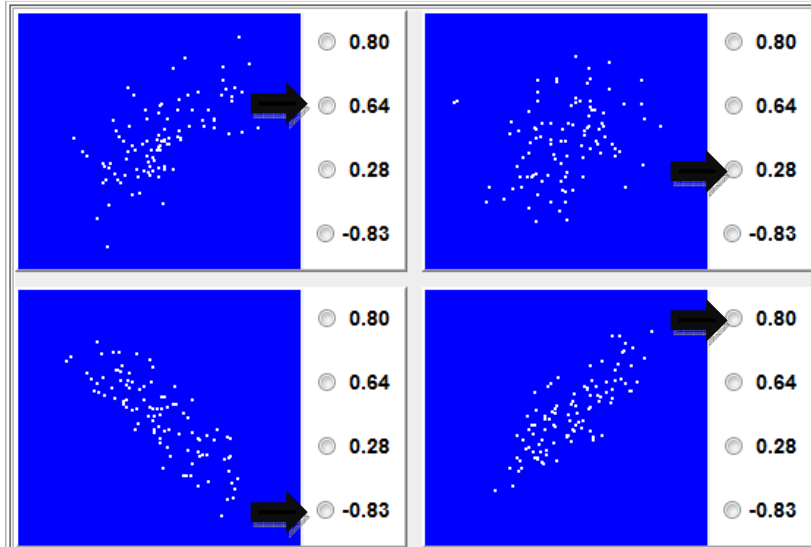


# ΑΣΚΗΣΕΙΣ ΣΤΗ ΣΥΣΧΕΤΙΣΗ ΚΑΙ ΤΟΥΣ ΕΛΕΓΧΟΥΣ ΥΠΟΘΕΣΕΩΝ

## Άσκηση 1

Να αντιστοιχίσετε κάθε τιμή του δείκτη συσχέτισης  $r$  του Pearson με το κατάλληλο διάγραμμα διασποράς.



## Άσκηση 2

«Οι υψηλές επιδόσεις στη δοκιμασία σχεδιασμού των γραμμάτων συνοδεύονται από υψηλές επιδόσεις στη δοκιμασία της ποιότητας γραφής ( $r=0.426$ ). Η συσχέτιση αυτή βρέθηκε στατιστικά σημαντική σε ε.σ.  $\alpha = 0.05$  ( $p = 0.002$ )».

Με βάση το παραπάνω κείμενο να σχολιάσετε την τιμή του συντελεστή συσχέτισης και το αποτέλεσμα του στατιστικού ελέγχου.

### Απάντηση

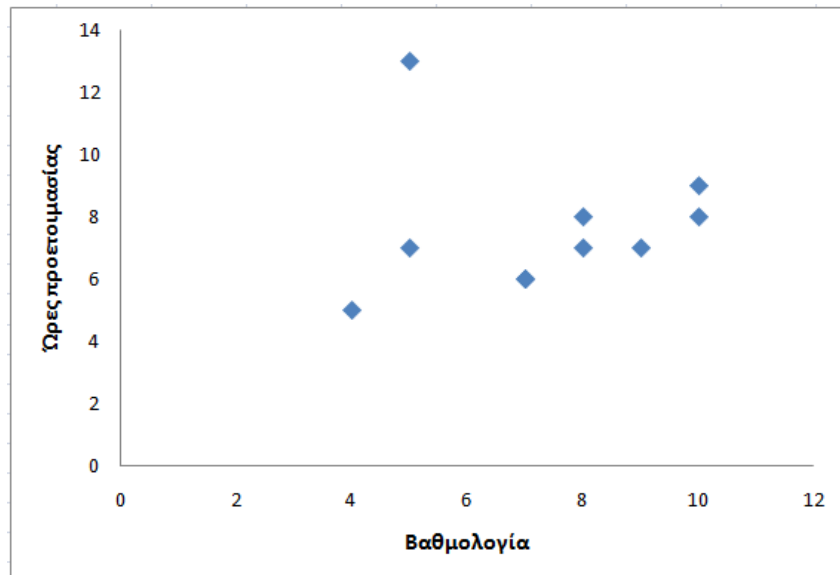
Σύμφωνα με τον δείκτη συσχέτισης  $r$  του Pearson, η συσχέτιση ανάμεσα στις δύο επιδόσεις μπορεί να χαρακτηριστεί ως μέτρια θετική. Η πιθανότητα, αυτή η τιμή του δείκτη συσχέτισης να προέκυψε τυχαία είναι 2% και είναι μικρότερη από τη στάθμη σημαντικότητας του 5% που έθεσε αρχικά ο ερευνητής.

## Άσκηση 3

Στον παρακάτω πίνακα δίνονται οι βαθμοί 10 μαθητών σε ένα σύντομο τεστ Ιστορίας και οι ώρες που οι ίδιοι οι μαθητές δήλωσαν ότι προετοιμάστηκαν για το τεστ αυτό στο σπίτι.

