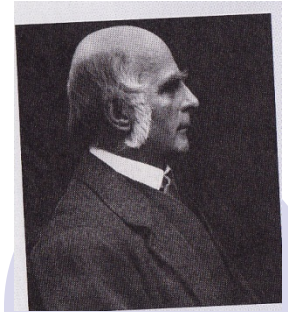
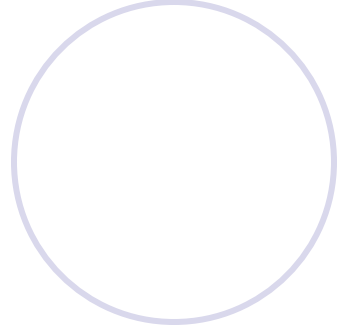
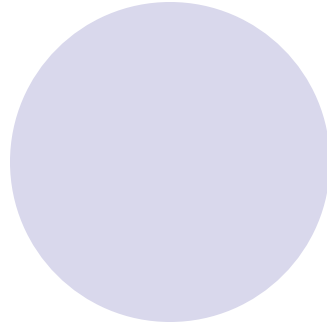
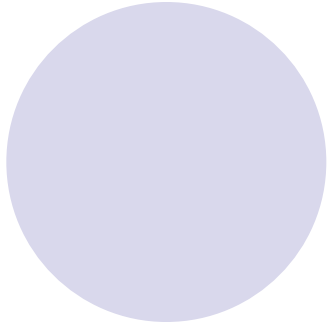
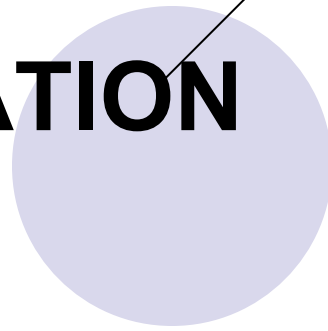
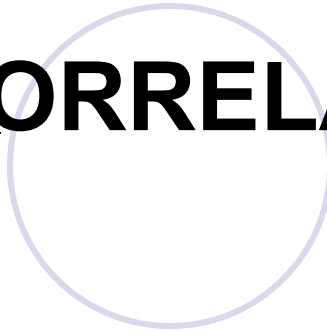
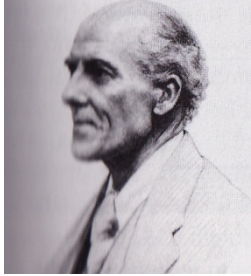


Francis Galton



CORRELATION

Karl Pearson



Correlation

- A correlation is a measure of the extent to which two variables are nearly related.

Is there a relationship between the drinks (mean) and the times of attendance (mean) of the lecture in statistics?



What we could do..



- Using descriptive statistics...
- Να παρουσιάσουμε την κατανομή των μεταβλητών (κανονική κατανομή?)
- Μ.Ο και Τ.Α της κάθε μεταβλητής ώστε να δούμε την πιο αντιπροσωπευτική τιμή αλλά και τον τρόπο που διασπείρονται οι τιμές γύρω απ' αυτήν
- Conclusion of the relationship???
- Answer = No.....

Άλλα παραδείγματα...

- Η σχέση μεταξύ της μόρφωσης/εκπαίδευσης (σε χρόνια) των γυναικών στις αναπτυσσόμενες χώρες και ο αριθμός των παιδιών που γεννούν. Περισσότερα χρόνια εκπαίδευσης λιγότερα παιδιά...
- Η σχέση μεταξύ της μέσης θερμοκρασίας και του αριθμού των απεργιών σε ένα μήνα. Όταν η θερμοκρασία είναι υψηλή, ο κόσμος είναι έξω, απολαμβάνει τον ήλιο, κάνει βόλτες... περισσότερες απεργίες...

Έρευνα του Aron et al. 2000

Aron, A., & Aron, E. N., & Coups, E. J. (2009).
Statistics for Psychology.
London. Pearson.

- Δείγμα: 113 παντρεμένα ζευγάρια
- Υπόθεση: Σχέση του να κάνω συναρπαστικά πράγματα με το σύντροφο μου με το επίπεδο ικανοποίησης από το γάμο μου
- Πόσο ενδιαφέροντα-συναρπαστικά είναι αυτά που κάνετε, στον ελεύθερο χρόνο σας, με το σύντροφό σας; 1 = «Καθόλου συναρπαστικά», 5 = «Πολύ συναρπαστικά» επίσης απάντησαν και σε μια κλίμακα ικανοποίησης από το γάμο, Γενικά, πόσο συχνά σκέφτεστε ότι ο γάμος σας πηγαίνει καλά; 10 = «Ποτέ», 60 = «Πολύ συχνά»

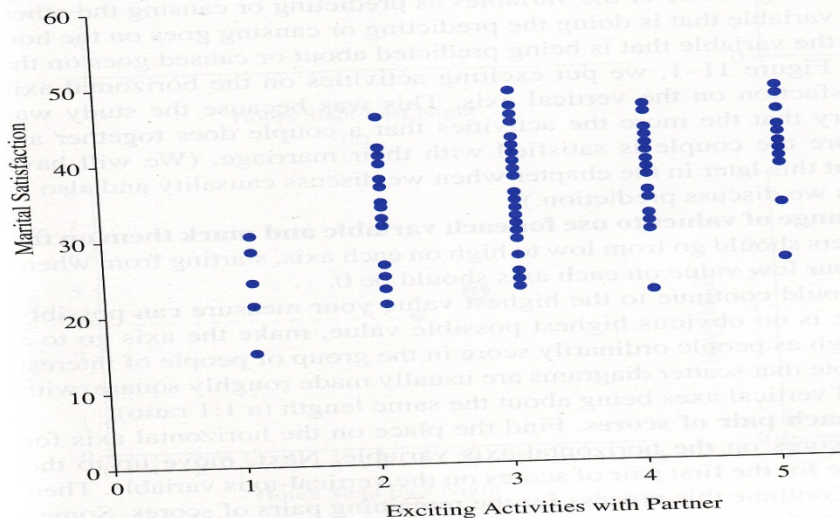


Figure 11-1 Scatter diagram showing the correlation for 113 married individuals between doing exciting activities with their partner and their marital satisfaction. (Data from Aron et al., 2000.)

