

Συντελεστής Συσχέτισης

Βασικές Έννοιες

- Μέχρι τώρα ασχοληθήκαμε με την ανάλυση μιας μόνο τυχαίας μεταβλητής, όπως είναι, π.χ., η ηλικία των εργαζομένων μιας επιχείρησης, οι βαθμολογίες των φοιτητών μιας σχολής κ.ά.
- Θα ασχοληθούμε τώρα με τις σχέσεις μεταξύ δύο ή περισσότερων τυχαίων μεταβλητών, όπως π.χ. τιμή και ζητούμενη ποσότητα ενός προϊόντος, κατανάλωση και εισόδημα κ.ά.

Η έννοια της Αλληλεξάρτησης

- Υπάρχουν πολλές έρευνες στις οποίες μας ενδιαφέρει να μελετήσουμε αν υπάρχει **αλληλεξάρτηση** μεταξύ δύο μεταβλητών.
- Δηλαδή, μας ενδιαφέρει να διαπιστώσουμε κατά πόσο οι τιμές που παίρνει μια μεταβλητή **επηρεάζονται** από τις τιμές που παίρνει η άλλη μεταβλητή.

Η έννοια της Αλληλεξάρτησης (2)

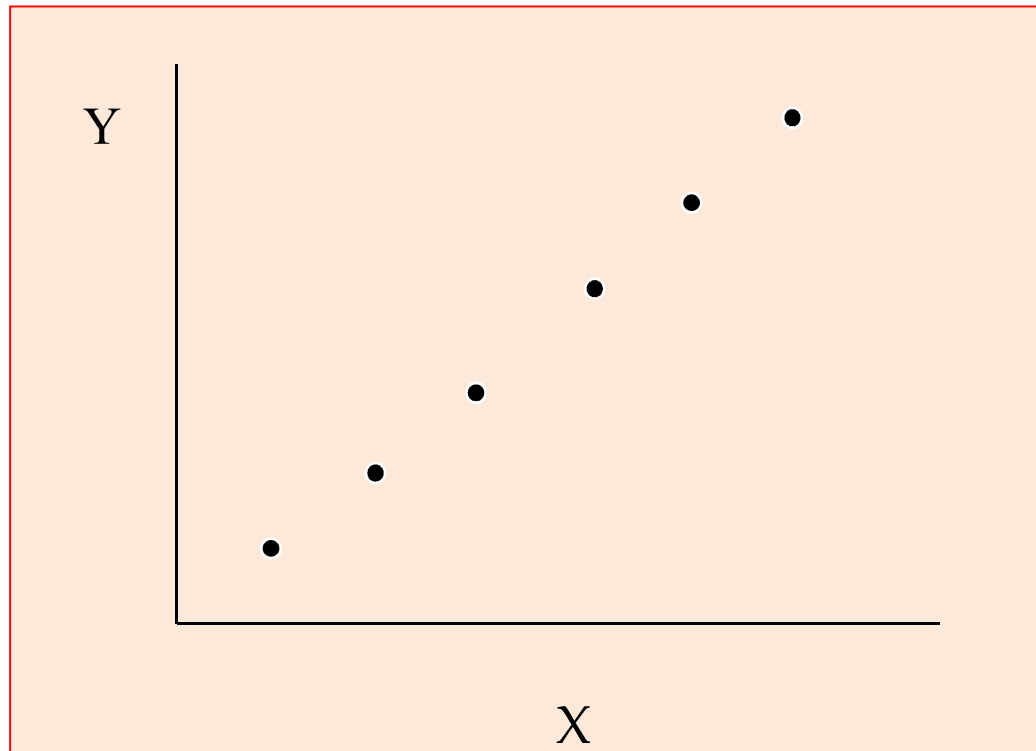
- Η πιο απλή περίπτωση αφορά στην σχέση μεταξύ δύο μεταβλητών. Η γενική μαθηματική έκφραση μιας τέτοιας σχέσης είναι:

$$Y = f(X)$$

- Η μεταβλητή Y της οποίας οι τιμές καθορίζονται από τις τιμές της X ονομάζεται εξαρτημένη μεταβλητή ενώ η X ονομάζεται ανεξάρτητη μεταβλητή.
- Προκειμένου να σχηματίσουμε μια πρώτη εικόνα γύρω από την σχέση μεταξύ δύο μεταβλητών κατασκευάζουμε ένα **διάγραμμα διασποράς**, το οποίο απεικονίζει όλα τα ζεύγη των παρατηρήσεων των μεταβλητών X και Y .

Τέλεια Σχέση

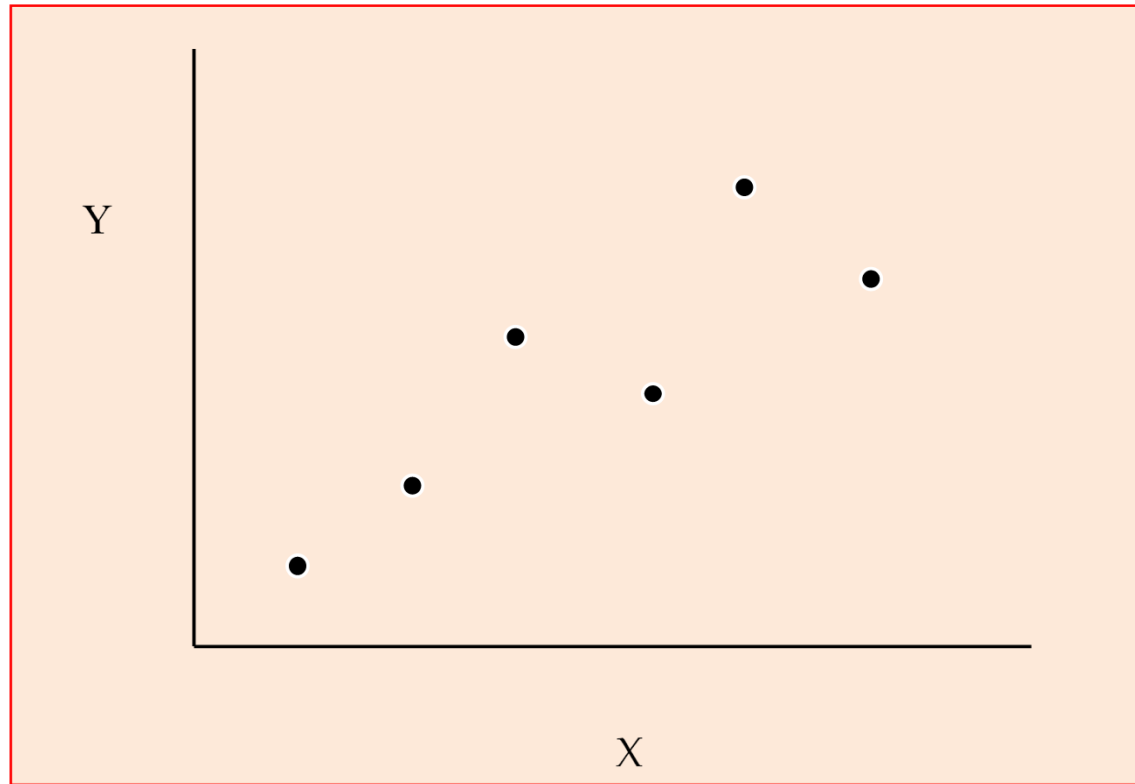
Όταν η σχέση μεταξύ των μεταβλητών Y και X είναι τέλεια τότε μιλάμε για **μαθηματική σχέση**:



Στις οικονομικές και τις διοικητικές επιστήμες όμως οι μαθηματικές σχέσεις είναι πολύ σπάνιες ή ανύπαρκτες και δεν θα μας απασχολήσουν.

