Μαθηματικά Β' Λυκείου - Πετρίδης Κωνσταντίνος
Στατιστική

1. Μια εταιρεία αποτελείται από τέσσερις τομείς. Οι υπάλληλοι σε κάθε τομέα ρωτήθηκαν αν θέλουν να μετακινηθεί το ωράριο εργασίας κατά μισή ώρα. Οι απαντήσεις τους δίνονται στον παρακάτω πίνακα.

	1	2	3	4	5	6	7	8	9	10	11	12
Τομέας Α	Ναι	Όχι	Ναι	Ναι	Ναι	Όχι	Ναι	Ναι	Ναι	Όχι	Ναι	Όχι
Τομέας Β	Όχι	Όχι	Ναι	Όχι	Ναι							
Τομέας Γ	Ναι	Ναι	Όχι	Όχι	Όχι	Ναι	Ναι	Όχι				
Τομέας Δ	Όχι	Ναι	Όχι	Ναι	Ναι	Όχι	Όχι	Ναι	Ναι	Όχι	Όχι	

i. Να υπολογίσετε την επικρατούσα τιμή των απαντήσεων σε κάθε τομέα.

ii. Να αναφέρετε αν η εταιρεία πρέπει να αλλάξει το ωράριο εργασίας.

Λύση: (Ασκ: 1/191)

i. Μετράμε «Ναι» και «Όχι» ανά τομέα:

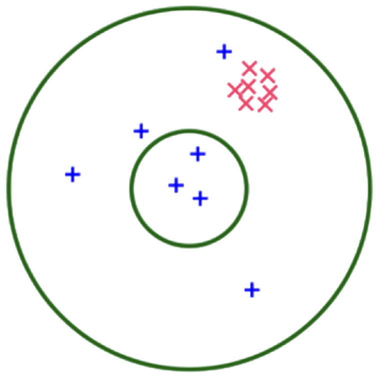
Τομέας	Ναι	Όχι	Επικρατούσα τιμή
Α	8	4	Ναι
Β	2	3	Όχι
Γ	4	4	Καμία (Ναι, Όχι)
Δ	5	6	Όχι

ii. Συνολικά, η επικρατούσα τιμή των απαντήσεων είναι το «Όχι» $\Rightarrow$ Η εταιρεία δεν πρέπει να αλλάξει το ωράριο εργασίας.

2. Δύο τοξότες ρίχνουν βέλη στον διπλανό στόχο. Με κόκκινο σημειώνονται οι βολές του τοξότη A και με μπλε οι βολές του τοξότη B.

i. Να βρείτε τον τοξότη, του οποίου οι βολές έχουν τη μικρότερη διασπορά.

ii. Ποιος είναι ο καλύτερος τοξότης;



Λύση: (Ασκ: 2/191)

i. Οι βολές του A είναι πολύ «μαζεμένες» μεταξύ τους (σχηματίζουν σφιχτή συστάδα), άρα ο A έχει μικρότερη διασπορά (μεγαλύτερη ακρίβεια - precision).

ii. Ο καλύτερος τοξότης είναι ο B, επειδή το κέντρο των βολών του βρίσκεται πιο κοντά στο κέντρο του στόχου (μεγαλύτερη ορθότητα -accuracy). Ο A, παρότι ακριβής, παρουσιάζει συστηματική μετατόπιση, μακριά από το κέντρο.

3. Σε δύο δείγματα A και B δίνονται $\bar{x}_A = 50,\ s_A = 4,\ \bar{x}_B = 70,\ s_B = 6$.

i. Να υπολογίσετε τον συντελεστή μεταβολής για το καθένα από τα παραπάνω δείγματα.

ii. Να βρείτε ποιο από τα δείγματα A και B παρουσιάζει μεγαλύτερη ομοιογένεια.

Λύση: (Ασκ: 3/191)

Ο συντελεστής μεταβλητότητας ορίζεται από $CV = \dfrac{s}{|\bar{x}|}$.

i. Για το δείγμα A: $CV_A = \dfrac{s_A}{|\bar{x}_A|} = \dfrac{4}{50} = 0{,}08\ (= 8\%)$.

Για το δείγμα B: $CV_B = \dfrac{s_B}{|\bar{x}_B|} = \dfrac{6}{70} = \dfrac{3}{35} = 0{,}0857\ (\approx 8{,}57\%)$.

ii. Εφόσον μικρότερο $CV \Rightarrow$ μεγαλύτερη ομοιογένεια και $CV_A < CV_B$, το δείγμα A παρουσιάζει μεγαλύτερη ομοιογένεια.

Παρατήρηση: Και τα δύο δείγματα έχουν $CV < 10\%$, άρα είναι ομοιογενή.

4. Δίνονται η μέση τιμή και η τυπική απόκλιση των βαθμολογιών στην τελική εξέταση στο μάθημα των Μαθηματικών τεσσάρων τμημάτων της Β' Λυκείου.

Τμήμα	Μέση Τιμή	Τυπική Απόκλιση
B_1	14,1	1,5
B_2	14,2	2,3
B_3	12,9	1,2
B_4	13,7	2,7

i. Ποιο τμήμα έχει τη μεγαλύτερη μέση τιμή και ποιο τμήμα έχει τη μικρότερη τυπική απόκλιση;

ii. Ποιο είναι, κατά τη γνώμη σας, το καλύτερο από τα τέσσερα τμήματα;

iii. Ποιο από τα πιο πάνω τμήματα παρουσιάζει μεγαλύτερη ομοιογένεια;

Λύση: (Ασκ: 4/191)

i. Μεγαλύτερη μέση τιμή: B_2 (14,2). Μικρότερη τυπική απόκλιση: B_3 (1,2).

ii. Καλύτερο τμήμα (επίδοση): B_1 — έχει τη μεγαλύτερη μέση τιμή, συμπεριλαμβανομένης της τυπικής απόκλισης, 14.1 ± 1.5.

iii. Υπολογίζουμε τον συντελεστή μεταβλητότητας $CV = \dfrac{s}{|\bar{x}|}$ για κάθε τμήμα:

Τμήμα	$\bar{x}$	s	$CV = \dfrac{s}{\lvert\bar{x}\rvert}$
B_1	14,1	1,5	0,106 (= 10,6%)
B_2	14,2	2,3	0,162 (= 16,2%)
B_3	12,9	1,2	0,093 (= 9,3%)
B_4	13,7	2,7	0,197 (= 19,7%)

Μικρότερο CV έχει το B_3, άρα το B_3 παρουσιάζει μεγαλύτερη ομοιογένεια.

Παρατήρηση: Μόνο το B_3 έχει $CV < 10\%$ (πλήρης ομοιογένεια με βάση το κριτήριο $CV < 10\%$).