	Βαθμολογία	Ώρες Προετοιμασίας
1	8	7
2	7	6
3	5	13
4	9	7
5	8	8
6	10	9
7	7	6
8	4	5
9	10	8
10	5	7

(α) Κατασκευάστε το διάγραμμα διασποράς βαθμολογίας-ώρες και σχολιάστε τη σχέση των δύο μεταβλητών.



Από το διάγραμμα διασποράς διαπιστώνουμε ότι η σχέση μεταξύ των δύο μεταβλητών είναι θετική και σε μεγάλο βαθμό γραμμική. Από το πρότυπο αυτό φαίνεται ότι αποκλίνει μόνο το σημείο (5,13) που αντιστοιχεί στον 3<sup>ο</sup> μαθητή.

**(β) Αντικατοπτρίζει η τιμή του δείκτη  $r$  του Pearson την πραγματική σχέση ανάμεσα στις δύο μεταβλητές; Να δικαιολογήσετε την απάντησή σας.**

Γνωρίζουμε ότι ο δείκτης  $r$  του Pearson είναι κατάλληλος για την περιγραφή του είδους και της έντασης γραμμικών σχέσεων μεταξύ μεταβλητών. Επιπλέον, ο δείκτης είναι ευαίσθητος στην παρουσία ακραίων τιμών, οι οποίες μπορεί να αλλοιώσουν σημαντικά την τιμή του. Κατά συνέπεια, η παρουσία του ακραίου σημείου (5,13) αναμένουμε να επηρεάσει σημαντικά την τιμή του δείκτη. Στην περίπτωση αυτή  $r=0,05$  κάτι που δείχνει απουσία γραμμικής σχέσης, ενώ αυτή υπάρχει. Αν αφαιρέσουμε το σημείο (5,13) από τους υπολογισμούς έχουμε  $r=0,78$ , που δείχνει υψηλή θετική συσχέτιση και συνάδει με τα όσα διαπιστώσαμε στο (α).

#### Άσκηση 4

Ο παρακάτω πίνακας συνοψίζει τις μηνιαίες αποδοχές 10 δημοσίων υπαλλήλων και τα έτη προϋπηρεσίας τους:

	Αποδοχές	Έτη εργασίας
1	1000	1
2	1100	2
3	1200	4
4	1300	7
5	1400	10
6	1450	15
7	1450	17
8	1450	17
9	1450	25
10	1500	25

(α) Ποια είναι η σχέση ανάμεσα στις δύο μεταβλητές σύμφωνα με τον συντελεστή συσχέτισης  $r$  του Pearson;

	Αποδοχές (X)	Έτη εργασίας (Y)	X x Y
1	1000	1	1000
2	1100	2	2200
3	1200	4	4800
4	1300	7	9100
5	1400	10	14000

6	1450	15	21750
7	1450	17	24650
8	1450	17	24650
9	1450	25	36250
10	1500	25	37500
Σύνολο	<b>13300</b>	<b>123</b>	<b>175900</b>

Υπολογίζουμε και τις τυπικές αποκλίσεις των δύο μεταβλητών με τη χρήση του κατάλληλου τύπου (βλέπε Τυπολόγιο στο τέλος)  $s_x=173,5$  και  $s_y=8,88$ . Έτσι έχουμε:

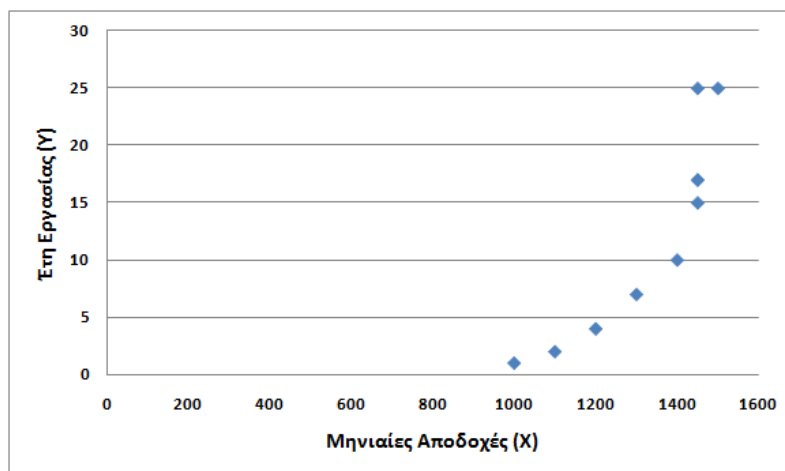
$$COV_{XY} = \frac{1}{n-1} \left( \sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n} \right) = \frac{1}{9} \left( 175900 - \frac{13300 \times 123}{10} \right) = 1367,77$$

$$r_{XY} = \frac{COV_{XY}}{s_x s_y} = \frac{1367,77}{173,5 \times 8,88} = \frac{1367,77}{1540,68} = 0,88$$

Σύμφωνα με την τιμή του συντελεστή γραμμικής συσχέτισης, η σχέση ανάμεσα στις δύο μεταβλητές είναι θετική και υψηλή.