Ατελής Σχέση

Το πιο συνηθισμένο είδος σχέσης που συναντάμε είναι οι ατελείς ή **στατιστικές σχέσεις**:



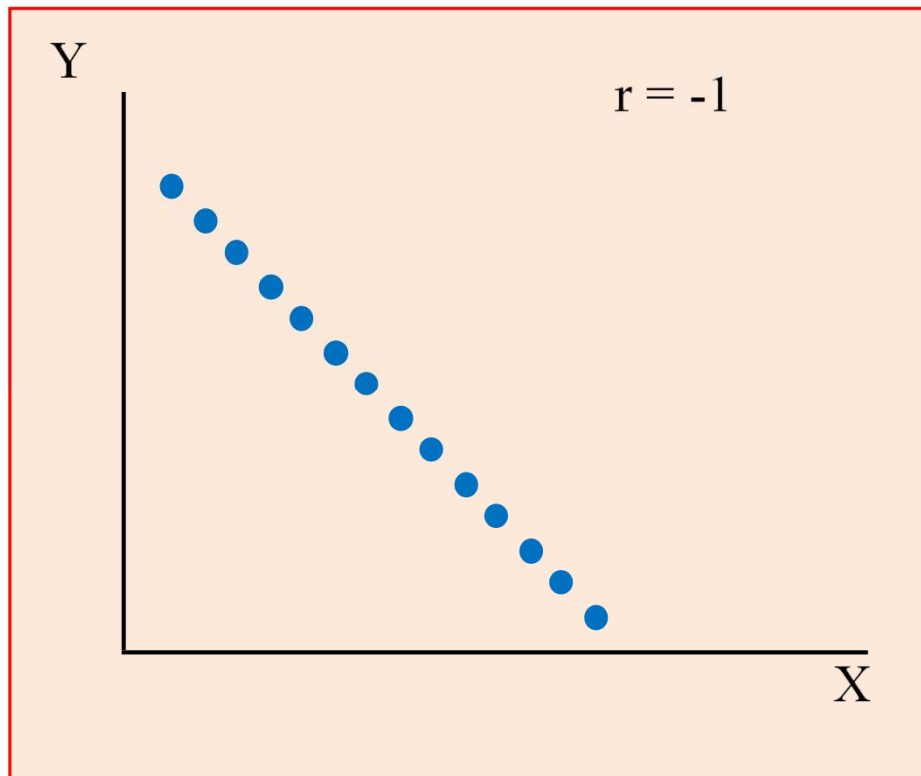
Συσχέτιση

- Η **Συσχέτιση** μεταξύ δύο μεταβλητών δείχνει την ένταση και την κατεύθυνση της **γραμμικής σχέσης** που υπάρχει μεταξύ των μεταβλητών αυτών.
- Η μέτρηση του βαθμού της γραμμικής σχέσης πραγματοποιείται με την χρήση ενός δείκτη που ονομάζεται **συντελεστής συσχέτισης**.
- Ο Συντελεστής Συσχέτισης εκφράζει την κατεύθυνση (δηλαδή το αν η σχέση είναι θετική ή αρνητική) και την ένταση της συμμεταβολής δύο μεταβλητών, χρησιμοποιείται ως μέτρο του βαθμού της γραμμικής σχέσης τους και συμβολίζεται με το γράμμα r .
- Ο Συντελεστής Συσχέτισης μπορεί να λάβει οποιαδήποτε τιμή μεταξύ -1 και 1 .

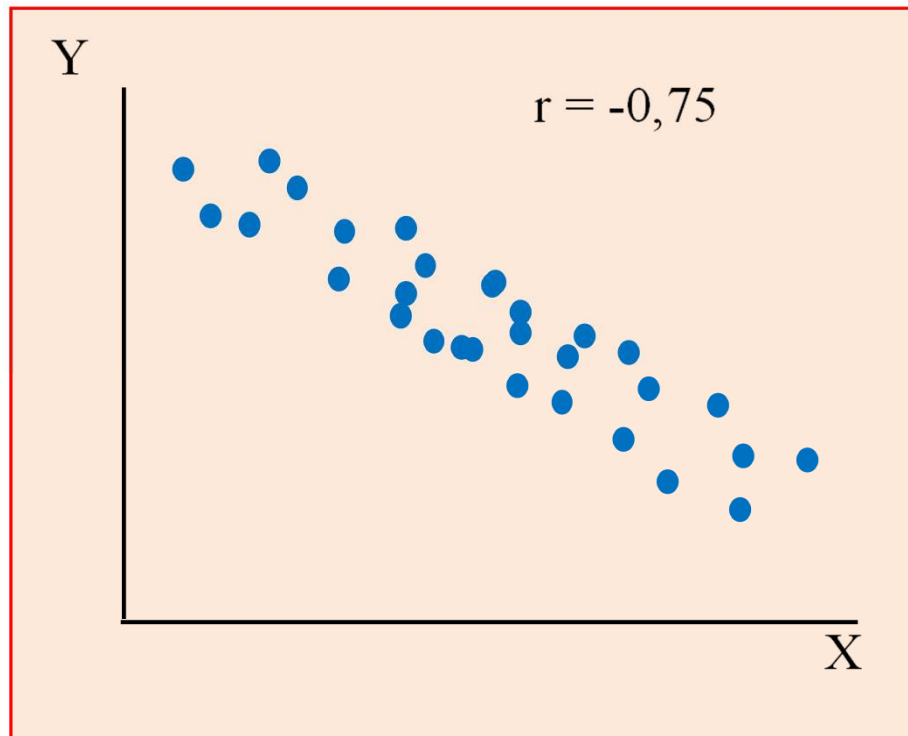
Συσχέτιση (συν.)