Τώρα ή ποτέ...επιτέλους καιρός να τη ρίξω από τα ψηλά βουνά

Μ' αγαπάει όσο τα ψηλά βουνά





Όμως..

- Σε μερικές περιπτώσεις ακόμη και αν υπάρχει συσχέτιση μεταξύ δυο μεταβλητών αυτό δεν πρέπει να λαμβάνεται ως απόδειξη ύπαρξης *αιτιώδους σχέσης – causal relationship*.
Π.χ Αν η τιμή μιας μεταβλητής αυξάνεται όσο και η τιμή της άλλης, **αυτό δεν σημαίνει ότι η μια μεταβλητή εξηγεί την άλλη ή μπορεί η μια να προβλέψει την άλλη...**

Diamond, I., & Jefferies, J. (2001).
Beginning statistics. An introduction
for social scientists. London: Sage.

Π.χ. Σχέση καπνίσματος με Καρκίνου του Πνεύμονα

- Συσχέτιση = συμμεταβολή δυο μεταβλητών όχι αιτιώδης σχέση. Δηλαδή αν κάθε μεταβολή των τιμών της μιας μεταβλητής συνοδεύεται από συστηματική μεταβολή στις τιμές της άλλης.
- Αιτιώδης σχέση = Χρονικά η αιτία προηγείται του αποτελέσματος. Αίτιο = κάπνισμα, αποτέλεσμα = καρκίνος. Με τη συσχέτιση = **δεν προϋποθέτουμε καμιά χρονική σειρά.**

Π.χ. Σχέση καπνίσματος με Καρκίνου του Πνεύμονα

- Αν έχουμε συσχέτιση τότε:
- Η X είναι αιτία της Y
- Η Y είναι αιτία της X
- Μια τρίτη μεταβλητή Z είναι η αιτία της συσχέτισης των μεταβλητών X και Y
- Αν όμως χρησιμοποιούμε μόνο το δείκτη συσχέτισης τότε δεν μπορούμε να βρούμε τη μεταβλητή Z (π.χ κληρονομικότητα)
- Εντούτοις παράγοντες με αιτιώδης σχέση έχουν και υψηλή συσχέτιση...

Δείκτης συσχέτισης (Correlation Coefficient ή r)

- Με την εφαρμογή του δείκτη συσχέτισης:
- Αν οι μεταβλητές σχετίζονται
- Είδος συσχέτισης, θετική ή αρνητική
- Βαθμό συσχέτισης, υψηλή, χαμηλή ή μέτρια
- Οπότε... ανάλογα με το είδος των μεταβλητών (συνεχείς) έχουμε και τον ανάλογο δείκτη συσχέτισης. Εμείς θα χρησιμοποιούμε τον **Pearson r** ...

Παρουσίαση συσχέτισης με διάγραμμα σκεδασμού – scatter diagram

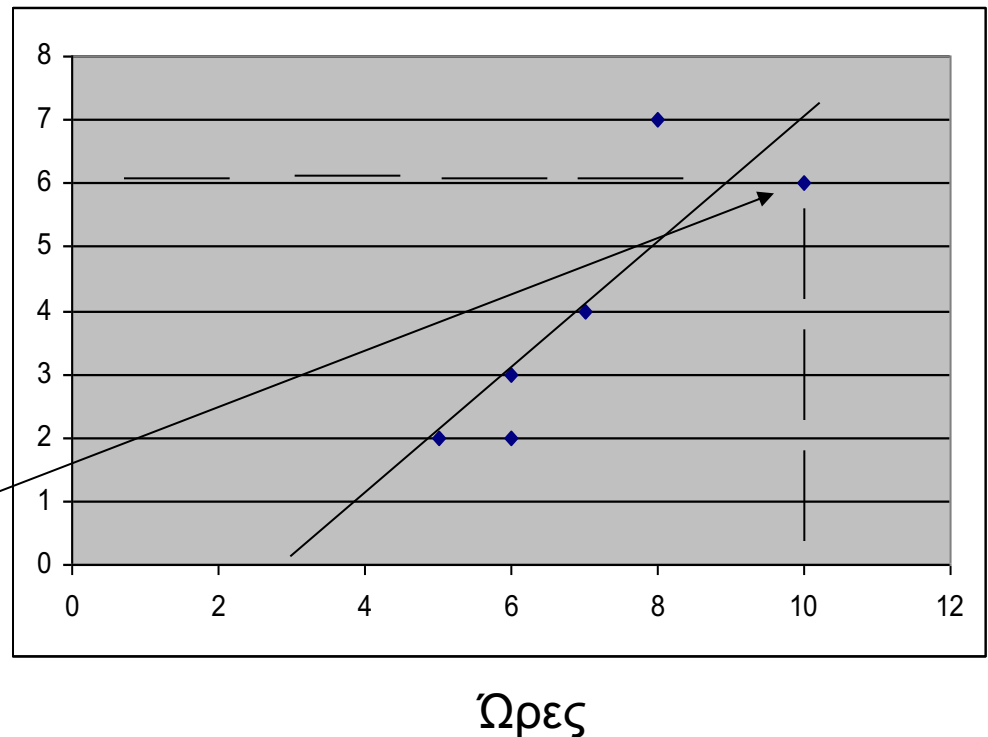
- Σχεδιάζουμε τους άξονες. Συνήθως η Y είναι στον κατακόρυφο και η X μεταβλητή στον οριζόντιο.
- Ορίζουμε την κλίμακα μέτρησης για κάθε μεταβλητή και τη γράφουμε στους άξονες
- Ορίζουμε με τελεία το κάθε ζευγάρι των τιμών

