

ΣΥΝΤΕΛΕΣΤΕΣ ΣΥΣΧΕΤΙΣΗΣ

**Βοηθητικές Σημειώσεις για Εργαστηριακή
Άσκηση 4 Μέρος Α**

PhDc Chasapi Maria
Konstantina

PhDc Chasapi Lamprini

Περιεχόμενα μαθήματος

Θεωρία Συσχέτισης

Συντελεστές συσχέτισης Pearson & Spearman

Ερμηνεία αποτελεσμάτων ενός ενδεικτικού
παραδείγματος στο SPSS για συσχέτιση

- ▶ Αν ενδιαφερόμαστε για το εάν δύο μεταβλητές σχετίζονται, τότε ενδιαφερόμαστε για το αν οι αλλαγές σε μια μεταβλητή συσχετίζονται με παρόμοιες ή αντίθετες αλλαγές στην άλλη μεταβλητή
- ▶ Επομένως, όταν μια μεταβλητή αποκλίνει από τη μέση τιμή της, αναμένουμε από την άλλη μεταβλητή να αποκλίνει από τη μέση τιμή με ανάλογο τρόπο

Μελετώντας τις σχέσεις μεταξύ μεταβλητών

Συσχέτιση (ή συνάφεια)

Συνάφεια (correlation): Ο βαθμός συμμεταβολής δύο μεταβλητών. Ο βαθμός στον οποίο οι τιμές δύο μεταβλητών αλλάζουν ταυτόχρονα.

*Οι όροι συνάφεια και συσχέτιση χρησιμοποιούνται εναλλακτικά

Μια συσχέτιση bivariate είναι μια συσχέτιση μεταξύ δύο μεταβλητών

- ▶ **Θετική συσχέτιση (positive correlation):** Όταν οι αλλαγές στις τιμές των δύο μεταβλητών έχουν την ίδια κατεύθυνση. Όταν αυξάνονται (ή μειώνονται) οι τιμές της μιας μεταβλητής, αυξάνονται (ή μειώνονται) και οι τιμές της άλλης μεταβλητής.
- ▶ **Αρνητική συσχέτιση (negative correlation):** Όταν οι αλλαγές στις τιμές των δύο μεταβλητών έχουν αντίθετη κατεύθυνση. Όταν αυξάνονται (ή μειώνονται) οι τιμές της μιας μεταβλητής, μειώνονται (ή αυξάνονται) οι τιμές της άλλης μεταβλητής.



Είδη συσχέτισης

Συντελεστής συσχέτισης I

Συντελεστής ή δείκτης συσχέτισης (correlation coefficient): Το στατιστικό κριτήριο με το οποίο υπολογίζεται ο βαθμός συσχέτισης μεταξύ δύο μεταβλητών

Η τιμή του συντελεστή συσχέτισης υποδηλώνει την κατεύθυνση (το είδος) και την ισχύ (το μέγεθος) της συνάφειας μεταξύ δύο μεταβλητών

Η τιμή ενός συντελεστή συσχέτισης κυμαίνεται μεταξύ - 1 και + 1

Το πρόσημο του συντελεστή συσχέτισης υποδηλώνει την κατεύθυνση της συνάφειας μεταξύ δύο μεταβλητών. Θετικό πρόσημο δείχνει θετική συσχέτιση και αρνητικό πρόσημο δείχνει αρνητική συσχέτιση.

Η απόλυτη τιμή του συντελεστή συνάφειας υποδηλώνει την ισχύ της συνάφειας μεταξύ δύο μεταβλητών. Για παράδειγμα, οι τιμές συνάφειας .7 και -.7 δείχνουν ίδιο μέγεθος συσχέτισης, αλλά αντίθετο είδος συνάφειας.

Όσο πιο υψηλή ή χαμηλή είναι η τιμή του συντελεστή συσχέτισης, τόσο πιο ισχυρή η συνάφεια μεταξύ των δύο μεταβλητών.

Όσο πιο κοντά στο μηδέν είναι η τιμή του συντελεστή συσχέτισης, τόσο πιο ανίσχυρη η συνάφεια μεταξύ των δύο μεταβλητών.

Συντελεστής συσχέτισης II

Απόλυτη θετική συσχέτιση (absolute positive correlation): Ένας συντελεστής συνάφειας με τιμή $+ 1$ υποδηλώνει ότι οι δύο μεταβλητές είναι απολύτως θετικά συσχετιζόμενες (ότι υπάρχει απόλυτη θετική συνάφεια). Καθώς οι τιμές της μιας μεταβλητής αλλάζουν, οι τιμές της άλλης μεταβλητής μεταβάλλονται ισόποσα προς την ίδια κατεύθυνση.

Απόλυτη αρνητική συσχέτιση (absolute negative correlation): Ένας συντελεστής συνάφειας με τιμή $- 1$ υποδηλώνει ότι οι δύο μεταβλητές είναι απολύτως αρνητικά συσχετιζόμενες (ότι υπάρχει απόλυτη αρνητική συνάφεια). Καθώς οι τιμές της μιας μεταβλητής αλλάζουν, οι τιμές της άλλης μεταβλητής μεταβάλλονται ισόποσα προς την αντίθετη κατεύθυνση.

Μηδενική συσχέτιση (zero correlation): Ένας συντελεστής συνάφειας με τιμή μηδέν υποδηλώνει ότι υπάρχει μηδενική συνάφεια. Καθώς οι τιμές της μιας μεταβλητής αλλάζουν, οι τιμές της άλλης μεταβλητής παραμένουν σταθερές.

Συντελεστής συσχέτισης και συνάφεια

Ένας συντελεστής συσχέτισης με απόλυτη τιμή μεγαλύτερη του μηδενός, αλλά μικρότερη του .3, υποδηλώνει ότι υπάρχει **ανίσχυρη (χαμηλή) συνάφεια** μεταξύ των μεταβλητών

Ένας συντελεστής συσχέτισης με απόλυτη τιμή ίση ή μεγαλύτερη του .3, αλλά μικρότερη του .7, υποδηλώνει ότι υπάρχει **σχετικά ισχυρή/ανίσχυρη (μέτρια) συνάφεια** μεταξύ των μεταβλητών

Ένας συντελεστής συσχέτισης με απόλυτη τιμή ίση ή μεγαλύτερη του .7, αλλά μικρότερη του 1, υποδηλώνει ότι υπάρχει **ισχυρή (υψηλή) συνάφεια** μεταξύ των μεταβλητών

Σημείωση: Τα παραπάνω κριτήρια δεν είναι απόλυτα

Είδη συντελεστών συσχέτισης

- ▶ **Συντελεστής συσχέτισης του Pearson (*Pearson's correlation coefficient*)**: Παραμετρικό τεστ που εφαρμόζεται όταν οι μεταβλητές είναι ίσων διαστημάτων ή αναλογικές και οι τιμές της κάθε μεταβλητής σχηματίζουν κανονική κατανομή.

Σύμβολο: r

- ▶ **Συντελεστής συσχέτισης του Spearman (*Spearman's correlation coefficient*)**: Μη παραμετρικό τεστ που εφαρμόζεται είτε όταν οι μεταβλητές είναι (ή έχουν μετατραπεί σε) ιεραρχικές είτε όταν οι μεταβλητές είναι ίσων διαστημάτων ή αναλογικές, αλλά οι τιμές της κάθε μεταβλητής δεν σχηματίζουν κανονική κατανομή.

Σύμβολο: ρ ή ρ

Συνάφεια - Διάγραμμα διασποράς

Διάγραμμα διασποράς ή σκεδασμού (scatter plot): Η γραφική αναπαράσταση της συσχέτισης μεταξύ δύο μεταβλητών

Απεικονίζει κατά προσέγγιση την κατεύθυνση, την ισχύ και τη μορφή της συνάφειας μεταξύ δύο μεταβλητών

Σημείωση 1: Ανάλογα με τη μορφή τους, οι συσχετίσεις διακρίνονται σε γραμμικές ή ευθύγραμμες και καμπυλόγραμμες

Σημείωση 2: Το κριτήριο r και το κριτήριο ρ εφαρμόζονται για τον υπολογισμό της γραμμικής (ευθύγραμμης) συσχέτισης (linear correlation) μεταξύ δύο μεταβλητών και όχι της καμπυλόγραμμης συσχέτισης (curvilinear correlation) τους

Συνάφεια - Διάγραμμα διασποράς

Κάθε μία από τις δύο μεταβλητές αναπαριστάται σε έναν από τους δύο άξονες του διαγράμματος και κάθε σημείο αντιπροσωπεύει ζεύγη τιμών από τις δύο μεταβλητές

Όταν το σχήμα των σημείων έχει κατεύθυνση από αριστερά και κάτω προς τα δεξιά και πάνω, τότε ενδεχομένως να υπάρχει θετική συνάφεια μεταξύ των δύο μεταβλητών

Όταν το σχήμα των σημείων έχει κατεύθυνση από αριστερά και πάνω προς τα δεξιά και κάτω, τότε ενδεχομένως να υπάρχει αρνητική συνάφεια μεταξύ των δύο μεταβλητών

Όσο πιο μεγάλη διασπορά παρουσιάζουν τα σημεία του διαγράμματος, τόσο λιγότερες οι πιθανότητες να υπάρχει γραμμική συσχέτιση ανάμεσα στις δύο μεταβλητές

Συνάφεια - Διάγραμμα διασποράς

Γραμμή παλινδρόμησης (regression line): Η γραμμή που περνά πιο κοντά από τα περισσότερα σημεία του διαγράμματος διασποράς

Αν η απόλυτη τιμή του συντελεστή συσχέτισης ισούται με 1, τότε όλα τα σημεία του διαγράμματος θα εφάπτονται της γραμμής παλινδρόμησης

Αν η τιμή του συντελεστή συσχέτισης ισούται με μηδέν, τότε τα σημεία του διαγράμματος θα κατανέμονται με τυχαίο τρόπο γύρω από τη γραμμή παλινδρόμησης

Σημείωση 1: Η γραμμή παλινδρόμησης είναι ευθεία στις γραμμικές συσχετίσεις και καμπύλη στις καμπυλόγραμμες συσχετίσεις