Ανάλογα με την τιμή που λαμβάνει ο συντελεστής συσχέτισης, η σχέση μεταξύ των δύο μεταβλητών χαρακτηρίζεται ως εξής:

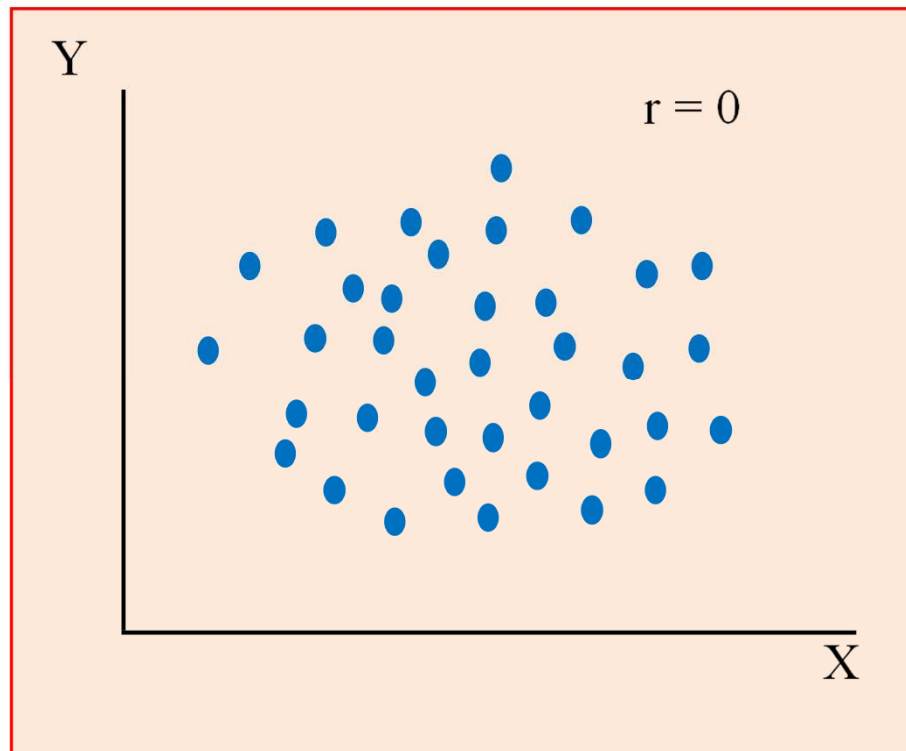
α) $r = -1$. Απόλυτα αρνητική γραμμική συσχέτιση.



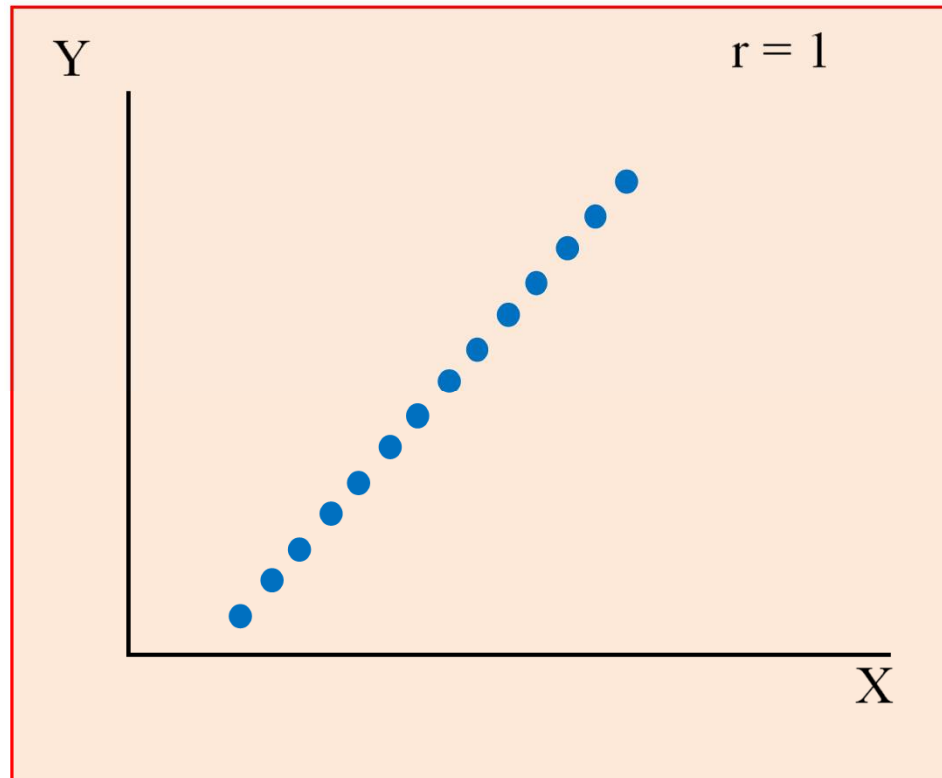
β) $-1 < r < 0$. Αρνητική γραμμική συσχέτιση.



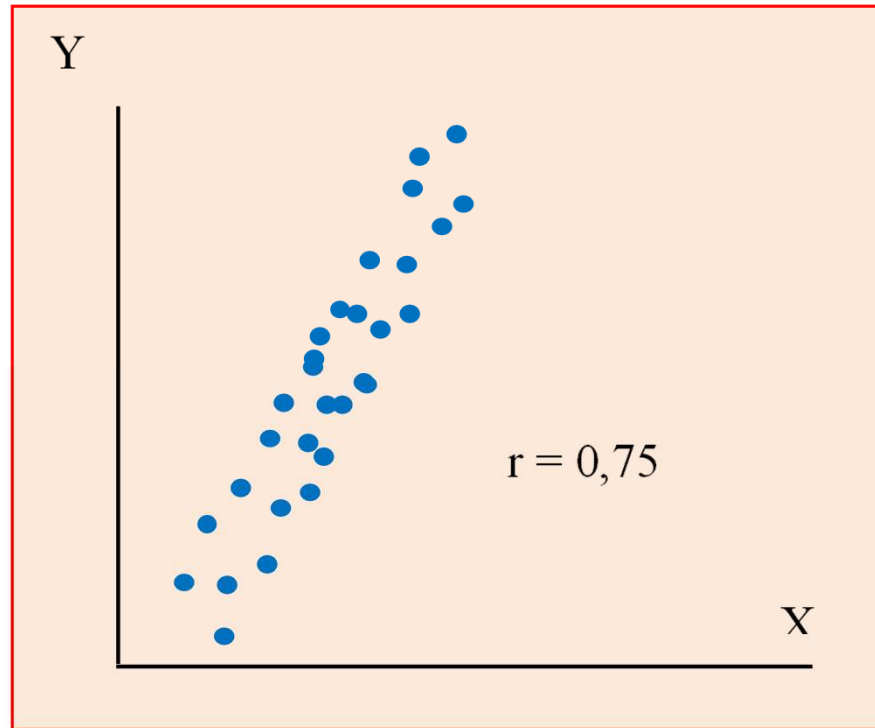
γ) $r = 0$. Δεν υπάρχει γραμμική συσχέτιση.



