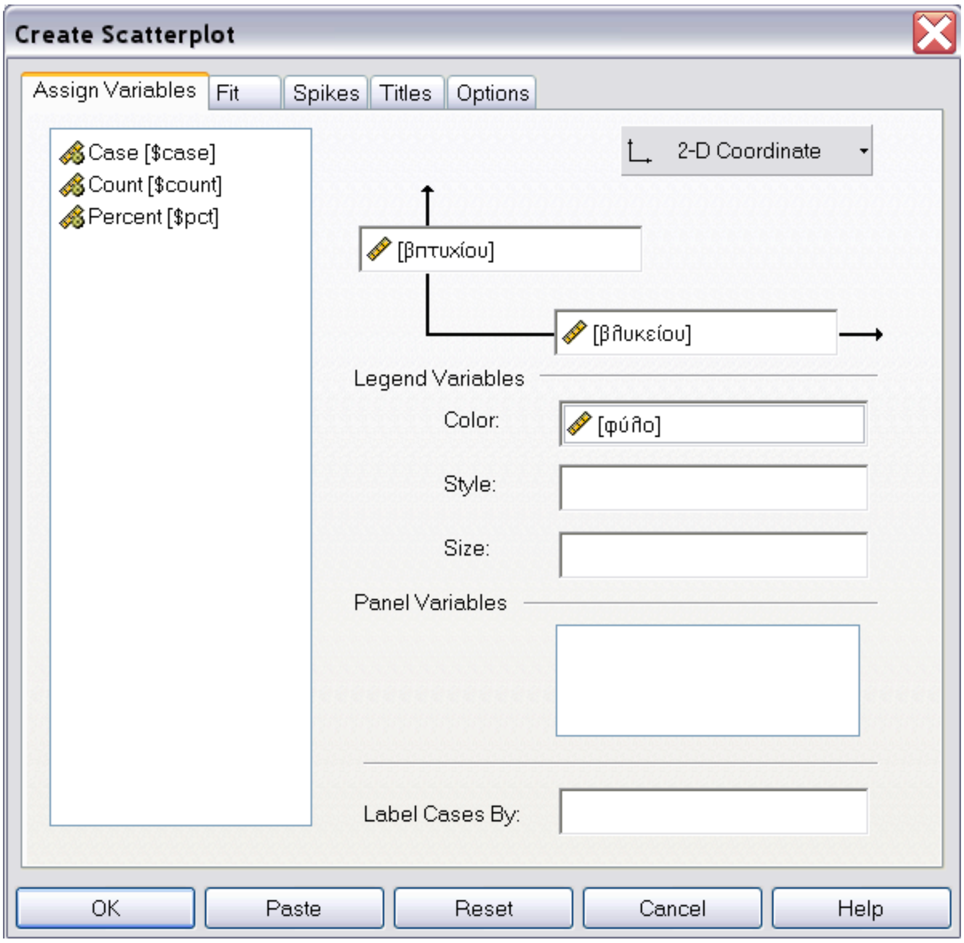
Για να δείξουμε την κατασκευή ενός διαγράμματος διασποράς με αλληλεπίδραση θα χρησιμοποιήσουμε το αρχείο βαθμοί.sav. Στο αρχείο αυτό περιέχονται ο βαθμός του απολυτηρίου Λυκείου, ο βαθμός πτυχίου και το φύλο κάποιων αποφοίτων ενός Πανεπιστημίου (υποθετικά δεδομένα). Θα εξετάσουμε αν ο βαθμός του πτυχίου ενός αποφοίτου του Πανεπιστημίου εξαρτάται από το βαθμό του απολυτηρίου του Λυκείου. Για να δημιουργήσουμε ένα interactive διάγραμμα διασποράς επιλέγουμε διαδοχικά από το μενού (Εικόνα 5):

Graphs → Interactive → Scatterplot...



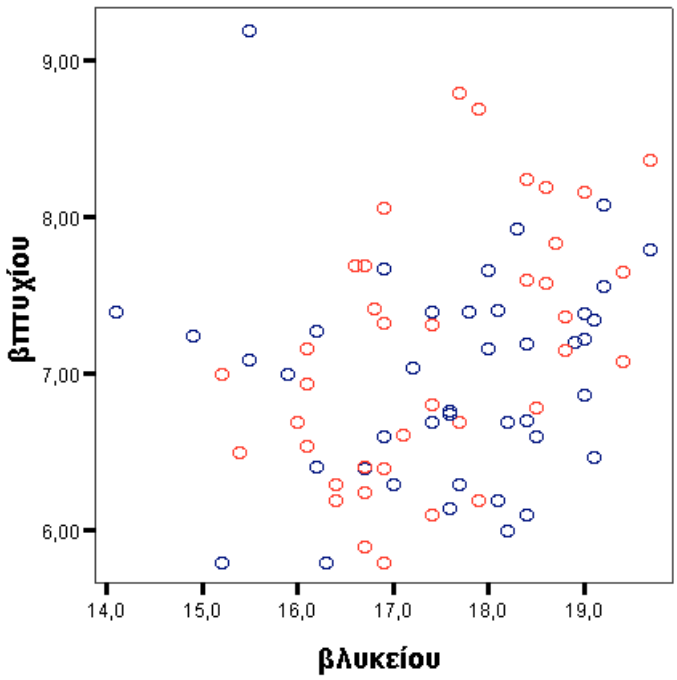
Εικόνα 5: Κατασκευή διαγράμματος διασποράς με αλληλεπίδραση

Τοποθετούμε τις μεταβλητές στους άξονες με την τεχνική του συρσίματος και στο πλαίσιο Color τοποθετούμε τη μεταβλητή φύλο. Με αυτόν τον τρόπο ζητάμε να παρασταθούν ξεχωριστά στο διάγραμμα (χρησιμοποιώντας διαφορετικά χρώματα), οι δύο ομάδες (άντρες/γυναίκες). Έτσι, θα έχουμε μια εικόνα του συνολικού δείγματος, αλλά και των δύο ομάδων ξεχωριστά.



