

## 2.7. Γραμμική συσχέτιση ποσοτικών μεταβλητών και διαγράμματα διασποράς

### Διδακτικοί στόχοι

- Πώς σχετίζονται γραμμικά δύο ποσοτικές μεταβλητές
- Τι είναι το διάγραμμα διασποράς και που χρησιμοποιείται
- Τι είναι ο συντελεστής γραμμικής συσχέτισης δύο ποσοτικών μεταβλητών
- Πώς χαράζουμε μια ευθεία στο διάγραμμα διασποράς

### Θεωρία και Μεθοδολογία

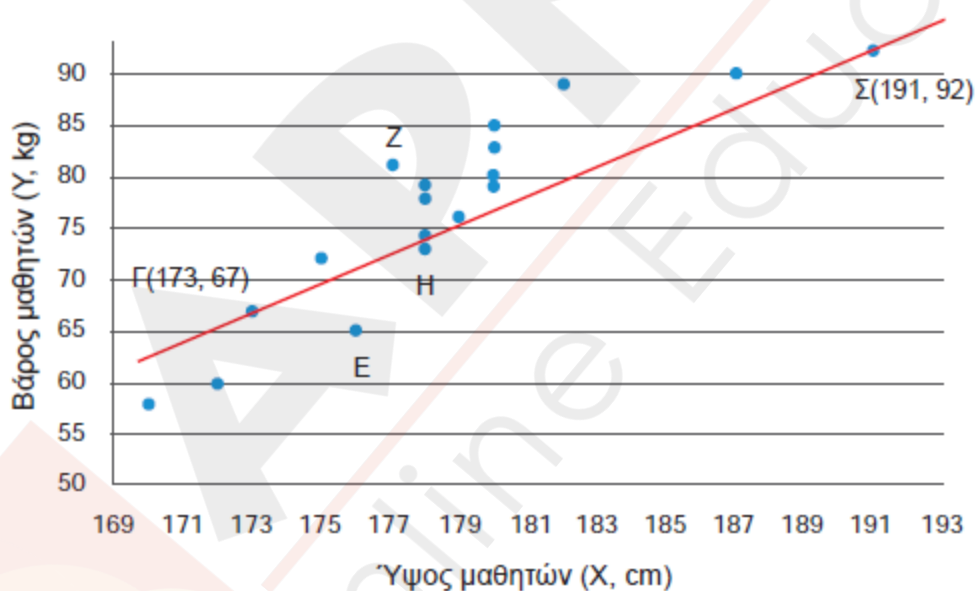
#### ➤ Διάγραμμα διασποράς

Είναι η γραφική παράσταση των σημείων σε ένα ορθογώνιο σύστημα αξόνων, που αποτελείται από τα ζεύγη τιμών  $(x_i, y_i)$  των συνεχών μεταβλητών  $X$  και  $Y$ .

**Εφαρμογή 1 :** Καταγράφουμε το ύψος ( $X$  σε cm) και το βάρος ( $Y$  σε kg) 18 μαθητών της Β' Λυκείου. Θα δούμε τι συμβαίνει με το βάρος των παιδιών όταν αλλάζει το ύψος τους ή τι συμβαίνει με το ύψος των παιδιών όταν αλλάζει το βάρος τους. Αν μας ενδιαφέρει η πρώτη περίπτωση, θεωρούμε ως ανεξάρτητη μεταβλητή το ύψος των μαθητών και ως εξαρτημένη μεταβλητή το βάρος των μαθητών. Στη δεύτερη περίπτωση συμβαίνει το αντίθετο.

*Απλά και Κατανοητά η Γνώση!*

Μαθητής	Ύψος (X)	Βάρος (Y)	Μαθητής	Ύψος (X)	Βάρος (Y)
A	170	58	Κ	178	79
B	172	60	Λ	179	76
Γ	173	67	Μ	180	79
Δ	175	72	Ν	180	80
Ε	176	65	Ξ	180	83
Z	177	81	Ο	180	85
Η	178	73	Π	182	89
Θ	178	74	Ρ	187	90
Ι	178	78	Σ	191	92



Διάγραμμα Διασποράς Ύψους και Βάρους 18 μαθητών της Β' λυκείου και ευθεία προσαρμοσμένη «με το μάτι» για τα δεδομένα του πίνακα 1

Στο διάγραμμα παρατηρούμε μια διασπορά των σημείων που αντιστοιχούν στους μαθητές που εξετάζουμε. Από το διάγραμμα διασποράς μπορούμε να εντοπίσουμε την ύπαρξη συσχέτισης που ενδεχομένως να υπάρχει μεταξύ των μεταβλητών που εξετάζουμε. Η γενική

*Απλά και Κατανοητά η Γνώση!*

εικόνα που παίρνουμε από το διάγραμμα διασποράς είναι ότι οι ψηλότεροι μαθητές είναι συνήθως και βαρύτεροι. Για παράδειγμα, ο μαθητής Z είναι ψηλότερος και βαρύτερος από τον μαθητή E, ο Η είναι ψηλότερος και βαρύτερος από τον E, αλλά υπάρχουν και εξαιρέσεις, όπως στο ότι ο Η είναι είναι ψηλότερος από τον Z αλλά ο Z είναι βαρύτερος από τον Η. Το άπλωμα των σημείων στο διάγραμμα θα μπορούσαμε να πούμε ότι γίνεται κοντά σε μια νοητή ευθεία γραμμή, οπότε η σχέση θεωρείται γραμμική.

➤ **Συντελεστής γραμμικής συσχέτισης δύο ποσοτικών μεταβλητών**

Είναι ένα μέτρο που μας δίνει το μέγεθος της γραμμικής σχέσης των δύο ποσοτικών μεταβλητών X και Y. Ορίζεται με βάση ένα δείγμα n ζευγών παρατηρήσεων  $(x_i, y_i)$ ,  $i = 1, 2, \dots, n$ , συμβολίζεται με  $r(X, Y)$  ή απλά r και δίνεται από τον τύπο:

$$r = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{n s_x s_y}$$

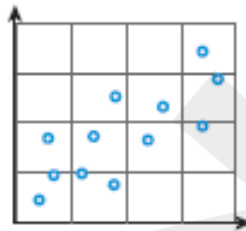
όπου  $\bar{x}$ ,  $s_x$  η δειγματική μέση τιμή και δειγματική τυπική απόκλιση της μεταβλητής X και  $\bar{y}$ ,  $s_y$  η δειγματική μέση τιμή και δειγματική τυπική απόκλιση της μεταβλητής Y.