5. Να χαρακτηρίσετε ΣΩΣΤΟ ή ΛΑΘΟΣ την καθεμιά από τις πιο κάτω προτάσεις και να αιτιολογήσετε την απάντησή σας.

i. Όταν έχουμε τέλεια θετική γραμμική συσχέτιση ο συντελεστής συσχέτισης είναι 1.

ii. Αν για τον συντελεστή γραμμικής συσχέτισης ισχύει $r = 0$, τότε οι μεταβλητές X και Y δεν έχουν γραμμική συσχέτιση.

iii. Ο συντελεστής συσχέτισης $r = 0{,}1$ δείχνει πιο ισχυρή γραμμική συσχέτιση των δύο μεταβλητών από ό,τι ο συντελεστής συσχέτισης $r = -0{,}8$.

iv. Όταν ερευνούμε τη «μάζα των μαθητών μιας τάξης» σε σχέση με «το ύψος τους», τότε η ανεξάρτητη μεταβλητή είναι η μάζα και η εξαρτημένη μεταβλητή το ύψος.

v. Δύο μεταβλητές έχουν αρνητική γραμμική συσχέτιση όταν, όσο αυξάνονται οι τιμές της μίας μεταβλητής, τείνουν να μειώνονται οι τιμές της άλλης μεταβλητής.

vi. Όταν ο συντελεστής συσχέτισης είναι $r = -0{,}99$, τότε οι δύο μεταβλητές έχουν ισχυρή θετική γραμμική συσχέτιση.

Λύση: (Ασκ: 1/201)

i. ΣΩΣΤΟ. Τέλεια θετική γραμμική συσχέτιση $\Rightarrow r = 1$.

ii. ΣΩΣΤΟ. $r = 0 \Rightarrow$ μηδενική *γραμμική* συσχέτιση (μπορεί να υπάρχει μη γραμμική σχέση).

iii. ΛΑΘΟΣ. Η *ισχύς* της συσχέτισης μετριέται από $|r|$. Εφόσον $|-0{,}8| = 0{,}8 > 0{,}1$, το $r = -0{,}8$ δηλώνει ισχυρότερη (αρνητική) συσχέτιση.

iv. ΛΑΘΟΣ. Συνήθως θεωρούμε ανεξάρτητο το ύψος και εξαρτημένη τη μάζα (η μάζα εξαρτάται από το ύψος).

v. ΣΩΣΤΟ. Ορισμός αρνητικής γραμμικής συσχέτισης: καθώς αυξάνεται η μία μεταβλητή, η άλλη τείνει να μειώνεται.

vi. ΛΑΘΟΣ. $r = -0{,}99$ υποδηλώνει πολύ ισχυρή αρνητική γραμμική συσχέτιση, όχι θετική.

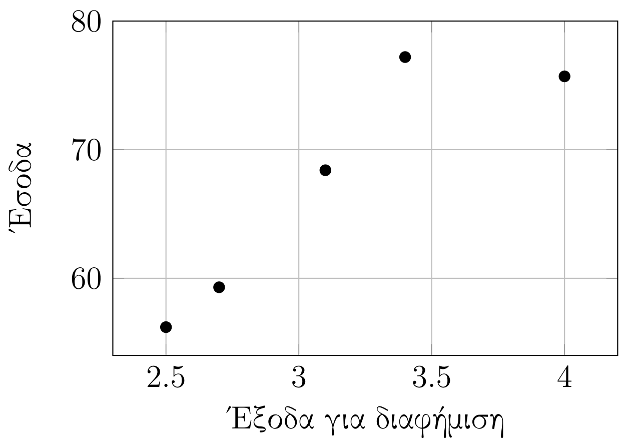
6. Μια εταιρεία ενδιαφέρεται να ερευνήσει την επίδραση των εξόδων για διαφήμιση στους τελευταίους μήνες στα έσοδα από τις πωλήσεις. Τα δεδομένα δίνονται στον πίνακα:

Έξοδα για διαφήμιση	Έσοδα
2,5	56,2
3,4	77,2
2,7	59,3
3,1	68,4
4,0	75,7

i. Να κατασκευάσετε το διάγραμμα διασποράς.
ii. Να χαρακτηρίσετε τη συσχέτιση μεταξύ των δύο μεταβλητών.

Λύση: (Ασκ: 2/201)

i. Διάγραμμα διασποράς:



ii. Συντελεστής συσχέτισης:

$$r = \frac{\sum xy - \nu\,\bar{x}\bar{y}}{\nu\, s_x s_y}.$$

Από τα δεδομένα έχουμε:

$$\bar{x} = 3{,}14, \quad \bar{y} = 67{,}36, \quad s_x \approx 0{,}53, \quad s_y \approx 8{,}45 \quad \sum xy = 1077{,}93, \quad \nu = 5$$

Άρα

$$r = \frac{1077{,}93 - 5 \cdot 3{,}14 \cdot 67{,}36}{5 \cdot 0{,}53 \cdot 8{,}45} = 0{,}908 \rightarrow \text{ισχυρή θετική γραμμική συσχέτιση.}$$

7. Να διατάξετε τις πιο κάτω τιμές του r σε αύξουσα τάξη του βαθμού γραμμικής συσχέτισης δύο μεταβλητών:

$$-0{,}6, \quad 0{,}9, \quad -0{,}7, \quad 0{,}2, \quad -1.$$

Λύση: (Ασκ: 3/201)

Ο βαθμός (ισχύς) της συσχέτισης μετριέται από την απόλυτη τιμή $|r|$ (το πρόσημο δηλώνει μόνο τη διεύθυνση). Υπολογίζουμε:

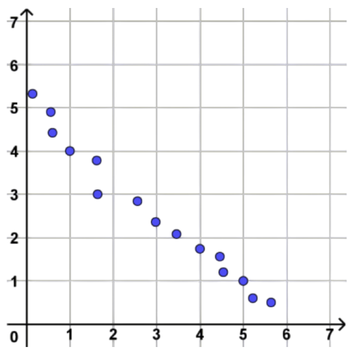
$$|-0{,}6| = 0{,}6, \quad |0{,}9| = 0{,}9, \quad |-0{,}7| = 0{,}7, \quad |0{,}2| = 0{,}2, \quad |-1| = 1.$$

Άρα, σε αύξουσα τάξη βαθμού (από την πιο ασθενή στη ισχυρότερη συσχέτιση):

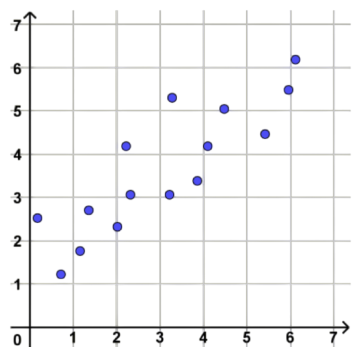
$$0{,}2 \;<\; -0{,}6 \;<\; -0{,}7 \;<\; 0{,}9 \;<\; -1$$

Σημείωση: τα $r = \pm 1$ έχουν τον ίδιο βαθμό (τέλεια συσχέτιση), αλλά αντίθετη διεύθυνση.