Εικόνα 6: Εισαγωγή μεταβλητών

Το αποτέλεσμα που παίρνουμε είναι το εξής:



Εικόνα 7: Διάγραμμα διασποράς με αλληλεπίδραση

Παρατηρούμε ότι και για τους άντρες και για τις γυναίκες, άρα και για το σύνολο των δεδομένων μας τα σημεία είναι πολύ διεσπαρμένα στο επίπεδο, οπότε δεν μπορούμε να θεωρήσουμε ότι οι δύο μεταβλητές έχουν κάποια σχέση. Δεν μπορούμε, δηλαδή, να ισχυριστούμε ότι ο βαθμός του πτυχίου ενός απόφοιτου (είτε αυτός είναι άντρας είτε γυναίκα) έχει κάποια σχέση με το βαθμό του απολυτηρίου του Λυκείου.

Στη συνέχεια θα ασχοληθούμε με την απλή γραμμική συσχέτιση και την παλινδρόμηση, που εμπίπτουν στην περιοχή της ανάλυσης δύο μεταβλητών. Πολύ συχνά συμβαίνει να θέλουμε να συγκρίνουμε δύο συνεχείς μεταβλητές μεταξύ τους και να απαντήσουμε σε ερευνητικές ερωτήσεις του τύπου:

- Σχετίζεται το ύψος ενός ατόμου με το βάρος του;
- Σχετίζονται τα έτη εκπαίδευσης ενός ατόμου με το εισόδημα του;

Μπορεί επίσης να θέλουμε να προβλέψουμε την τιμή μιας μεταβλητής από μια άλλη, για παράδειγμα:

- Αν ένα άτομο έχει ύψος 180 cm, ποιο προβλέπεται να είναι το βάρος του;
- Αν κάποιος έχει σπουδάσει 12 χρόνια, ποιο προβλέπεται να είναι το εισόδημα του;

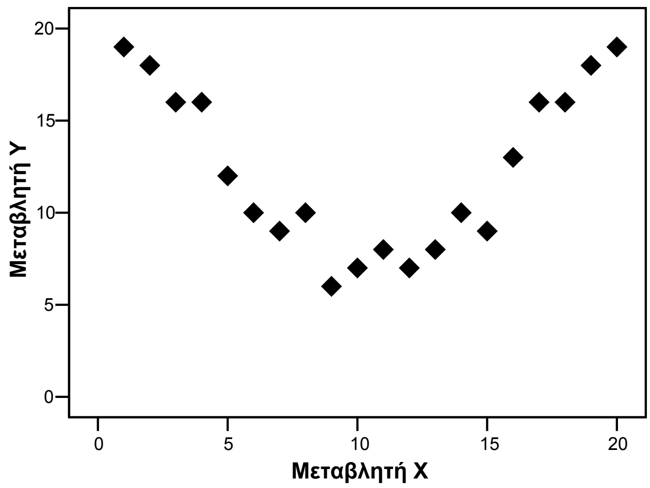
Σε αυτού του τύπου τις ερωτήσεις, απαντούμε χρησιμοποιώντας τη Συσχέτιση και την Παλινδρόμηση:

Η **Συσχέτιση** (Correlation) μετρά την ένταση της σχέσης δύο συνεχών μεταβλητών.

Παλινδρόμηση (Regression) είναι η πρόβλεψη των τιμών μιας μεταβλητής από τις τιμές μιας άλλης μεταβλητής.

ΣΥΣΧΕΤΙΣΗ

Η συσχέτιση μετρά όπως αναφέραμε και προηγουμένως την **ένταση της σχέσης** δύο συνεχών μεταβλητών. Η κατασκευή ενός διαγράμματος διασποράς προηγείται πάντα της ανάλυσης της συσχέτισης, δεδομένου ότι από το διάγραμμα μπορεί να προκύψει ότι δεν υπάρχει σχέση (τα σημεία είναι εντελώς διεσπαρμένα στο επίπεδο), ή υπάρχει μια καμπυλόγραμμη σχέση μεταξύ των μεταβλητών, οπότε τα δεδομένα δεν είναι κατάλληλα για ανάλυση γραμμικής συσχέτισης (Εικόνα 8).



Εικόνα 8: Καμπυλόγραμμη σχέση μεταξύ των μεταβλητών X και Y

Ο συντελεστής συσχέτισης

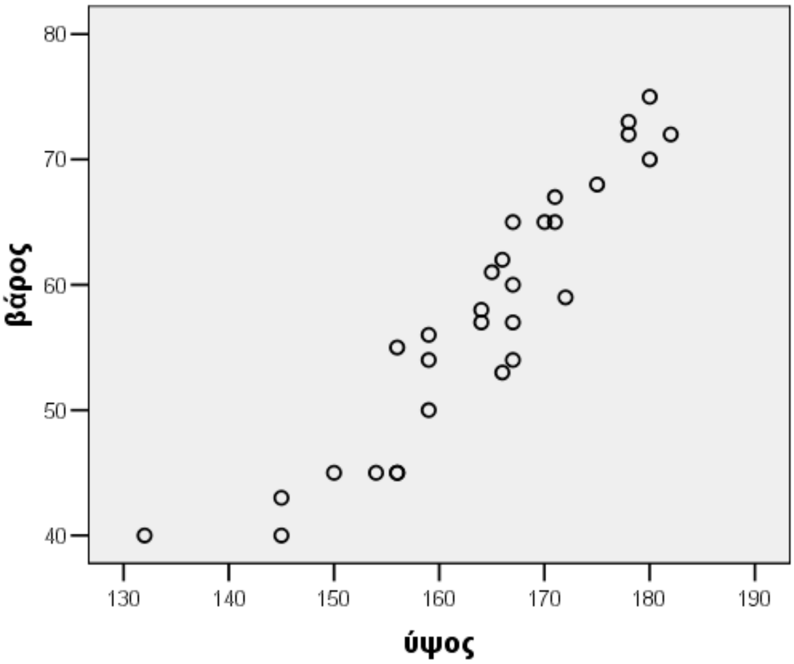
Για να ποσοτικοποιήσουμε μια συσχέτιση - με άλλα λόγια, για να μετρήσουμε την ένταςή της αριθμητικά - χρησιμοποιούμε το **συντελεστή συσχέτισης του Pearson**. Ο συντελεστής συσχέτισης r παίρνει τιμές μεταξύ του $+1$ και του -1. Πιο συγκεκριμένα, μπορούμε να πούμε ότι οι διαφορετικές τιμές του r υποδηλώνουν τα εξής:

- $r=+1$:τέλεια θετική συσχέτιση (όλα τα σημεία βρίσκονται σε μια ευθεία γραμμή).
- r μεταξύ 0 και 1: θετική αλλά όχι τέλεια συσχέτιση.
- $r=0$: καμία συσχέτιση μεταξύ των δύο μεταβλητών.
- r μεταξύ -1 και 0: αρνητική αλλά όχι τέλεια συσχέτιση.
- $r=-1$:τέλεια αρνητική συσχέτιση (όλα τα σημεία βρίσκονται σε μια ευθεία γραμμή).