Ο συντελεστής γραμμικής συσχέτισης δύο ποσοτικών μεταβλητών X και Y είναι καθαρός αριθμός, δηλαδή δεν εκφράζεται σε συγκεκριμένες μονάδες μέτρησης, επομένως είναι ανεξάρτητος των χρησιμοποιούμενων μονάδων μέτρησης των μεταβλητών X και Y. Ισχύει πάντοτε:  $-1 \leq r \leq 1$

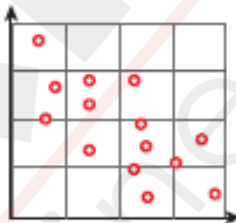
*Απλά και Κατανοητά η Γνώση!*

**Πιο συγκεκριμένα, όταν:**

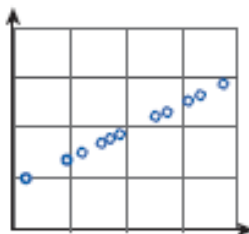
- $0 < r < 1$ , τότε οι  $X$  και  $Y$  είναι θετικά γραμμικά συσχετισμένες. Αυτό σημαίνει ότι, όταν οι τιμές της μεταβλητής  $X$  αυξάνονται, οι τιμές της  $Y$  τείνουν να αυξάνονται. Όσο ο συντελεστής πλησιάζει το 1 τόσο πιο συγκεντρωμένες γύρω από μια νοητή ευθεία είναι οι παρατηρήσεις. Επομένως, τόσο πιο ισχυρή είναι η θετική γραμμική συσχέτιση.



- $-1 < r < 0$ , τότε οι  $X$  και  $Y$  είναι αρνητικά γραμμικά συσχετισμένες. Αυτό σημαίνει ότι, όταν οι τιμές της μεταβλητής  $X$  αυξάνονται, οι τιμές της  $Y$  τείνουν να μειώνονται. Όσο ο συντελεστής πλησιάζει το -1 τόσο πιο συγκεντρωμένες γύρω από μια νοητή ευθεία είναι οι παρατηρήσεις. Επομένως, τόσο πιο ισχυρή είναι η αρνητική γραμμική συσχέτιση.

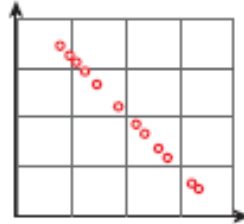


- $r = 1$ , τότε οι  $X$  και  $Y$  είναι τέλεια θετικά γραμμικά συσχετισμένες και όλα τα σημεία βρίσκονται πάνω σε μια ευθεία με θετική κλίση.

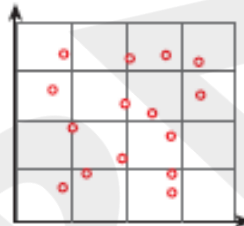


*Απλά και Κατανοητά η Γνώση!*

- $r = -1$ , τότε οι  $X$  και  $Y$  είναι τέλεια αρνητικά γραμμικά συσχετισμένες και όλα τα σημεία βρίσκονται πάνω σε μια ευθεία με αρνητική κλίση.



- $r = 0$ , τότε οι  $X$  και  $Y$  είναι γραμμικά ασυσχέτιστες, χωρίς αυτό να σημαίνει ότι δεν μπορεί να έχουν κάποια άλλη σχέση, μη γραμμική.



➤ **Χάραξη ευθείας στο διάγραμμα διασποράς**

Είναι η ευθεία  $y = \alpha + \beta x$ , όπου όπου  $\alpha$  και  $\beta$  είναι παράμετροι τις οποίες θέλουμε να υπολογίσουμε ή, όπως λέμε, «να εκτιμήσουμε», έτσι ώστε η ευθεία που θα προκύψει από τη να περιγράφει την αναμενόμενη τιμή  $y$  της μεταβλητής  $Y$ , όταν η τιμή της μεταβλητής  $X$  είναι ίση με  $x$ .

Η παράμετρος  $\alpha$  μας δίνει το σημείο  $(0, \alpha)$  όπου η ευθεία αυτή τέμνει τον άξονα  $y'g$ , ενώ η παράμετρος  $\beta$  παριστάνει τον **συντελεστή διεύθυνσης** της ευθείας. Ο πιο εύκολος τρόπος χάραξης της ευθείας είναι αυτός που γίνεται «με το μάτι».