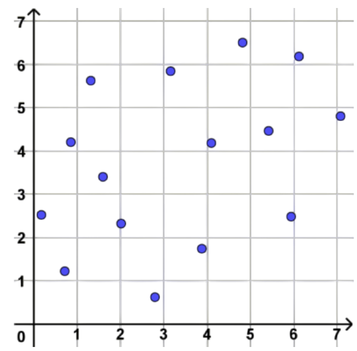
8. Να περιγράψετε το είδος της συσχέτισης των δύο μεταβλητών με τους χαρακτηρισμούς (τέλεια θετική γραμμική συσχέτιση, τέλεια αρνητική γραμμική συσχέτιση, μηδενική συσχέτιση, ισχυρή θετική γραμμική συσχέτιση, ασθενής θετική γραμμική συσχέτιση, ισχυρή αρνητική γραμμική συσχέτιση, ασθενής αρνητική γραμμική συσχέτιση) στα πιο κάτω διαγράμματα:



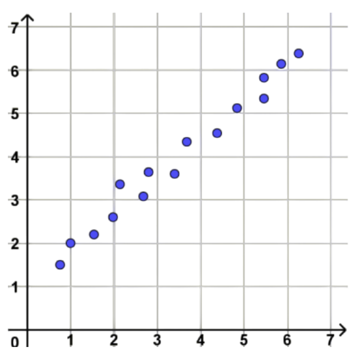
i.



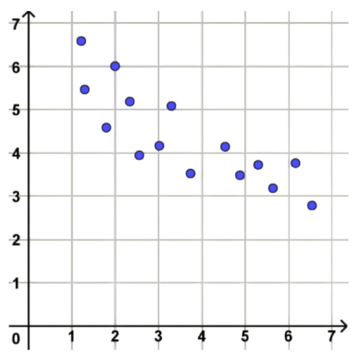
ii.



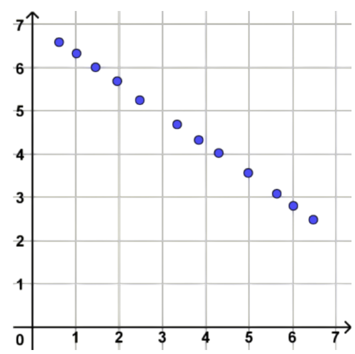
iii.



iv.



v.



vi.

Λύση: (Ασκ: 4/202)

i. Ισχυρή αρνητική γραμμική συσχέτιση.

ii. Ασθενής θετική γραμμική συσχέτιση.

iii. Μηδενική συσχέτιση.

iv. Ισχυρή θετική γραμμική συσχέτιση.

v. Ασθενής αρνητική γραμμική συσχέτιση.

vi. Τέλεια αρνητική γραμμική συσχέτιση.

9. Δίνονται οι βαθμοί 8 μαθητών σε δύο διαγωνίσματα.

Διαγώνισμα A (x)	Διαγώνισμα B (y)
13	17
11	14
16	18
15	13
20	20
13	14
9	12
18	16

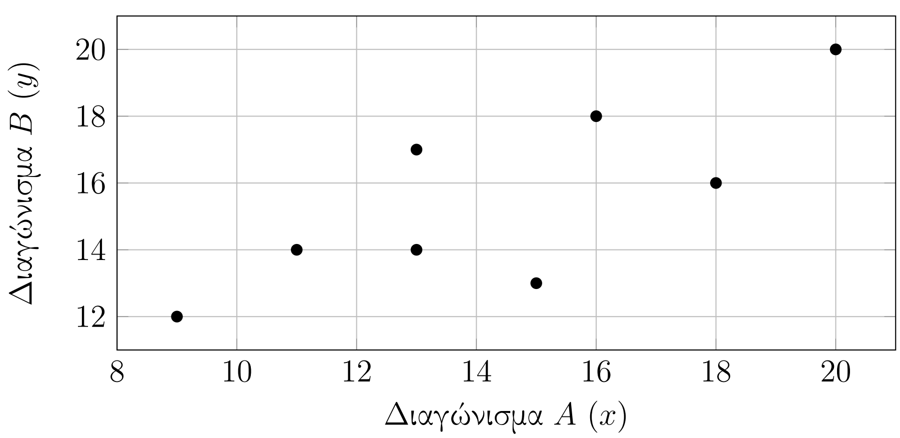
i. Να κατασκευάσετε το διάγραμμα διασποράς.

ii. Να υπολογίσετε τον γραμμικό συντελεστή συσχέτισης.

iii. Να χαρακτηρίσετε το είδος της συσχέτισης μεταξύ των δύο μεταβλητών.

Λύση: (Ασκ: 5/203)

i. Διάγραμμα διασποράς:



ii. Ο συντελεστής συσχέτισης:

$$r = \frac{\sum xy - n\,\bar{x}\bar{y}}{\sqrt{\left(\sum x^2 - n\,\bar{x}^2\right)\left(\sum y^2 - n\,\bar{y}^2\right)}}.$$

Με $n=8$, $\sum x = 115 \Rightarrow \bar{x} = 14{,}375$, $\sum y = 124 \Rightarrow \bar{y} = 15{,}5$, $\sum x^2 = 1745$, $\sum y^2 = 1974$, $\sum xy = 1836$. Άρα

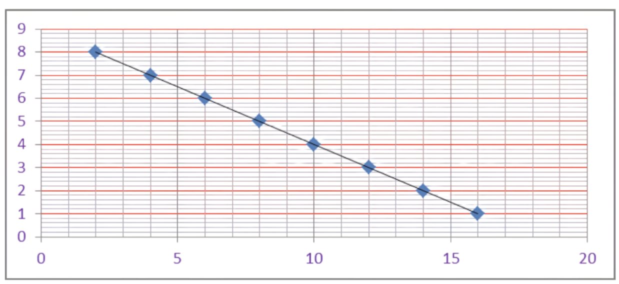
$$r = \frac{1836 - 8 \cdot 14{,}375 \cdot 15{,}5}{\sqrt{(1745 - 8 \cdot 14{,}375^2)\,(1974 - 8 \cdot 15{,}5^2)}} = \frac{53{,}5}{69{,}119} = 0{,}774.$$

iii. Επειδή $r \approx 0{,}774$, οι δύο μεταβλητές εμφανίζουν θετική γραμμική συσχέτιση.

10. Δίνεται το πιο κάτω διάγραμμα διασποράς μεταξύ δύο μεταβλητών.

i. Να υπολογίσετε τον συντελεστή συσχέτισης.

ii. Να χαρακτηρίσετε τη συσχέτιση μεταξύ των μεταβλητών.



Λύση: (Ασκ: 6/203)

Από το διάγραμμα διαβάζουμε τα σημεία:

$$(3,8),\ (5,7),\ (7,6),\ (9,5),\ (11,4),\ (13,3),\ (15,2),\ (17,1).$$

Έχουμε $n=8$, $\sum x = 80$, $\sum y = 36$, $\sum x^2 = 968$, $\sum y^2 = 204$, $\sum xy = 276$.