Π.χ. Σχέση ωρών ύπνου με διάθεση

- A) πόσες ώρες κοιμήθηκες χτες;
- B) πόσο όμορφα αισθάνεσαι τώρα σε κλίμακα από 0 = «Καθόλου χαρούμενος/η», 8 = «Υπερβολικά χαρούμενος»

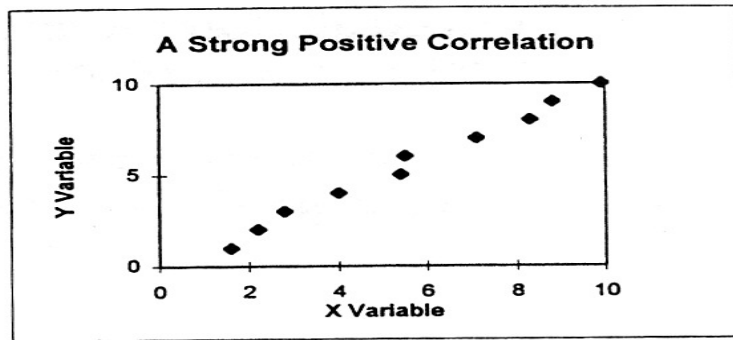
Ώρες ύπνου	Διάθεση
7	4
5	2
8	7
6	2
6	3
10	6