Σημείωση 2: Συνήθως η γραμμή παλινδρόμησης δεν χρειάζεται να χαραχθεί, καθώς η κλίση της γίνεται αντιληπτή από την κατεύθυνση που ακολουθούν τα σημεία του διαγράμματος

Συνάφεια - Διάγραμμα διασποράς και Ακραίες τιμές

Ένας παράγοντας που επηρεάζει σημαντικά την κατεύθυνση, την ισχύ και τη μορφή της συνάφειας μεταξύ δύο μεταβλητών είναι η ύπαρξη ακραίων τιμών

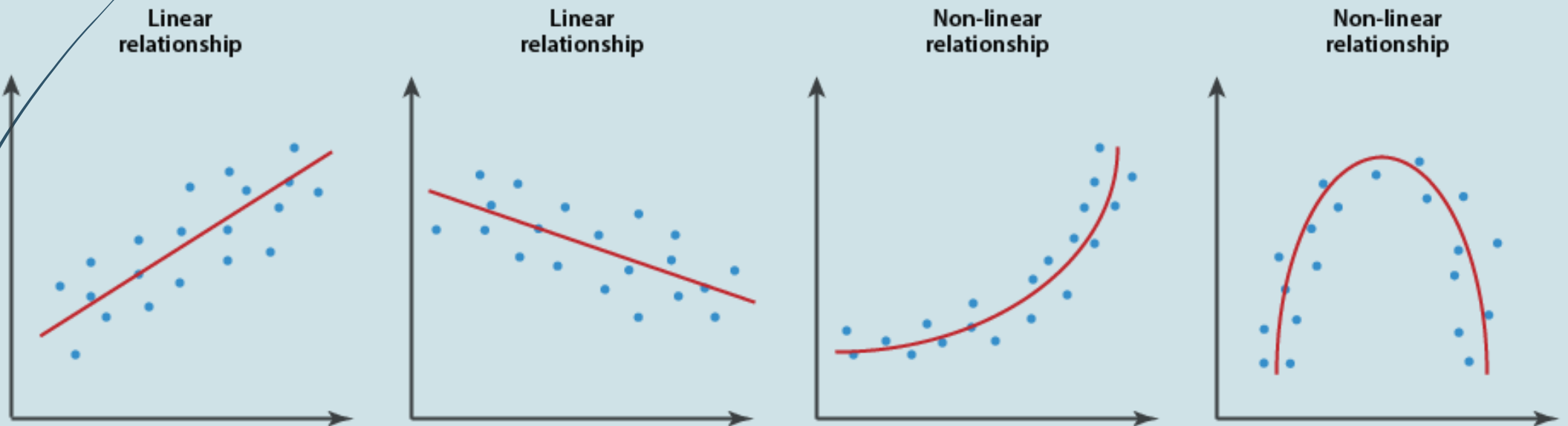
Ακραίες τιμές ή απόμακρες παρατηρήσεις (outliers): Μία ή περισσότερες τιμές οι οποίες διαφέρουν σημαντικά από τα υπόλοιπα συλλεγμένα δεδομένα

Τα διαγράμματα διασποράς χρησιμεύουν στον εντοπισμό ακραίων τιμών, καθώς τα σημεία που τις αντιπροσωπεύουν απέχουν σημαντικά από τα υπόλοιπα σημεία του διαγράμματος

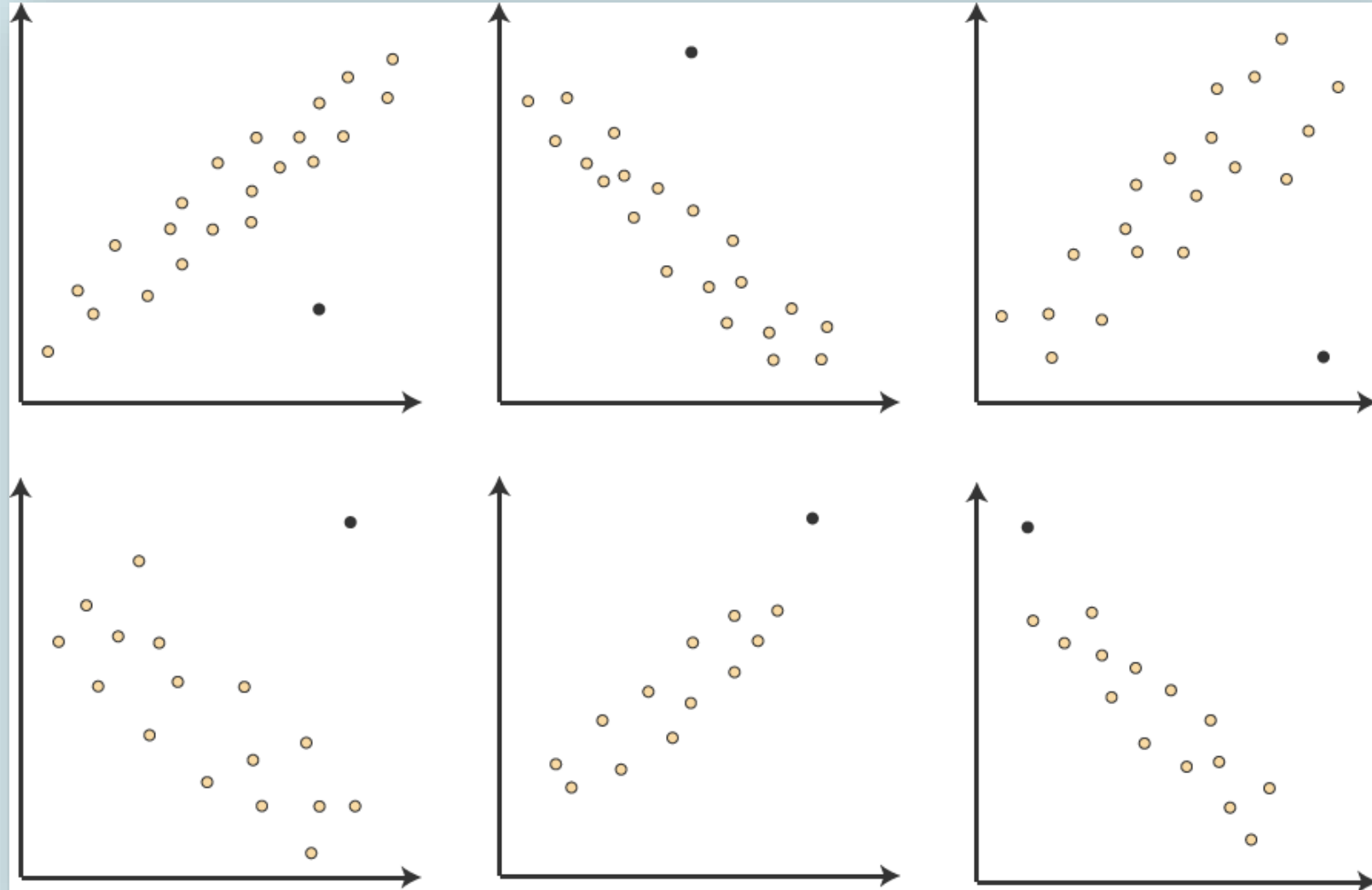
Σημείωση: Σύμφωνα με τους ερευνητικούς κανόνες δεοντολογίας, οι ακραίες τιμές μπορούν να αποκλειστούν από τη στατιστική ανάλυση των δειγματικών δεδομένων αλλά αυτό πρέπει πάντα να καταγράφεται στα αποτελέσματα

Η συσχέτιση του Pearson είναι κατάλληλη μόνο όταν υπάρχει γραμμική σχέση μεταξύ των δύο μεταβλητών σας. Σε αυτό το παράδειγμα, μια γραμμική σχέση μεταξύ "χρόνου που αφιερώθηκε παρακολουθώντας τηλεόραση" και "συγκέντρωσης χοληστερόλης"

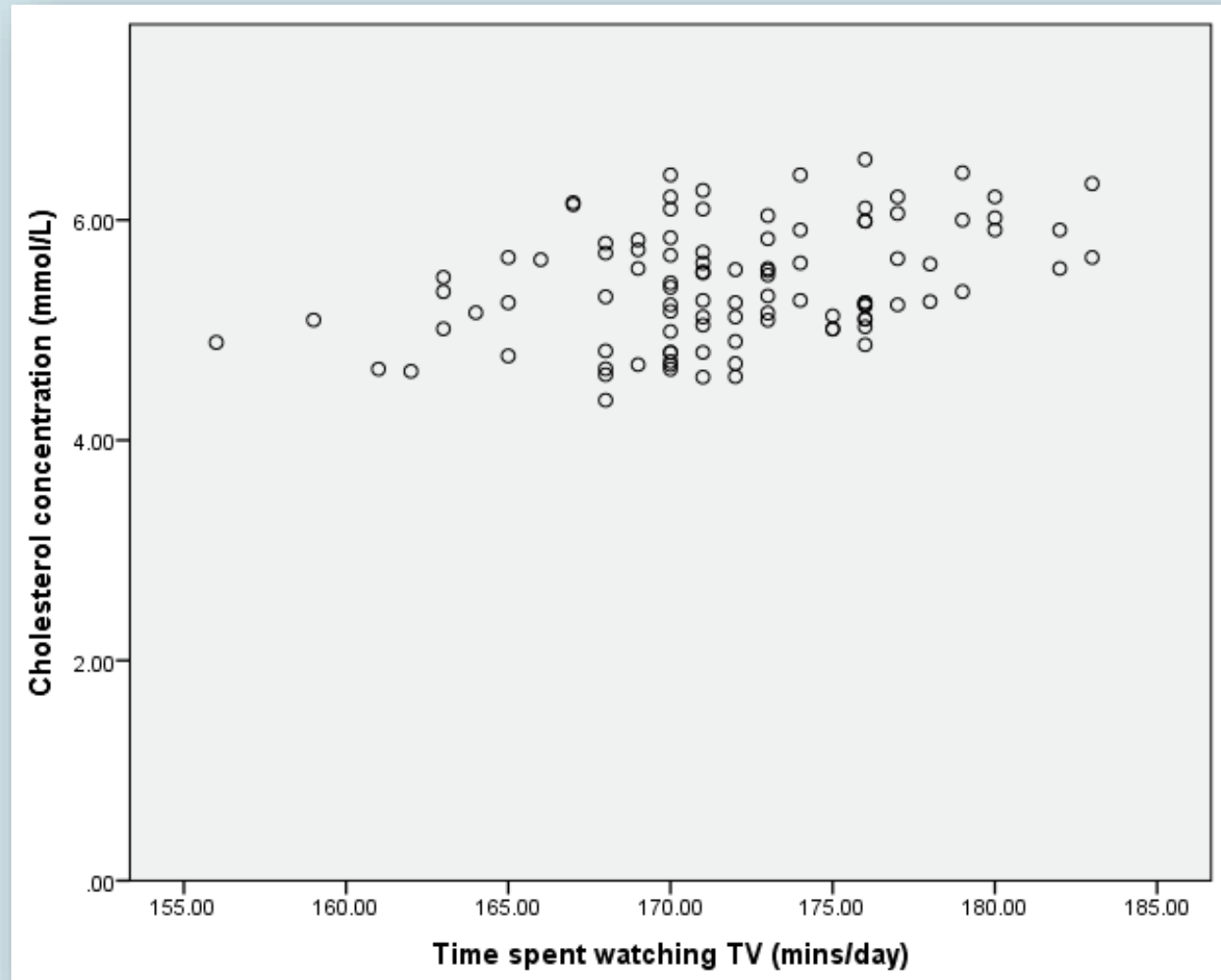
Γραμμική σχέση: Εάν η σχέση ακολουθεί περίπου μια ευθεία γραμμή, έχετε μια γραμμική σχέση.