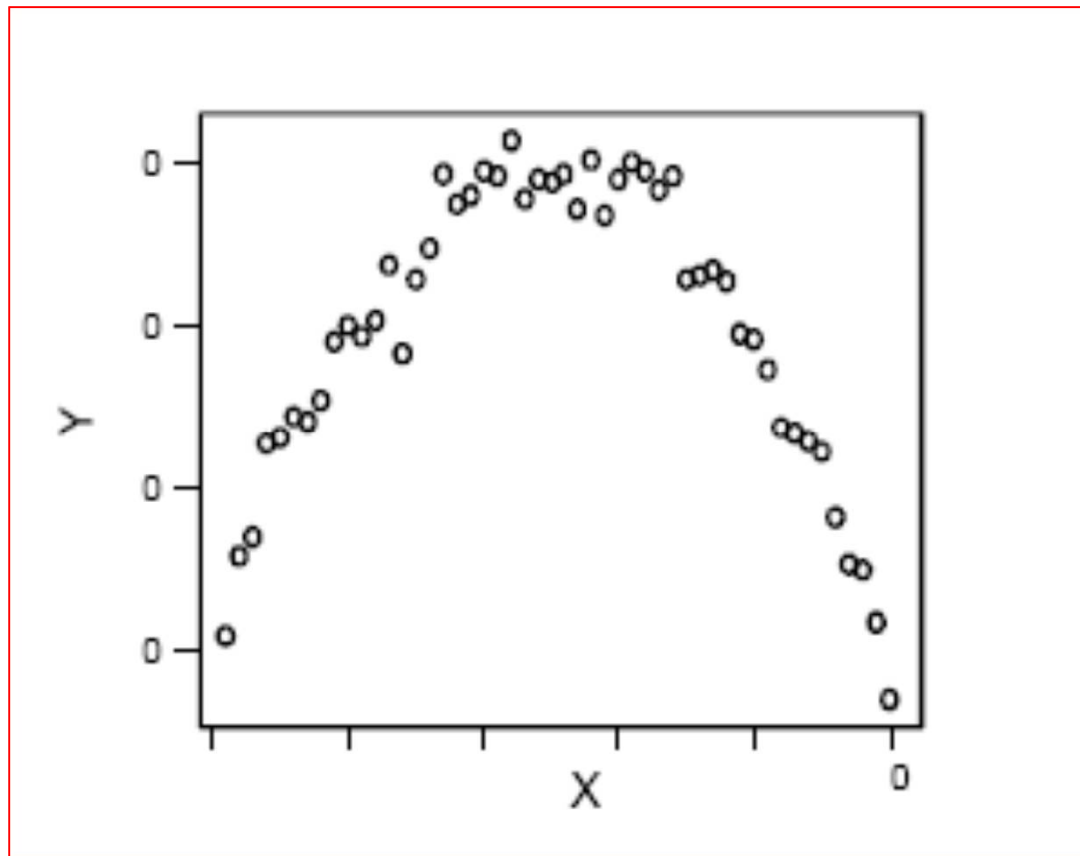
δ) $r = 1$. Απόλυτα θετική γραμμική συσχέτιση.



ε) $0 < r < 1$. Θετική γραμμική συσχέτιση.



$r =$; Μη γραμμική συσχέτιση



Παράδειγμα

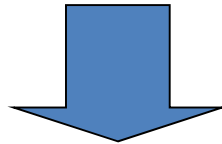
- Θέλουμε να μελετήσουμε τη σχέση που υπάρχει ανάμεσα στην εκπαίδευση και την ικανότητα του ατόμου να επεξεργάζεται αριθμητικές έννοιες

ΖΗΤΑΜΕ ΝΑ ΔΙΑΠΙΣΤΩΣΟΥΜΕ:

- Κατά πόσο οι μεταβολές στο χρόνο **ΕΚΠΑΙΔΕΥΣΗΣ** επιφέρουν αλλαγές στην επίδραση της **ΑΡΙΘΜΗΤΙΚΗΣ ΙΚΑΝΟΤΗΤΑΣ**

**Ποιο στατιστικό κριτήριο θα πρέπει να
χρησιμοποιήσουμε;**

Το στατιστικό κριτήριο που χρησιμοποιούμε για να
διαπιστώσουμε αν υπάρχει **γραμμική σχέση** μεταξύ
δύο μεταβλητών, ονομάζεται



**ΣΥΝΤΕΛΕΣΤΗΣ ΣΥΣΧΕΤΙΣΗΣ (r)
(CORRELATION COEFFICIENT)**

Πληροφορίες που παίρνουμε από τον συντελεστή συσχέτισης

- **Αν υπάρχει αλληλεξάρτηση (συσχέτιση)**
- **Το είδος της συσχέτισης**
- **Το βαθμό της συσχέτισης**

Πώς παρουσιάζουμε την συσχέτιση;

Η συσχέτιση μεταξύ δύο μεταβλητών μπορεί να παρουσιαστεί με δύο τρόπους:

- **αριθμητικά**
- **γραφικά**

Η παρουσίαση της συσχέτισης με αριθμητικό τρόπο

-1

Αρνητική
Συσχέτιση

0

Καθόλου
Συσχέτιση

+1

Θετική
Συσχέτιση

Είδος Συσχέτισης

- **Θετική Συσχέτιση (positive correlation)**

Όταν αυξάνεται η μια μεταβλητή, αυξάνεται και η άλλη

- **Αρνητική Συσχέτιση (negative correlation)**

Όταν αυξάνεται η μία μεταβλητή, μειώνεται η άλλη

Η παρουσίαση με γραφικό τρόπο

- Για να παρουσιάσουμε γραφικά την συσχέτιση χρησιμοποιούμε το **ΔΙΑΓΡΑΜΜΑ ΔΙΑΣΠΟΡΑΣ**