Διάθεση

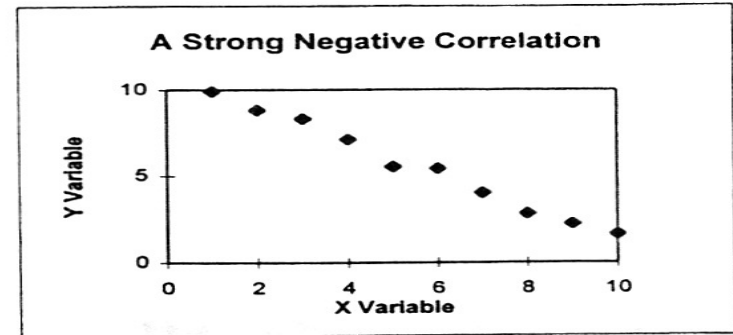


Διαφορετικοί τύποι συσχέτισης σε γραφήματα σκεδασμού

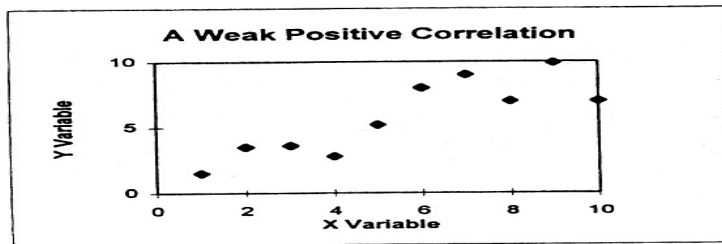
Υψηλή Θετική Συσχέτιση



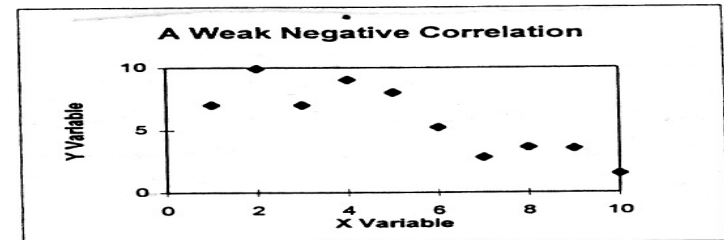
Υψηλή Αρνητική Συσχέτιση



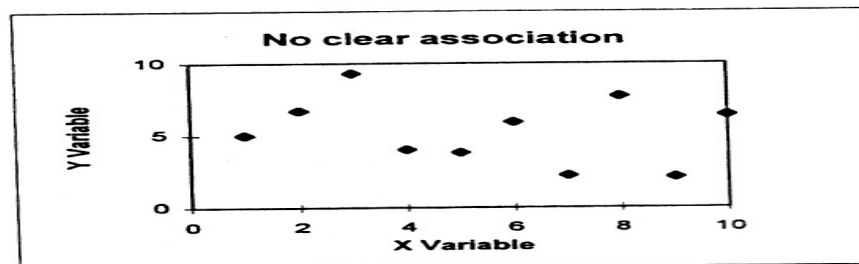
Χαμηλή Θετική Συσχέτιση



Χαμηλή Αρνητική Συσχέτιση



Δεν υπάρχει συσχέτιση – Μηδέν συσχέτιση

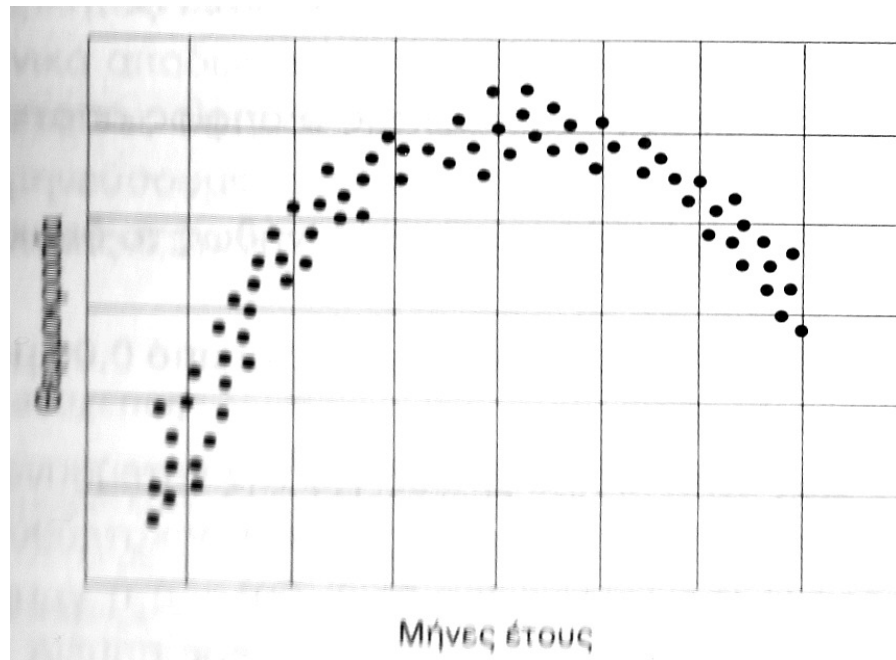


Σχέσεις μεταξύ δυο μεταβλητών που δεν ακολουθούν μια ευθύγραμμη κατεύθυνση

- Μέχρι τώρα όταν αυξάνει η μια μεταβλητή αυξάνει και η άλλη (θετική συσχέτιση) ή όταν αυξάνει η μια μειώνεται η άλλη (αρνητική συσχέτιση), **ευθύγραμμη σχέση-
Linear relationship**

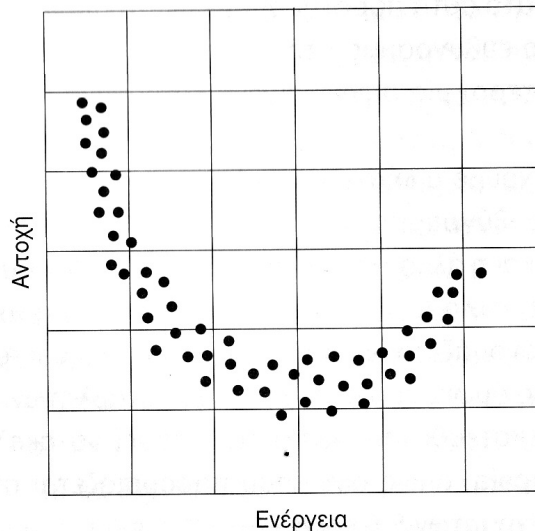
Καμπυλόγραμμες σχέσεις – curvilinear relationships

- Μήνες του χρόνου με θερμοκρασία. Στην αρχή αυξάνονται οι τιμές της X και της Y ενώ στη συνέχεια ενώ αυξάνονται της X της Y μειώνονται = **Θετική καμπυλόγραμμη σχέση – positive curvilinear relationship**



Π.χ. Σχέση Αντοχής με ενέργεια

- Ενέργεια που καταναλώνεται κατά την άσκηση με την αντοχή. Αρχικά όταν αυξάνονται οι τιμές της X (ενέργεια), μειώνονται οι τιμές της Y (αντοχή) μέχρι ένα σημείο, ενώ στη συνέχεια αυξάνονται και της X αλλά και της Y , καθώς με την εξάσκηση το άτομο αποκτά περισσότερη αντοχή = **αρνητική καμπυλόγραμμη σχέση – negative curvilinear relationship**



Ερμηνεία του δείκτη συσχέτισης

- Ο δείκτης αποτελείται από τρία στοιχεία:
 - A) Από ένα **θετικό** ή **αρνητικό** πρόσημο (συνήθως το θετικό παραλείπεται)
 - B) Αριθμητική τιμή που κυμαίνεται από **-1.00 έως 1.00**
 - Γ) Αν είναι στατιστικά σημαντική $p < .05$ ή $p < .001$

Ερμηνεία δείκτη συσχέτισης

- Κατεύθυνση της συσχέτισης = γραμμή παλινδρόμησης (regression line), δηλαδή η γραμμή που περνά όσο το δυνατόν πλησιέστερα από τα περισσότερα σημεία του διαγράμματος.
- **+** = η κλίση έχει πορεία από κάτω αριστερά προς πάνω δεξιά, **-** = από πάνω αριστερά προς κάτω δεξιά
- **Τιμές r** όπως (π.χ 0.42 & 0.64) μας πληροφορούν πόσο κοντά από τα σημεία του διαγράμματος σκεδασμού περνά η γραμμή παλινδρόμησης
- **Η τιμή 1** σημαίνει ότι όλα τα σημεία εφάπτονται της γραμμής παλινδρόμησης
- **Η τιμή 0** τυχαία κατανομή γύρω από τη γραμμή παλινδρόμησης
- Σε γενικές γραμμές αν τα σημεία του διαγράμματος είναι κοντά στη γραμμή τόσο μεγαλύτερος ο βαθμός συσχέτισης

Ερώτηση: Πόσο υψηλή πρέπει να είναι μια συσχέτιση για να υποστηρίξουμε ότι δυο μεταβλητές εμφανίζουν υψηλή συσχέτιση

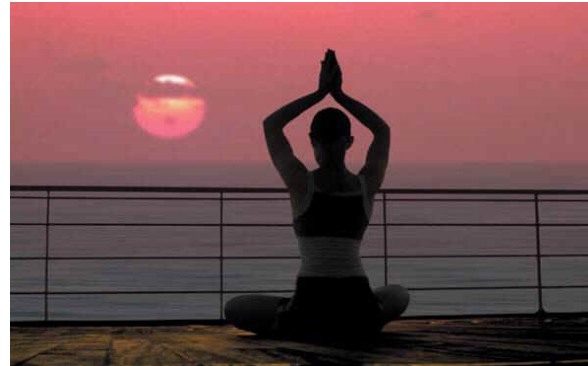
Αν ο δείκτης είναι μικρότερος ή ίσος ± 0.20	Δεν υπάρχει συσχέτιση
Αν ο δείκτης κυμαίνεται μεταξύ του ± 0.20 και ± 0.30	Χαμηλή προς μέτρια συσχέτιση
Αν ο δείκτης κυμαίνεται μεταξύ του ± 0.30 και ± 0.50	Μέτρια συσχέτιση
Αν ο δείκτης κυμαίνεται μεταξύ του ± 0.50 και ± 0.60	Μέτρια προς Υψηλή συσχέτιση
Αν ο δείκτης κυμαίνεται μεταξύ του ± 0.60 και ± 0.99	Υψηλή συσχέτιση

Ο δείκτης συσχέτισης correlation coefficient

- Βαθιά αναπνοή....