Ο συντελεστής συσχέτισης είναι

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{\left(n\sum x^2 - (\sum x)^2\right)\left(n\sum y^2 - (\sum y)^2\right)}} = \frac{8 \cdot 276 - 80 \cdot 36}{\sqrt{(8 \cdot 968 - 80^2)\,(8 \cdot 204 - 36^2)}} = \frac{-672}{\sqrt{451\,584}} = -1.$$

ii. Συνεπώς υπάρχει τέλεια αρνητική γραμμική συσχέτιση.

11. Να χαρακτηρίσετε ΣΩΣΤΟ ή ΛΑΘΟΣ την καθεμιά από τις πιο κάτω προτάσεις και να αιτιολογήσετε την απάντησή σας.

i. Ο συντελεστής μεταβλητότητας εκφράζεται με την ίδια μονάδα μέτρησης που εκφράζονται οι παρατηρήσεις.

ii. Ένα δείγμα έχει μεγαλύτερη ομοιογένεια από ένα άλλο, αν έχει μεγαλύτερο συντελεστή μεταβλητότητας.

iii. Όταν μεγαλώνει η απόλυτη τιμή του συντελεστή συσχέτισης, τότε έχουμε πιο ισχυρή συσχέτιση.

iv. Όταν $r > 0$, τότε υπάρχει τέλεια θετική γραμμική συσχέτιση.

v. Ένας συντελεστής συσχέτισης $r_1 = 0{,}6$ δείχνει μεγαλύτερη γραμμική συσχέτιση από έναν άλλον $r_2 = -0{,}9$.

vi. Η διάμεσος είναι μέτρο διασποράς.

vii. Ο λόγος της μέσης τιμής προς την τυπική απόκλιση λέγεται συντελεστής μεταβολής και είναι καθαρός αριθμός.

Λύση: (Ασκ: 1/206)

i. ΛΑΘΟΣ. Ο συντελεστής μεταβλητότητας $CV = \dfrac{s}{|\bar{x}|}$ είναι αδιάστατος.

ii. ΛΑΘΟΣ. Μεγαλύτερη ομοιογένεια $\Leftrightarrow$ μικρότερο CV.

iii. ΣΩΣΤΟ. Η ισχύς της γραμμικής συσχέτισης μετριέται από $|r|$.

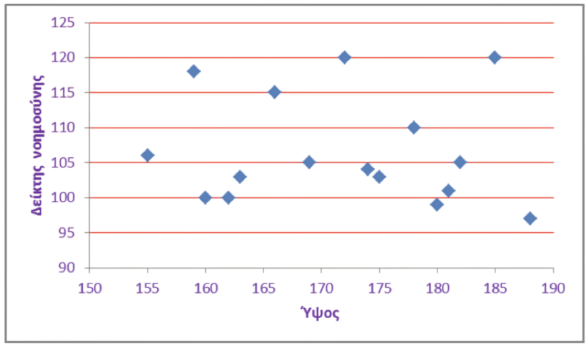
iv. ΛΑΘΟΣ. $r > 0$ δηλώνει θετική, όχι κατ' ανάγκην τέλεια συσχέτιση.

v. ΛΑΘΟΣ. $|-0{,}9| = 0{,}9 > 0{,}6 \Rightarrow$ το $r = -0{,}9$ είναι ισχυρότερη (αρνητική) συσχέτιση.

vi. ΛΑΘΟΣ. Η διάμεσος είναι μέτρο θέσης (κεντρικής τάσης), όχι διασποράς.

vii. ΛΑΘΟΣ. Ο συντελεστής μεταβλητότητας είναι $CV = \dfrac{s}{|\bar{x}|}$ (όχι $\bar{x}/s$). Είναι πράγματι καθαρός αριθμός, αλλά ο ορισμός που δίνεται είναι αντεστραμμένος.

12. Δίνεται το πιο κάτω διάγραμμα διασποράς. Να εξετάσετε αν και πόσο υπάρχει συσχέτιση μεταξύ των δύο μεταβλητών, αιτιολογώντας την απάντησή σας.



Λύση: (Ασκ: 2/206)

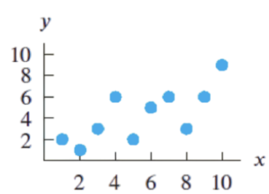
Τα σημεία είναι διάσπαρτα γύρω από μια σχεδόν οριζόντια ζώνη χωρίς σαφή ανιούσα ή κατιούσα κλίση και χωρίς να συγκεντρώνονται γύρω από κάποια ευθεία.
Επομένως η γραμμική συσχέτιση είναι μηδενική ή πολύ ασθενής (δηλαδή ο συντελεστής συσχέτισης r είναι κοντά στο 0).

13. Να αντιστοιχίσετε τους συντελεστές συσχέτισης

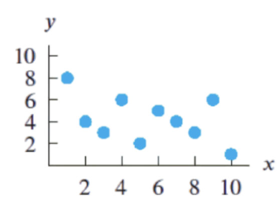
$$r=-0{,}98, \quad r=-0{,}5, \quad r=-0{,}25, \quad r=0, \quad r=0{,}7, \quad r=1$$

με τα αντίστοιχα διαγράμματα διασποράς $(\alpha),(\beta),(\gamma),(\delta),(\varepsilon),(\sigma\tau)$.

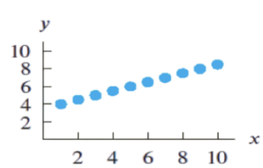
(α)



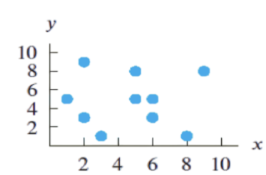
(β)



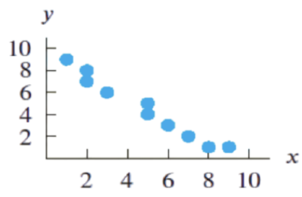
(γ)



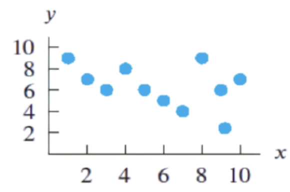
(δ)



(ε)



(στ)



Λύση: (Ασκ: 3/207)

$(\alpha) \;\rightarrow\; r = 0{,}7$

$(\beta) \;\rightarrow\; r = -0{,}5$

$(\gamma) \;\rightarrow\; r = 1$

$(\delta) \;\rightarrow\; r = 0$

$(\varepsilon) \;\rightarrow\; r = -0{,}98$

$(\sigma\tau) \;\rightarrow\; r = -0{,}25$

14. Σε ένα διαγώνισμα, η μέση τιμή της βαθμολογίας για το τμήμα B_1 ήταν 13,5 και η τυπική απόκλιση 1,4.

i. Να υπολογίσετε τον συντελεστή μεταβλητότητας.

ii. Να συγκρίνετε την ομοιογένεια των βαθμών του τμήματος B_2 που έχει μέσο όρο 16,5 και τυπική απόκλιση 3,2.