Ακραίες τιμές

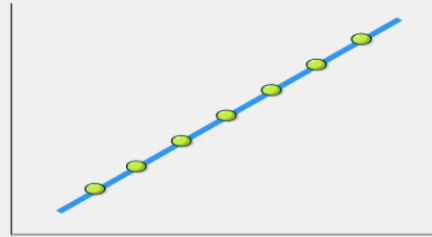


Απουσία ακραίων τιμών

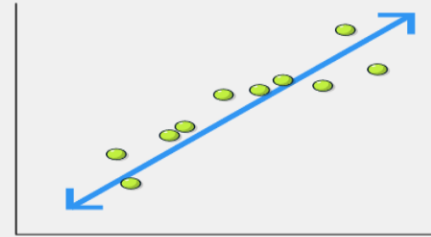


Είδη ΣΥΣΧΕΤΙΣΕΩΝ

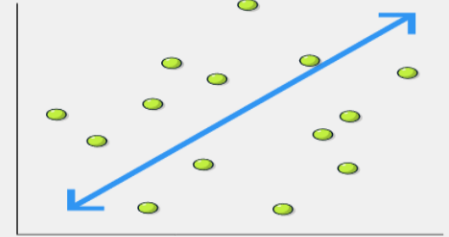
Scatter Plots and Correlation Examples



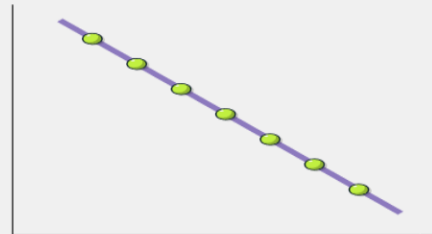
Perfect Positive Correlation



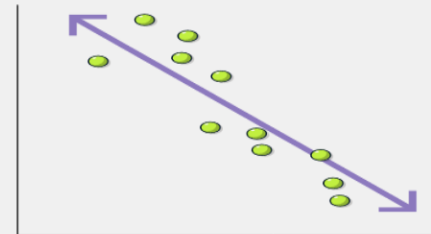
Highly Positive Correlation



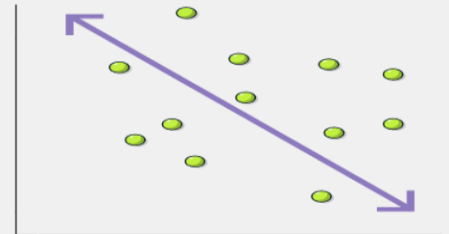
Low Positive Correlation



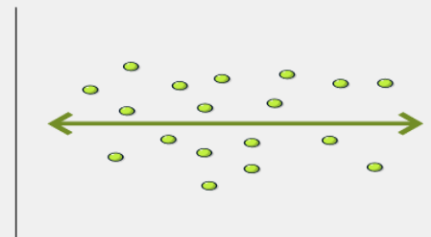
Perfect Negative Correlation



Highly Negative Correlation



Low Negative Correlation



No Correlation

Οι συντελεστές συσχέτισης δεν υποδηλώνουν σχέση αιτιότητας μεταξύ των υπό έρευνα μεταβλητών

->correlation is not causation!!!

- Μπορεί να υπάρχει αιτιώδης σχέση μεταξύ των μεταβλητών που μελετήθηκαν, αλλά η χρήση συντελεστών συσχέτισης δεν επιτρέπει αυτό το συμπέρασμα

Πιθανές ερμηνείες ύπαρξης συνάφειας μεταξύ των μεταβλητών A και B:

- Η μεταβλητή A μπορεί να προκαλεί αλλαγές στη μεταβλητή B
- Η μεταβλητή B μπορεί να προκαλεί αλλαγές στη μεταβλητή A
- Μια τρίτη μεταβλητή μπορεί να προκαλεί αλλαγές στις μεταβλητές A και B

Ερμηνεία

Προϋποθέσεις Pearson

- ▶ Κανονικότητα
- ▶ Απουσία ακραίων τιμών
- ▶ Συνέχεια
- ▶ Ανεξαρτησία των παρατηρήσεων
- ▶ Και οι δύο μεταβλητές πρέπει να είναι παραμετρικές
- ▶ Οι μεταβλητές να είναι διαστημικές ή αναλογικές

Ενδεικτικό Παράδειγμα Pearson

Πλησιάζει η άνοιξη και το κυκείο του Πανεπιστημίου έβαλε ψυγείο παγωτών. Ένας ερευνητής ενδιαφέρεται να μελετήσει εάν με την αλλαγή της θερμοκρασίας πωλούνται περισσότερα χωνάκια παγωτού.

Ρωτάμε τον πωλητή, πόσα χωνάκια παγωτού έχει πουλήσει κάθε μέρα σε ένα διάστημα 20 ημερών.

Μηδενική Υπόθεση (H_0)

Οι πωλήσεις του παγωτού δε θα αυξάνονται όσο αυξάνεται και η θερμοκρασία.

Εναλλακτική Υπόθεση (H_1)

Οι πωλήσεις του παγωτού θα αυξάνονται όσο αυξάνεται και η θερμοκρασία (μονής κατεύθυνσης)

Βήματα SPSS: Pearson

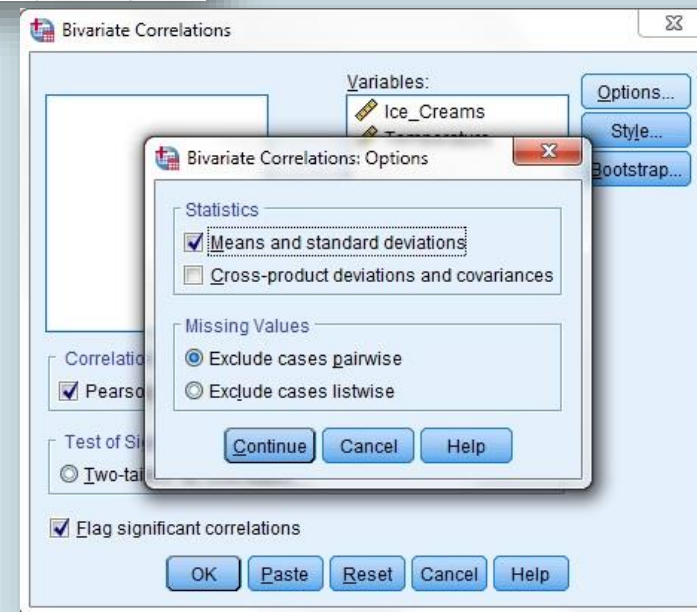
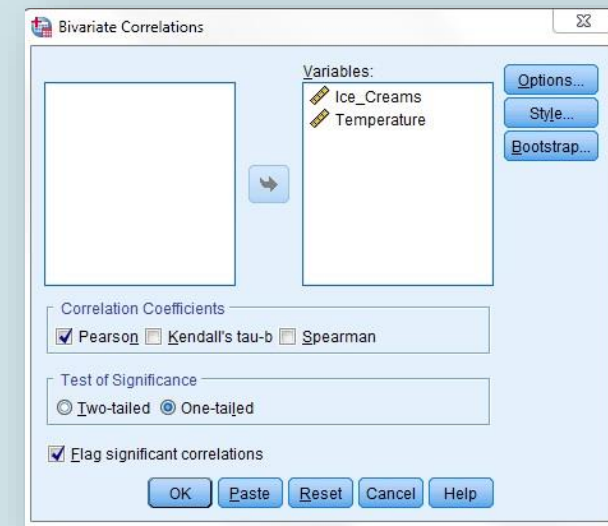
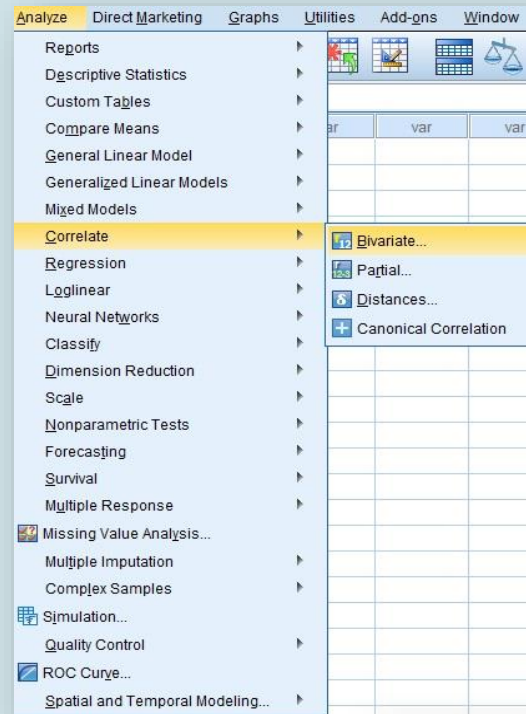
Analyze-> Correlate -> Bivariate

Στο αναδυόμενο παράθυρο μεταφέρω τις μεταβλητές μου στο κουτί Variables.