Τύπος

$$r = \frac{\sum [(X_i - \bar{X})(Y_i - \bar{Y})]}{\sqrt{[\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2]}}$$



Δεν είναι τόσο δύσκολος όσο φαίνεται
... ωστόσο μπορεί να μας φάει αρκετό
χρόνο στον υπολογισμό του...

Βήματα

Country	Total fertility rate, 1994	% of women using contraception (any method), 1987-94
UK	1.8	82
USA	2.1	71
Gambia	5.4	12
Indonesia	2.8	55
Mexico	3.0	53
Brazil	2.3	66
Uganda	7.1	15
Slovakia	1.8	74
Niger	7.3	4
Botswana	4.7	33

Source: UNDP, 1997

$$r = \frac{\sum [(X_i - \bar{X})(Y_i - \bar{Y})]}{\sqrt{[\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2]}}$$

- ΒΗΜΑ 1: Γράφουμε όλες τις παρατηρήσεις των δυο μεταβλητών σε δυο στήλες (A και B). Να θυμάστε ποια στήλη αντιπροσωπεύει τη X και ποια τη Y
- ΒΗΜΑ 2: Για κάθε μεταβλητή προσθέτουμε όλες τις τιμές και τις διαιρούμε με το σύνολο των τιμών για να πάρουμε το μέσο όρο της X και της Y
- ΒΗΜΑ 3: Στις στήλες C και D, υπολογίζουμε τα υπόλοιπα (residuals) αφαιρώντας το μέσο όρο από κάθε τιμή.
- ΒΗΜΑ 4: Στις στήλες E και F, υψώστε στο τετράγωνο τις τιμές που πήρατε από τις στήλες C και D
- ΒΗΜΑ 5: Στη στήλη G, πολλαπλασιάστε τις τιμές από τη στήλη C με τις τιμές από τη στήλη D
- ΒΗΜΑ 6: Για τις στήλες E, F, και G προσθέστε όλες τις τιμές σε κάθε στήλη για να πάρετε το σύνολο στο τέλος από κάθε στήλη. Αυτά είναι τα νούμερα που θα χρησιμοποιήσετε για την εξίσωση

A	B	C	D	E	F	G
X	Y	(X - \bar{X})	(Y - \bar{Y})	(X - \bar{X}) ²	(Y - \bar{Y}) ²	(X - \bar{X})(Y - \bar{Y})
82	1.8	35.5	-2.03	1260.25	4.12	-72.07
71	2.1	24.5	-1.73	600.25	2.99	-42.39
12	5.4	-34.5	1.57	1190.25	2.46	-54.17
55	2.8	8.5	-1.03	72.25	1.06	-8.76
53	3.0	6.5	-0.83	42.25	0.69	-5.40
66	2.3	19.5	-1.53	380.25	2.34	-29.84
15	7.1	-31.5	3.27	992.25	10.69	-103.01
74	1.8	27.5	-2.03	756.25	4.12	-55.83
4	7.3	-42.5	3.47	1806.25	12.04	-147.48
33	4.7	-13.5	0.87	182.25	0.76	-11.75
$\bar{X} = 46.5$	$\bar{Y} = 3.83$		Totals	7282.50	41.27	-530.70
				$\sum(X - \bar{X})^2$	$\sum(Y - \bar{Y})^2$	$\sum(X - \bar{X})(Y - \bar{Y})$

$$r = \frac{\sum [(X_i - \bar{X})(Y_i - \bar{Y})]}{\sqrt{[\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2]}}$$

$$r = \frac{-530.70}{\sqrt{(7282.50 \times 41.27)}} = -0.97 \text{ (to two decimal places)}$$

Σχέση μεγέθους, σημαντικότητας και δείγματος

- Αν μια σχέση είναι στατιστικά σημαντική εξαρτάται από το μέγεθος του συντελεστή συσχέτισης και το μέγεθος του δείγματος
- Όσο πιο μεγάλος (κοντά στο ± 1) είναι ο συντελεστής συσχέτισης, τόσο πιο πιθανό είναι η συσχέτιση να είναι στατιστικά σημαντική,

ΩΣΤΟΣΟ

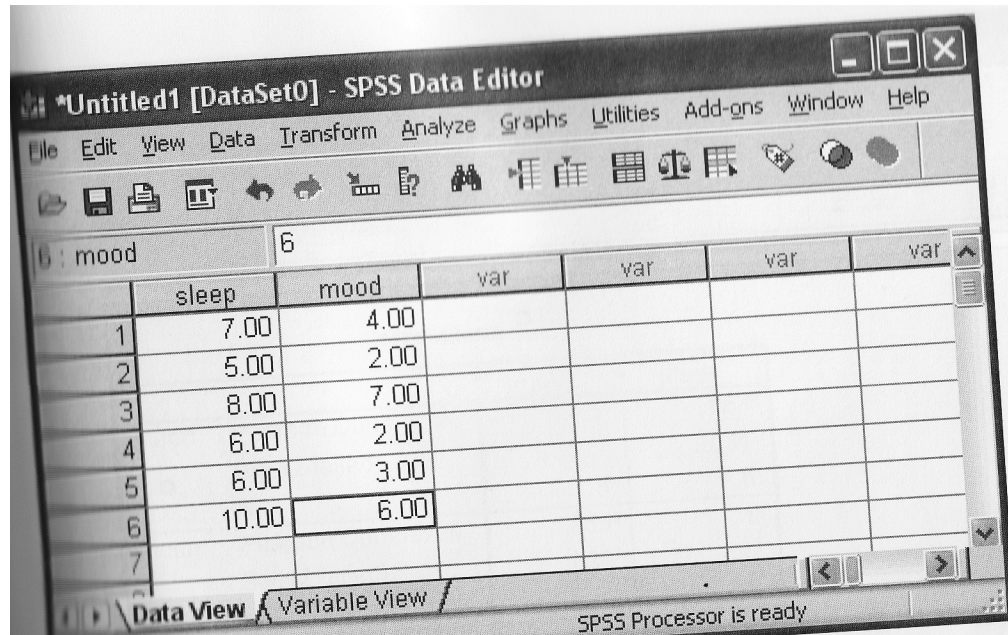
Στατιστικά σημαντική σχέση ΔΕΝ σημαίνει ΜΕΓΑΛΗ σχέση επειδή...