Λύση: (Ασκ: 4/207)

Ο συντελεστής μεταβλητότητας είναι $CV = \dfrac{s}{|\bar{x}|}$.

i. Για το B_1: $CV_{B_1} = \dfrac{1{,}4}{13{,}5} = 0{,}1037 \; (= 10{,}37\%)$.

ii. Για το B_2: $CV_{B_2} = \dfrac{3{,}2}{16{,}5} = 0{,}1939 \; (= 19{,}39\%)$.

Επειδή $CV_{B_1} < CV_{B_2}$, το τμήμα B_1 παρουσιάζει μεγαλύτερη ομοιογένεια από το B_2.

15. Μια εταιρεία εξετάζει τη διάρκεια ζωής δύο ειδών μπαταριών A και B. Παίρνει τυχαία 7 μπαταρίες από το κάθε είδος και καταγράφει τις ώρες λειτουργίας τους (σε χιλιάδες):

Είδος A	22	20	22	26	24	22	18
Είδος B	24	26	32	24	19	23	20

i. Να υπολογίσετε τη μέση τιμή και την τυπική απόκλιση για τη διάρκεια ζωής των μπαταριών κάθε είδους.

ii. Να βρείτε ποιο από τα δύο είδη παρουσιάζει μεγαλύτερη ομοιογένεια ως προς τη διάρκεια ζωής του.

Λύση: (Ασκ: 5/207)

Θεωρούμε *τυπική απόκλιση δείγματος*:

$$s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}, \qquad n = 7.$$

i. Για το είδος A: $\sum x_i = 154 \Rightarrow \bar{x}_A = 22$, $\sum (x_i - \bar{x}_A)^2 = 40$.

$$s_A = \sqrt{\frac{40}{6}} = 2{,}58.$$

Για το είδος B: $\sum x_i = 168 \Rightarrow \bar{x}_B = 24$, $\sum (x_i - \bar{x}_B)^2 = 110$.

$$s_B = \sqrt{\frac{110}{6}} = 4{,}28.$$

ii. Εφόσον μικρότερο s (ή μικρότερο CV) σημαίνει μεγαλύτερη ομοιογένεια, και $s_A < s_B$, το *είδος* A έχει περισσότερη ομοιογένεια.

16. Σε ένα τμήμα Α' Λυκείου στο μάθημα των Μαθηματικών έγιναν 5 διαγωνίσματα. Ο καθηγητής έδωσε στους μαθητές δύο τρόπους υπολογισμού της συνολικής γραπτής βαθμολογίας και ζήτησε να επιλέξουν τον τρόπο που θα ήθελαν: ο πρώτος τρόπος είναι να μη ληφθεί υπόψη ο μικρότερος από τους βαθμούς στα 5 διαγωνίσματα και να υπολογιστεί ο μέσος όρος των υπολοίπων 4 βαθμών, ενώ ο δεύτερος τρόπος είναι να ληφθούν όλα τα διαγωνίσματα υπόψη και να υπολογιστεί η διάμεσος των βαθμών και των 5 διαγωνισμάτων. Αν οι βαθμοί του Σόλωνα στα διαγωνίσματα είναι 8, 15, 18, 18, 19, να βρείτε ποιον από τους δύο τρόπους πρέπει να επιλέξει, για να υπολογιστεί η καλύτερη γραπτή βαθμολογία του.

Λύση: (Ασκ: 6/208)

Ταξινόμηση βαθμών: 8, 15, 18, 18, 19.

i. Μέσος όρος χωρίς τον μικρότερο: $\bar{x}_{(4)} = \dfrac{15+18+18+19}{4} = \dfrac{70}{4} = 17{,}5.$

ii. Διάμεσος όλων των 5 βαθμών: μεσαίος όρος $=$ 18.

iii. Σύγκριση: $18 > 17{,}5 \Rightarrow$ συμφέρει ο 2ος τρόπος (διάμεσος).

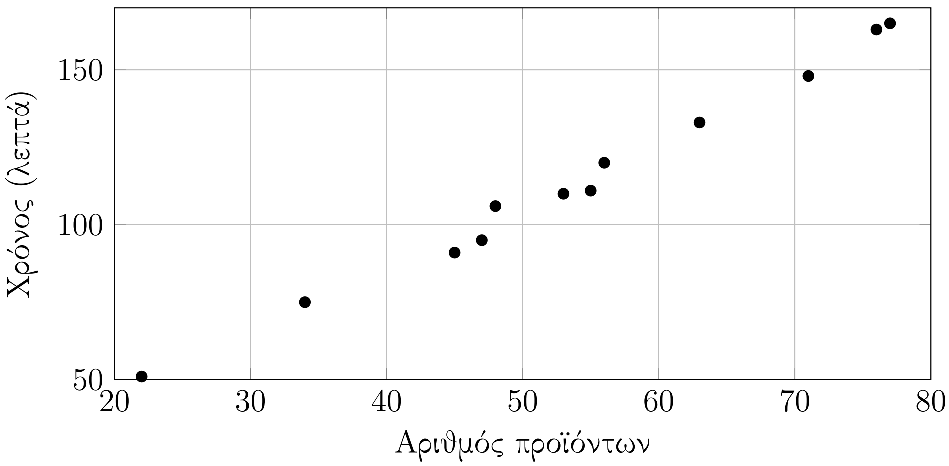
17. Στον πιο κάτω πίνακα παρουσιάζεται ο αριθμός των προϊόντων που πώλησε μια εταιρεία για 12 εβδομάδες και ο χρόνος που χρειάστηκε για να κατασκευαστούν (σε λεπτά).

Αριθμός προϊόντων	Χρόνος (λεπτά)
22	51
48	106
45	91
77	165
71	148
63	133
34	75
56	120
47	95
53	110
55	111
76	163

i. Να κατασκευάσετε το διάγραμμα διασποράς.
ii. Να υπολογίσετε τον συντελεστή συσχέτισης.
iii. Να χαρακτηρίσετε τη συσχέτιση του αριθμού των προϊόντων με τον χρόνο που χρειάστηκε για να κατασκευαστούν.

Λύση: (Ασκ: 7/208)

i. Διάγραμμα διασποράς:



ii. Συντελεστής συσχέτισης (Pearson):

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{\left(n\sum x^2 - (\sum x)^2\right)\left(n\sum y^2 - (\sum y)^2\right)}}.$$

Με $n = 12$, $\sum x = 647$, $\sum y = 1368$, $\sum x^2 = 37863$, $\sum y^2 = 168976$, $\sum xy = 79955$,

$$r = \frac{12 \cdot 79955 - 647 \cdot 1368}{\sqrt{(12 \cdot 37863 - 647^2)(12 \cdot 168976 - 1368^2)}} = \frac{74364}{74745{,}081} = 0{,}995.$$

iii. Η συσχέτιση είναι πολύ ισχυρή θετική γραμμική (σχεδόν τέλεια).