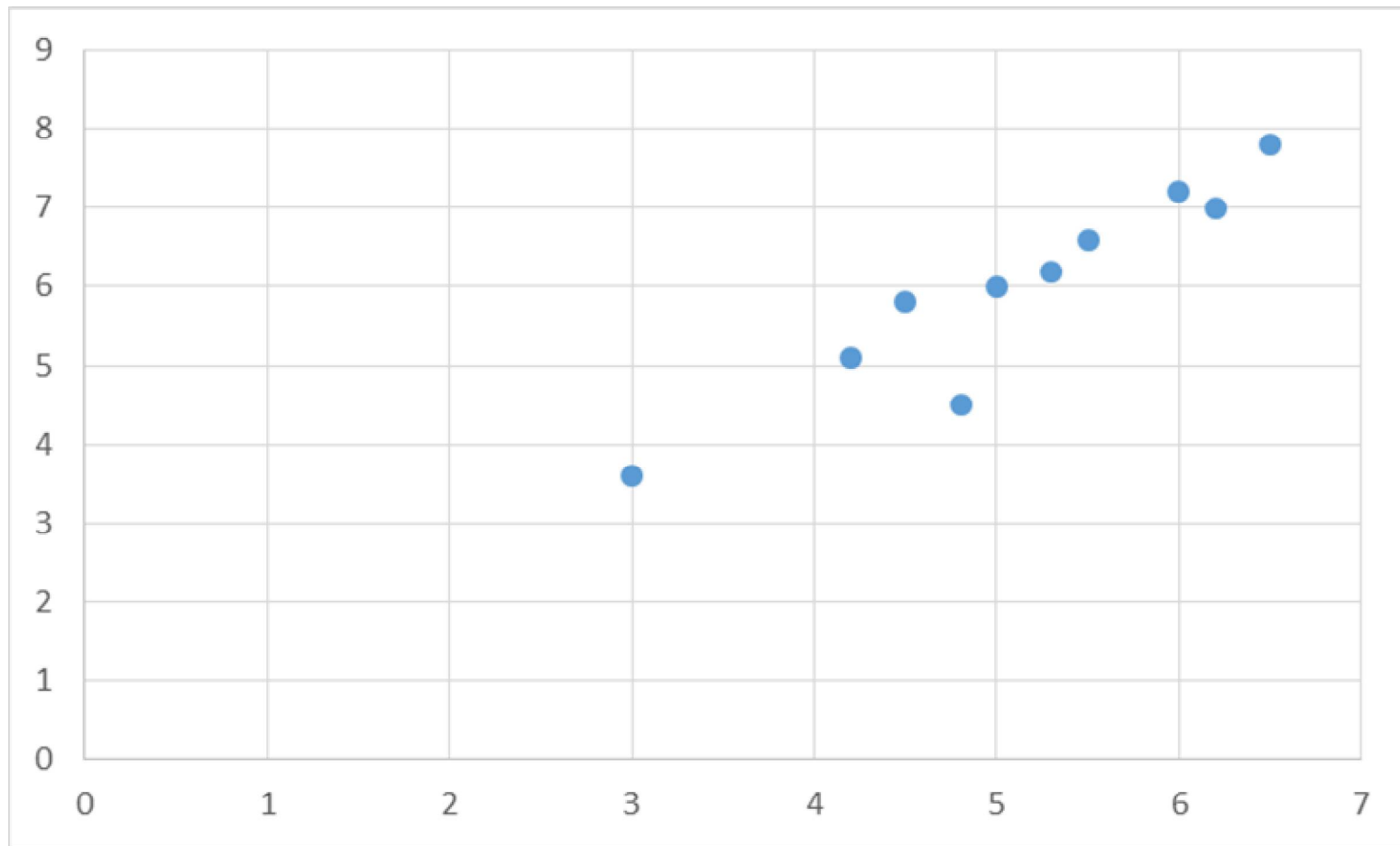
- **Παράδειγμα:**

Θέλουμε να δούμε αν υπάρχει συσχέτιση ανάμεσα στις επιδόσεις των φοιτητών σε 4 μαθήματα: **Μικροοικονομική, Μακροοικονομική, Ποσοτικές Μεθόδους και Δίκαιο**

Οι επιδόσεις των φοιτητών στα τέσσερα μαθήματα

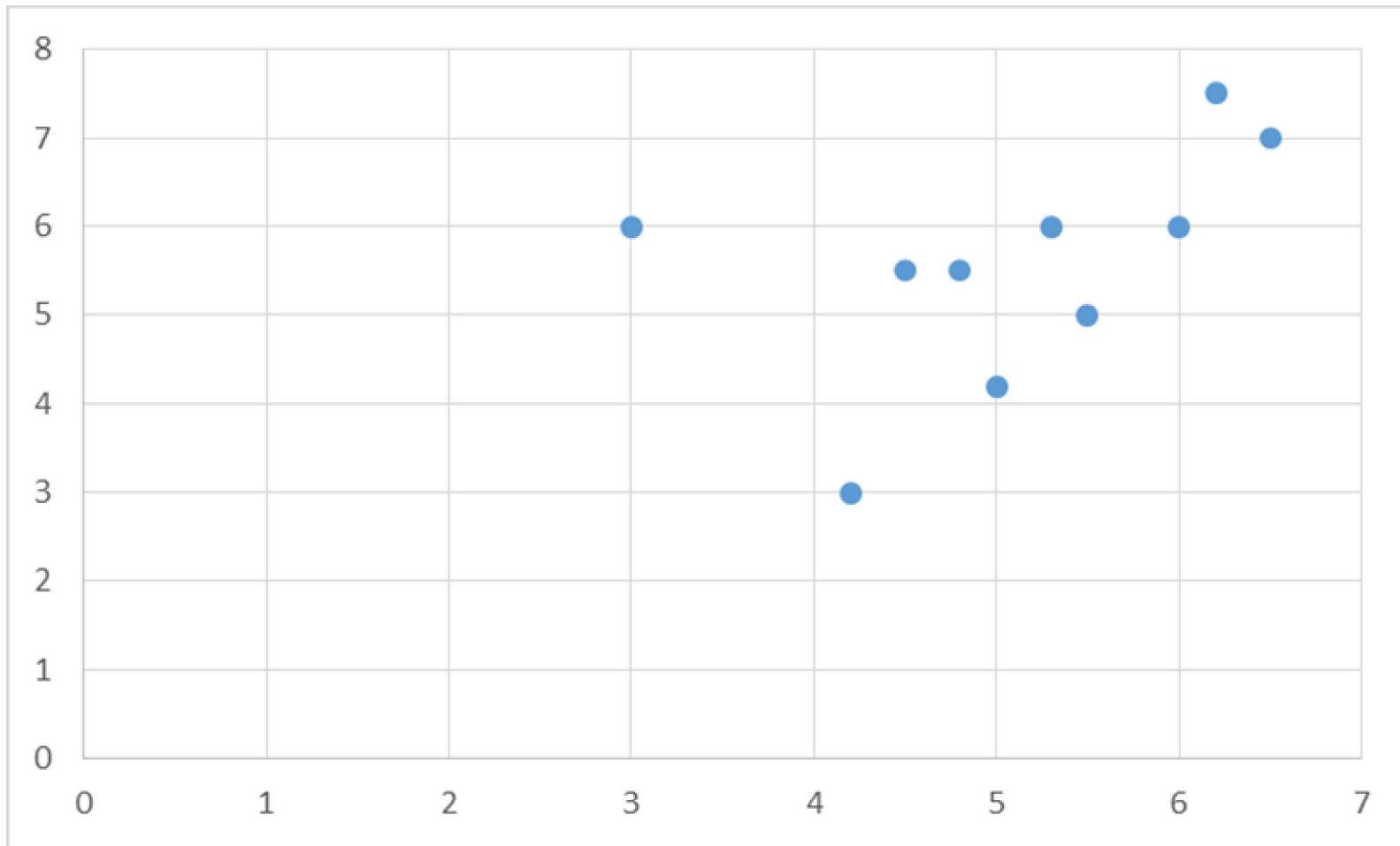
Φοιτητής	Μικροοικονομική	Μακροοικονομική	Ποσοτικές Μέθοδοι	Δίκαιο
A	6,5	7,8	7	6
B	6	7,2	6	5
Γ	6,2	7	7,5	8
Δ	5,5	6,6	5	6
E	5,3	6,2	6	6,5
ΣΤ	5	6	4,2	7
Z	4,8	4,5	5,5	5
H	4,5	5,8	5,5	4
Θ	4,2	5,1	3	9
I	3	3,6	6	5