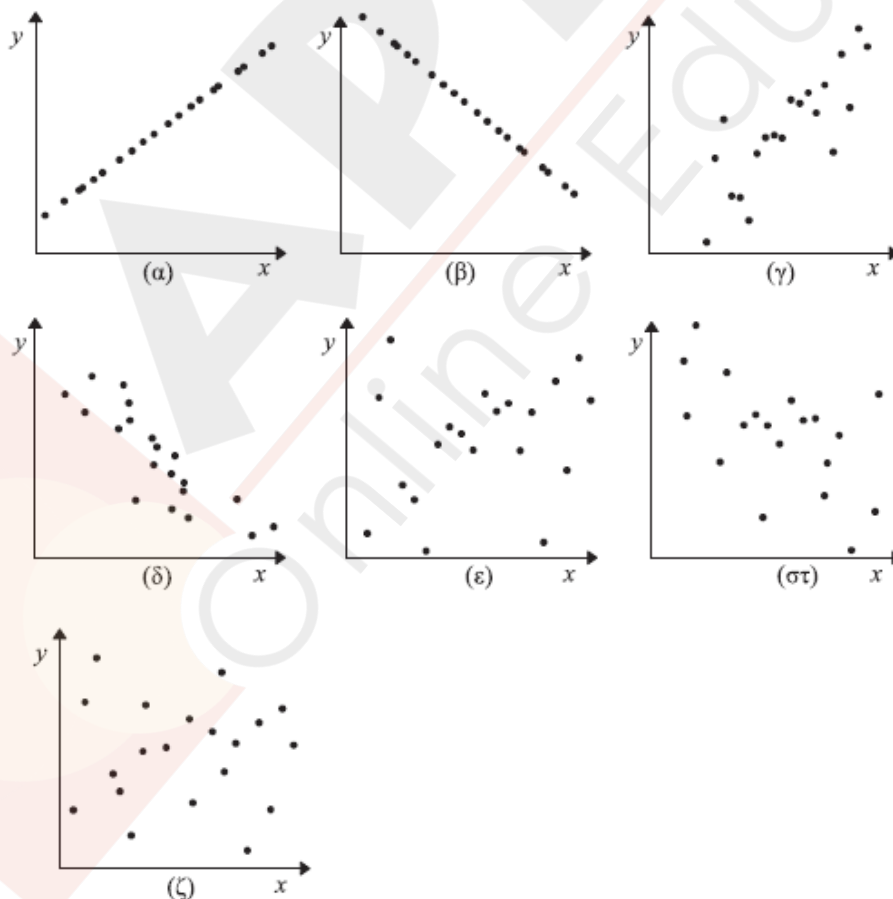
*Απλά και Κατανοητά η Γνώση!*

**Εφαρμογή 2 :** Έχοντας το διάγραμμα από την Εφαρμογή 1, για να βρούμε τα  $\alpha$  και  $\beta$  επιλέγουμε δύο σημεία, έστω τα  $\Gamma(173,67)$  και  $\Sigma(191,92)$  πάνω στην ευθεία που φέραμε «με το μάτι» και καταλήγουμε στην εξίσωση:  $y = -173,28 + 1,39x$ . Επομένως, η ευθεία που προσαρμόζεται καλύτερα στα σημεία του διαγράμματος διασποράς διέρχεται από το σημείο  $(0, -173,28)$  και έχει συντελεστή διεύθυνσης 1,39.

### Λυμένες Ασκήσεις

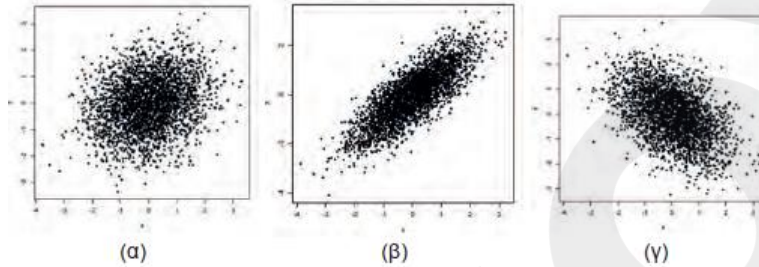
#### Άσκηση 1 :

A) Για δύο ποσοτικές μεταβλητές  $X$  και  $Y$  να αντιστοιχίσετε τους συντελεστές γραμμικής συσχέτισης  $r_1 = -1, r_2 = -0,8, r_3 = -0,2, r_4 = 0, r_5 = 0,2, r_6 = 0,8$  και  $r_7 = 1$ , με τα αντίστοιχα διαγράμματα διασποράς.



Απλά και Κατανοητά η Γνώση!

Β) Για δύο ποσοτικές μεταβλητές  $X$  και  $Y$  να αντιστοιχίσετε τους συντελεστές γραμμικής συσχέτισης  $r_1=0,8$ ,  $r_2=0,2$  και  $r_3= -0,5$ , με τα αντίστοιχα διαγράμματα διασποράς.



**Λύση :**

Α) Σύμφωνα με τις τιμές που παίρνει ένας συντελεστής γραμμικής συσχέτισης μεταξύ δύο ποσοτικών μεταβλητών προκύπτει ότι η τιμή του  $r_1 = -1$  αντιστοιχεί στο διάγραμμα (β) , η τιμή του  $r_2 = -0,8$  στο διάγραμμα (δ), η τιμή του  $r_3 = -0,2$  στο διάγραμμα (στ), η τιμή του  $r_4 = 0$  στο διάγραμμα (ζ), η τιμή του  $r_5 = 0,2$  στο διάγραμμα (ε), η τιμή του  $r_6 = 0,8$  στο διάγραμμα (γ) και η τιμή του  $r_7 = 1$  στο διάγραμμα (α).

Β) Σε καθένα από αυτά τα διαγράμματα διασποράς βλέπουμε ένα σμήνος σημείων σε αντίθεση με τα διαγράμματα διασποράς του Α που είχαμε λίγα σημεία. Το διάγραμμα διασποράς (α) παρουσιάζει ασθενή θετική γραμμική συσχέτιση με συντελεστή γραμμικής συσχέτισης  $r_2=0,2$ . Το διάγραμμα διασποράς (β) παρουσιάζει ισχυρή θετική γραμμική συσχέτιση με συντελεστή γραμμικής συσχέτισης  $r_1=0,8$ . Τέλος, το διάγραμμα διασποράς (γ) παρουσιάζει αρνητική γραμμική συσχέτιση με συντελεστή γραμμικής συσχέτισης  $r_3 = -0,5$ .

*Απλά και Κατανοητά η Γνώση!*

**Άσκηση 2 :**

Στον Πίνακα δίνονται τα προσδόκιμα ζωής γυναικών (Π.Ζ.Γ.) και ανδρών (Π.Ζ.Α.) του έτους 2005 για 30 τυχαία επιλεγμένες χώρες.

Χώρα	Π.Ζ.Γ.	Π.Ζ.Α.	Χώρα	Π.Ζ.Γ.	Π.Ζ.Α.
Αίγυπτος	63	60	Λευκορωσία	76	66
Αυστρία	79	73	Λιθουανία	77	68
Αφγανιστάν	44	45	Μαλαισία	72	66
Βέλγιο	79	73	Μποτσουάνα	66	60
Βολιβία	64	59	Νησιά Μπαρμπάντος	78	73
Δομινικανή Δημοκρατία	70	66	Νιγηρία	57	54
Ελ Σαλβαδόρ	69	64	Νικαράγουα	67	61
Ελλάδα	80	75	Ολλανδία	81	75
Ζάμπια	45	44	Περου	67	63
Ινδία	59	58	Ρωσία	74	64
Ισημερινός	73	67	Σενεγάλη	58	55
Καμερούν	58	55	Σομαλία	55	54
Καναδάς	81	74	Τανζανία	45	41
Κίνα	69	67	Τσεχία	77	69
Κουβέιτ	78	73	Χιλή	78	71