Όσο πιο μεγάλο είναι το δείγμα, τόσο πιο πιθανό είναι να είναι μια συσχέτιση στατιστικά σημαντική

Παράγοντες που επηρεάζουν το δείκτη συσχέτισης

- Η καμπυλόγραμμη σχέση μεταξύ των μεταβλητών
- Η ύπαρξη ακραίων τιμών (outliers)

Συσχέτιση με τη χρήση του SPSS

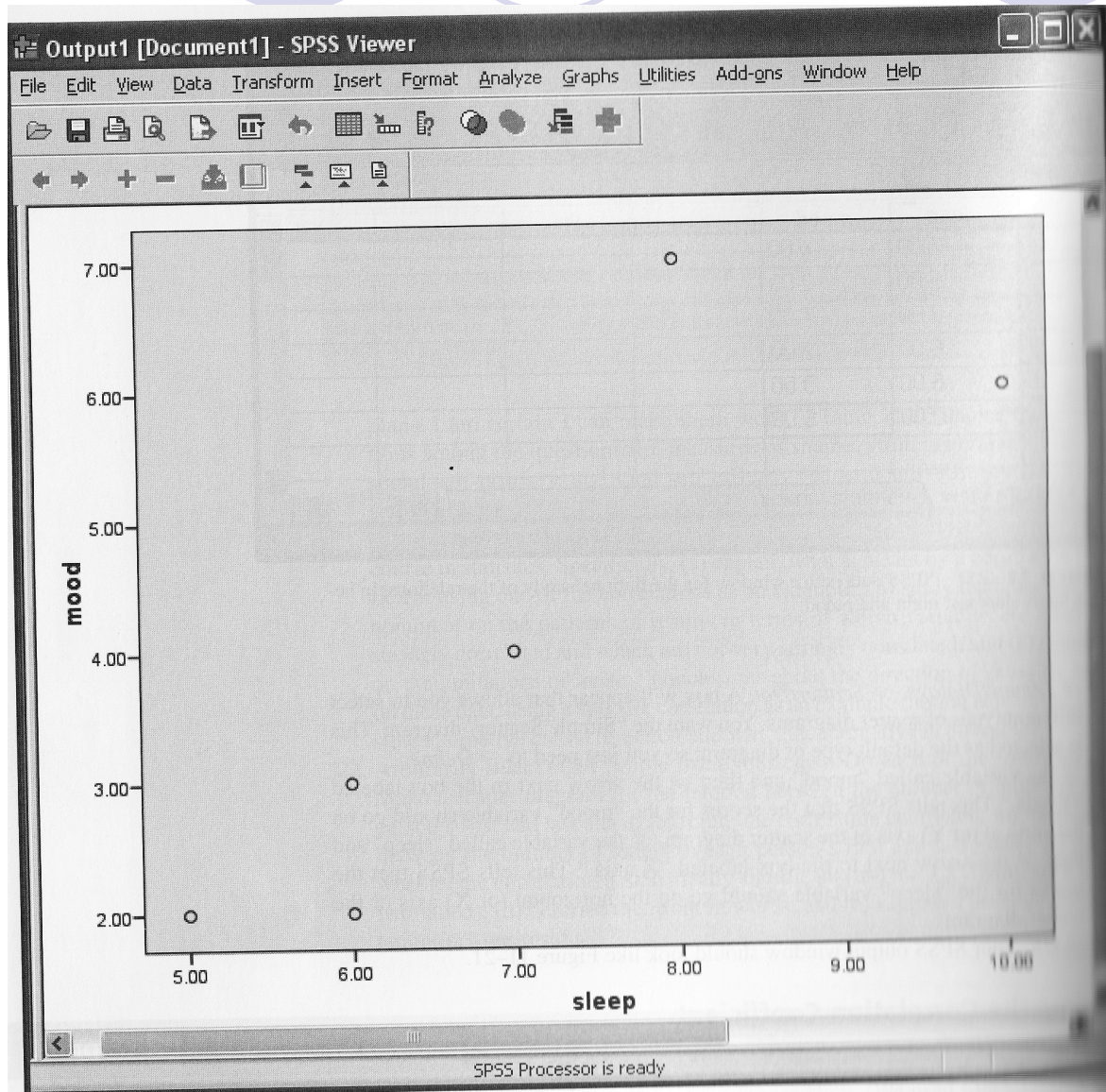


The screenshot shows the SPSS Data Editor window titled '*Untitled1 [DataSet0] - SPSS Data Editor'. The window displays a data table with two columns: 'sleep' and 'mood'. The 'sleep' column contains values 7.00, 5.00, 8.00, 6.00, 6.00, 10.00, and an empty cell. The 'mood' column contains values 4.00, 2.00, 7.00, 2.00, 3.00, 6.00, and an empty cell. The interface includes a menu bar (File, Edit, View, Data, Transform, Analyze, Graphs, Utilities, Add-ons, Window, Help) and a toolbar with various icons. The status bar at the bottom indicates 'SPSS Processor is ready'.

	sleep	mood	var	var	var	var
1	7.00	4.00				
2	5.00	2.00				
3	8.00	7.00				
4	6.00	2.00				
5	6.00	3.00				
6	10.00	6.00				
7						

Aron, A., & Aron, E. N., & Coups, E. J. (2009). *Statistics for Psychology*. London. Pearson.