Το διάγραμμα διασποράς για τους βαθμούς στην Μικροοικονομική και τη Μακροοικονομική



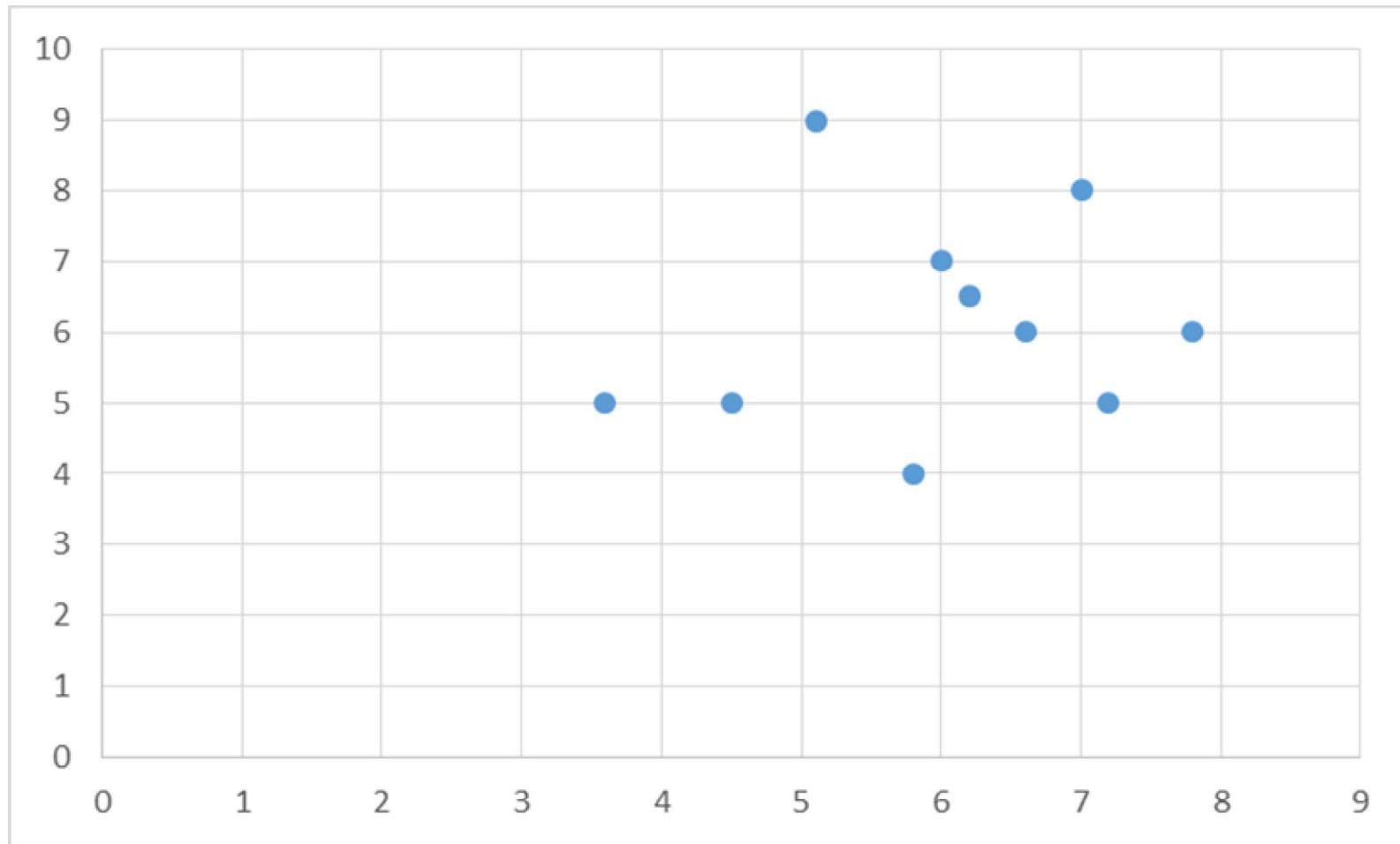
Υψηλή Θετική Συσχέτιση $r = + 0,939$

Το διάγραμμα διασποράς για τους βαθμούς στην Μικροοικονομική και τις Ποσοτικές Μεθόδους



