

# Ανάλυση δεδομένων στο περιβάλλον του SPSS

Λαβίδας Κωνσταντίνος

Μαθηματικός

[lavidas@upatras.gr](mailto:lavidas@upatras.gr)

Διερεύνηση σχέσεων

Μεταξύ:

**Δύο ποσοτικών  
μεταβλητών**

# Παραδείγματα συσχέτισης δύο ποσοτικών μεταβλητών

- Ύψος των παιδιών μέχρι τα 10 έτη και βάρος τους
- Η επίδοση στα μαθηματικά και η επίδοση στη φυσική
- Η εργασιακή εμπειρία και το ύψος των αποδοχών
- Το πλήθος των διαφημίσεων ενός προϊόντος και το σύνολο των πωλήσεων του προϊόντος
- Χρόνου παρακολούθησης τηλεόρασης και εισοδήματος
- Ελεύθερος χρόνος και επίδοση στη στατιστική
- Χρόνος προετοιμασίας στις εξετάσεις και βαθμός επίδοσης των παιδιών

# Μέτρο προσδιορισμού της σχέσης δύο ποσοτικών μεταβλητών

- Συμβάλει στην ανάδειξη της **συμμεταβολής** (συσχέτισης) μεταξύ δύο ποσοτικών μεταβλητών
- Η συμμεταβολή μετριέται με την **συνδιακύμανση** (covariance) δύο ποσοτικών μεταβλητών.
  - Η συνδιακύμανση είναι μέτρο προσδιορισμού της γραμμικής τους σχέσης

# Συμμεταβολή

των μεταβλητών: Μέγεθος διαφήμισης και Ποσότητα προϊόντων που πουλήθηκε

Αντίθετη Κατεύθυνση

Υποκείμενα	1	2	3	4	5	Μ.Τ.	Τ.Α.
Μέγεθος Διαφήμισης	5	4	4	6	8	5,4	1,67
Απόκλιση (Τιμή - Μ.Τ.)	-0,4	-1,4	-1,4	0,6	2,6		
Ποσότητα Προϊόντων	8	9	12	13	15	11,4	2,88
Απόκλιση (Τιμή - Μ.Τ.)	-3,4	-2,4	0,6	1,6	3,6		

Ίδια Κατεύθυνση

Συνδιακύμανση στο δείγμα

$$Cov(x, y) = \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{n - 1} = \frac{14,2}{9} \cong 1,57$$

# Συνδιακύμανση

28/5/2024

Διακύμανση  
της  
**μεταβλητής 1**



Διακύμανση  
της  
μεταβλητής 2

Ο κοινός τόπος των  
δύο περιοχών  
αναπαριστά τη  
Συνδιακύμανση των  
δύο Μεταβλητών

## Συνδιακύμανση: Δίνει πληροφορίες για την ένταση (βαθμός) της σχέσης αλλά και τη κατεύθυνση της

- Υποδεικνύει κάποιο μοντέλο σχέσης, το οποίο φαίνεται όταν οι τιμές των δύο μεταβλητών μεταβάλλονται ταυτόχρονα:
  - Όταν οι τιμές δύο μεταβλητών κινούνται προς την ίδια κατεύθυνση, η συνδιακύμανση είναι (μεγάλη) και **θετική**.
  - Όταν οι τιμές δύο μεταβλητών κινούνται προς την αντίθετη κατεύθυνση, η συνδιακύμανση είναι (μεγάλη) και **αρνητική**.
  - Όταν δεν φαίνεται έντονα κάποιο από τα παραπάνω, η συνδιακύμανση είναι (μικρή), θεωρητικά κοντά στο **μηδέν**.
- Κατεύθυνση: Αύξουσα ή θετική, Φθίνουσα ή αρνητική

# Συνδιακύμανση: εξαρτάται από τη μονάδα μέτρησης

- Δεν μπορεί να χρησιμοποιηθεί για τη σύγκριση δύο διαφορετικών σχέσεων
  - Για τη σύγκριση δύο ή περισσότερων γραμμικών σχέσεων, πρέπει να χρησιμοποιηθεί ένας δείκτης ανεξάρτητος της μονάδας μέτρησης.
- Μέτρο προσδιορισμού της γραμμικής σχέσης δύο μεταβλητών είναι ο συντελεστής συσχέτισης **Pearson**.