Όσο πιο κοντά είναι το r στο +1 ή στο -1, τόσο ισχυρότερη είναι η σχέση μεταξύ των δύο μεταβλητών. Τιμές κοντά στο μηδέν δείχνουν μια πολύ ασθενή σχέση.
Γενικά, αν η απόλυτη τιμή του r, $|r|$, είναι:

- $0 < |r| < 0.2$: ασθενή συσχέτιση.
- $0.2 \le |r| < 0.4$: μέτρια συσχέτιση.
- $0.4 \le |r| < 0.7$: σημαντική συσχέτιση.
- $0.7 \le |r| < 1$: ισχυρή συσχέτιση.

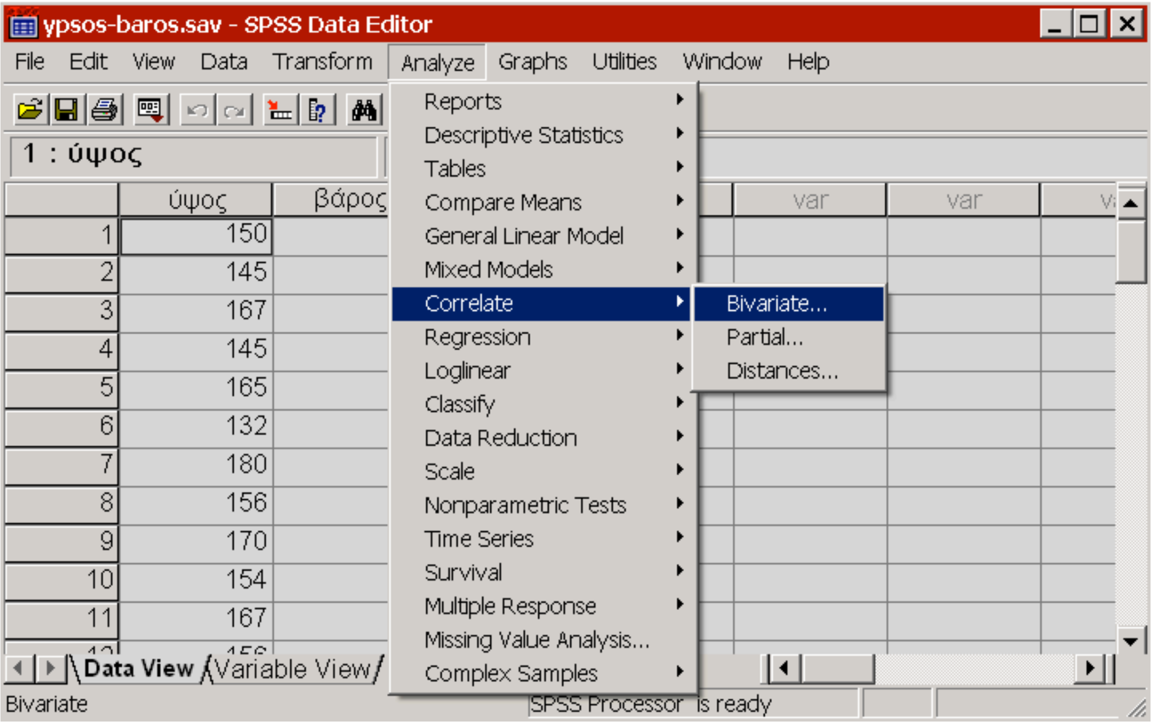
Για να εξετάσουμε τη συσχέτιση δύο μεταβλητών θα χρησιμοποιήσουμε τα δεδομένα του αρχείου ypsos-baros.sav. Κατασκευάζουμε αρχικά το αντίστοιχο διάγραμμα διασποράς, το οποίο δίνεται στην Εικόνα 9.



Εικόνα 9: Διάγραμμα διασποράς των μεταβλητών ύψος και βάρος

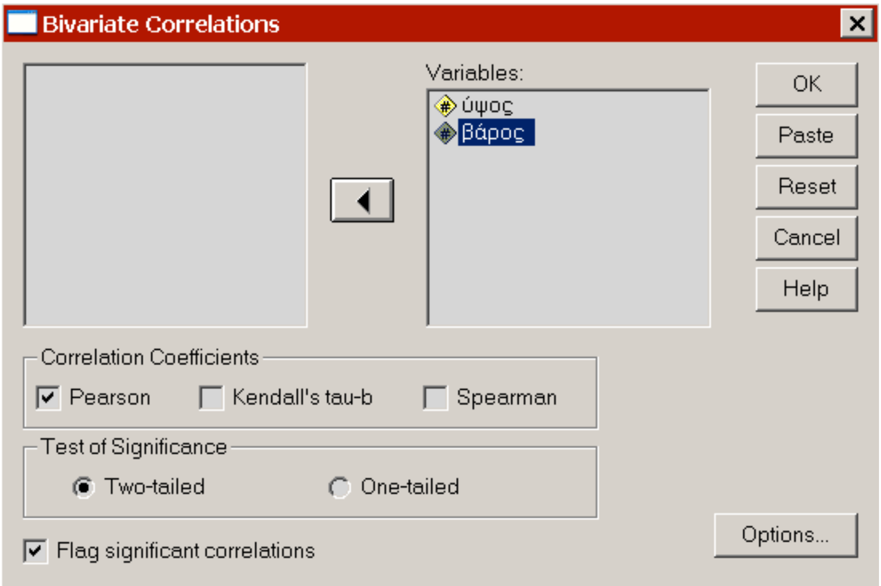
Από το διάγραμμα διασποράς φαίνεται να υπάρχει σχέση μεταξύ των δύο μεταβλητών και πιο συγκεκριμένα φαίνεται να υπάρχει μια αρκετά ισχυρή, θετική συσχέτιση. Τα δεδομένα είναι λοιπόν κατάλληλα για ανάλυση απλής γραμμικής συσχέτισης. Από το μενού των εντολών επιλέγουμε διαδοχικά:

Analyze → Correlate → Bivariate …



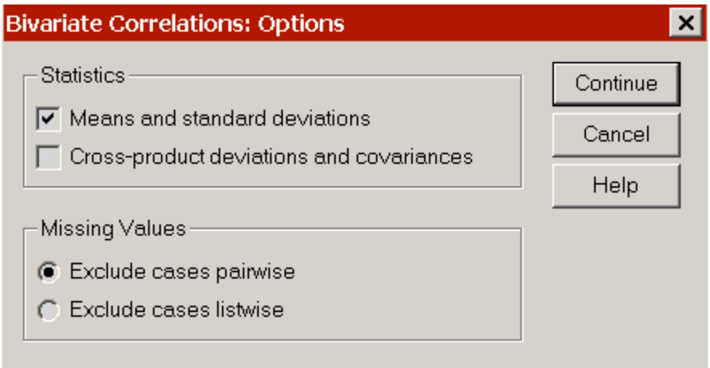
Εικόνα 10: Απλή γραμμική συσχέτιση

Επιλέγουμε τις μεταβλητές μας και τις τοποθετούμε στο πλαίσιο Variables, όπως φαίνεται στην Εικόνα 11. Αν έχουμε συνεχή ποσοτικά δεδομένα επιλέγουμε το Pearson (που θα μας δώσει τον συντελεστή συσχέτισης του Pearson). Αν έχουμε διατάξιμα δεδομένα επιλέγουμε το Kendall's tau-b ή το συντελεστή Spearman.



Εικόνα 11: Παράθυρο διαλόγου: Bivariate Correlations

Από το Options επιλέγουμε το Means and standard deviations και πατάμε Continue.



Εικόνα 12: Παράθυρο διαλόγου Bivariate Correlations: Options

Τα αποτελέσματα που παίρνουμε είναι τα εξής:

Descriptive Statistics

	Mean	Std. Deviation	N
ύψος	164,03	11,598	30
βάρος	57,70	10,323	30

Correlations

		ύψος	βάρος
ύψος	Pearson Correlation	1	,929(**)
	Sig. (2-tailed)		,000
	N	30	30
βάρος	Pearson Correlation	,929(**)	1
	Sig. (2-tailed)	,000	
	N	30	30

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

Ο συντελεστής συσχέτισης του Pearson είναι ίσος με 0.929, που αντιστοιχεί σε πολύ ισχυρή θετική γραμμική συσχέτιση. Η αντίστοιχη τιμή Sig. είναι ίση με 0.000<0.05 που σημαίνει ότι υπάρχει στατιστικά σημαντική σχέση μεταξύ των δύο μεταβλητών.

ΠΑΛΙΝΔΡΟΜΗΣΗ: Πρόβλεψη της εξαρτημένης μεταβλητής από την ανεξάρτητη

Όταν περιγράφουμε μια μόνο μεταβλητή μπορούμε να συνοψίσουμε τα δεδομένα μας χρησιμοποιώντας μέσες τιμές, διαμέσους, κλπ. Όταν έχουμε δύο μεταβλητές συνοψίζουμε τη σχέση τους, χρησιμοποιώντας κάποια εξίσωση.

Αν X και Y είναι δύο μεταβλητές οι οποίες συνδέονται γραμμικά, τότε η εξίσωση τους είναι της μορφής:

$$Y = a + bX$$

όπου a είναι η τιμή του Y , για $X = 0$, και b είναι η κλίση της ευθείας.

Αν υπολογίσουμε τις τιμές του a και του b, αυτή η εξίσωση μπορεί να χρησιμοποιηθεί για να προβλέψει την τιμή της μεταβλητής Y για οποιαδήποτε τιμή της μεταβλητής X. Το πρόβλημα είναι το εξής: πώς μπορούμε να προσαρμόσουμε την καλύτερη ευθεία γραμμή σε ένα σύνολο δεδομένων; Αν οι δύο μεταβλητές μας συνδέονται γραμμικά, η ευθεία βέλτιστης προσαρμογής, είναι γνωστή ως **ευθεία γραμμικής παλινδρόμησης.** Οι τιμές των a και b για την ευθεία αυτή δίνονται από τις παρακάτω σχέσεις:

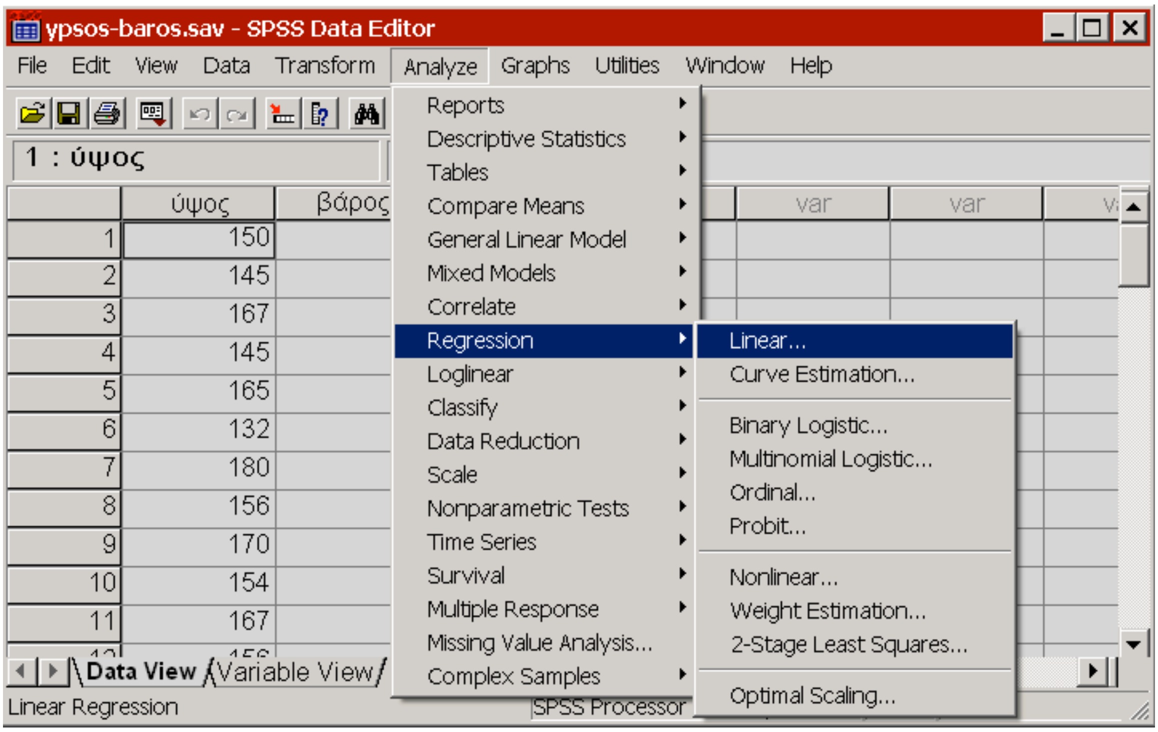
$$b = \frac{\sum\left(x_i - \bar{X}\right)\left(y_i - \bar{Y}\right)}{\sum\left(x_i - \bar{X}\right)^2}$$