Γράφημα σκεδασμού



Graphs

Legacy/Dialogs

Scatter/Dot

Simple Scatter (selected)

Define

Click στο Mood

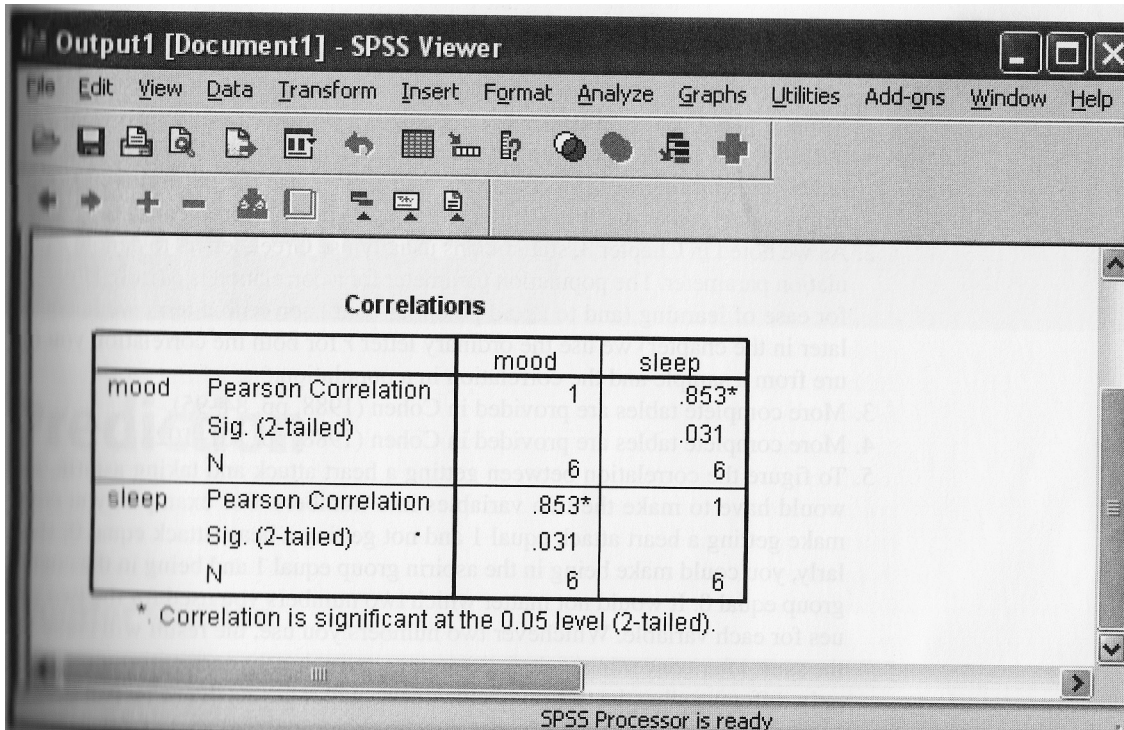
Click στο Y axis

Click στο Sleep

Click στο X axis

Aron, A., & Aron, E. N., & Coups, E. J. (2009).
Statistics for Psychology.
London. Pearson.

Output



The screenshot shows the SPSS Viewer window titled "Output1 [Document1] - SPSS Viewer". The menu bar includes File, Edit, View, Data, Transform, Insert, Format, Analyze, Graphs, Utilities, Add-ons, Window, and Help. The toolbar contains various icons for file operations and analysis. The main content area displays a table titled "Correlations" with the following data:

		mood	sleep
mood	Pearson Correlation	1	.853*
	Sig. (2-tailed)		.031
	N	6	6
sleep	Pearson Correlation	.853*	1
	Sig. (2-tailed)	.031	
	N	6	6

*. Correlation is significant at the 0.05 level (2-tailed).

SPSS Processor is ready

Analyze

Correlate

Bivariate

Click στο mood

Click στο βελάκι δίπλα από το variables

Click στο sleep

Click στο βελάκι δίπλα από το variables

Πως αναφέρεται η συσχέτιση σε ένα άρθρο

Huck, S.W. (2000). *Reading Statistics and Research*. New York: Longman

Excerpts 3.20–3.22 Pearson's Product-Moment Correlation

The study discussed in this article examined the relationship between the degree of involvement with clients with severe mental illness and social workers' job satisfaction and burnout. . . . The four research hypotheses were investigated by using Pearson's product-moment correlation coefficient.

Source: G. A. Acker. (1999). The impact of clients' mental illness on social workers' job satisfaction and burnout. *Health and Social Work, 24*(2), p. 112.

First, relationships among personality, psychosocial, and health behaviors were examined using Pearson correlations.

Source: G. R. Marks and S. K. Lutendorf. (1999). Perceived health competence and personality factors differentially predict health behaviors in older adults. *Journal of Aging and Health, 11*(2), p. 221.

Ratings regarding thoroughness of training correlated moderately with both amount of supervision received ($r = .54$) and satisfaction with supervision ($r = .65$).

Source: C. A. Pietz, R. L. DeMier, R. D. Dienst, J. B. Green, and B. Scully. (1999). Psychology internship training in a correctional facility. *Criminal Justice and Behavior, 25*(1), p. 99.