# Συντελεστής συσχέτισης Pearson (correlation coefficient Pearson)

- Ο συντελεστής συσχέτισης Pearson ορίζεται ως ο λόγος της συνδιακύμανσης προς το γινόμενο των τυπικών αποκλίσεων των δύο μεταβλητών

- $$r = \frac{s_{xy}}{s_x s_y}$$

- Ανεξάρτητος από μονάδα μέτρησης

- Προϋποθέσεις χρήσης του συντελεστή Pearson:

- Κλίμακα τουλάχιστον ισοδιαστημική

- Κανονικότητα των κατανομών των μεταβλητών

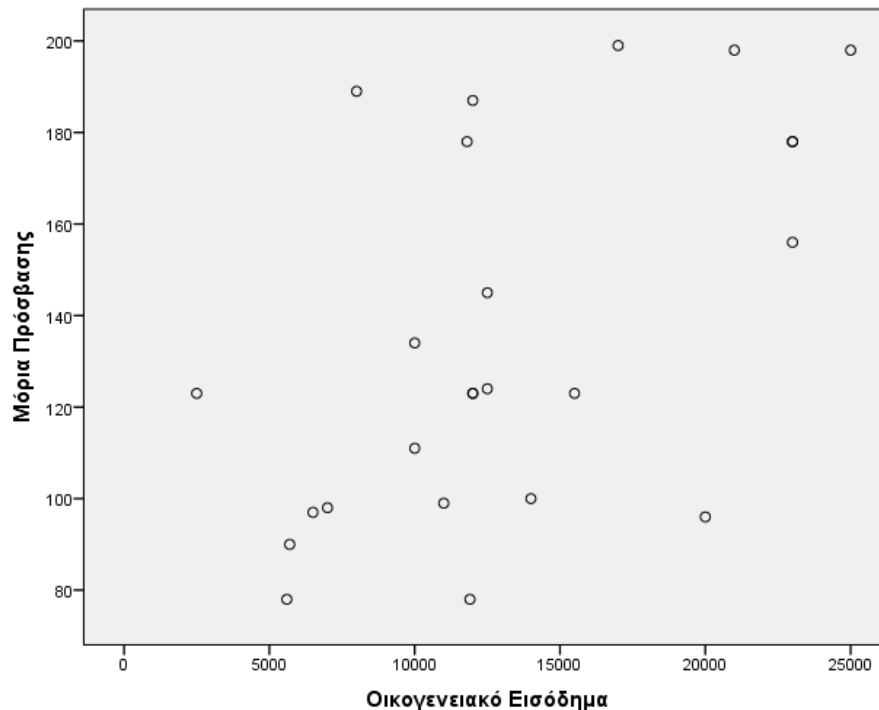
- Εξαρτημένα δείγματα

- Γραμμικότητα της σχέσης των μεταβλητών

- Η μεταβλητότητα των τιμών και στις δύο μεταβλητές να είναι περίπου ίδια (ομοσκεδαστικότητα)

# Διάγραμμα διασποράς ή σκεδασμού (scatter plot)

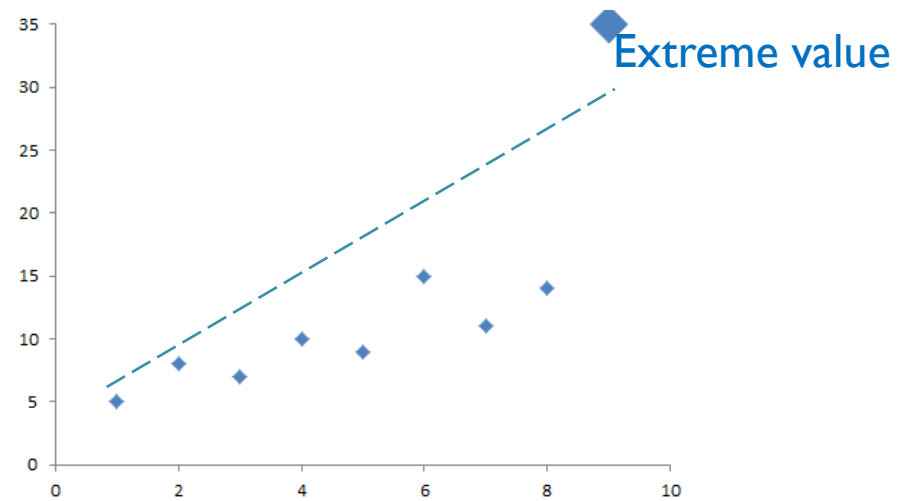
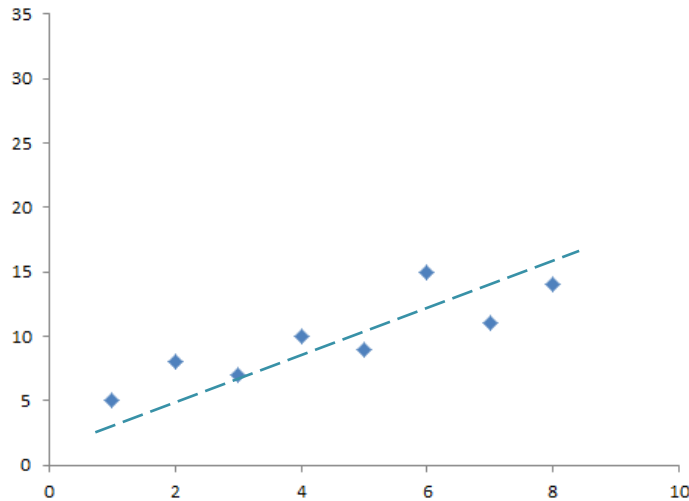
- Μεταξύ δύο ποσοτικών μεταβλητών
  - Αναπαριστά τα ζεύγη τιμών των δύο μεταβλητών με σημεία σε ορθοκανονικό σύστημα αξόνων