$$a = \bar{Y} - b\bar{X}$$

Τα a και b είναι γνωστά ως **συντελεστές παλινδρόμησης**. Το a είναι η τιμή του Y, όταν το $X = 0$. Το $X = 0$ όμως δεν είναι πάντα μια λογική δυνατότητα. Πάντα, λοιπόν, πρέπει να ελέγχουμε το πλαίσιο των δεδομένων μας και να μην κάνουμε προβλέψεις έξω από τα όρια αυτών. Το b είναι η κλίση της ευθείας παλινδρόμησης. Γενικά, μπορούμε να πούμε ότι το b είναι η μεταβολή του Y, όταν το X μεταβάλλεται κατά μια μονάδα. Για να δούμε το πόσο καλά προσαρμόζεται η ευθεία παλινδρόμησης στα δεδομένα, υπολογίζουμε το τετράγωνο του r και το χρησιμοποιούμε σαν μέτρο προσαρμογής. Μια τιμή $r^2 = 1$ δείχνει μια τέλεια προσαρμογή, ενώ μια τιμή $r^2 = 0.8$ θα ήταν μια άριστη προσαρμογή για τις κοινωνικές επιστήμες. Ακόμη και οι τιμές γύρω από $r^2 = 0.5$ θεωρούνται για πολλά δεδομένα κοινωνικών επιστημών, ότι δείχνουν καλή προσαρμογή. Γενικά, το r^2 είναι το ποσοστό της μεταβλητότητας της Y που εξηγείται από το X. Αν π.χ. $r^2 = 0.8$, τότε 80% της μεταβλητότητας της Y εξηγείται από τη X.

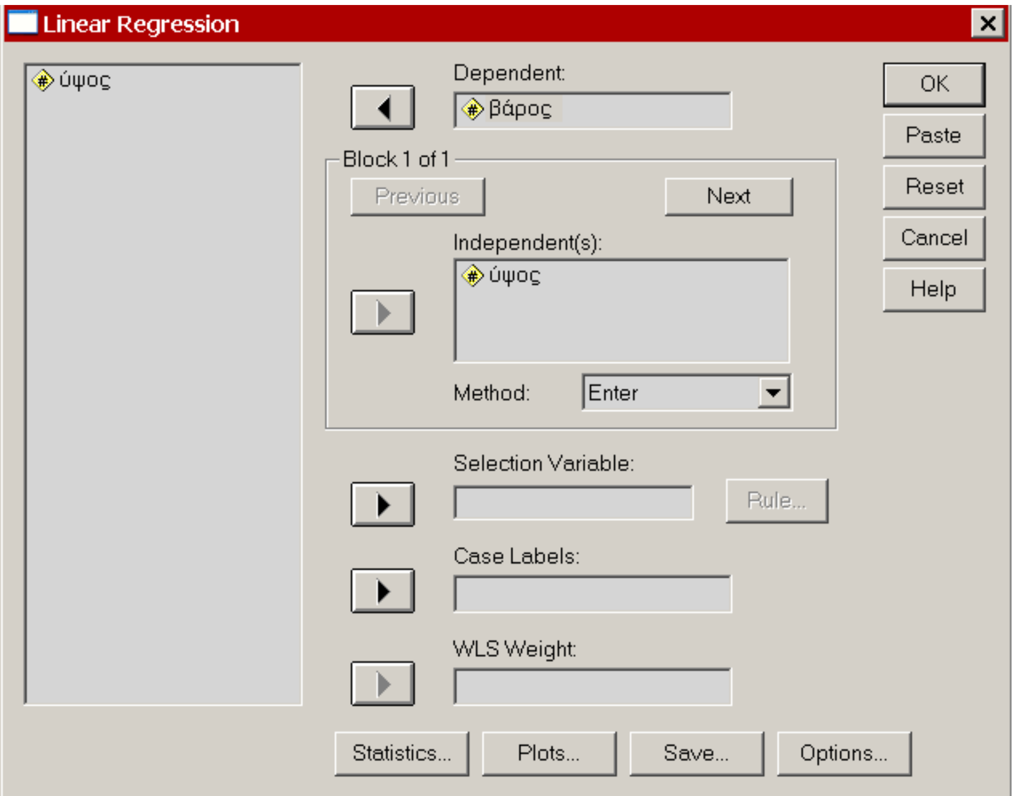
Ας δούμε στη συνέχεια την απλή γραμμική παλινδρόμηση, χρησιμοποιώντας τα ίδια δεδομένα του αρχείου ypsos-baros.sav. Επιλέγουμε διαδοχικά από το μενού των εντολών:

Analyze → Regression → Linear...