18. Στον πιο κάτω πίνακα παρουσιάζεται η ηλικία και τα επίπεδα γλυκόζης 6 ασθενών.

Ηλικία	Επίπεδα γλυκόζης
43	99
21	65
25	79
42	75
57	87
59	81

i. Να υπολογίσετε τον συντελεστή συσχέτισης.
ii. Να χαρακτηρίσετε τη συσχέτιση της ηλικίας με το επίπεδο γλυκόζης.

Λύση: (Ασκ: 8/208)

i. Χρησιμοποιούμε τον τύπο

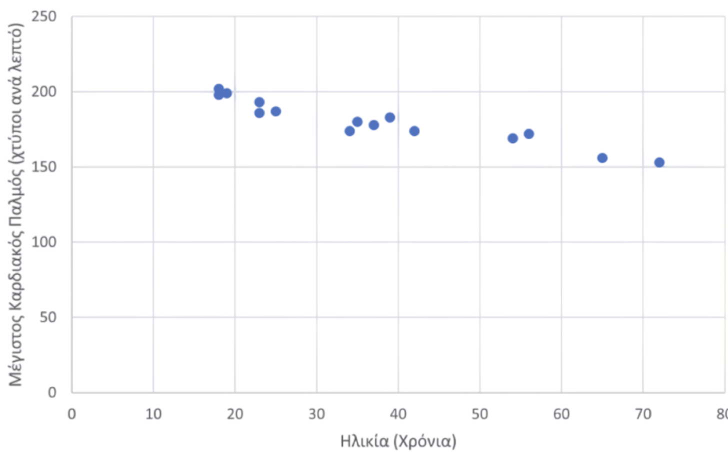
$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{\left(n\sum x^2 - (\sum x)^2\right)\left(n\sum y^2 - (\sum y)^2\right)}}.$$

Με $n = 6$, $\sum x = 247$, $\sum y = 486$, $\sum x^2 = 11409$, $\sum y^2 = 40022$, $\sum xy = 20485$ προκύπτει

$$r = \frac{6 \cdot 20485 - 247 \cdot 486}{\sqrt{(6 \cdot 11409 - 247^2)\,(6 \cdot 40022 - 486^2)}} \approx 0{,}530.$$

ii. Επομένως η σχέση ηλικίας–γλυκόζης είναι μέτρια θετική γραμμική συσχέτιση.

19. Θέλουμε να διερευνήσουμε αν υπάρχει συσχέτιση μεταξύ του Μέγιστου Καρδιακού Παλμού (MHR) και της ηλικίας στους ενήλικες. Δίνονται τα δεδομένα για 15 τυχαία επιλεγμένους ενήλικες:



Ηλικία	MHR	Ηλικία	MHR	Ηλικία	MHR
18	202	34	174	18	198
23	186	56	172	39	183
25	187	72	153	37	178
35	180	19	199		
65	156	23	193		
54	169	42	174		

i. Να χαρακτηρίσετε τη συσχέτιση του MHR ενός ενήλικα με την ηλικία του.
ii. Τι συμπέρασμα μπορείτε να εξάγετε;

Λύση: (Ασκ: 9/209)

i. Υπολογίζουμε τον συντελεστή συσχέτισης

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{\left(n\sum x^2 - (\sum x)^2\right)\left(n\sum y^2 - (\sum y)^2\right)}}.$$

Με $n = 15$, $\sum x = 560$, $\sum y = 2704$, $\sum x^2 = 25188$, $\sum y^2 = 490438$, $\sum xy = 97534$,

$$r = \frac{15 \cdot 97534 - 560 \cdot 2704}{\sqrt{(15 \cdot 25188 - 560^2)\,(15 \cdot 490438 - 2704^2)}} = -0{,}953.$$

Άρα υπάρχει *πολύ ισχυρή αρνητική γραμμική συσχέτιση.*

ii. Από τα δεδομένα προκύπτει ότι όσο αυξάνει η ηλικία, τόσο μειώνεται ο μέγιστος καρδιακός παλμός. Η ευθεία παλινδρόμησης εκτιμάται (με ελάχιστα τετράγωνα) ως

$$\widehat{\text{MHR}} \approx 210{,}05 - 0{,}80 \cdot \text{Ηλικία},$$

δηλαδή ο MHR μειώνεται περίπου κατά 0,8 παλμούς ανά έτος.

20. Ο πιο κάτω πίνακας παρουσιάζει τους βαθμούς του Β' Τετραμήνου και τους βαθμούς στις Παγκύπριες εξετάσεις των 20 μαθητών του τμήματος Γ_1 στα Μαθηματικά Κατεύθυνσης.

Β' Τ.	14	19	19	19	17	20	18	19	17	16	14	12	13	20	18	14	20	19	12	13
Παγκ.	9	16	17	12	6	17	17	16	12	14	7	3	11	18	14	12	16	16	6	10

i. Να επιλέξετε κατάλληλες στατιστικές μεθόδους (μέτρα θέσης – διασποράς, γραφικά διαγράμματα) για να συγκρίνετε τους δύο πίνακες βαθμών.

ii. Τι συμπεράσματα μπορείτε να εξάγετε;

iii. Υπάρχει γραμμική συσχέτιση για τις δύο μεταβλητές;

Λύση: (Ασκ: 1/210)

Τύποι:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad Me = \text{διάμεσος}, \qquad Q_1 = \frac{n+1}{4}\text{-οστός όρος}, \quad Q_3 = \frac{3(n+1)}{4}\text{-οστός όρος},$$

$$IQR = Q_3 - Q_1,$$
$$s = \sqrt{\frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^{\,2}}{n-1}}, \qquad CV = \frac{s}{|\bar{x}|},$$
$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{\Big(n\sum x^2 - (\sum x)^2\Big)\Big(n\sum y^2 - (\sum y)^2\Big)}}.$$

Υπολογισμοί για Β' Τετράμηνο:

$$\sum x = 333, \quad \bar{x}_{\mathrm{B}} = \frac{333}{20} = 16{,}65.$$
$$\sum x^2 = 5701, \quad s_{\mathrm{B}} = \sqrt{\frac{5701 - 20\cdot(16{,}65)^2}{19}} = \sqrt{\frac{5701 - 5544{,}45}{19}} = \sqrt{8{,}24} \approx 2{,}87.$$

Ταξινόμηση δίνει:

$$Me_{\mathrm{B}} = \frac{17+18}{2} = 17{,}5, \quad Q_{1,\mathrm{B}} = 14, \quad Q_{3,\mathrm{B}} = 19, \quad IQR_{\mathrm{B}} = 5.$$
$$CV_{\mathrm{B}} = \frac{2{,}87}{16{,}65} \approx 0{,}172.$$