Πολύ  
χρήσιμο για  
να ελέγξω  
κατά πόσο  
είναι  
γραμμική η  
σχέση μεταξύ  
των δύο  
μεταβλητών

# Επίδραση των παράτυπων τιμών στην σχέση δύο ποσοτικών μεταβλητών

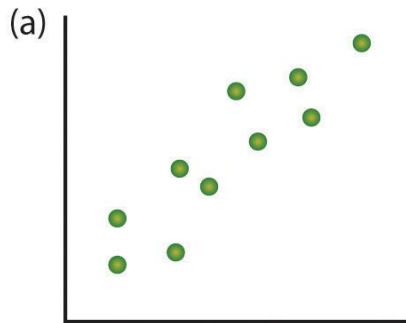
Προσοχή στις **ακραίες** και στις **παράτυπες τιμές**: μπορεί να επηρεάσουν το βαθμό της συσχέτισης



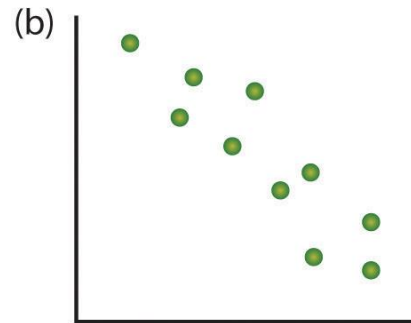
# Ερμηνεία του συντελεστή συσχέτισης

- Ο συντελεστής συσχέτισης:
  - Υποδεικνύει την **ένταση** και τη **φύση** της γραμμικής σχέσης μεταξύ των δύο μεταβλητών
    - **Όχι την κλίση της ευθείας**
- Κυμαίνεται πάντα μεταξύ -1 και +1.
- Αν  $r = \pm 1$  υπάρχει τέλεια γραμμική συσχέτιση.
- Αν  $-0,2 < r < 0,2$  **δεν** υπάρχει **γραμμική** συσχέτιση “αμελητέα σχέση”.
  - **Προσοχή**, μπορεί να υπάρχει άλλου είδους συσχέτιση μεταξύ των δύο μεταβλητών.
- Αν  $-0,3 < r < -0,2$  ή  $0,2 < r < 0,3$  : υπάρχει ασθενής γραμμική συσχέτιση.
- Αν  $-0,5 < r < -0,3$  ή  $0,3 < r < 0,5$  : υπάρχει μέτρια γραμμική συσχέτιση.
- Αν  $-0,8 < r < -0,5$  ή  $0,5 < r < 0,8$  υπάρχει ισχυρή γραμμική συσχέτιση.
- Αν  $-1 < r < -0,8$  ή  $0,8 < r < 1$  υπάρχει πολύ ισχυρή γραμμική συσχέτιση.
- Προσοχή για να **συγκρίνουμε** συντελεστές συσχέτισης συγκρίνουμε τις απόλυτες τιμές των συντελεστών (χωρίς τα πρόσημα)

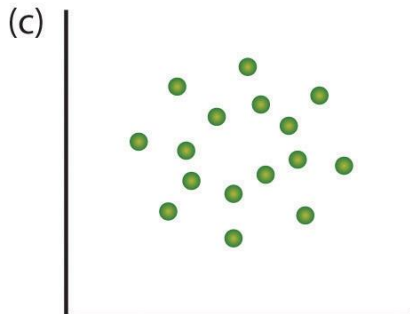
# Συντελεστές συσχέτισης και γραφική αναπαράσταση της σχέσης



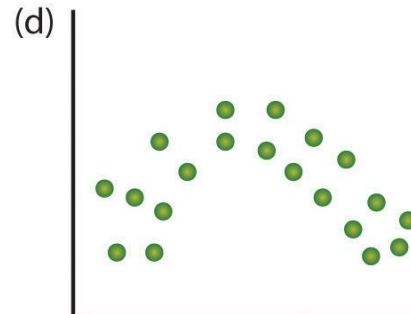
Positive linear  
 $r = +.82$



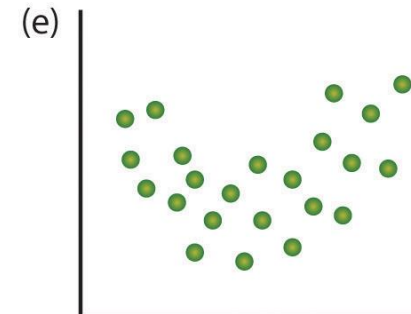
Negative linear  
 $r = -.70$



Independent  
 $r = 0.00$



Curvilinear  
 $r = 0.00$



Curvilinear  
 $r = 0.00$

## Συντελεστής συσχέτισης Spearman ( $\rho$ )

- Δίνει το μέγεθος της γραμμικής συσχέτισης ποιοτικών μεταβλητών **διάταξης (ordinal)**.
  - Προσδιορίζοντας τη σχέση της σειράς κατάταξης των τιμών των μεταβλητών
- Ενδείκνυται και στην περίπτωση που δεν ικανοποιούνται οι προϋποθέσεις χρήσης του Pearson.