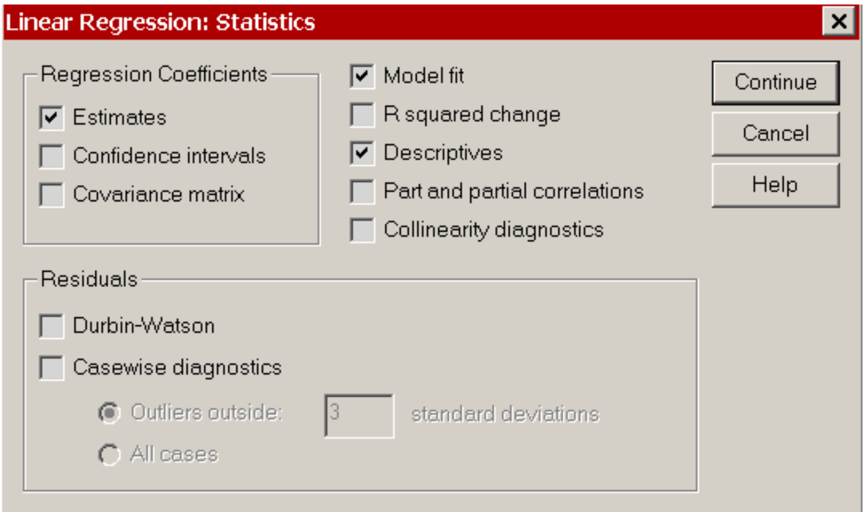


Εικόνα 13: Απλή γραμμική παλινδρόμηση

Τοποθετούμε την εξαρτημένη μεταβλητή στο πλαίσιο Dependent και την ανεξάρτητη μεταβλητή στο πλαίσιο Independent(s) και επιλέγουμε το Statistics.



Εικόνα 14: Παράθυρο διαλόγου: Linear Regression



Εικόνα 15: Παράθυρο διαλόγου: Linear Regression: Statistics

Εκεί κάνουμε κλικ στο Descriptives και πατάμε Continue.

Τα πρώτα αποτελέσματα που παίρνουμε είναι τα εξής:

Descriptive Statistics

	Mean	Std. Deviation	N
βάρος	57,70	10,323	30
ύψος	164,03	11,598	30

Correlations

		βάρος	ύψος
Pearson Correlation	βάρος	1,000	,929
	ύψος	,929	1,000
Sig. (1-tailed)	βάρος	.	,000
	ύψος	,000	.
N	βάρος	30	30
	ύψος	30	30

Στον πρώτο πίνακα παίρνουμε τη μέση τιμή και την τυπική απόκλιση για τις δύο μεταβλητές.

Στο δεύτερο Πίνακα βλέπουμε το συντελεστή συσχέτισης του Pearson, ο οποίος έχει τιμή ίση με 0.929. Αυτή είναι πολύ μεγάλη τιμή για το συντελεστή συσχέτισης και υποδηλώνει μια ισχυρή θετική συσχέτιση για τις δύο μεταβλητές. Η συσχέτιση είναι στατιστικά σημαντική και δεν μπορούμε να πούμε ότι εξαρτάται από τον παράγοντα τύχη.

Στον επόμενο Πίνακα βλέπουμε τις μεταβλητές που έχουμε εισάγει και τη μέθοδο που χρησιμοποιήσαμε (ENTER).

ΜΕΘΟΔΟΣ ENTER

Όλες οι επιλεγμένες ανεξάρτητες μεταβλητές εισέρχονται στη διαδικασία της παλινδρόμησης ανεξάρτητα από το αν κάποια από αυτές δεν συνεισφέρει στατιστικά σημαντικά στην πρόβλεψη ή στην ερμηνεία της εξαρτημένης μεταβλητής.

Variables Entered/Removed[b]

Model	Variables Entered	Variables Removed	Method
1	ύψος[a]	.	Enter

a. All requested variables entered.

b. Dependent Variable: βάρος

Στον επόμενο Πίνακα βλέπουμε τους συντελεστές a και b και το συντελεστή συσχέτισης του Pearson.

Coefficients[a]

Συντελεστής συσχέτισης του Pearson

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	-77,880	10,260		-7,590	,000
	ύψος	,827	,062	,929	13,246	,000

a. Dependent Variable: βάρος

Οι συντελεστές συσχέτισης λοιπόν είναι: $a = -77.88$ και $b = 0.827$.

Δηλαδή:

$$Y = -77.88 + 0.827X$$

ή αλλιώς:

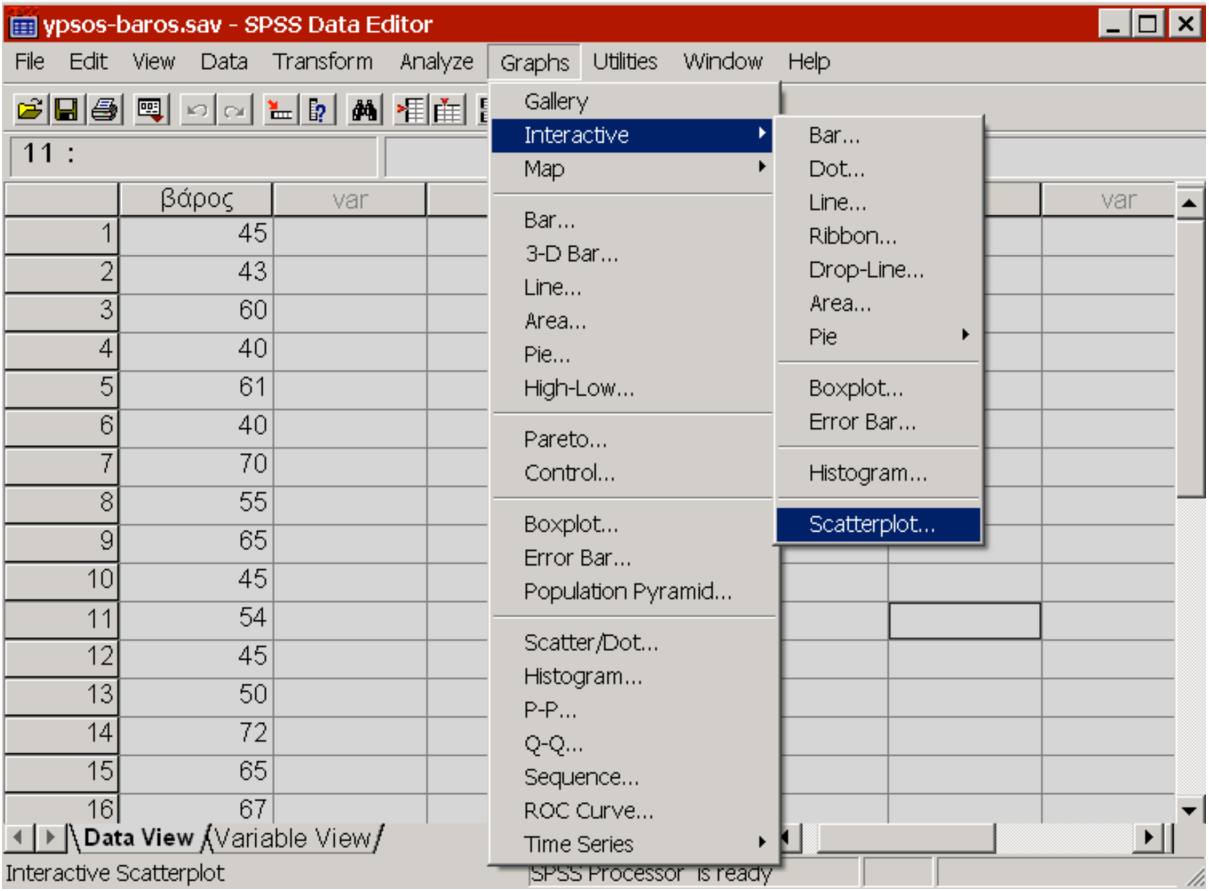
$$\text{βάρος} = -77.88 + 0.827 \cdot \text{ύψος}$$