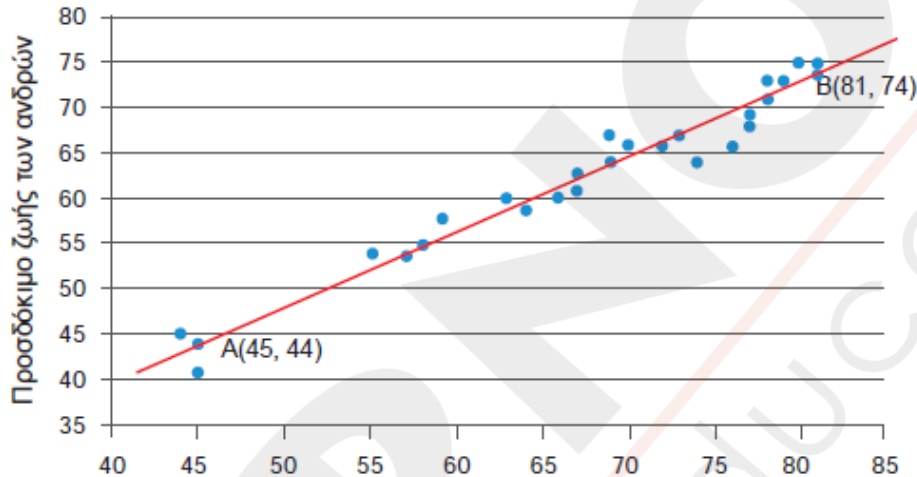
- α) Να κατασκευάσετε το διάγραμμα διασποράς μεταξύ των προσδόκιμων ζωής γυναικών και ανδρών και να περιγράψετε το είδος της σχέσης που φαίνεται να έχουν οι δύο μεταβλητές.
- β) Να υπολογίσετε και να ερμηνεύσετε τον συντελεστή γραμμικής συσχέτισης μεταξύ των προσδόκιμων ζωής γυναικών και ανδρών.
- γ) Να σχεδιάσετε «με το μάτι» την ευθεία που φαίνεται να προσαρμόζεται καλύτερα στα δεδομένα.
- δ) Αν για την Αλβανία, το προσδόκιμο ζωής των γυναικών είναι τα 75 έτη, μπορείτε να εκτιμήσετε το προσδόκιμο ζωής των ανδρών;

*Απλά και Κατανοητά η Γνώση!*



**Λύση :**

α) Θεωρούμε ως ανεξάρτητη μεταβλητή (X) το προσδόκιμο ζωής των γυναικών και ως εξαρτημένη μεταβλητή (Y) το προσδόκιμο ζωής των ανδρών για 30 χώρες. Βέβαια, στην περίπτωση μας είναι τέτοια η φύση των μεταβλητών που θα μπορούσε να είναι και ανάποδα.



Από το διάγραμμα διασποράς φαίνεται ότι τα σημεία του διαγράμματος είναι κοντά σε μια νοητή ευθεία. Παρατηρούμε ότι, καθώς αυξάνεται το προσδόκιμο ζωής των γυναικών, φαίνεται να αυξάνεται και το προσδόκιμο ζωής των ανδρών.

β) Υπολογίζουμε πρώτα τη μέση τιμή και την τυπική απόκλιση για κάθε ποσοτική μεταβλητή:

$\bar{x} = 67,97$  ,  $\bar{y} = 63,1$ ,  $s_x = 10,93$ ,  $s_y = 9,16$  και το άθροισμα  $\sum_{i=1}^{30} x_i y_i = 131611$  . Οπότε,  $r=0,98$ .

Αυτό σημαίνει ότι το προσδόκιμο ζωής των γυναικών και των ανδρών στο δείγμα μας είναι ισχυρά θετικά γραμμικά συσχετισμένα.

γ) Τα σημεία  $(x_i , y_i), i = 1, 2, \dots, 30$  είναι συγκεντρωμένα γύρω από μια ευθεία, όπως φαίνεται στο σχήμα .Επιλέγουμε δύο σημεία, έστω τα A(45,44) και B(81,74) πάνω στην ευθεία που φέραμε «με το μάτι».

*Απλά και Κατανοητά η Γνώση!*

Αντικαθιστούμε τις συντεταγμένες  $(x, y)$  των σημείων αυτών, οπότε προκύπτει το σύστημα:

$$\begin{cases} y_A = \alpha + \beta x_A \\ y_B = \alpha + \beta x_B \end{cases} \Leftrightarrow \begin{cases} 44 = \alpha + 45\beta \\ 74 = \alpha + 81\beta \end{cases}$$

Επιλύοντας το σύστημα αυτό βρίσκουμε  $\alpha = -173,28$  και  $\beta = 1,39$ , οπότε η εξίσωση της ευθείας γίνεται:  $y = 6,5 + 0,83x$

Επομένως, η ευθεία που προσαρμόζεται καλύτερα στα σημεία του διαγράμματος διασποράς διέρχεται από το σημείο  $(0, 6,5)$  και έχει συντελεστή διεύθυνσης  $0,83$ .

δ) Αν για την Αλβανία, το προσδόκιμο ζωής των γυναικών είναι τα 75 έτη, χρησιμοποιώντας την παραπάνω ευθεία μπορούμε να εκτιμήσουμε το προσδόκιμο ζωής των ανδρών της Αλβανίας. Έτσι,  $y = 6,5 + 0,83 \cdot 75 = 68,75$ , δηλαδή 69 περίπου έτη. Η θετική συσχέτιση μεταξύ Π.Ζ.Γ και Π.Ζ.Α. δεν μπορεί να αποδοθεί σε αιτιακή σχέση μεταξύ των Π.Ζ.Γ και Π.Ζ.Α., αλλά στην ύπαρξη συγχυτικών παραγόντων όπως, το επίπεδο διαβίωσης, το ΑΕΠ, την ποιότητα των υπηρεσιών υγείας κάθε χώρας, κτλ., που επηρεάζουν και τα δύο προσδόκιμα ζωής.

### Ασκήσεις για Διδασκαλία

#### Άσκηση 1 :

Να διατάξετε τις παρακάτω τιμές του  $r$  σε αύξουσα τάξη του βαθμού γραμμικής συσχέτισης δύο ποσοτικών μεταβλητών  $X$  και  $Y$ :  $-0,6, 0,9, -0,7, 0,2, 0, -1$ .

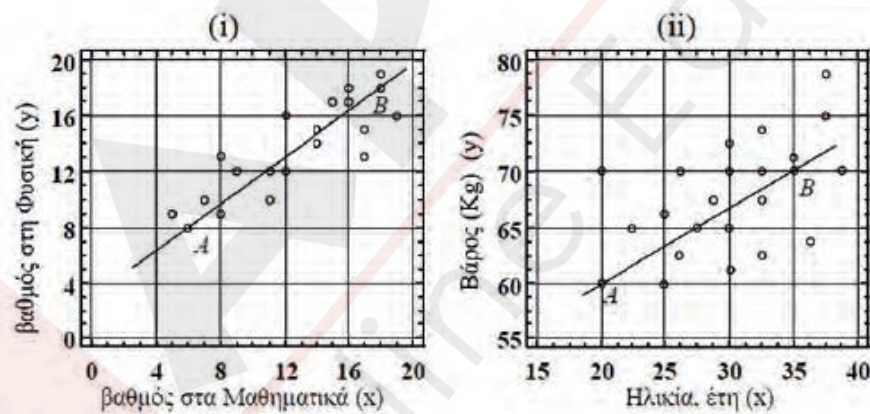
*Απλά και Κατανοητά η Γνώση!*

**Άσκηση 2 :**

Ο συντελεστής γραμμικής συσχέτισης δύο μεταβλητών  $X$  και  $Y$  είναι  $0,96$ , ενώ ο συντελεστής γραμμικής συσχέτισης δύο μεταβλητών  $Z$  και  $\Phi$  είναι  $-0,96$ . Ποια είναι η διαφορά τους;

**Άσκηση 3 :**

Δίνονται δυο διαγράμματα διασποράς με χαραγμένες «με το μάτι» δύο ευθείες από έναν μαθητή.