# Μεταβλητές διάταξης ( $x, y$ ) και σειρές κατάταξης (rank) αυτών

Θέση αρχική	$x$	$y$	rank $x$	rank $y$
1	1	1	2	2
2	1	1	2	2
3	1	2	2	6
4	2	2	5	6
5	2	2	5	6
6	2	3	5	11
7	3	2	8	6
8	3	2	8	6
9	3	1	8	2
10	4	3	11	11
11	4	3	11	11
12	4	3	11	11
13	5	4	14,5	15
14	5	3	14,5	11
15	5	4	14,5	15
16	5	4	14,5	15

# Συντελεστής συσχέτισης Spearman ή Kendall W;

- Για να υπολογίσουμε τον συντελεστή spearman, πρέπει να βρούμε την κατάταξη των τιμών των μεταβλητών
- Αν στην κατάταξη έχουμε ισοβαθμίες δίνουμε ως θέση κατάταξης τη μέση τιμή της κατάταξή τους.
- Όταν υπάρχουν πολλές ισοβαθμίες ο συντελεστής rho είναι λιγότερο αξιόπιστος. Σε αυτή την περίπτωση ενδείκνυται ο δείκτης Kendall W.

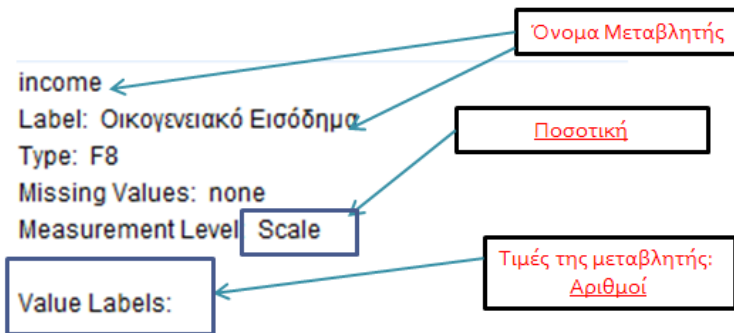


# Υποθέσεις ελέγχου για το συντελεστή συσχέτισης ( $\rho$ ) δύο ποσοτικών μεταβλητών

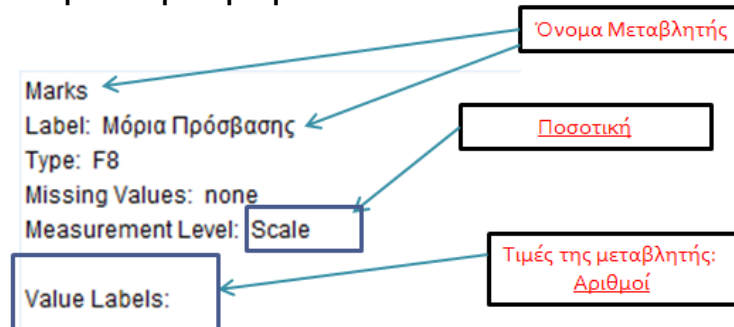
- Αμφίδρομος ή αμφίπλευρος έλεγχος (Two-tailed)
  - $H_0: \rho = 0$  (δεν υπάρχει συσχέτιση μεταξύ των δύο μεταβλητών)
  - $H_1: \rho \neq 0$  (υπάρχει μη μηδενική συσχέτιση μεταξύ των δύο μεταβλητών)
- Μονόπλευρος ή έλεγχος μιας ουράς (One-tailed)
  - $H_0: \rho = 0$  (δεν υπάρχει συσχέτιση μεταξύ των δύο μεταβλητών)
  - $H_1: \rho > 0$  (υπάρχει θετική συσχέτιση μεταξύ των δύο μεταβλητών)
  - ή
  - $H_1: \rho < 0$  (υπάρχει αρνητική συσχέτιση μεταξύ των δύο μεταβλητών)

# Παράδειγμα

- Μας ενδιαφέρει να διερευνήσουμε τη σχέση μεταξύ του οικογενειακού εισοδήματος και των μορίων πρόσβασης των μαθητών στην τριτοβάθμια εκπαίδευση
- 1<sup>η</sup> μεταβλητή



- 2<sup>η</sup> μεταβλητή



# Παράδειγμα 10.1

- Πρέπει να παρουσιαστεί ο επόμενος πίνακας

Η περιγραφή της «συνάφειας» των δύο μεταβλητών

Η περιγραφή της «συνάφειας» των δύο μεταβλητών βασίζεται κυρίως στην παρουσίαση της τιμής π.χ.  $r=0,792$