Διατύπωση αποτελεσμάτων σύμφωνα με
την Αμερικανική Ψυχολογική Εταιρεία
(APA)

$$r = .54, p < .05$$

ή

$$r(8) = .54, p < .05$$

Πίνακας συσχέτισης

Table 11-12 Zero-Order Correlations for Study Variables

Variable	1	2	3	4	5	6	7	8	9	10
1. Women's report of stress	—									
2. Men's report of women's stress	.17	—								
3. Partner Support 1	-.28*	-.18	—							
4. Partner Support 2	-.27*	-.18	.44***	—						
5. Depressed Mood 1	.23*	.10	-.34**	-.17	—					
6. Depressed Mood 2	.50***	.14	-.42***	-.41***	.55***	—				
7. Women's age	.06	.16	.04	-.24*	-.35*	-.09	—			
8. Women's ethnicity	-.19	-.09	-.16	-.14	.11	.13	-.02	—		
9. Women's marital status	-.18	.01	.12	.24*	-.04	-.20	.05	-.34**	—	
10. Parity	.19	.13	-.11	-.17	.10	.16	.26*	.31*	-.12	—

* $p < .05$, ** $p < .01$, *** $p < .001$.

Source: Chapman, H. A., Hobfoll, S. E., & Ritter, C. (1997). Partners' stress underestimations lead to women's distress: A study of pregnant inner-city women. *Journal of Personality and Social Psychology*, 73, 418–425. Published by the American Psychological Association. Reprinted with permission.

Aron, A., & Aron, E. N., & Coups, E. J. (2009). *Statistics for Psychology*. London. Pearson.

Φάση 3 (N = 507)	Περιγραφικά Στατιστικά		α	Συσχετίσεις							
	M.O.	T.A.		1	2	3	4	5	6	7	
1 Εμφύχωση	3.09	.69	.75	-							
2 Έλεγχος άγχους	2.75	.78	.64	.31**	-						
3 Αυτοπεποίθηση	2.78	.72	.76	.51**	.26**	-					
4 Συγκέντρωση	2.87	.78	.79	.46**	.34**	.55**	-				
5 Ανησυχία	1.44	.80	.85	-.29**	.01	-.30**	-.05	-			
6 Αποφυγή Προσπάθειας	1.01	.89	.84	-.34**	-.06	-.22**	-.09*	.63**	-		
7 Σωματική κούραση	1.53	.74	.74	-.15**	-.03	-.13**	.01	.58**	.60**	-	
8 Μη σχετικές σκέψεις	1.23	.95	.75	-.22**	.03	-.06	.04	.53**	.59**	.50**	-

* $p < .05$, ** $p < .01$

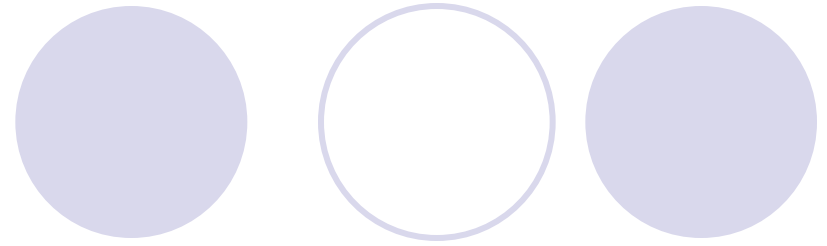
Συγχρονική και διαχωριστική εγκυρότητα

Zourbanos, Hatzigeorgiadis, Chroni, Theodorakis, & Papaioannou, (in press). Automatic Self-Talk Questionnaire for Sports (ASTQS): Development and preliminary validation of a measure identifying the structure of athletes' self-talk. *The Sport Psychologist*.

ΚΑΑ	ΑΑΠ			CSAI-2R	ASI			ΚΣΑΑ	
	ΓΑ	ΣΑ	ΔΣ	ΑΠ	ΕΝ	ΕΝΤ	ΑΝ	ΘΣ	ΑΣ
Εμφύχωση	-.12	.04	-.08	.32**	.26**	-.16*	-.26**	.56**	-.33**
Αυτοπεποίθηση	-.34**	.00	-.10	.53**	.44**	-.36**	-.35**	.63**	-.29*
Συγκέντρωση	-.05	.13*	-.09	.35**	.37**	-.23**	-.24**	.48**	-.13
Έλεγχος Άγχους	.06	.24**	-.01	.25**	.29**	-.12	-.22**	.24*	-.03
Ανησυχία	.50**	.25**	.14*	-.30**	-.54**	.56**	.47**	-.37**	.66**
Σωματική Κούραση	.38**	.28**	.30**	-.16	-.45**	.51**	.50**	-.12	.16
Αποφυγή Προσπάθειας	.40**	.29**	.32**	-.12	-.42**	.39**	.40**	-.25*	.46**
Μη σχετικές Σκέψεις	.14*	.16*	.23**	-.19	-.40**	.48**	.49**	-.03	.14
* $p < 0.05$; ** $p < 0.01$									

ΓΑ: γνωστικό άγχος, ΣΑ: σωματικό άγχος, ΔΣ: διατάραξη αυτοσυγκέντρωσης, ΑΠ: αυτοπεποίθηση, ΕΝ: ενεργητικότητα, ΕΝΤ: ένταση, ΑΝ: ανία, ΘΑ: θετική σκέψη, ΑΣ: αρνητική σκέψη

Συνοψίζοντας...



- Αλληλοεξάρτηση/συσχέτιση δυο μεταβλητών X και Y (συνεχείς μεταβλητές)
- Μεταβολές στις τιμές της μιας συνοδεύονται από μεταβολές στις τιμές της άλλης
- Θετική ή αρνητική κατεύθυνση
- Τιμές από -1.00 έως $+1.00$ (απόλυτη θετική συσχέτιση)
- Βαθμός: χαμηλή, μέτρια, υψηλή
- Ευθύγραμμες και καμπυλόγραμμες (μορφή)
- Στατιστικά σημαντική $p < .05$, $p < .001$

Έρθε η ώρα για SPSS