- α) Χρησιμοποιώντας τα σημεία A και B να βρείτε τις εξισώσεις των δύο ευθειών.
- β) Πώς θα μπορούσατε να χρησιμοποιήσετε τις ευθείες του ερωτήματος α);

*Απλά και Κατανοητά η Γνώση!*

**Άσκηση 4 :**

Μια εταιρεία διαφημίσεων παρουσίασε τον επόμενο πίνακα:

Αριθμός διαφημίσεων	Έσοδα από πωλήσεις
10	20000
18	28000
24	35000
32	44000
35	48000
37	50000
42	55000

- α) Να κατασκευάσετε το διάγραμμα διασποράς και να εκτιμήσετε από αυτό τον συντελεστή γραμμικής συσχέτισης μεταξύ του αριθμού διαφημίσεων της εταιρείας και των εσόδων της από τις πωλήσεις.
- β) Να υπολογίσετε και να ερμηνεύσετε τον συντελεστή γραμμικής συσχέτισης μεταξύ του αριθμού διαφημίσεων της εταιρείας και των εσόδων της από τις πωλήσεις.
- γ) Να σχεδιάσετε «με το μάτι» την ευθεία που φαίνεται να προσαρμόζεται καλύτερα στα δεδομένα.
- δ) Αν το πλήθος των διαφημίσεων που αγόραζε ένα κανάλι ήταν 30, πόσο εκτιμάτε ότι θα ήταν τα έσοδα της εταιρείας διαφημίσεων;
- ε) Αν το πλήθος των διαφημίσεων που αγόραζε ένα κανάλι ήταν 60, θα ήταν ασφαλές να εκτιμήσετε τα έσοδα της εταιρείας διαφημίσεων, όπως στο δ);

*Απλά και Κατανοητά η Γνώση!*

**Άσκηση 5 :**

Τα δεδομένα του επόμενου πίνακα παριστάνουν τους βαθμούς (στην κλίμακα του 100) 10 μαθητών/τριών της Β' τάξης του Γενικού Λυκείου στα μαθήματα της Φυσικής (X) και των Μαθηματικών (Y) κορμού σε μια γραπτή αξιολόγηση.

Βαθμός- X	Βαθμός- Y	Βαθμός- X	Βαθμός- Y
67	63	81	85
74	67	93	89
67	70	81	89
78	74	96	96
89	81	89	100

α) Να κατασκευάσετε το διάγραμμα διασποράς και να εκτιμήσετε από αυτό τον συντελεστή γραμμικής συσχέτισης μεταξύ των βαθμών της Φυσικής (X) και των Μαθηματικών (Y) των 10 μαθητών/τριών της Β' τάξης του Γενικού Λυκείου.

β) Να υπολογίσετε και να ερμηνεύσετε τον συντελεστή γραμμικής συσχέτισης μεταξύ των βαθμών της Φυσικής (X) και των Μαθηματικών (Y) των 10 μαθητών/τριών της Β' τάξης του Γενικού Λυκείου.

γ) Να σχεδιάσετε «με το μάτι» την ευθεία που φαίνεται να προσαρμόζεται καλύτερα στα δεδομένα.

δ) Πώς θα μπορούσατε να εκτιμήσετε τον βαθμό των Μαθηματικών ενός μαθητή της Β' Λυκείου, εάν γνωρίζατε ότι στη Φυσική έγραψε 70;

*Απλά και Κατανοητά η Γνώση!*

### Άσκηση 6 :

Στον παρακάτω πίνακα δίνονται οι ηλικίες και οι πιέσεις αίματος 10 γυναικών.

Ηλικία (x)	56	42	72	36	63	47	55	49	38	60
Πίεση αίματος (y)	17	12	14	10	13	9	11	8	11	15

α) Να σημειώσετε σε ορθοκανονικό σύστημα αξόνων τα σημεία  $(x, y)$  σύμφωνα με τον παραπάνω πίνακα, όπου  $x$  είναι η ηλικία των γυναικών σε έτη και  $y$  είναι η πίεση αίματος των γυναικών σε cm Hg.

β) Να υπολογίσετε και να ερμηνεύσετε τον συντελεστή γραμμικής συσχέτισης μεταξύ των μεταβλητών της ηλικίας των γυναικών σε έτη ( $x$ ) και της πίεσης τους σε cm Hg ( $y$ ).

### Άσκηση 7 :

Στον ακόλουθο πίνακα παρουσιάζονται στην πρώτη γραμμή οι τιμές (σε €) για διαφορετικά κράνη ποδηλασίας και στη δεύτερη γραμμή η βαθμολογία ποιότητάς τους που έγινε από ειδικούς (σε μια κλίμακα από 0 έως 100, όπου όσο μεγαλύτερη είναι η τιμή τόσο πιο ποιοτικό είναι το κράνος).

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

α) Να κατασκευάσετε το διάγραμμα διασποράς.

β) Υπάρχει γραμμική συσχέτιση ανάμεσα στην τιμή και τη βαθμολογία ποιότητας;

γ) Θα μπορούσαμε να πούμε με βεβαιότητα ότι αν αγοράσουμε πιο φθινό κράνος θα έχει πιο χαμηλή ποιότητα;

δ) Να σχεδιάσετε «με το μάτι» στο διάγραμμα διασποράς μια ευθεία που θα μπορούσε να περιγράψει τη σχέση του αναμενόμενου βαθμού ποιότητας ενός ποδηλατικού κράνους με την τιμή του.

*Απλά και Κατανοητά η Γνώση!*

### Άσκηση 8 :

Στον ακόλουθο πίνακα παρουσιάζονται οι συντελεστές γραμμικής συσχέτισης των γραπτών βαθμολογιών στις εξετάσεις Ιουνίου σε 5 μαθήματα ενός τμήματος Β΄ τάξης γενικού λυκείου.