Επιλέγω το κουμπί Pearson και το One-tailed μιας και έχω μονής κατεύθυνσης υπόθεσης.

Επιλέγω το κουμπί Options και στις επιλογές Statistics επιλέγω το Means and standards deviations.

5. Έπειτα το κουμπί Continue και OK.



Output SPSS

Descriptive Statistics			
	Mean	Std. Deviation	N
Ice_Creams	787.80	197.687	20
Temperature	20.35	4.308	20

Correlations			
		Ice_Creams	Temperature
Ice_Creams	Pearson Correlation	1	.893**
	Sig. (1-tailed)		.000
	N	20	20
Temperature	Pearson Correlation	.893**	1
	Sig. (1-tailed)	.000	
	N	20	20

** . Correlation is significant at the 0.01 level (1-tailed).

- Η σχέση μεταξύ πωλήσεων παγωτού ($M = 787.80$, $SD = 197.68$) και θερμοκρασίας ($M = 20.35$, $SD = 4.30$) διερευνήθηκε σε ημέρες ($N = 20$) χρησιμοποιώντας ανάλυση συσχέτισης Pearson's correlation. Υπήρξε μια ισχυρή θετική συσχέτιση μεταξύ των δύο μεταβλητών, $r(20) = 0.89$, $p = .000$.

Ο δείκτης προσδιορισμού

- ▶ Είναι ο δείκτης που μας δείχνει τι ποσοστό της συνολικής διακύμανσης της μεταβλητής X οφείλεται στη μεταβλητή Y , και συμβολίζεται με το γράμμα r^2 .
- ▶ Τον υπολογίζουμε εάν υψώσουμε στο τετράγωνο το δείκτη συσχέτισης συσχέτισης των δύο μεταβλητών που μελετάμε και στη συνέχεια πολλαπλασιάζοντάς τον με το 100.

Πως τον υπολογίζουμε;

- ▶ Στο προηγούμενο παράδειγμα είδα ότι ο δείκτης συσχέτισης ήταν $r=0,89$, **τότε** $r^2= 0,89^2 = 0,79 \times 100 = 79\%$.
- ▶ Ερμηνεία: Η πώληση παγωτού συνδέθηκε έντονα με τη θερμοκρασία. Όσο αυξάνεται η θερμοκρασία, αυξάνεται και η πώληση των παγωτών. Το r του $0,89$ έδειξε ότι το 79% της διακύμανσης στις πωλήσεις παγωτού οφείλεται στη διακύμανση της θερμοκρασίας.
- ▶ Μένει ένα 21% της μεταβλητότητας για να αποδοθεί σε άλλες μεταβλητές. Μπορεί η πώληση παγωτού να μοιράζεται το 79% της διακύμανσης της θερμοκρασίας, αλλά αυτό δεν σημαίνει ότι απαραίτητα αποτελεί το αίτιο αυτής της διακύμανσης.

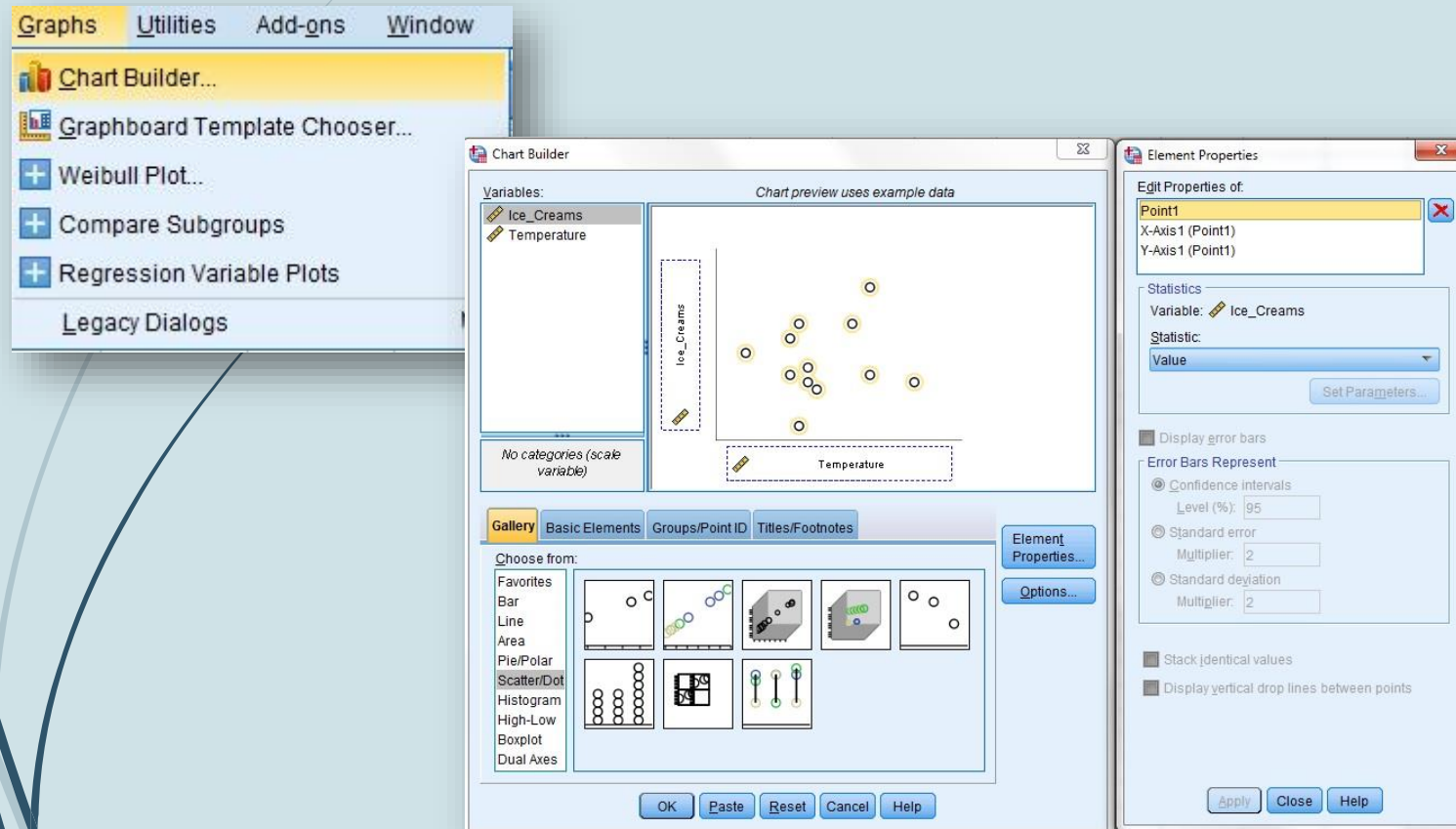
Βήματα στο SPSS

Επιλέγω Graphs →
Chart Builder

Επιλέγω το Scatter/dot
και το πρώτο γράφημα
(από τη λίστα στην
περιοχή Choose from:)
και το σέρνω στο
παράθυρο
προεπισκόπησης
γραφήματος

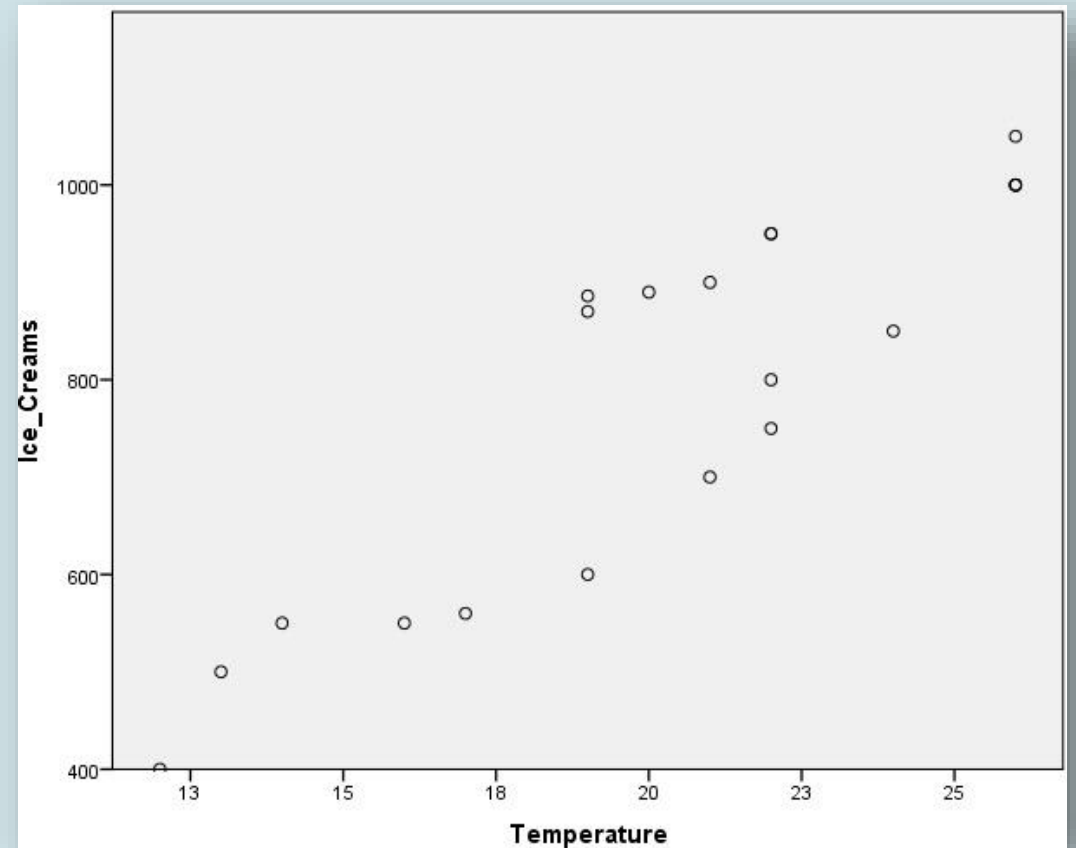
Μεταφέρω στον άξονα
Y το Ice Cream και
στον άξονα X το
Temperature

Πάτάω OK



Output SPSS

Μπορούμε να δούμε ότι η θερμοκρασία και ο αριθμός των παγωτού που πωλούνται σχετίζονται. Προφανώς δεν είναι τέλειος συσχετισμός, αλλά κοιτάζοντας τα δεδομένα μπορούμε να το δούμε είναι θετικό.