Στη συνέχεια, αυτή η ευθεία μπορεί να χρησιμοποιηθεί για να κάνουμε προβλέψεις. Για παράδειγμα, ποια θα ήταν η τιμή του βάρους, αν το ύψος ενός ατόμου είναι ίσο με 177 cm.

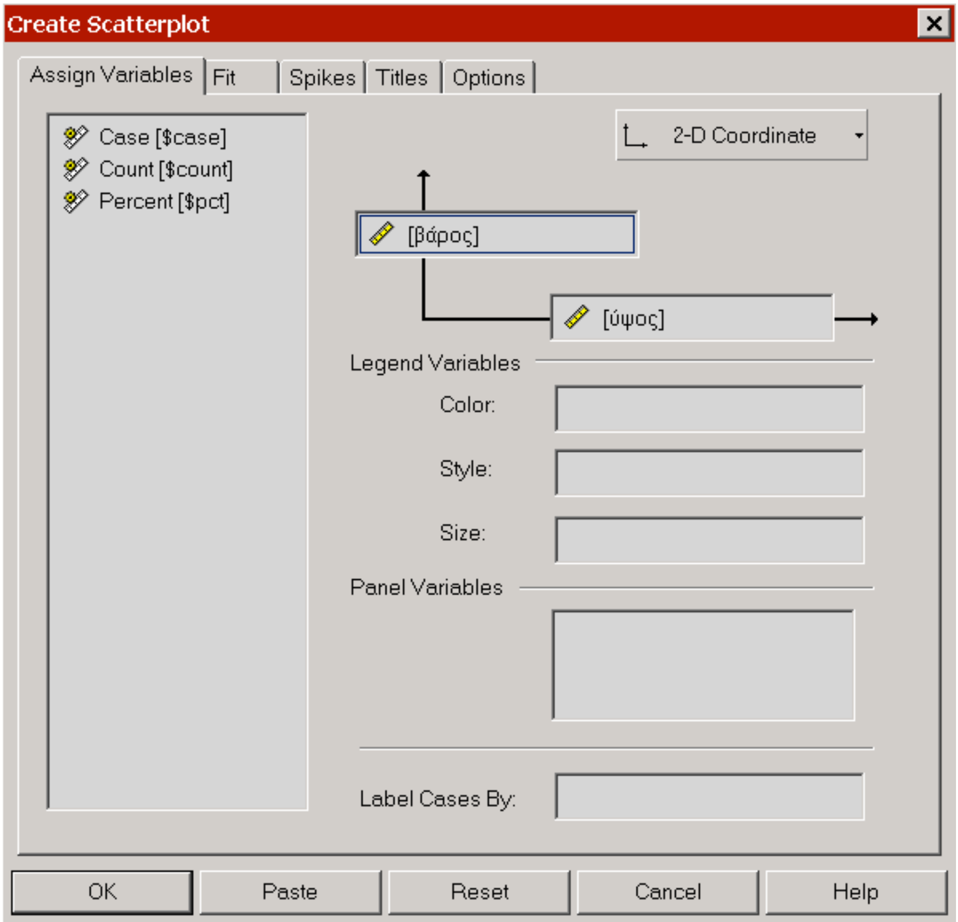
Αν, λοιπόν, ύψος $= 177$ μια εκτίμηση για το βάρος θα ήταν:

$$\text{βάρος} = -77.88 + 0.827 \cdot 177 = 68.5$$

Στην απλή γραμμική παλινδρόμηση μπορούν να μας βοηθήσουν και τα διαγράμματα διασποράς με αλληλεπίδραση (τα Interactive Scatterplots).

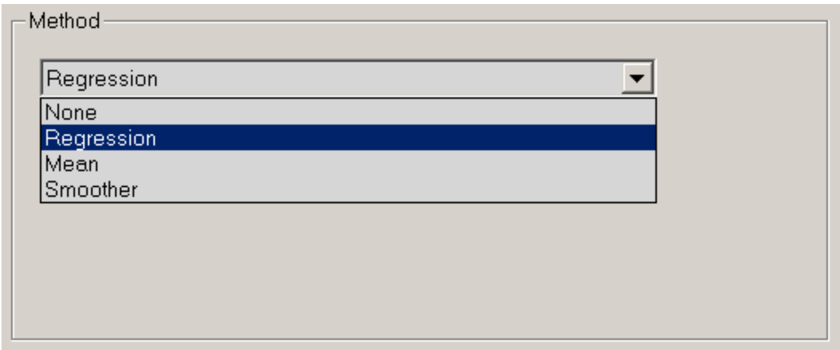


Εικόνα 16: Δημιουργία διαγράμματος διασποράς με αλληλεπίδραση



Εικόνα 17: Παράθυρο διαλόγου Create Scatterplot

Κάνουμε κλικ στο Fit και στο Method επιλέγουμε το Regression (Εικόνα 18).