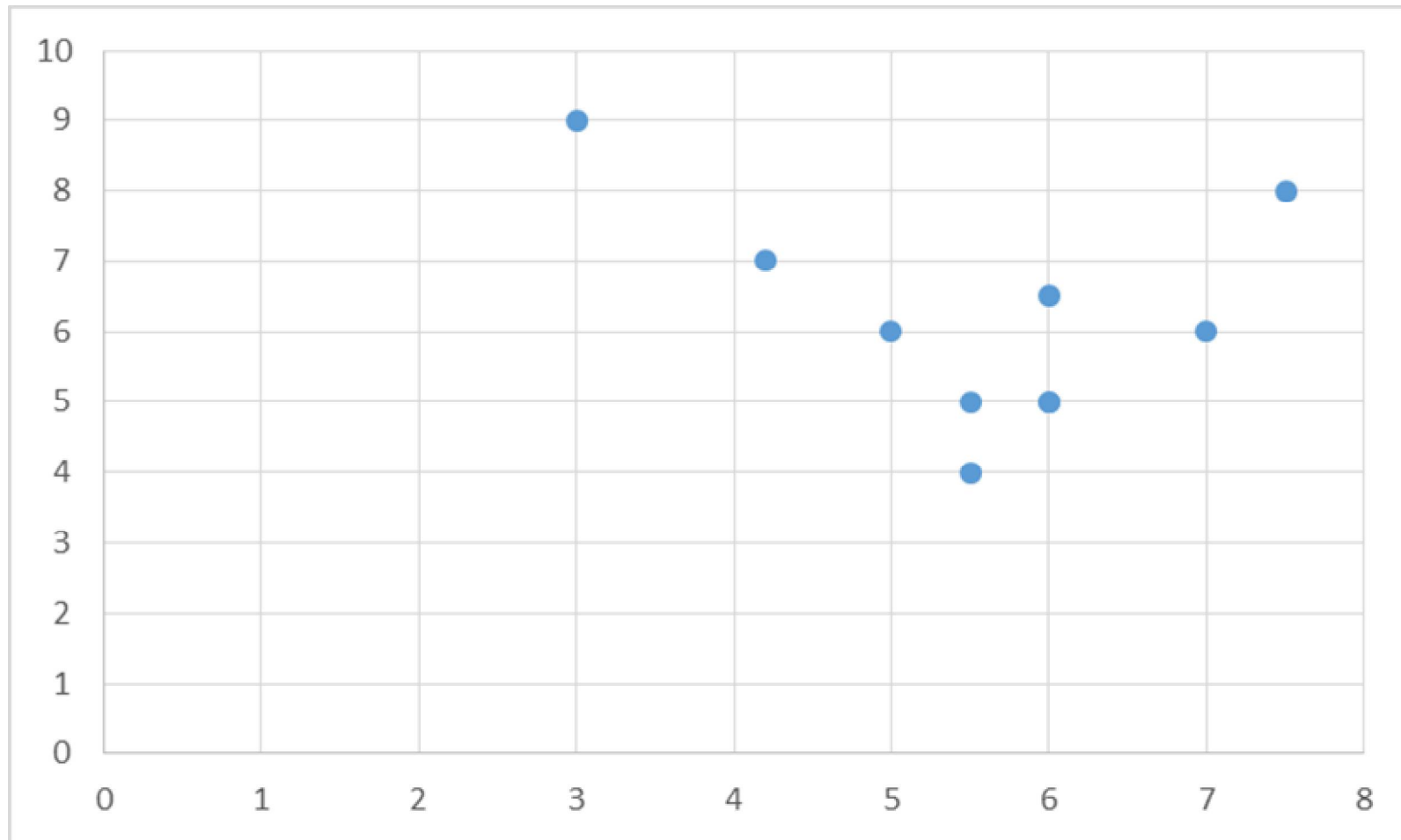
Μέτρια Θετική Συσχέτιση $r = + 0,489$

Το διάγραμμα διασποράς για τους βαθμούς στην Μακροοικονομική και το Δίκαιο



Σχεδόν Μηδενική Συσχέτιση $r = + 0,142$

Το διάγραμμα διασποράς για τους βαθμούς στις Ποσοτικές Μεθόδους και το Δίκαιο



Χαμηλή Αρνητική Συσχέτιση $r = - 0,317$

Η Ερμηνεία του Συντελεστή Συσχέτισης

Ο συντελεστής συσχέτισης αποτελείται από δύο στοιχεία:

- **ένα πρόσημο**
(πληροφορίες για την κατεύθυνση)
- **μια αριθμητική τιμή από 0 ως 1**
(πληροφορίες για το βαθμό)

Το μέγεθος του Συντελεστή Συσχέτισης

Αν ο δείκτης είναι μικρότερος του ± 0.30

Δεν υπάρχει συσχέτιση

Αν ο δείκτης κυμαίνεται μεταξύ $\pm 0.30 - 0.49$

Χαμηλή συσχέτιση

Αν ο δείκτης κυμαίνεται μεταξύ $\pm 0.50 - 0.69$

Μέτρια συσχέτιση

Αν ο δείκτης κυμαίνεται μεταξύ $\pm 0.70 - 0.79$

Υψηλή συσχέτιση

Αν ο δείκτης κυμαίνεται μεταξύ $\pm 0.80 - 0.99$

Πολύ υψηλή συσχέτιση

Παράγοντες που επηρεάζουν τον συντελεστή συσχέτισης

Η ομοιογένεια της ομάδας

Εάν η ομάδα του πληθυσμού που μελετάται είναι πολύ **ομοιογενής** (εάν έχει τα ίδια χαρακτηριστικά), είναι δυνατόν να βρούμε χαμηλή ή και καθόλου συσχέτιση **χωρίς αυτό να είναι αλήθεια**

Παράγοντες που επηρεάζουν τον συντελεστή συσχέτισης (2)

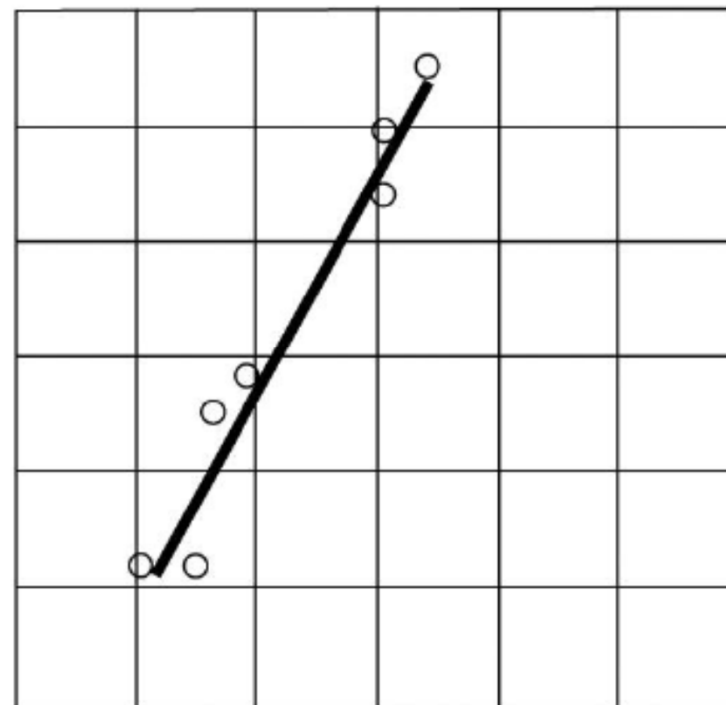
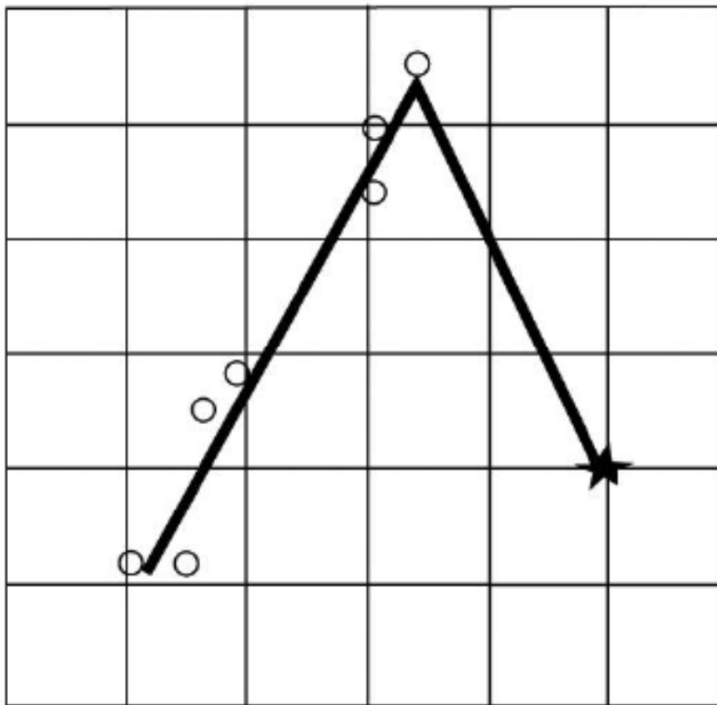
Ο βαθμός γραμμικότητας της σχέσης μεταξύ των μεταβλητών

Εάν υπάρχει καμπυλόγραμμη αντί για ευθύγραμμη σχέση ανάμεσα στις δύο μεταβλητές που μελετάμε είναι πιθανόν να βρεθεί χαμηλή ή και καθόλου συσχέτιση αλλά αυτό να μην ισχύει για **όλο το μήκος** της σχέσης