Κριτήριο rho/ Spearman

- Το κριτήριο rho εφαρμόζεται όταν τα δεδομένα είναι ιεραρχημένα
- Αν οι μεταβλητές είναι ποσοτικές, η διάταξη των δεδομένων στην κάθε μεταβλητή γίνεται κατά φθίνουσα σειρά με βάση το μέγεθος των τιμών της. Η υψηλότερη τιμή παίρνει την πρώτη θέση, η δεύτερη υψηλότερη τιμή παίρνει τη δεύτερη θέση κ.ο.κ. Οι αποστάσεις μεταξύ των τιμών χάνουν τη σημασία τους καθώς αξιολογείται η θέση των τιμών στη μια μεταβλητή σε σχέση με την αντίστοιχη θέση τους στην άλλη μεταβλητή.
- Σημείωση: Τιμές που ισοβαθούν παίρνουν τη μέση τιμή της θέσης τους
- Μικρή διαφορά ανάμεσα στις θέσεις των τιμών της κάθε μεταβλητής αποτελεί ένδειξη θετικής συνάφειας μεταξύ των δύο μεταβλητών, ενώ μεγάλη διαφορά ανάμεσα στις θέσεις των τιμών της κάθε μεταβλητής αποτελεί ένδειξη αρνητικής συνάφειας μεταξύ των δύο μεταβλητών.
- Η μόνη προϋπόθεση του μοντέλου είναι η μεταβλητή να είναι ordinal, δε μπορεί να είναι κατηγορική.

Ενδεικτικό Παράδειγμα Κριτήριο rho/Spearman

Θέλουμε να μελετήσουμε τη σχέση μεταξύ της αντίληψης για την ποιότητα του ύπνου και της διάθεσης.

Μηδενική υπόθεση (H_0): Δεν θα υπάρξει αρνητική συσχέτιση μεταξύ της αντίληψης για την ποιότητα ύπνου και της διάθεσης.

Εναλλακτική υπόθεση (H_1): Θα υπάρξει αρνητική συσχέτιση μεταξύ της αντίληψης για την ποιότητα ύπνου και της διάθεσης

Για να ελέγξουμε τη μηδενική υπόθεση πρέπει να εφαρμόσουμε το συντελεστή συσχέτισης rho.

Βήματα στο SPSS

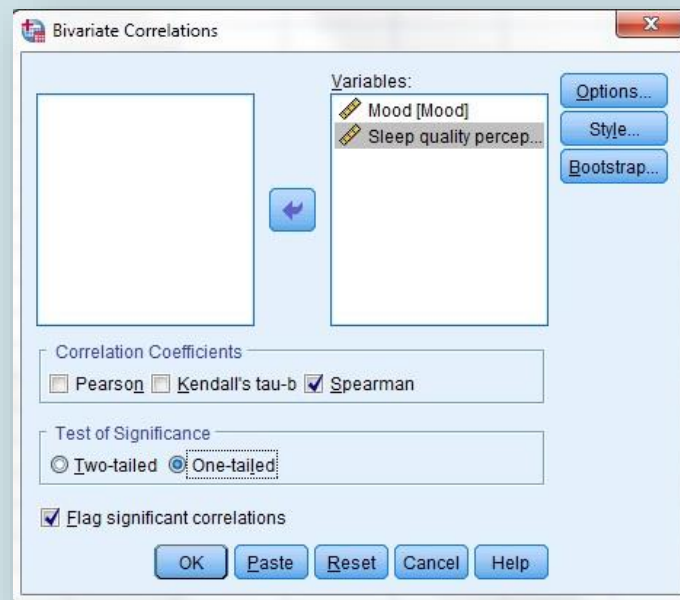
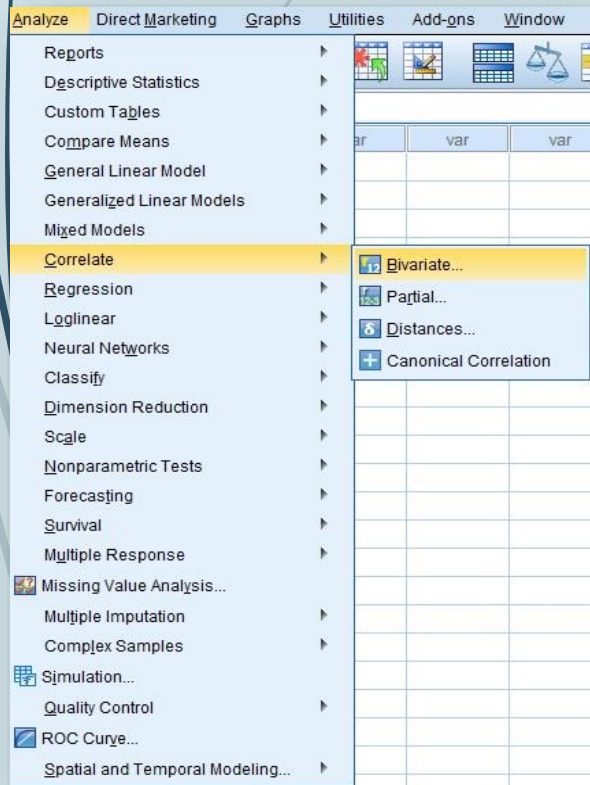
Spearman

Analyze-> Correlate -> Bivariate

Στο αναδυόμενο παράθυρο μεταφέρω τις μεταβλητές μου στο κουτί Variables.

Επιλέγω το κουμπι Spearman και το One-tailed μιας και έχω μονής κατεύθυνσης υπόθεσης.

Πατάω το κουμπι OK.



Output SPSS

Το Output επιβεβαιώνει μια μετρίως αρνητική, αλλά μη σημαντική, συσχέτιση: $r_s(13) = -.353, p = .098$. Συνεπώς δε μπορούμε να απορρίψουμε τη μηδενική υπόθεση.

Το ερευνητικό συμπέρασμα είναι ότι δεν υπάρχει αρνητική συσχέτιση ανάμεσα στην αντίληψη για την ποιότητα του ύπνου και της διάθεσης.

Για να εντοπίσουμε την κατάλληλη κρίσιμη τιμή, πρέπει πρώτα να υπολογίσουμε τους βαθμούς ελευθερίας (df). Οι βαθμοί ελευθερίας του κριτηρίου rho είναι ίσοι με το συνολικό αριθμό των συμμετεχόντων μείον δύο. Για την παρούσα έρευνα, οι βαθμοί ελευθερίας υπολογίζονται ως εξής: $df = N - 2 = 15 - 2 = 13$.

Correlations				
			Mood	Sleep quality perceptions
Spearman's rho	Mood	Correlation Coefficient	1.000	-.353
		Sig. (1-tailed)	.	.098
		N	15	15
	Sleep quality perceptions	Correlation Coefficient	-.353	1.000
		Sig. (1-tailed)	.098	.
		N	15	15

Βασικά σημεία της σημερινής διάλεξης

- Ο συντελεστής συσχέτισης του **Pearson r** είναι ένα παραμετρικό κριτήριο που απαιτεί δεδομένα διαστήματος και για τις δύο μεταβλητές, ενώ, ο συντελεστής συσχέτισης του **Spearman rs** είναι ένα μη παραμετρικό κριτήριο που απαιτεί μόνο μεταβλητές είτε ποσοτικές είτε ποιοτικές.
- Οι τιμές των συντελεστών βρίσκονται μεταξύ του **-1** και του **+1**
- Ένας συντελεστής με τιμή **+1** υποδηλώνει μια **Θετική σχέση**, ενώ ένας συντελεστής με τιμή **-1** υποδηλώνει μια εντελώς **αρνητική σχέση**.
- Ένας συντελεστής με τιμή **0** υποδηλώνει ότι δεν υπάρχει καμία γραμμική σχέση.
- Ο **συντελεστής συσχέτισης** χρησιμοποιείται ως μέτρο του μεγέθους μιας επίδρασης.