Υπολογισμοί για Παγκύπριες:

$$\sum y = 249, \quad \bar{y}_{\Pi} = \frac{249}{20} = 12{,}45.$$
$$\sum y^2 = 3471, \quad s_{\Pi} = \sqrt{\frac{3471 - 20\cdot(12{,}45)^2}{19}} = \sqrt{\frac{3471 - 3098{,}05}{19}} = \sqrt{19{,}63} \approx 4{,}42.$$

Ταξινόμηση δίνει:

$$Me_{\Pi} = \frac{12+13}{2} = 13, \quad Q_{1,\Pi} = 9{,}5, \quad Q_{3,\Pi} = 16, \quad IQR_{\Pi} = 6{,}5.$$
$$CV_{\Pi} = \frac{4{,}42}{12{,}45} \approx 0{,}355.$$

Συσχέτιση:

$$n = 20, \ \sum x = 333, \ \sum y = 249, \ \sum x^2 = 5701, \ \sum y^2 = 3471, \ \sum xy = 4344.$$
$$r = \frac{20\cdot 4344 - (333)(249)}{\sqrt{(20\cdot 5701 - 333^2)(20\cdot 3471 - 249^2)}} = \frac{86880 - 82917}{\sqrt{(114020 - 110889)(69420 - 62001)}}.$$

$$r = \frac{3963}{\sqrt{3131 \cdot 7419}} = \frac{3963}{4819{,}6} \approx 0{,}822.$$

ii. Οι βαθμοί του Β' Τετραμήνου έχουν μεγαλύτερο μέσο όρο και μικρότερη διασπορά, άρα είναι πιο ομοιογενείς.
iii. Υπάρχει ισχυρή θετική γραμμική συσχέτιση ($r \approx 0{,}822$).

21. Στον διπλανό πίνακας παρουσιάζονται οι βαθμολογίες των Παγκύπριων Εξετάσεων στα Νέα Ελληνικά των μαθητών των τμημάτων $\Gamma_1, \Gamma_2, \Gamma_3$ ενός σχολείου.

i. Ποιο τμήμα έχει τον πιο ψηλό μέσο όρο;
ii. Ποιες οι διαφορές των μέσων όρων και ποιες των διασπορών των βαθμών;
iii. Υπάρχουν διαφορές ως προς το ενδοτεταρτημοριακό εύρος των βαθμών των τριών τμημάτων;
iv. Να επιλέξετε άλλες κατάλληλες στατιστικές μεθόδους για να συγκρίνετε τους βαθμούς των τριών τμημάτων. Να διατυπώσετε τα συμπεράσματά σας.

Γ_1	Γ_2	Γ_3
13	10	16
18	12	4
12	15	11
15	12	6
11	14	7
13	14	16
17	13	12
16	12	13
17	7	12
16	14	8
12	14	12
9	14	8
8	11	11
18	13	8
16	11	12
14	10	12
17	18	8
17	15	7
11	11	12
12	11	9
16	14	10
13	16	10
15	14	10

Λύση: (Ασκ: 2/210)

Τύποι που θα χρησιμοποιηθούν

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad s = \sqrt{\frac{\sum x_i^2 - n\bar{x}^2}{n-1}}, \qquad CV = \frac{s}{|\bar{x}|},$$

$$Me = \tfrac{n+1}{2}), \qquad Q_1 = \tfrac{n+1}{4}, \qquad Q_3 = \tfrac{3(n+1)}{4}, \qquad IQR = Q_3 - Q_1.$$

Δεδομένα και αθροίσματα

$$\Gamma_1: \ n = 23, \ \sum x = 326, \ \sum x^2 = 4800 \quad \Rightarrow \quad \bar{x}_1 = \frac{326}{23} = 14{,}1739,$$

$$s_1 = \sqrt{\frac{4800 - 23\cdot(14{,}1739)^2}{22}} = \sqrt{\frac{4800 - 4635{,}65}{22}} = \sqrt{7{,}120} \approx 2{,}855, \quad CV_1 = \frac{2{,}855}{14{,}174} \approx 0{,}201.$$

Ταξινομημένα Γ_1 : $8, 9, 11, 11, 12, 12, 12, 13, 13, 13, 14, 15, 15, 16, 16, 16, 16, 17, 17, 17, 17, 18, 18.$

$$Me_1 = 12\text{η τιμή} = 15, \quad Q_{1,1} = 6\text{η} = 12, \quad Q_{3,1} = 18\text{η} = 17, \quad IQR_1 = 17 - 12 = 5.$$

$$\Gamma_2: \ n = 23, \ \sum x = 295, \ \sum x^2 = 3905 \quad \Rightarrow \quad \bar{x}_2 = \frac{295}{23} = 12{,}8261,$$

$$s_2 = \sqrt{\frac{3905 - 23\cdot(12{,}8261)^2}{22}} = \sqrt{\frac{3905 - 3791{,}30}{22}} = \sqrt{3{,}186} \approx 2{,}348, \quad CV_2 = \frac{2{,}348}{12{,}826} \approx 0{,}183.$$

Ταξινομημένα Γ_2 : $7, 10, 10, 11, 11, 11, 11, 12, 12, 12, 13, 13, 14, 14, 14, 14, 14, 14, 14, 15, 15, 16, 18.$

$$Me_2 = 12\text{η} = 13, \quad Q_{1,2} = 6\text{η} = 11, \quad Q_{3,2} = 18\text{η} = 14, \quad IQR_2 = 14 - 11 = 3.$$

$$\Gamma_3: \ n = 23, \ \sum x = 234, \ \sum x^2 = 2574 \quad \Rightarrow \quad \bar{x}_3 = \frac{234}{23} = 10{,}1739,$$

$$s_3 = \sqrt{\frac{2574 - 23\cdot(10{,}1739)^2}{22}} = \sqrt{\frac{2574 - 2380{,}60}{22}} = \sqrt{9{,}781} \approx 2{,}964, \quad CV_3 = \frac{2{,}964}{10{,}174} \approx 0{,}291.$$

Ταξινομημένα Γ_3 : $4, 6, 7, 7, 8, 8, 8, 8, 9, 10, 10, 10, 11, 11, 12, 12, 12, 12, 12, 12, 13, 16, 16.$

$$Me_3 = 12\text{η} = 10, \quad Q_{1,3} = 6\text{η} = 8, \quad Q_{3,3} = 18\text{η} = 12, \quad IQR_3 = 12 - 8 = 4.$$

i. Γ_1 έχει τον υψηλότερο μέσο όρο ($\bar{x}_1 \approx 14{,}17$), έπειτα Γ_2 και τελευταίο Γ_3.

ii. *Διαφορές μέσων:*

$$\bar{x}_1 - \bar{x}_2 \approx 1{,}348, \qquad \bar{x}_1 - \bar{x}_3 \approx 3{,}999, \qquad \bar{x}_2 - \bar{x}_3 \approx 2{,}652.$$