**(β) Είναι ο συντελεστής συσχέτισης  $r$  του Pearson κατάλληλος δείκτης για την περιγραφή της παραπάνω σχέσης; Δικαιολογήστε την απάντησή σας.**

Κατασκευάζουμε το διάγραμμα διασποράς που περιγράφει την παραπάνω σχέση.



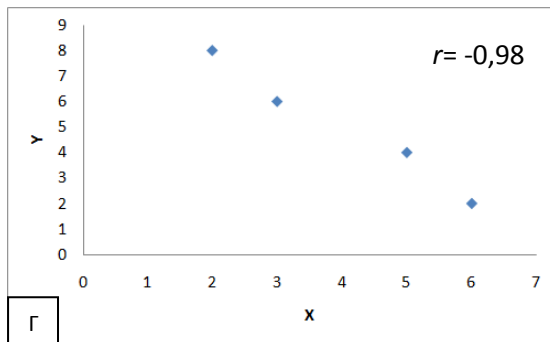
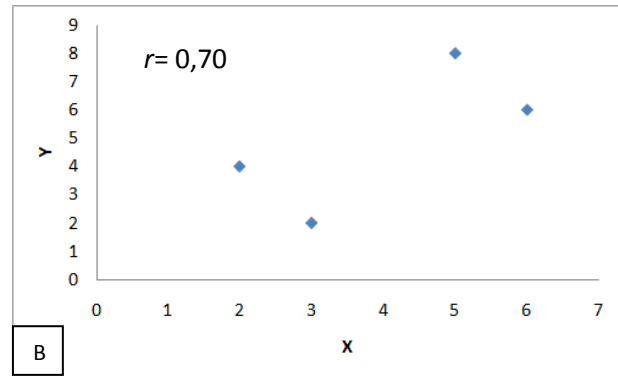
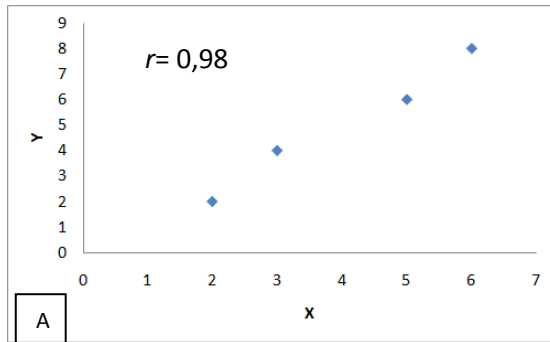
Από το διάγραμμα διαπιστώνουμε ότι αρχικά οι αποδοχές αυξάνονται καθώς αυξάνονται τα έτη εργασίας, αλλά μετά το έτη (Y)=15 οι αποδοχές σταθεροποιούνται. Συνεπώς, η σχέση μεταξύ των μεταβλητών είναι καμπυλόγραμμη, άρα ο συντελεστής γραμμικής συσχέτισης επηρεάζεται και δεν είναι ο πιο κατάλληλος δείκτης για την περιγραφή της.

### Άσκηση 5

Κατασκευάστε τα διαγράμματα διασποράς για τα παρακάτω ζεύγη μεταβλητών X και Y και υπολογίστε τους αντίστοιχους συντελεστές συσχέτισης.

X	Y	X	Y	X	Y
2	2	2	4	2	8
3	4	3	2	3	6
5	6	5	8	5	4
6	8	6	6	6	2

Με ποιόν τρόπο πρέπει να αναδιαταχθούν οι τιμές της μεταβλητής Y έτσι ώστε σε κάθε σύνολο δεδομένων να έχουμε την μικρότερη δυνατή αρνητική συσχέτιση;



Η μέγιστη δυνατή αρνητική συσχέτιση είναι αυτή που αντιστοιχεί στο σύνολο δεδομένων Γ, δηλαδή  $-0,98$ . Για να επιτευχθεί η συσχέτιση αυτή και για τα σύνολα Α και Β, πρέπει να τοποθετηθούν οι τιμές των στηλών Y για τα σύνολα αυτά σε φθίνουσα διάταξη  $\{8, 6, 4, 2\}$ , όπως ακριβώς δηλαδή οι τιμές της στήλης Y στο σύνολο Γ.

### Τυπολόγιο

$$COV_{XY} = \frac{1}{n-1} \left( \sum_{i=1}^n x_i y_i - \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n} \right) \quad s = \sqrt{\frac{1}{n-1} \sum_{i=1}^n n_i (X_i - \bar{X})^2} \quad r_{XY} = \frac{COV_{XY}}{s_X s_Y}$$

Όπου:  $n$  = μέγεθος του δείγματος,  $X, Y$  = οι τυχαίες μεταβλητές  $X$  και  $Y$ ,  $s$  = η δειγματική τυπ. απόκλιση και  $r$  = ο δειγματικός συντελεστής συσχέτισης,  $COV_{XY}$  = η συνδιακύμανση των μεταβλητών  $X$  και  $Y$ .