Πάμε τώρα στην άσκησή μας...
...Εργαστηριακή άσκηση 4
Μέρος Α'

Επίλυση Εργαστηριακής άσκησης 4

Μέρος Α'

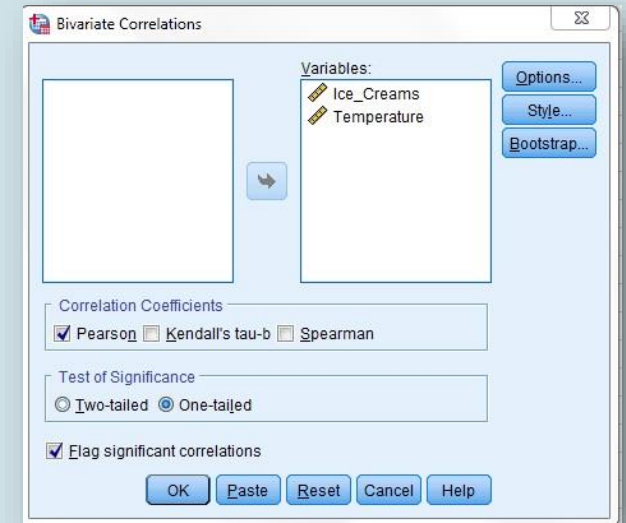
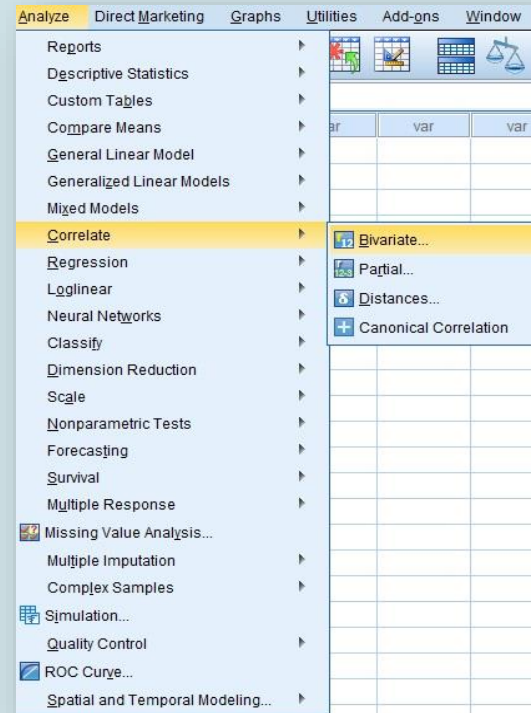
Τα ερωτήματα 1,2,3 έχουν συγγραφεί εντός του SPSS.

Ερώτημα 4: Για τις μεταβλητές Ηλικία και ΗΒ υπολογίστε τον συντελεστή συσχέτισης Pearson. Βεβαιωθείτε ότι καταλαβαίνετε την σημασία όλων των αποτελεσμάτων στον πίνακα που προέκυψε.

Λύση ερωτήματος 4

SPSS: Pearson

1. Analyze-> Correlate -> Bivariate
2. Στο αναδυόμενο παράθυρο μεταφέρω τις μεταβλητές μου στο κουτί Variables (**Ηλικία, HB**).
3. Επιλέγω το κουμπί **Pearson** και το One-tailed μιας και έχω μονής κατεύθυνσης υπόθεση.
4. Έπειτα το κουμπί OK.



Επίλυση Εργαστηριακής άσκησης 4

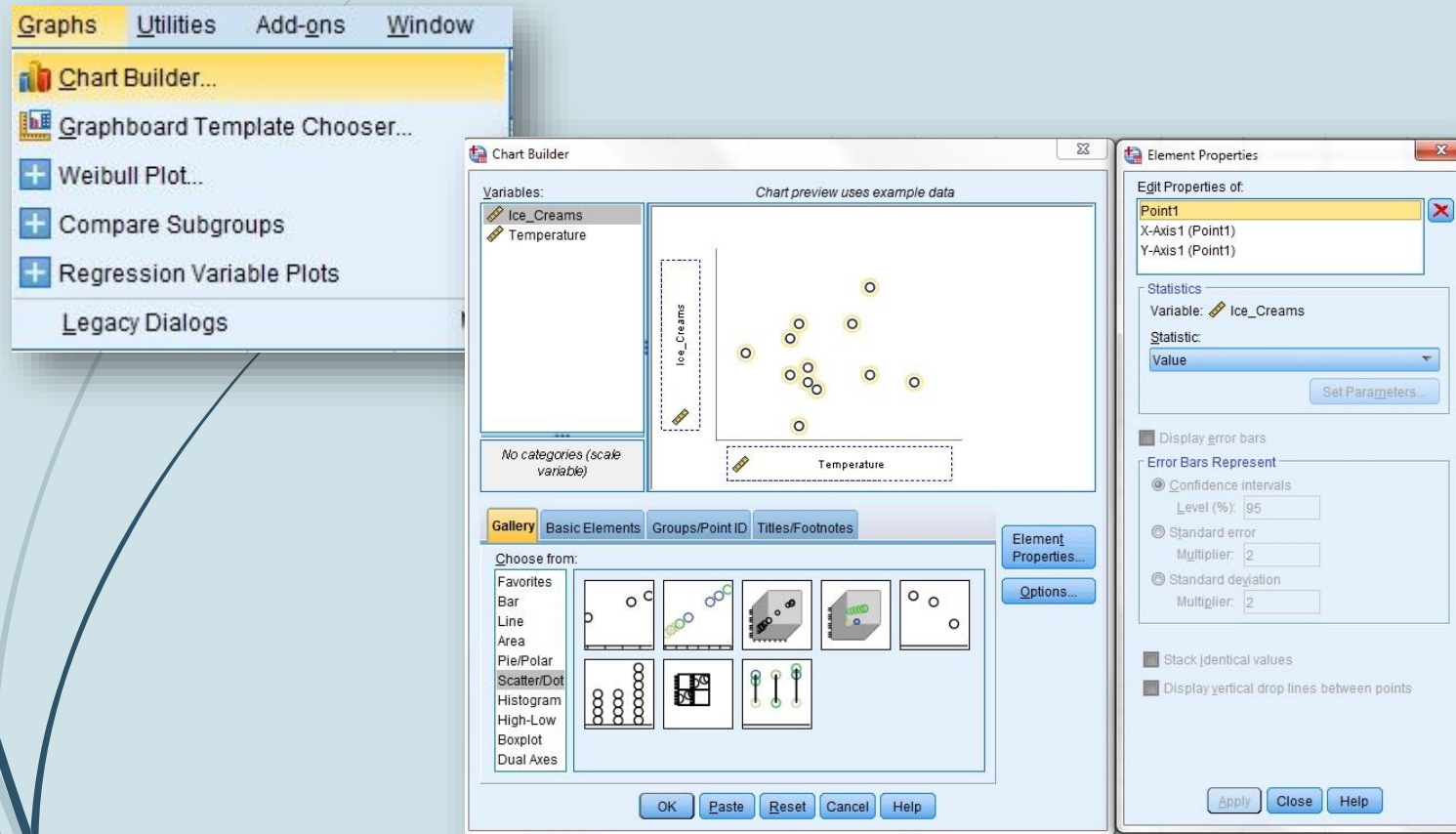
Μέρος Α'

Ερώτημα 5: Για τις παραπάνω μεταβλητές, δημιουργήστε το διάγραμμα διασποράς. Επιβεβαιώστε ότι το αποτέλεσμα που προέκυψε ταυτίζεται με αυτό του ερωτήματος 4.

Λύση ερωτήματος 5

Βήματα στο SPSS

1. Επιλέγω Graphs → Legacy Dialogs
2. Επιλέγω το Scatter/dot και το γράφημα simplescatter (από τη λίστα στην περιοχή Choose from:) και το σέρνω στο παράθυρο προεπισκόπησης γραφήματος
3. Μεταφέρω στον άξονα Y την Ηλικία και στον άξονα X το ΗΒ
4. Πατάω OK



Επίλυση Εργαστηριακής άσκησης

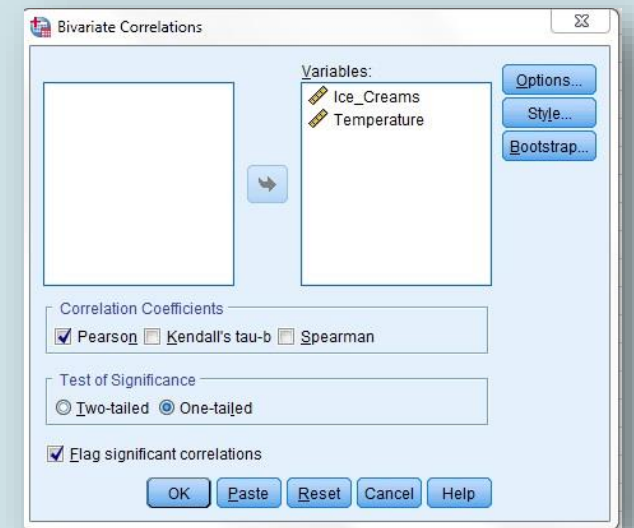
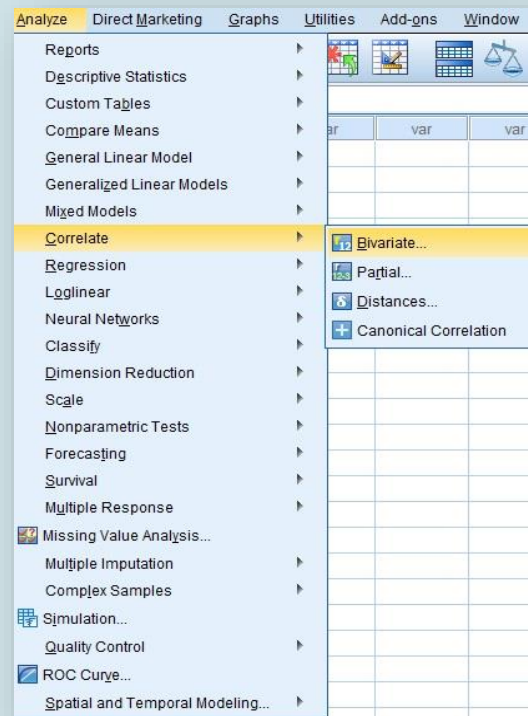
4 Μέρος Α'

Ερώτημα 6: Για τις μεταβλητές PCV και HB υπολογίστε τον συντελεστή συσχέτισης Pearson. Στον πίνακα που προέκυψε βεβαιωθείτε ότι καταλαβαίνετε την σημασία όλων των αποτελεσμάτων.

Λύση ερωτήματος 6

SPSS: Pearson

1. Analyze-> Correlate -> Bivariate
2. Στο αναδυόμενο παράθυρο μεταφέρω τις μεταβλητές μου στο κουτί Variables (**PCV, HB**).
3. Επιλέγω το κουμπί **Pearson** και το One-tailed μιας και έχω μονής κατεύθυνσης υπόθεση.
4. Έπειτα το κουμπί OK.