Διασπορές (δειγματικές):

$$s_1^2 \approx 8{,}150, \quad s_2^2 \approx 5{,}514, \quad s_3^2 \approx 8{,}787.$$

Το Γ_2 έχει τη μικρότερη διασπορά, το Γ_3 τη μεγαλύτερη.

iii. Ενδοτεταρτημοριακά εύρη: $IQR_1 = 5,\ IQR_2 = 3,\ IQR_3 = 4$. Υπάρχουν διαφορές: το Γ_2 είναι πιο «συμπαγές» στο μεσαίο 50% των τιμών.

iv. *Πρόσθετες συγκρίσεις:* χρησιμοποιούμε *boxplots* για τα τρία τμήματα, ιστόγραμμα/πολυγωνικές συχνότητας, καθώς και τον CV για σχετική μεταβλητότητα:

$$CV_1 \approx 0{,}201, \qquad CV_2 \approx 0{,}183, \qquad CV_3 \approx 0{,}291.$$

Συμπέρασμα: το Γ_1 παρουσιάζει υψηλότερες επιδόσεις κατά μέσο όρο, το Γ_2 τη μεγαλύτερη ομοιογένεια, ενώ το Γ_3 χαμηλότερες επιδόσεις και τη μεγαλύτερη σχετική διασπορά.

22. Στον πιο κάτω πίνακα παρουσιάζονται στην πρώτη γραμμή οι τιμές σε διαφορετικά κράνη ποδηλασίας και στη δεύτερη γραμμή η βαθμολογία ποιότητάς τους που έγινε από ειδικούς.

Τιμή (€)	35	22	33	42	50	23	29	18	39	28	20	25
Βαθμολογία	64	60	58	55	54	45	47	43	42	41	40	32

i. Υπάρχει γραμμική συσχέτιση ανάμεσα στην τιμή και τη βαθμολογία ποιότητας;
ii. Να εξετάσετε κατά πόσο ισχύει ο ισχυρισμός ότι «αν αγοράσουμε πιο ακριβό κράνος θα έχει πιο υψηλή ποιότητα».

Λύση: (Ασκ: 3/211)

Θέτουμε x=τιμή, y=βαθμολογία, $n = 12$. Υπολογίζουμε τα αθροίσματα:

$$\sum x = 364, \quad \sum y = 581, \quad \sum x^2 = 12086, \quad \sum y^2 = 29153, \quad \sum xy = 18042.$$

i. Συντελεστής συσχέτισης (Pearson).

$$r = \frac{n\sum xy - \sum x \sum y}{\sqrt{\left(n\sum x^2 - (\sum x)^2\right)\left(n\sum y^2 - (\sum y)^2\right)}}.$$

Αριθμητής:

$$n\sum xy - \sum x \sum y = 12 \cdot 18042 - 364 \cdot 581 = 216\,504 - 211\,484 = 5\,020.$$

Παρονομαστής:

$$\sqrt{(12 \cdot 12086 - 364^2)\,(12 \cdot 29153 - 581^2)} = \sqrt{(145\,032 - 132\,496)\,(349\,836 - 337\,561)}$$

$$= \sqrt{12\,536 \cdot 12\,275} = \sqrt{153\,879\,400} \approx 12\,404{,}81.$$

Άρα

$$r = \frac{5\,020}{12\,404{,}81} = 0{,}405.$$

Συνεπώς υπάρχει *ασθενής έως μέτρια θετική γραμμική συσχέτιση* μεταξύ τιμής και ποιότητας.

ii. Ο συντελεστής προσδιορισμού είναι

$$R^2 = r^2 \approx (0{,}405)^2 \approx 0{,}164,$$

δηλαδή περίπου το 16,4% της μεταβλητότητας της βαθμολογίας εξηγείται γραμμικά από την τιμή. Ο ισχυρισμός «πιο ακριβό ⇒ πιο ποιοτικό» *υποστηρίζεται μόνο μέτρια* και δεν είναι καθολικός. (Υπάρχουν αντίπαραδείγματα, π.χ. $x = 22$ με $y = 60$ έναντι $x = 39$ με $y = 42$.)

(Προαιρετικά, η ευθεία ελάχιστων τετραγώνων είναι $\widehat{y} = 36{,}27 + 0{,}400\,x$: αύξηση τιμής κατά 1 συνδέεται, κατά μέσο όρο, με αύξηση ποιότητας κατά $\approx 0{,}40$ μονάδες.)

23. Στον πιο κάτω πίνακα παρουσιάζονται οι γραμμικές συσχετίσεις των γραπτών βαθμολογιών στις εξετάσεις Ιουνίου στα 5 εξεταζόμενα μαθήματα του τμήματος A_4.

	Μαθηματικά	Βιολογία	Νέα Ελληνικά	Φυσική	Χημεία
Μαθηματικά	1,00	0,54	0,76	0,70	0,41
Βιολογία	0,54	1,00	0,81	0,73	0,80
Νέα Ελληνικά	0,76	0,81	1,00	0,71	0,67
Φυσική	0,70	0,73	0,71	1,00	0,66
Χημεία	0,41	0,80	0,67	0,66	1,00

Να εξετάσετε κατά πόσο υπάρχει, ισχυρή ή όχι, γραμμική συσχέτιση ανάμεσα στις βαθμολογίες στα 5 εξεταζόμενα μαθήματα των μαθητών αυτών.

Λύση: (Ασκ: 4/211)

Ο συντελεστής γραμμικής συσχέτισης (Pearson) δύο μεταβλητών X και Y ορίζεται από τον τύπο:

$$r_{XY} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{x})^2}\,\sqrt{\sum_{i=1}^{n}(y_i - \overline{y})^2}}$$

ή ισοδύναμα:

$$r_{XY} = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{\left(n\sum x_i^2 - (\sum x_i)^2\right)\left(n\sum y_i^2 - (\sum y_i)^2\right)}}.$$

Οι τιμές r είναι ήδη υπολογισμένες και δίνονται στον πίνακα. Για την ερμηνεία τους χρησιμοποιούμε:

$$0 \le r < 0{,}50 \;\Rightarrow\; \text{ασθενής}, \qquad 0{,}50 \le r < 0{,}70 \;\Rightarrow\; \text{μέτρια}, \qquad 0{,}70 \le r \le 1 \;\Rightarrow\; \text{ισχυρή}.$$

Ισχυρές θετικές συσχετίσεις:

Μαθηματικά–Νέα Ελληνικά (0,76), Μαθηματικά–Φυσική (0,70),
Βιολογία–Νέα Ελληνικά (0,81), Βιολογία–Φυσική (0,73),
Βιολογία–Χημεία (0,80), Νέα Ελληνικά–Φυσική (0,71).

Μέτριες συσχετίσεις:

Μαθηματικά–Βιολογία (0,54), Νέα Ελληνικά–Χημεία (0,67), Φυσική–Χημεία (0,66).

Ασθενής συσχέτιση:

Μαθηματικά–Χημεία (0,41).