	Άλγεβρα	Βιολογία	Γλώσσα	Φυσική	Χημεία
Άλγεβρα	1,00				
Βιολογία	0,54	1,00			
Γλώσσα	0,76	0,81	1,00		
Φυσική	0,70	0,73	0,71	1,00	
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/τριών αυτών.

### Άσκηση 9 :

Το πλήθος  $x$  των οχημάτων σε εκατομμύρια και ο αριθμός  $y$  των ατυχημάτων σε εκατοντάδες, σε 15 διαφορετικές χώρες, δίνονται από τον επόμενο πίνακα:

Χώρα	A	B	Γ	Δ	E	Z	H	Θ	I	K	Λ	M	N	Ξ	O
x	8,6	13,4	12,8	9,3	1,3	9,4	13,1	4,9	13,5	9,6	7,5	9,8	23,3	21	19,4
y	33	51	30	48	12	23	46	18	36	50	34	35	95	99	69

α) Να κατασκευάσετε το διάγραμμα διασποράς μεταξύ των μεταβλητών του πλήθους των οχημάτων και του αριθμού των οχημάτων για τις 15 χώρες.

β) Να υπολογίσετε και να ερμηνεύσετε τον συντελεστή γραμμικής συσχέτισης μεταξύ των μεταβλητών του πλήθους των οχημάτων και του αριθμού των οχημάτων για τις 15 χώρες.

*Απλά και Κατανοητά η Γνώση!*

### Άσκηση 10 :

Από 8 γάμους που έγιναν σε μια εκκλησία ενός χωριού κατά τη διάρκεια ενός μηνός οι ηλικίες των ανδρόγυνων ήταν:

Ηλικία γαμπρού	20	22	24	25	28	30	33	38
Ηλικία νύφης	20	20	22	27	24	25	28	34

- Να κατασκευάσετε το διάγραμμα διασποράς μεταξύ των ηλικιών της νύφης ( $Y$ ) και του γαμπρού ( $X$ ) και να περιγράψετε το είδος της σχέσης που φαίνεται να έχουν οι δύο μεταβλητές.
- Να υπολογίσετε και να ερμηνεύσετε τον συντελεστή γραμμικής συσχέτισης μεταξύ των ηλικιών νύφης και γαμπρού.
- Να σχεδιάσετε «με το μάτι» την ευθεία που φαίνεται να προσαρμόζεται καλύτερα στα δεδομένα.
- Να βρείτε την αναμενόμενη ηλικία της νύφης για έναν υποψήφιο γαμπρό ετών 34.

### Άσκηση 11 :

Δίνεται δείγμα  $n$  ζευγών παρατηρήσεων  $(x_i, y_i)$ ,  $i=1,2,\dots,n$  δύο μεταβλητών  $X$  και  $Y$  και έστω  $r$ , ο συντελεστής γραμμικής συσχέτισης των μεταβλητών  $X$  και  $Y$ . Να αποδείξετε ότι αν όλα τα παραπάνω σημεία βρίσκονται πάνω σε μια ευθεία με εξίσωση:  $y_i = \alpha + \beta x_i$ ,  $i = 1,2,\dots,n$ , τότε να δείξετε ότι  $r = 1$  αν  $\alpha > 0$  και  $r = -1$  αν  $\alpha < 0$ .

Απλά και Κατανοητά η Γνώση!



**Άσκηση 12 :**

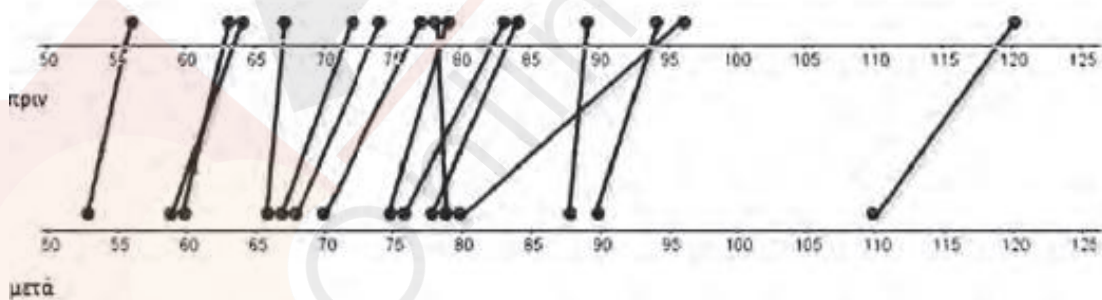
Σε μια έρευνα που έγινε με σκοπό να διερευνηθεί η αποτελεσματικότητα μιας δίαιτας, μετρήθηκε το βάρος 15 ατόμων πριν και μετά τη δίαιτα.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Πριν	89	83	78	56	96	120	67	63	72	74	79	94	84	64	77
Μετά	88	76	79	53	80	110	66	60	67	67	75	90	78	59	70

α) Να υπολογίσετε τη μέση τιμή και την τυπική απόκλιση της μεταβλητής του βάρους των 15 ατόμων πριν ( $x_{\text{πριν}}$ ) και μετά ( $x_{\text{μετά}}$ ) τη δίαιτα. Να σχολιάσετε τα αποτελέσματα.

β) Να κατασκευάσετε τα θηκογράμματα για τη μεταβλητή του βάρους των 15 ατόμων πριν και μετά τη δίαιτα. Να τα συγκρίνετε και να διατυπώσετε την άποψή σας για το αν υπάρχει διαφορά πριν και μετά και τη δίαιτα.

γ) Δίνονται τα επόμενα σημειογράμματα για τη μεταβλητή του βάρους των 15 ατόμων πριν και μετά τη δίαιτα, στα οποία έχουν ενωθεί τα βάρη των 15 ατόμων πριν και μετά τη δίαιτα. Τι παρατηρείτε από το γράφημα; Επιβεβαιώνεται η παρατήρησή σας, συγκρίνοντας με τα αποτελέσματα των προηγούμενων ερωτημάτων;



*Απλά και Κατανοητά η Γνώση!*

**Άσκηση 13 :**

Έχοντας τα δεδομένα της άσκησης 12:

α) Να δημιουργήσετε τη μεταβλητή  $Z = x_{\text{πριν}} - x_{\text{μετά}}$  για κάθε άτομο και να κατασκευάσετε το θηκόγραμμά της. Πώς μπορείτε να αναδείξετε τη διαφορά των τιμών του βάρους των 15 ατόμων πριν και μετά τη δίαιτα από αυτό το θηκόγραμμα;

β) Να κατασκευάσετε τα διαγράμματα διασποράς ανάμεσα στις μεταβλητές του βάρους ( $x_{\text{πριν}}$ ) και ( $x_{\text{μετά}}$ ) των 15 ατόμων.