Επίλυση Εργαστηριακής άσκησης

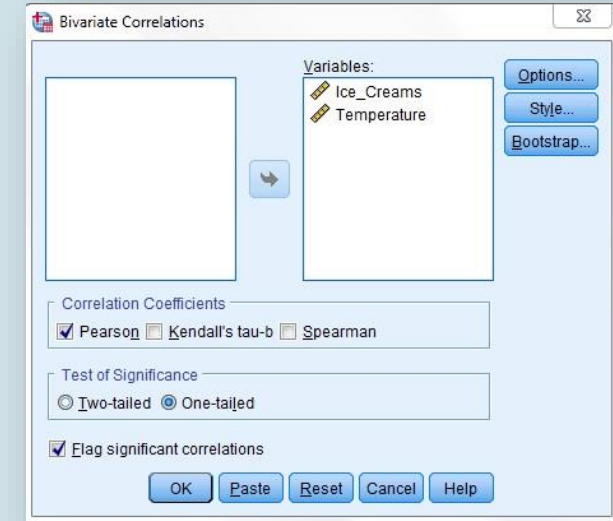
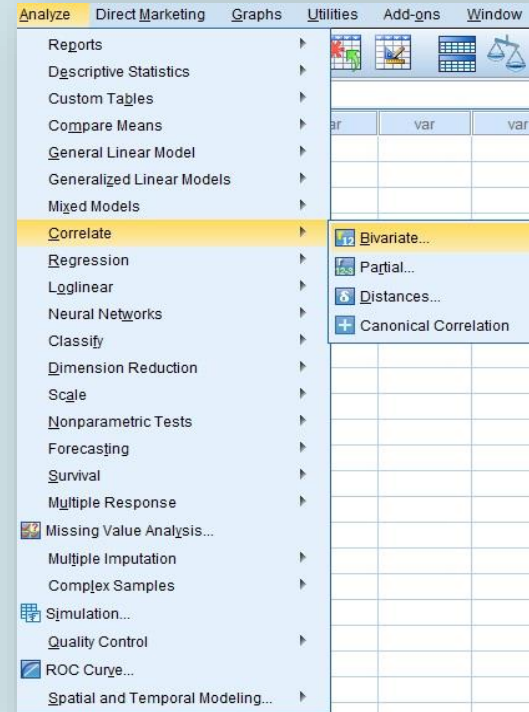
4 Μέρος Α'

Ερώτημα 7: Για τις μεταβλητές PCV και HB υπολογίστε τους συντελεστές συσχέτισης Kendall's tau-b και Spearman. Στον πίνακα που προέκυψε βεβαιωθείτε ότι καταλαβαίνετε την σημασία όλων των αποτελεσμάτων. Επίσης βεβαιωθείτε ότι καταλαβαίνετε πότε ενδείκνυται η χρήση μη παραμετρικών συσχετίσεων.

Λύση ερωτήματος 7

SPSS: Kendall's tau-b και Spearman

1. Analyze-> Correlate -> Bivariate
2. Στο αναδυόμενο παράθυρο μεταφέρω τις μεταβλητές μου στο κουτί Variables (**PCV, HB**).
3. Επιλέγω το κουμπί **Kendall's tau-b και Spearman** και το One-tailed μιας και έχω μονής κατεύθυνσης υπόθεση.
4. Έπειτα το κουμπί OK.



Σημείωση: Όταν δεν πληρούνται οι προϋποθέσεις Παραμετρικών Συσχετίσεων που αναλύσαμε νωρίτερα στις διαφάνειες, κάνουμε χρήση συντελεστή συσχέτισης Spearman ή Kendall's tau-b! Σε οποιαδήποτε άλλη περίπτωση χρησιμοποιούμε συντελεστή συσχέτισης Pearson.



ΠΑΛΙΝΔΡΟΜΗΣΗ
Βοηθητικές Σημειώσεις για
Εργαστηριακή Άσκηση 4 Μέρος Β

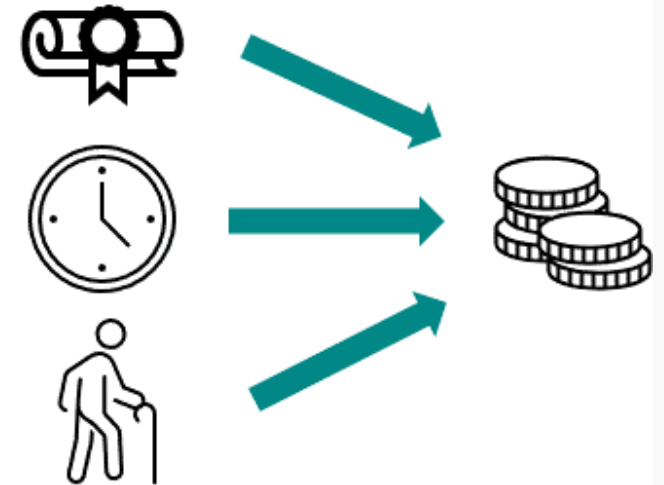
Απλή και πολλαπλή παλινδρόμηση

- ➔ Η ανάλυση παλινδρόμησης είναι ένας τρόπος για να προβλέψουμε την τιμή **μιας μεταβλητής** από τις τιμές **μίας** (απλή παλινδρόμηση) ή **περισσότερων** γνωστών μεταβλητών (πολλαπλή παλινδρόμηση).

Simple Linear Regression



Multiple Linear Regression



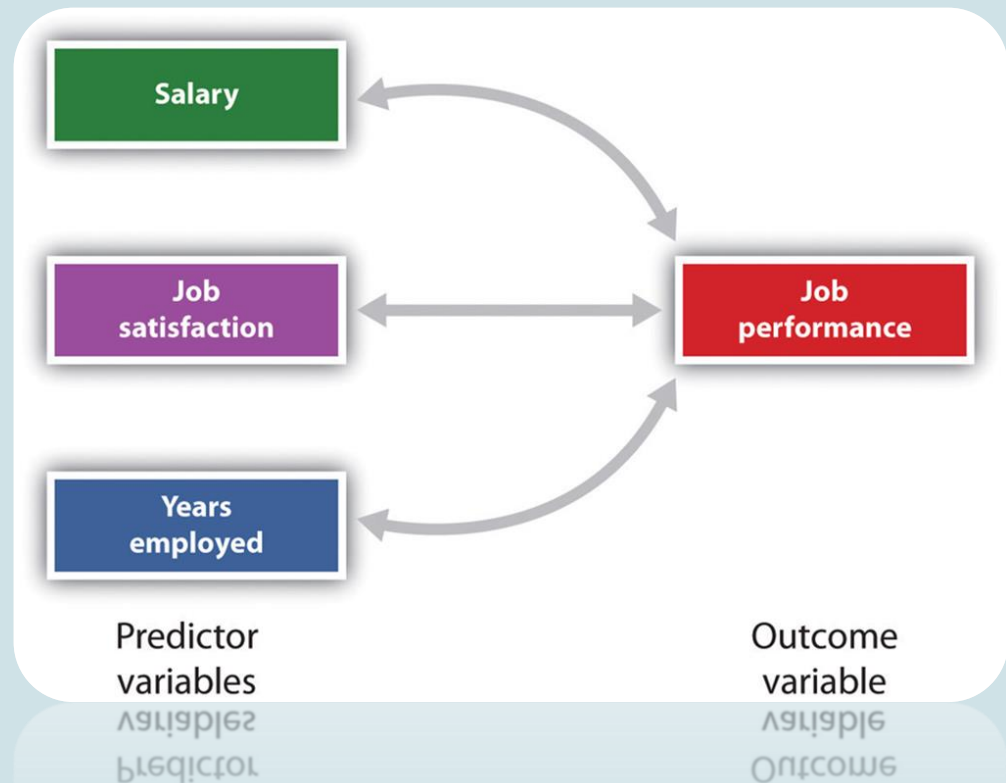
Απλή γραμμική παλινδρόμηση

Απλή ονομάζεται η γραμμική παλινδρόμηση κατά την οποία χρησιμοποιούμε τις τιμές **μίας** μόνο μεταβλητής (ονομάζεται ερμηνευτική ή προβλεπτική μεταβλητή) για να προβλέψουμε τη μεταβλητή κριτήριο.

Προσοχή στην ορολογία!

Προβλεπτική/ερμηνευτική
μεταβλητή (~ανεξάρτητη,
predictive)

Μεταβλητή κριτήριο
(~εξαρτημένα,
response/outcome)



Μπορεί να χρησιμοποιηθεί για να ποσοτικοποιήσει την κατεύθυνση και την ένταση της γραμμικής σχέσης μεταξύ δύο αριθμητικών μεταβλητών

Προβλέπει πώς μια επεξηγηματική μεταβλητή (x) προκαλεί την αλλαγή της μεταβλητής κριτηρίου (y).

Και οι δύο μεταβλητές θεωρούνται τυχαίες μεταβλητές

Οι δύο μεταβλητές είναι εναλλάξιμες, δηλαδή λαμβάνουν πανομοιότυπα αποτελέσματα εάν αλλάξει το "x" και "y"

Συσχέτιση



Παλινδρόμηση



Προϋποθέσεις απλής παλινδρόμησης

Οι μεταβλητές πρέπει να είναι ποσοτικές.

Ανεξαρτησία των παρατηρήσεων

Γραμμικότητα των μεταβλητών, έλεγχος της σχέσης μεταξύ της ανεξάρτητης και της εξαρτημένης μεταβλητής (Q-Q plot residuals)

Κανονικότητα μεταβλητών

*Ομοσκεδαστικότητα Σφαλμάτων, η διακύμανση των καταλοίπων να παραμένει σταθερή.