		Marks Μόρια Πρόσβασης	income Οικογενειακό Εισόδημα
Marks Μόρια Πρόσβασης	Pearson Correlation	1	,792**
	Sig. (2-tailed)		,000
	N	25	25
income Οικογενειακό Εισόδημα	Pearson Correlation	,792**	1
	Sig. (2-tailed)	,000	
	N	25	25

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

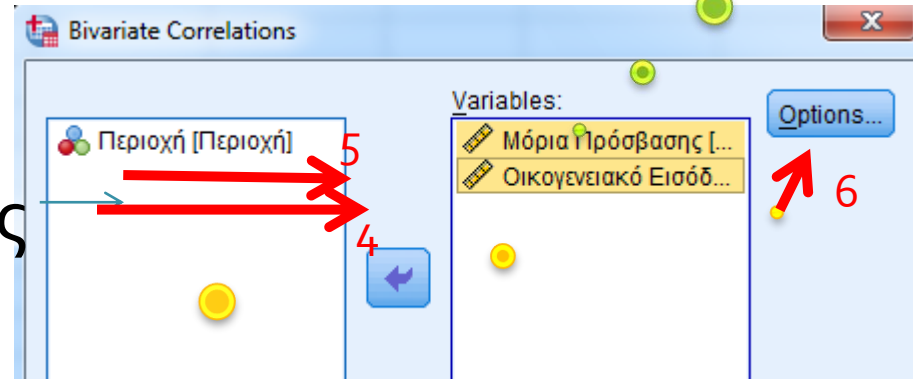
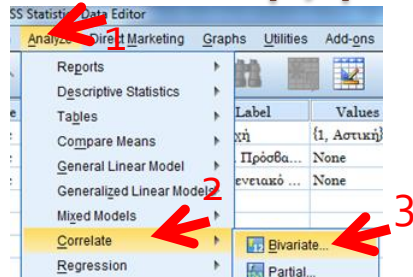
Marks	income
78	5600
78	11900
90	5700
96	200
97	6500
98	7000
99	11000
100	9000
111	12000
123	10000
123	17000
123	12000
124	15000
134	10000
145	17000

- Έχουμε την βάση δεδομένων με τις τιμές των μεταβλητών

Αρχείο: «correlation example.sav»

# Σύνθετη παρουσίαση ποσοτικής και ποσοτικής μεταβλητής, με το SPSS

Ποσοτική  
Μεταβλητή



- Σέρνω μεταβλητές

Ποσοτικές  
Μεταβλητές  
(τουλάχιστον 2)

Στην επιλογή: options, μπορώ να διαλέξω να εμφανιστούν μέτρα θέσης, διασποράς και η συνδιακύμανση των δύο μεταβλητών.

# Στατιστική συμπερασματολογία

- Στην ίδια διαδικασία που εκτελέσαμε για να δούμε την όποια γραμμική σχέση
- Στον πίνακα κάτω από τον συντελεστή Pearson ή Spearman, ελέγχουμε το **sign.**

**Correlations**

		Marks Μόρια Πρόσβασης	income Οικογενειακό Εισόδημα
Marks Μόρια Πρόσβασης	Pearson Correlation	1	,792**
	Sig. (2-tailed)		,000
	N	25	25
income Οικογενειακό Εισόδημα	Pearson Correlation	,792**	1
	Sig. (2-tailed)	,000	
	N	25	25