γ) Να υπολογίσετε και να ερμηνεύσετε τον συντελεστή γραμμικής συσχέτισης μεταξύ των μεταβλητών του βάρους ( $x_{\text{πριν}}$ ) και ( $x_{\text{μετά}}$ ) των 15 ατόμων.

δ) Να σχεδιάσετε «με το μάτι» την ευθεία που φαίνεται να προσαρμόζεται καλύτερα στα δεδομένα.

ε) Θα μπορούσατε να εκτιμήσετε το βάρος ενός ατόμου που πρόκειται να ακολουθήσει αυτή τη δίαιτα, εάν το αρχικό του βάρος ήταν 91 κιλά;

*Απλά και Κατανοητά η Γνώση!*

**Άσκηση 14 :**

Αποδείξτε ότι ο τύπος της γραμμικής συσχέτισης, είναι ισοδύναμος με:

$$r = \frac{\sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i\right) \left(\sum_{i=1}^n y_i\right)}{\sqrt{\sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i\right)^2} \sqrt{\sum_{i=1}^n y_i^2 - \left(\sum_{i=1}^n y_i\right)^2}}$$

**Άσκηση 15 :**

Να υπολογιστεί και να ερμηνευτεί ο συντελεστής συσχέτισης  $r$  μεταξύ των μεταβλητών  $X$  και  $Y$  με βάση τις παρακάτω τιμές:

$x$	10	13	17	21	25	28	30
$y$	21	24	29	25	36	33	40

Απλά και Κατανοητά η Γνώση!

**Άσκηση 16 :**

Από τα διαγράμματα διασποράς των παρακάτω ζευγών  $(x_i, y_i)$ , να εκτιμήσετε (χωρίς πράξεις) εάν η γραμμική συσχέτιση μεταξύ των μεταβλητών  $X$  και  $Y$  είναι θετική ή αρνητική και επιπλέον αν είναι μικρή, μέτρια ή μεγάλη:

α)	$x$	1	3	5	7	9	10	12	13
	$y$	-2	0	1	3	5	6	8	10
β)	$x$	1	3	5	7	9	10	12	13
	$y$	4	5	1	6	4	3	8	10
γ)	$x$	1	3	5	7	9	10	12	13
	$y$	10	8	3	4	6	3	2	5

**Άσκηση 17 :**

Να υπολογίσετε τους συντελεστές γραμμικής συσχέτισης για τα ζεύγη τιμών  $(x_i, y_i)$  της προηγούμενης άσκησης και να τους συγκρίνετε με τις αντίστοιχες εκτιμήσεις σας.

Απλά και Κατανοητά η Γνώση!

**Άσκηση 18 :**

Για τέσσερα ζεύγη παρατηρήσεων  $(x_i, y_i)$  έχουμε:

$$\bar{x} = 7, \bar{y} = 4,5, \sum x_i^2 = 210, \sum y_i^2 = 92, \sum x_i y_i = 138.$$

Να υπολογίσετε το συντελεστή συσχέτισης.

**Άσκηση 19 :**

Να συμπληρώσετε τον παρακάτω πίνακα και να υπολογίσετε το συντελεστή συσχέτισης  $(x, y > 0)$ :

$x$	$y$	$(x - \bar{x})^2$	$(y - \bar{y})^2$	$(x - \bar{x})(y - \bar{y})$
1	3			
2	6	9	1	3
-	1			
6	-			
8	8			
9	13			

Απλά και Κατανοητά η Γνώση!

### Άσκηση 20 :

Να βρείτε τους συντελεστές γραμμικής συσχέτισης για τα παρακάτω ζεύγη τιμών και να σχολιάσετε τα αποτελέσματα.

α)	<table style="border-collapse: collapse;"> <tr> <td style="border-right: 1px solid black; padding: 5px;"><math>x</math></td> <td style="padding: 5px;">1</td> <td style="padding: 5px;">2</td> <td style="padding: 5px;">3</td> <td style="padding: 5px;">4</td> <td style="padding: 5px;">5</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 5px;"><math>y</math></td> <td style="padding: 5px;">4</td> <td style="padding: 5px;">2</td> <td style="padding: 5px;">0</td> <td style="padding: 5px;">-2</td> <td style="padding: 5px;">-4</td> </tr> </table>	$x$	1	2	3	4	5	$y$	4	2	0	-2	-4	β)	<table style="border-collapse: collapse;"> <tr> <td style="border-right: 1px solid black; padding: 5px;"><math>x</math></td> <td style="padding: 5px;">1</td> <td style="padding: 5px;">2</td> <td style="padding: 5px;">3</td> <td style="padding: 5px;">4</td> <td style="padding: 5px;">5</td> </tr> <tr> <td style="border-right: 1px solid black; padding: 5px;"><math>y</math></td> <td style="padding: 5px;">-4</td> <td style="padding: 5px;">-2</td> <td style="padding: 5px;">0</td> <td style="padding: 5px;">2</td> <td style="padding: 5px;">4</td> </tr> </table>	$x$	1	2	3	4	5	$y$	-4	-2	0	2	4
$x$	1	2	3	4	5																						
$y$	4	2	0	-2	-4																						
$x$	1	2	3	4	5																						
$y$	-4	-2	0	2	4																						

### Ασκήσεις για Μελέτη

#### Άσκηση 1 :

Η κατά άτομο κατανάλωση (σε γαλιόνια) άπαχου (Y) και πλήρους (X) γάλακτος για τα έτη 1982-87 στις ΗΠΑ δίνεται στον παρακάτω πίνακα:

	Έτος					
	1982	1983	1984	1985	1986	1987
Πλήρες γάλα, $x$	15,6	15,2	14,7	14,3	13,4	12,8
Άπαχο γάλα, $y$	10,8	11,1	11,5	12,1	12,8	13,2

α) Να υπολογίσετε και να ερμηνεύσετε το συντελεστή συσχέτισης.