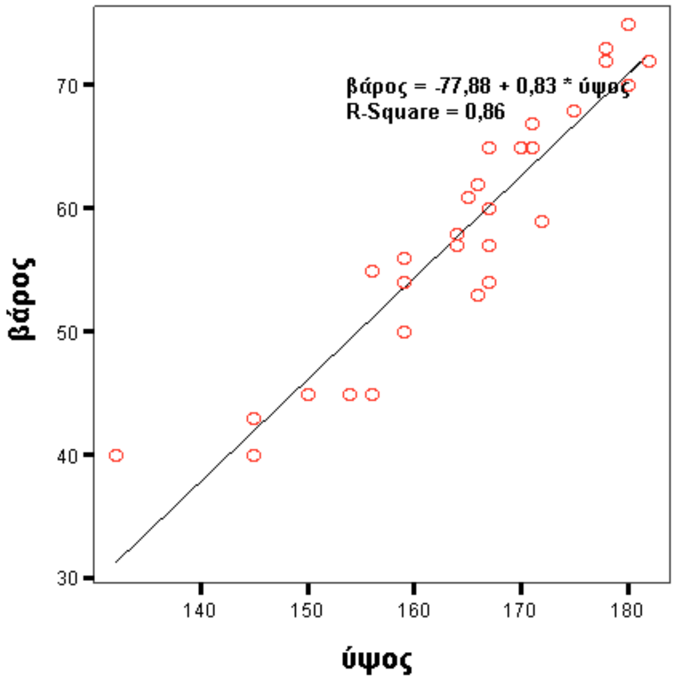


Εικόνα 18. Method: Regression

Στο διάγραμμα διασποράς (Εικόνα 19), δίνεται η ευθεία που προσαρμόζεται καλύτερα στα δεδομένα και αυτή είναι όπως φαίνεται η εξής:

$$\text{βάρος} = -77.88 + 0.827 \cdot \text{ύψος}$$

Επιπλέον, το r^2 είναι ίσο με 0.86 που σημαίνει ότι η ευθεία προσαρμόζεται πολύ καλά στα δεδομένα μας.



Εικόνα 19: Διάγραμμα διασποράς με αλληλεπίδραση

Ασκήσεις

1. Τα παρακάτω δεδομένα αντιστοιχούν στις μετρήσεις του ύψους και της αυτοεκτίμησης 20 ατόμων. Έστω ότι θέλουμε να εξετάσουμε αν οι μεταβλητές ύψος και αυτοεκτίμηση σχετίζονται γραμμικά μεταξύ τους. Χρησιμοποιείστε τη διαδικασία Regression για να ελέγξετε την ύπαρξη ή μη μιας γραμμικής σχέσης μεταξύ των μεταβλητών.

	ύψος	αυτοεκτίμηση	var
1	173	4,1	
2	181	4,6	
3	158	3,8	
4	191	4,4	
5	148	3,2	
6	153	3,1	
7	171	3,8	
8	173	4,1	
9	181	4,3	
10	175	3,7	
11	173	3,5	
12	171	3,2	
13	160	3,7	
14	157	3,3	
15	153	3,4	
16	160	4,0	
17	165	4,1	
18	170	3,8	
19	160	3,4	
20	155	3,6	
21			

2. Στα παρακάτω ζευγάρια ο πρώτος αριθμός σημαίνει ηλικία και ο δεύτερος σημαίνει βάρος σε kgr για 33 άτομα που μετρήθηκαν:

1	25	72	
2	35	74	
3	35	80	
4	45	88	
5	55	94	
6	25	70	
7	35	76	
8	45	70	
9	55	80	
10	65	86	
11	25	76	
12	35	70	
13	45	76	
14	55	84	
15	65	88	
16	25	74	
17	35	80	
18	45	82	
19	55	88	
20	65	88	
21	25	68	
22	35	84	
23	45	84	
24	55	90	
25	65	90	
26	25	72	
27	32	76	
28	45	84	
29	55	88	
30	65	98	
31	25	80	
32	35	80	
33	45	86	

Να γίνει το διάγραμμα διασποράς. Νομίζετε ότι υπάρχει γραμμικό πρότυπο που προσαρμόζεται στα δεδομένα μας;

3. Μετρήθηκε το ύψος και το βάρος σε 18 υπέρβαρα παιδιά:

	ΥΨΟΣ	ΒΑΡΟΣ	var
1	132,5	56,6	
2	148,5	84,3	
3	139,6	70,2	
4	121,5	47,2	
5	144,5	71,3	
6	133,6	71,4	
7	127,8	48,7	
8	124,7	52,7	
9	139,0	67,9	
10	136,0	61,3	
11	133,0	53,5	
12	146,2	82,7	
13	139,5	62,0	
14	135,3	62,3	
15	129,4	52,4	
16	148,8	61,5	
17	143,1	73,6	
18	129,2	62,1	
19			

Νομίζετε ότι υπάρχει γραμμική σχέση βάρους και ύψους; Αν ναι, να βρεθεί η καλύτερη ευθεία που προσαρμόζεται στα δεδομένα.

Εργασία 1:

Τα παρακάτω δεδομένα αντιστοιχούν στα έτη εκπαίδευσης 30 ανδρών (=1) και γυναικών (=2) και στον ετήσιο μισθό τους. Κατασκευάστε δύο διαφορετικά διαγράμματα διασποράς με αλληλεπίδραση (αφού πρώτα χρησιμοποιήσετε την ακολουθία εντολών Data → Split File), για να ελέγξετε αν ο μισθός ενός ατόμου εξαρτάται από τα έτη εκπαίδευσης του, ανά κατηγορία φύλου. Δώστε μια πρόβλεψη για το μισθό μιας γυναίκας με 18 χρόνια εκπαίδευσης. Επαναλάβετε το ίδιο για το μισθό ενός άντρα. Σχολιάστε τα αποτελέσματά σας.

	εκπαίδευση	φύλο	μισθός
1	12	1	25000
2	12	1	21000
3	6	1	17000
4	12	1	20000
5	21	1	44000
6	25	1	48000
7	20	1	38000
8	9	1	19000
9	12	1	24000
10	14	1	25000
11	6	1	16000
12	9	1	20000
13	20	1	39000
14	12	1	20000
15	6	1	15500
16	6	2	17000
17	6	2	16450
18	9	2	19000
19	16	2	26000
20	6	2	15000
21	6	2	14000
22	12	2	23450
23	12	2	24550
24	15	2	27000
25	12	2	18000
26	12	2	19000
27	9	2	16000
28	18	2	25000
29	19	2	26000
30	6	2	10500