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

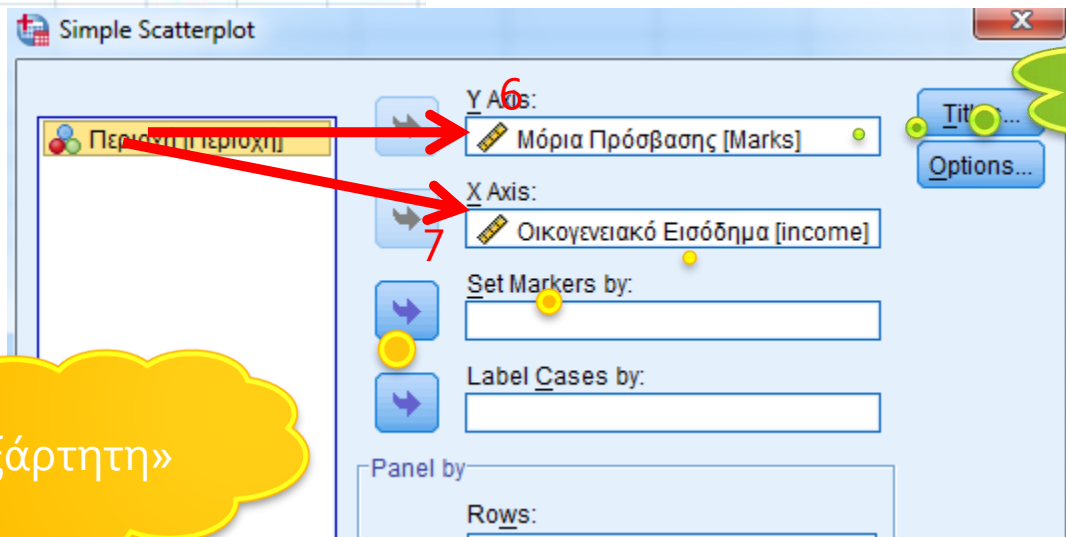
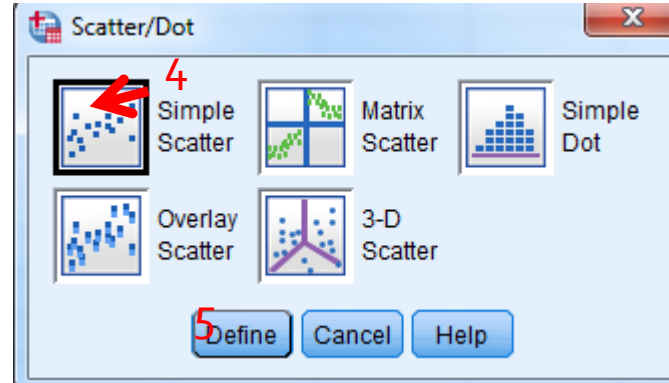
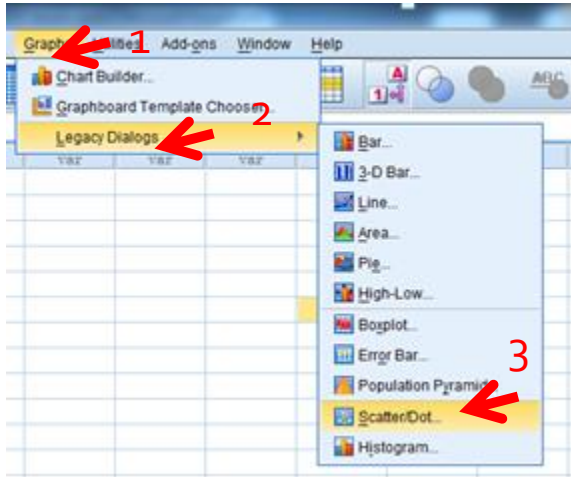
Ωστόσο μην ξεχνάμε ότι ο  
συντελεστής αυτός  
ερμηνεύει τη γραμμική  
σχέση μεταξύ δύο  
ποσοτικών μεταβλητών

## Γράφημα για να υποστηρίξουμε τη γραμμικότητα της σχέσης μεταξύ δύο ποσοτικών μεταβλητών

- Κατασκευάζουμε ένα διάγραμμα σκεδασμού (**scatterplot**), το οποίο σχεδιάζει την μία μεταβλητή σε σχέση με την άλλη.
- Η μία από τις δύο μεταβλητές «ανεξάρτητη» συμβολίζεται με  $x$  και συνήθως τοποθετείται στον οριζόντιο άξονα, ενώ η άλλη μεταβλητή «εξαρτημένη» συμβολίζεται με  $y$  και συνήθως τοποθετείται στον κάθετο άξονα.
  - Οι μεταβλητές ως ανεξάρτητες και εξαρτημένες συνήθως ορίζονται τυχαία

# Γράφημα σκεδασμού-Scatterplot

- Graphs – Legacy dialogs- Scatter/dot



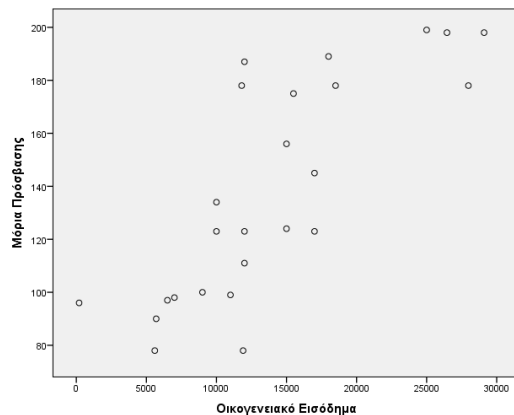
«εξαρτημένη»

«Ανεξάρτητη»



# Συμπεράσματα: παράδειγμα 10.1

- Ενδεικτικά θα μπορούσαμε να παρουσιάσουμε τα αποτελέσματα του πίνακα και του γραφημάτος ως εξής:
  - Στο γράφημα που παρουσιάζουμε φαίνεται η σχέση μεταξύ .... Και ... να είναι γραμμική.
  - Ο δείκτης pearson ( $r=...$ ) μας υποδεικνύει ότι οι μεταβλητές αυτές έχουν μέτρια θετική σχέση. Επομένως η επίδοση των μαθητών του δείγματός μας στις εξετάσεις εισαγωγής στην τριτοβάθμια εκπαίδευση, φαίνεται να σχετίζεται μέτρια και θετικά με το οικογενειακό εισόδημα των μαθητών . Συγκεκριμένα όσο αυξάνεται το οικογενειακό εισόδημα τα μόρια πρόσβασης στην τριτοβάθμια εκπαίδευσης είναι υψηλότερα. Η σχέση αυτή είναι και στατιστικά σημαντική ( $t(df=πλήθος\ υποκειμένων -2)=..., p=...$ )



- Προσοχή: Καλό θα είναι να εντάξουμε το γράφημα, ως συνοδευτικό τεκμήριο της γραμμικότητας της σχέσης, στο τελικό κείμενο.