β) Αν μετατραπούν οι ποσότητες γάλακτος σε λίτρα, ποια θα είναι η τιμή του συντελεστή συσχέτισης; (1 γαλιόνι  $\approx$  3,8 λίτρα)

*Απλά και Κατανοητά η Γνώση!*

**Άσκηση 2 :**

Τα παρακάτω δεδομένα παριστάνουν τους δείκτες ευφυΐας (I.Q.) 10 μητέρων (X) και των θυγατέρων τους (Y):

I.Q. μητέρας	I.Q. θυγατέρας	I.Q. μητέρας	I.Q. θυγατέρας
85	90	115	110
90	100	120	125
95	90	120	110
100	105	130	130
110	120	135	120

- α) Να κατασκευάσετε το διάγραμμα διασποράς
- β) Από το διάγραμμα διασποράς να εκτιμήσετε το συντελεστή συσχέτισης
- γ) Να υπολογίσετε και να ερμηνεύσετε το συντελεστή συσχέτισης

**Άσκηση 3 :**

- α) Να δείξετε ότι :

$$\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}$$

- β) Για επτά ζεύγη παρατηρήσεων  $(x_i, y_i)$ , έχουμε:

$$\sum (x_i - \bar{x})^2 = 28, \quad \sum (y_i - \bar{y})^2 = 112, \quad \sum x_i y_i = 308, \quad \bar{x} = 4, \quad \bar{y} = 9$$

Να υπολογίσετε το συντελεστή συσχέτισης.

*Απλά και Κατανοητά η Γνώση!*

**Άσκηση 4 :**

Σε μια εξέταση στα Μαθηματικά οκτώ μαθητών η βαθμολογία δύο εξεταστών Α, Β ήταν ως ακολούθως:

		Μαθητής							
		1	2	3	4	5	6	7	8
Εξεταστής	A	55	62	71	66	63	56	72	51
Εξεταστής	B	54	56	61	66	63	61	73	54

Να εξετάσετε εάν υπάρχει γραμμική συσχέτιση μεταξύ της βαθμολογίας των δύο εξεταστών.

**Άσκηση 5 :**

Οι κάτοικοι ανά τετραγωνικό χιλιόμετρο από το 1960 έως και το 1974 στην Ελλάδα ήταν:

Κάτοικοι, Y	63	64	64	64	66	65	65	66	66	67	67	67	67	68	68
Έτος, X	1960	61	62	63	64	65	66	67	68	69	1970	71	72	73	74

α) Να εκτιμήσετε την ευθεία γραμμικής παλινδρόμησης της Y στη X, και να την παραστήσετε στο διάγραμμα διασποράς.

β) Το 1976 είχαμε 69,5 κατοίκους/ τετραγωνικό χιλιόμετρο. Είναι αυτό αναμενόμενο;

(Υπόδειξη: Θεωρούμε ως έτος αναφοράς το 1960 με τιμή  $x = 1$ ).

*Απλά και Κατανοητά η Γνώση!*



**Άσκηση 6 :**

Ο συντελεστής γενικής θνησιμότητας της Ελλάδας για τα χρόνια 1931-1964 παρουσίασε την παρακάτω πορεία.

Έτος, $X$	1931	1936	1940	1950	1956	1961	1964
Σ.Γ.Θ.%, $Y$	17,7	15,1	12,8	7,9	7,4	7,6	8,2

α) Να χαράξετε “με το μάτι” την ευθεία γραμμικής παλινδρόμησης  $y = \alpha + \beta x$  στο διάγραμμα διασποράς και από την ευθεία αυτή να εκτιμήσετε το συντελεστή γενικής θνησιμότητας για το έτος 1965.

β) Χρησιμοποιώντας τη μέθοδο των δύο σημείων να υπολογίσετε την εξίσωση της ευθείας παλινδρόμησης και στη συνέχεια να εκτιμήσετε πάλι το συντελεστή γενικής θνησιμότητας για το έτος 1965. Συγκρίνετε με το προηγούμενο αποτέλεσμα.

**Άσκηση 7 :**

Η ποσότητα:

$$s_{xy} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$$

καλείται συνδιακύμανση των μεταβλητών  $X$  και  $Y$ . Αν καλέσουμε με  $s_x^2$ ,  $s_y^2$  τις διακυμάνσεις των  $X$  και  $Y$  αντίστοιχα, να δείξετε ότι ισχύουν οι σχέσεις:

$$\alpha) \hat{\beta} = \frac{s_{xy}}{s_x^2} \quad \beta) r = \hat{\beta} \frac{s_x}{s_y}$$

*Απλά και Κατανοητά η Γνώση!*

**Άσκηση 8 :**

Ένας μαθητής γνώριζε ότι η σχέση που συνδέει τους βαθμούς Φαρενάιτ (F) με τους βαθμούς Κελσίου (C) είναι γραμμική, δηλαδή  $F = \alpha + \beta C$ . Επειδή όμως δε θυμότανε τις σταθερές  $\alpha$ ,  $\beta$ , μέτρησε τη θερμοκρασία του δωματίου του σε πέντε διαφορετικές ώρες με δύο θερμόμετρα με κλίμακα σε F και C, αντίστοιχα, και πήρε τα παρακάτω ζεύγη τιμών:

°C	15	20	25	30	35
°F	59	68	77	86	95

Να βρείτε τη σχέση  $F = \alpha + \beta C$  που συνδέει τις δύο κλίμακες θερμοκρασίας.

**Άσκηση 9 :**

Δίνεται δείγμα  $n$  ζευγών  $(x_1, y_2), (x_2, y_2), \dots, (x_n, y_n)$  δύο μεταβλητών  $X$  και  $Y$  και έστω  $r(X, Y)$  ο συντελεστής συσχέτισης. Εάν  $Z = \lambda Y$  όπου  $\lambda$  θετική σταθερά, να δείξετε ότι ισχύει:

$r(X, Z) = r(X, Y)$ . Τι γίνεται, εάν  $\lambda < 0$ ;

**Άσκηση 10 :**

Ο αριθμός των παιδιών σε ένα δείγμα 80 οικογενειών μιας πόλης δίνεται στον παρακάτω πίνακα:

Αριθμός Παιδιών	0	1	2	3	4	5	6
Οικογένειες	10	25	20	12	6	5	2

α) Να βρείτε τη μέση τιμή, τη διάμεση τιμή, την επικρατούσα τιμή και την τυπική απόκλιση του αριθμού των παιδιών.

β) Να κατασκευάσετε το διάγραμμα σχετικών συχνοτήτων και το πολύγωνο αθροιστικών σχετικών συχνοτήτων.

γ) Από το πολύγωνο αθροιστικών σχετικών συχνοτήτων να εκτιμήσετε τα τρία τεταρτημόρια.

*Απλά και Κατανοητά η Γνώση!*