Παράγοντες που επηρεάζουν τον συντελεστή συσχέτισης (3)

Η ύπαρξη ακραίων τιμών

Μπορεί να επηρεάσουν τη **μορφή** της συσχέτισης



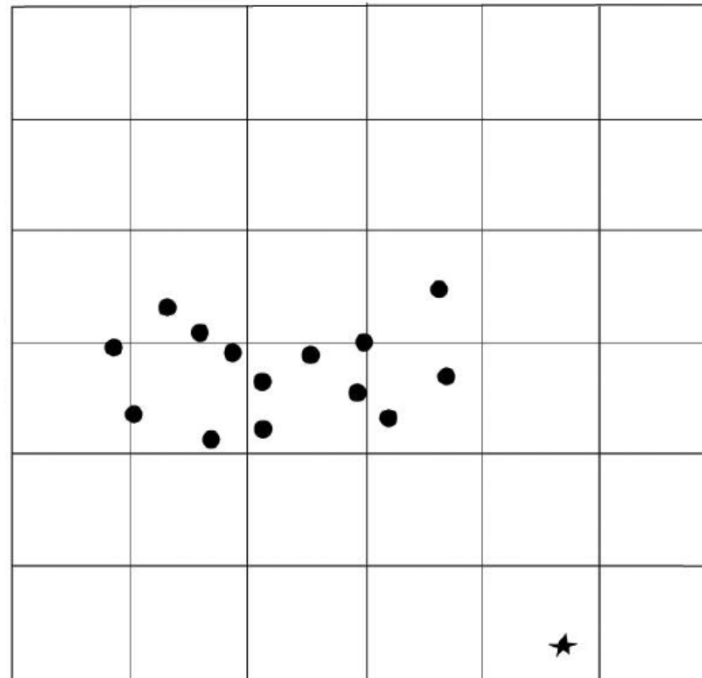
Παράγοντες που επηρεάζουν τον συντελεστή συσχέτισης (4)

Η ύπαρξη ακραίων τιμών

Μπορεί να επηρεάσουν το **βαθμό** της συσχέτισης

Με την
ακραία τιμή

$$r = -0,53$$



Χωρίς την
ακραία τιμή

$$r = 0,1$$

Συντελεστής Συσχέτισης και Αιτιότητα

- Ο συντελεστής συσχέτισης είναι ένα στατιστικό κριτήριο που μας πληροφορεί μόνο για τη **συμμεταβολή** των δύο μεταβλητών που μελετώνται και όχι για το εάν υπάρχει **αιτιώδης σχέση** μεταξύ τους.
- Η υψηλή συσχέτιση δεν δηλώνει σχέσεις **αιτίου** και **αποτελέσματος**. Μπορεί να οφείλεται σε μια **τρίτη μεταβλητή**, την οποία δεν έχουμε συμπεριλάβει στην έρευνά μας, και η οποία να λειτουργεί ως **αίτιο**.

Υπολογισμός Συντελεστή Συσχέτισης

Ο συντελεστής συσχέτισης υπολογίζεται εναλλακτικά από τις εξής σχέσεις:

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

$$r = \frac{\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}}{\sqrt{\left[\sum X_i^2 - \frac{(\sum X_i)^2}{n} \right] \left[\sum Y_i^2 - \frac{(\sum Y_i)^2}{n} \right]}}$$

Παράδειγμα

Ένα Ιατρικό Κέντρο δοκιμάζει μια νέα δίαιτα αδυνατίσματος. Ο παρακάτω πίνακας περιέχει, για 5 τυχαία επιλεγμένα άτομα (A – E) από αυτά που υποβλήθηκαν στη συγκεκριμένη δίαιτα, τον αριθμό των εβδομάδων (X) που την ακολούθησαν και το βάρος (Y) σε κιλά που έχασαν

Άτομα	A	B	Γ	Δ	E
X	2	3	5	6	8
Y	1	2	5	6	7

Με βάση τα παραπάνω στοιχεία και αν επιπλέον γνωρίζουμε ότι οι μεταβλητές X και Y συνδέονται με γραμμική σχέση να εκτιμηθεί ο συντελεστής συσχέτισης και να ερμηνευτεί.

Απάντηση

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

Με βάση τα δεδομένα της άσκησης δημιουργείται ο παρακάτω πίνακας:

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
X_i	Y_i	X_i^2	Y_i^2	$(X_i - \bar{X})$	$(Y_i - \bar{Y})$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$
2	1	4	1					
3	2	9	4					
5	5	25	25					
6	6	36	36					
8	7	64	49					
24	21	138	115					

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

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
X_i	Y_i	X_i^2	Y_i^2	$(X_i - \bar{X})$	$(Y_i - \bar{Y})$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$
2	1	4	1	-2,8	-3,2			
3	2	9	4	-1,8	-2,2			
5	5	25	25	0,2	0,8			
6	6	36	36	1,2	1,8			
8	7	64	49	3,2	2,8			
24	21	138	115	0	0			

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

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
X_i	Y_i	X_i^2	Y_i^2	$(X_i - \bar{X})$	$(Y_i - \bar{Y})$	$(X_i - \bar{X})(Y_i - \bar{Y})$	$(X_i - \bar{X})^2$	$(Y_i - \bar{Y})^2$
2	1	4	1	-2,8	-3,2	8,96	7,84	10,24
3	2	9	4	-1,8	-2,2	3,96	3,24	4,84
5	5	25	25	0,2	0,8	0,16	0,04	0,64
6	6	36	36	1,2	1,8	2,16	1,44	3,24
8	7	64	49	3,2	2,8	8,96	10,24	7,84
24	21	138	115	0	0	24,2	22,8	26,8

$$\sum (X_i - \bar{X})^2 = 22,8$$

$$\sum (Y_i - \bar{Y})^2 = 26,8$$

$$\sum (X_i - \bar{X})(Y_i - \bar{Y}) = 24,2$$

$$r = \frac{\sum [(Y_i - \bar{Y})(X_i - \bar{X})]}{\sqrt{\sum (Y_i - \bar{Y})^2 \sum (X_i - \bar{X})^2}} = 0,98$$

Εναλλακτικά:

$$\sum X_i = 24 \qquad \bar{X} = \frac{\sum X_i}{n} = \frac{24}{5} = 4,8$$

$$\sum Y_i = 21 \qquad \bar{Y} = \frac{\sum Y_i}{n} = \frac{21}{5} = 4,2$$

$$r = \frac{\sum X_i Y_i - \frac{\sum X_i \sum Y_i}{n}}{\sqrt{\left[\sum X_i^2 - \frac{(\sum X_i)^2}{n} \right] \left[\sum Y_i^2 - \frac{(\sum Y_i)^2}{n} \right]}} = 0,98$$

Η τιμή του r δείχνει ότι υπάρχει έντονη θετική συσχέτιση μεταξύ του χρόνου διαίτας ενός ατόμου και της απώλειας βάρους του.