# ? Ερώτημα

- Πως διαμορφώνεται το προηγούμενο συμπέρασμα, ανάλογα με το επίπεδο της παρεχόμενης εξωσχολικής υποστήριξης, που δήλωσαν ότι είχαν οι μαθητές;
  - Για να ελέγξετε τη σχέση αυτή στα διάφορα επίπεδα της μεταβλητής «παρεχόμενη εξωσχολική υποστήριξη»
    - Πραγματοποιήστε τον ίδιο έλεγχο, αφού πρώτα ζητήσετε ο έλεγχος να γίνει ξεχωριστά σε κάθε επίπεδο της μεταβλητής «παρεχόμενη εξωσχολική υποστήριξη». Επιλογές: `data-split files- compare groups` και στο πεδίο `grouped based on` τοποθετούμε τη συγκεκριμένη μεταβλητή

# Παράδειγμα «πολλών» συσχετίσεων.

- Αρχείο **«correlation first order.sav»**
- Ερευνητής ζητά να βαθμολογήσουν οι συμμετέχοντες, αυριανοί εκπαιδευτικοί, το περιβάλλον στο οποίο περιλαμβάνονται οι διδακτικές δραστηριότητες με αξιοποίηση των ΤΠΕ, που έχει κατασκευάσει για τη διδασκαλία των μαθηματικών. Τα χαρακτηριστικά των δραστηριοτήτων που θέλει οι συμμετέχοντες να αξιολογήσουν είναι:
  - Η εμφάνιση της διεπαφής
  - Οργάνωση και δόμηση του περιεχομένου
  - Βαθμός επίτευξης των στόχων

# Ανάλυση συσχετίσεων

- Θέλουμε να δούμε πως συνδέονται (αν) οι παραπάνω αξιολογήσεις.
  - Επομένως, να ελέγξετε το σύνολο των συσχετίσεων για όλα τα ζεύγη των μεταβλητών.
  - Στην περίπτωση ελέγχου όλων των δυνατών συσχετίσεων μεταξύ των  $(n)$  μεταβλητών αυξάνεται η πιθανότητα εμφάνισης λάθους τύπου I:
    - Να απορρίψουμε τη μηδενική υπόθεση ενώ αυτή είναι αληθινή.
    - Λύση στο πρόβλημα δίνει η αλλαγή του επιπέδου στατιστικής σημαντικότητας (Bonferroni) : Από 5% σε  $0,05/[(n-1)n/2]$ 
      - $n$  μεταβλητές, άρα το πλήθος των δυνατών συγκρίσεων  $= (n-1)n/2$ .

# Εκτελέστε τον έλεγχο όλων των δυνατών συσχετίσεων

- Αλλάξτε το επίπεδο σημαντικότητας και δείτε κατά πόσο οι συσχετίσεις είναι στατιστικά σημαντικές
- Πριν γράψετε τα συμπεράσματά σας ελέγξτε κατά πόσο τα διαγράμματα σκεδασμού υποδεικνύουν γραμμικές σχέσεις μεταξύ των μεταβλητών

# Προσοχή στις επίπλαστες σχέσεις!

- Ο ερευνητής κατά την ανάλυση υποψιάζεται ότι η διάταξη παρουσίασης των ερωτήσεων, στους συμμετέχοντες, διαμόρφωσε επίπλαστες σχέσεις.
  - Για παράδειγμα: «Η εμφάνιση της οθόνης» μάλλον λειτούργησε ως παρακινητής παρόμοιων βαθμολογιών και στις υπόλοιπες ερωτήσεις.

# Συσχέτιση δεν σημαίνει υποχρεωτικά και αιτιότητα

- Όταν σε μια μη πειραματική έρευνα: δύο μεταβλητές X και Y βρίσκονται συσχετισμένες αυτό σημαίνει μόνο ότι οι μεταβλητές αυτές συνδέονται με κάποια σχέση. Οι δύο μεταβλητές μπορεί βεβαίως να συνδέονται με σχέση αιτιότητας, μπορεί όμως όχι. Για παράδειγμα, μπορεί και οι δύο να επηρεάζονται από μια τρίτη μεταβλητή.
- Παράδειγμα **μη** σχέσης «αιτίου – αιτιατού»
  - Το ύψος των μαθητών ενός σχολείου, ηλικίας 6 έως 13 ετών, έχει ισχυρή θετική γραμμική συσχέτιση με την αντιληπτική ικανότητα των μαθητών. Προφανώς η αντιληπτική ικανότητα των μαθητών δεν επηρεάζεται από το ύψος τους. Απλώς η πνευματική ανάπτυξη επηρεάζεται από την **φυσική ανάπτυξη** και η δεύτερη επηρεάζει το ύψος.