Απουσία ακραίων τιμών

***Ομοσκεδαστικότητα: Μια ακολουθία ή ένα διάνυσμα τυχαίων μεταβλητών (τ.μ.) είναι ομοσκεδαστική αν όλες οι τ.μ. έχουν την ίδια πεπερασμένη διασπορα. Η μη ύπαρξη αυτής της ιδιότητας ονομάζεται ετεροσκεδαστικότητα. Σε περιπτώσεις όπου η ομοσκεδαστικότητα παραβιάζεται υπερβολικά, μπορεί να οδηγήσει σε λάθος συμπεράσματα.**

Συμπερασματικά

- ▶ Η απλή παλινδρόμηση είναι ένας τρόπος να προβλέψουμε τιμές μιας μεταβλητής από μια άλλη.
- ▶ Το οποίο παρουσιάζεται με τη μορφή ευθείας, όπου αυτή (γραμμή) συνοψίζει καλύτερα τη μορφή δεδομένων.
- ▶ Το R^2 μας λέει πως η διακύμανση εξηγείται στο μοντέλο σε σχέση με το πόση διακύμανση εξηγείται αρχικά. Είναι το ποσοστό διακύμανσης της εξαρτημένης μεταβλητής το οποίο εξηγείται με την ανεξάρτητη μεταβλητή.
- ▶ Ο λόγος F μας λέει πόση διακύμανση μπορεί να εξηγήσει το μοντέλο σε σχέση με το πόση δεν εξηγεί (αναλογικά μας λέει πόσο καλό είναι το μοντέλο σε σχέση με το πόσο κακό είναι)

Η διαφορά μεταξύ της πραγματικής τιμής Y και της προβλεπόμενης τιμής \hat{Y} που βρίσκουμε από την εξίσωση παλινδρόμησης αποτελεί το προβλεπτικό σφάλμα.

Field, 2016, σελ.362

Πάμε τώρα στην άσκησή μας...
...Εργαστηριακή άσκηση 4
Μέρος Β'

Επίλυση Εργαστηριακής άσκησης 4

Μέρος Β'

Ερώτημα 1: Για τις μεταβλητές Ηλικία και ΗΒ εφαρμόστε την τεχνική της απλής γραμμικής παλινδρόμησης για να βρείτε την ευθεία παλινδρόμησης.

Λύση ερωτήματος 1

- ▶ Analyze → Regression → Linear
- ▶ Στο αναδυόμενο παράθυρο μεταφέρω τις μεταβλητές μου στο dependent θέτω το HB και στο independent την Ηλικία
- ▶ Έπειτα το κουμπί OK.

Επίλυση Εργαστηριακής άσκησης 4

Μέρος Β'

Ερώτημα 2: Κάντε τις απαραίτητες ενέργειες, ώστε να εμφανίζονται τα διαστήματα εμπιστοσύνης. Βεβαιωθείτε ότι καταλαβαίνετε την σημασία όλων των αποτελεσμάτων που προέκυψαν.

Σημείωση: *Το διάστημα εμπιστοσύνης είναι ένα διάστημα αριθμών που πιστεύεται/εκτιμάται ότι εμπεριέχει μια άγνωστη παράμετρο (π.χ. μέσο, τυπική απόκλιση) του πληθυσμού. Ταυτόχρονα, είναι ένα μέτρο της εμπιστοσύνης για την άγνωστη παράμετρο.

Λύση ερωτήματος 2

1. Analyze → Descriptive statistics → Explore
2. Στο αναδυόμενο παράθυρο μεταφέρω τις μεταβλητές μου στο dependent list θέτω το HB και την Ηλικία.
3. Στο display επιλέγω Both
4. Έπειτα το κουμπί OK.

Επίλυση Εργαστηριακής άσκησης 4

Μέρος Β'

- ▶ **Ερώτημα 3:** Δημιουργήστε το διάγραμμα διασποράς για τις παραπάνω μεταβλητές και προσαρμόστε την ευθεία διασποράς ανάμεσα στα σημεία διασποράς. Ελέγξτε αν η ευθεία παλινδρόμησης που προκύπτει ταυτίζεται με την εξίσωση των προηγούμενων ερωτημάτων.

Λύση ερωτήματος 3

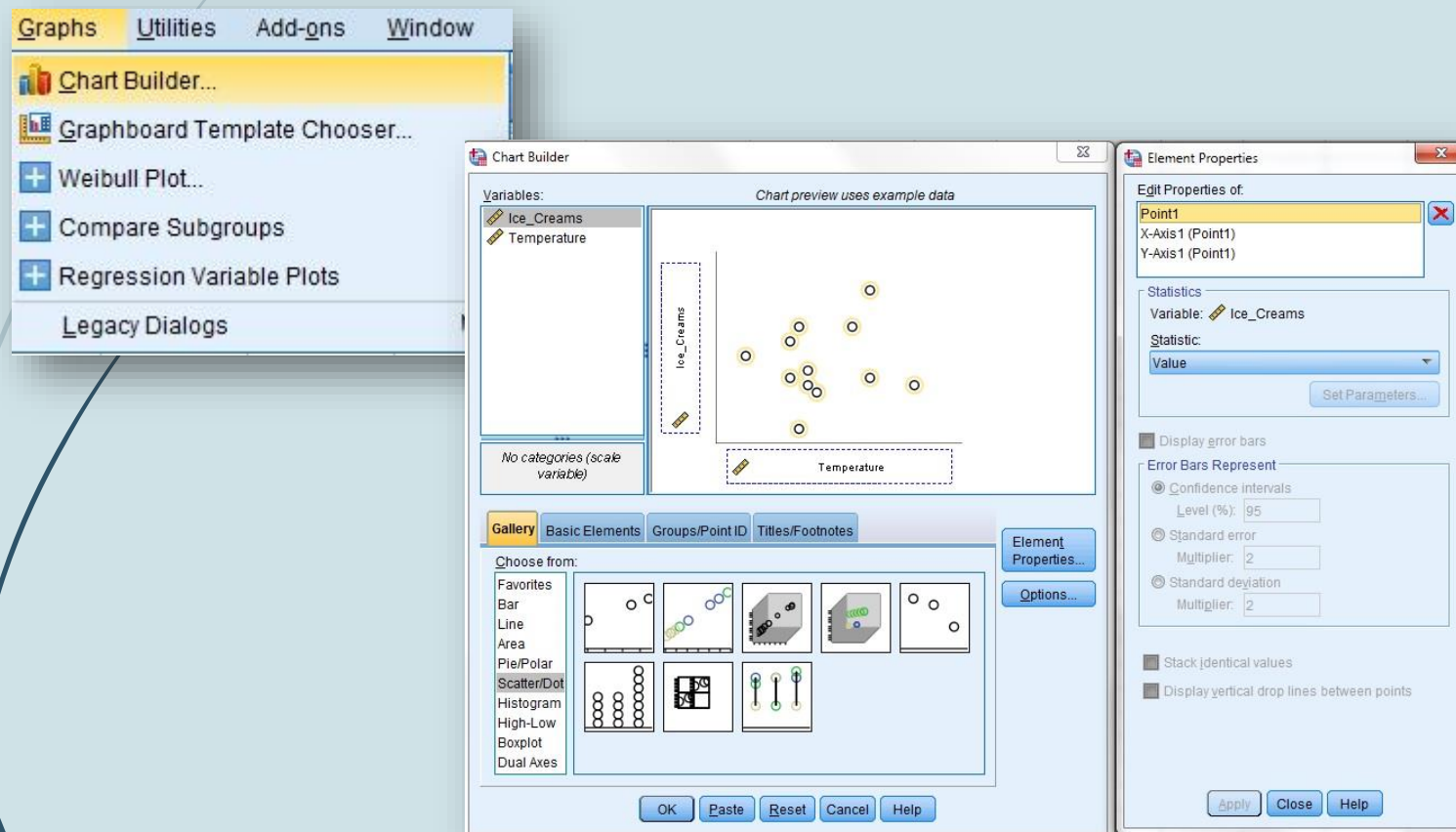
Βήματα στο SPSS

Επιλέγω Graphs →
Chart Builder

Επιλέγω το Scatter/dot
και το γράφημα simple
scatter with fit line (από
τη λίστα στην περιοχή
Choose from:) και το
σέρνω στο παράθυρο
προεπισκόπησης
γραφήματος

Μεταφέρω στον άξονα
Y το ΗΒ και στον άξονα
X την Ηλικία

Πατάω OK



Επίλυση Εργαστηριακής άσκησης 4

Μέρος Β'

- ▶ **Ερώτημα 4** (επανάληψη των ερωτημάτων 1,2,3 για τις μεταβλητές HB, PCV).
- ▶ **Ερώτημα 1:** Για τις μεταβλητές **HB, PCV** εφαρμόστε την τεχνική της απλής γραμμικής παλινδρόμησης για να βρείτε την ευθεία παλινδρόμησης.
- ▶ **Ερώτημα 2:** Κάντε τις απαραίτητες ενέργειες, ώστε να εμφανίζονται τα διαστήματα εμπιστοσύνης. Βεβαιωθείτε ότι καταλαβαίνετε την σημασία όλων των αποτελεσμάτων που προέκυψαν.
- ▶ **Ερώτημα 3:** Δημιουργήστε το διάγραμμα διασποράς για τις παραπάνω μεταβλητές και προσαρμόστε την ευθεία διασποράς ανάμεσα στα σημεία διασποράς. Ελέγξτε αν η ευθεία παλινδρόμησης που προκύπτει ταυτίζεται με την εξίσωση των προηγούμενων ερωτημάτων.

Λύσεις ερωτήματος 4

Ερώτημα 1

1. Analyze → Regression → Linear
2. Στο αναδυόμενο παράθυρο μεταφέρω τις μεταβλητές μου στο dependent θέτω το HB και στο independent το PCV
3. Έπειτα το κουμπί OK.

Ερώτημα 2

1. Analyze → Descriptive statistis → Explore
2. Στο αναδυόμενο παράθυρο μεταφέρω τις μεταβλητές μου στο dependent list θέτω το HB και το PCV.
3. Στο display επιλέγω Both
4. Έπειτα το κουμπί OK.

Ερώτημα 3

Επιλέγω Graphs →
Chart Builder

Επιλέγω το Scatter/dot
και το γράφημα simple
scatter with fit line (από
τη λίστα στην περιοχή
Choose from:) και το
σέρνω στο παράθυρο
προεπισκόπησης
γραφήματος

Μεταφέρω στον άξονα
Y το HB και στον άξονα
X το PCV

Πατάω OK



Σας ευχαριστώ!