# Μερική συσχέτιση (partial correlation)

- Σε κάποιες περιπτώσεις η συσχέτιση μεταξύ δύο μεταβλητών είναι πιθανό να επηρεάζεται από τη σχέση που καθεμία από αυτές τις δύο έχει με μια τρίτη μεταβλητή.
- Ο δείκτης μερικής συσχέτισης μπορεί να ελέγξει ή αν θέλετε να **αποκλείσει** τον όποιο ρόλο παίζει η τρίτη μεταβλητή στη σχέση των δύο άλλων.
- Κατά τον έλεγχο μιας σχέσης δύο μεταβλητών:
  - αν δεν ελέγγω την επίδραση μιας τρίτης, τότε αναφέρομαι σε **μηδενικής τάξης συσχέτιση (zero – order correlation)**
  - αν ελέγγω την επίδραση μιας τρίτης, τότε αναφέρομαι σε πρώτης τάξης μερική συσχέτιση (**first – order partial correlation**)
  - αν ελέγγω την επίδραση δύο μεταβλητών, τότε αναφέρομαι σε δεύτερης τάξης μερική συσχέτιση (**second– order partial correlation**)
- Ας εκτελέσουμε την αντίστοιχη μερική συσχέτιση σταθεροποιώντας (**ελέγχοντας**) την επίδραση της μεταβλητής: «εμφάνιση της οθόνης»
- Τι παρατηρείτε;



# Αποτελέσματα μηδενικής τάξης συσχέτισης και πρώτης τάξης μερικής συσχέτισης

Correlations

Control Variables			Οργάνωση_Δόμηση_Περιεχ	Επίτευξη_Στόχων	Εμφάνιση_Οθόνης
-none- <sup>a</sup>	Οργάνωση_Δόμηση_Περιεχ	Correlation	1,000	,953	,978
		Significance (2-tailed)	.	,000	,000
		df	0	13	13
	Επίτευξη_Στόχων	Correlation	,953	1,000	,950
		Significance (2-tailed)	,000	.	,000
		df	13	0	13
	Εμφάνιση_Οθόνης	Correlation	,978	,950	1,000
		Significance (2-tailed)	,000	,000	.
		df	13	13	0
Εμφάνιση_Οθόνης	Οργάνωση_Δόμηση_Περιεχ	Correlation	1,000	,374	
		Significance (2-tailed)	.	,188	
		df	0	12	
	Επίτευξη_Στόχων	Correlation	,374	1,000	
		Significance (2-tailed)	,188	.	
		df	12	0	

a. Cells contain zero-order (Pearson) correlations.

Τι παρατηρείτε;

# Συμπέρασμα

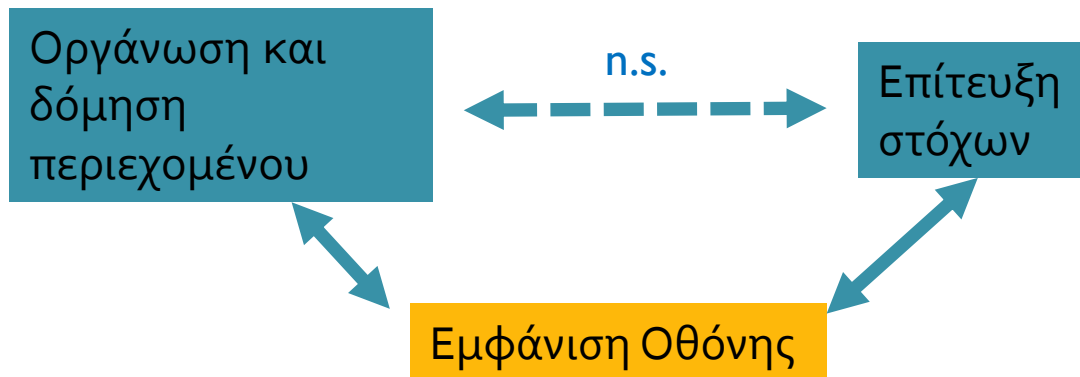
«Παγώνοντας» ή αδρανοποιώντας τη μεταβλητή Εμφάνιση Οθόνης

Η σχέση των μεταβλητών

α) Οργάνωση και δόμηση περιεχομένου και

β) Επίτευξη στόχων

«Εξαφανίστηκε»



**n.s.** =not significant  
Δεν είναι στατιστικά σημαντικό