ΜΑΘΗΜΑΤΙΚΑ
ΚΑΙ
ΣΤΟΙΧΕΙΑ ΣΤΑΤΙΣΤΙΚΗΣ

Γ΄ ΓΕΝΙΚΟΥ ΛΥΚΕΙΟΥ

Τόμος 3ος

ΣΤΟΙΧΕΙΑ ΑΡΧΙΚΗΣ ΕΚΔΟΣΗΣ

ΣΥΓΓΡΑΦΕΙΣ:

Αδαμόπουλος Λεωνίδας
Επ. Σύμβουλος Παιδαγωγικού Ινστιτούτου

Δαμιανού Χαράλαμπος
Αναπλ. Καθηγητής Παν/μίου Αθηνών

Σβέρκος Ανδρέας
Σχολικός Σύμβουλος

ΚΡΙΤΕΣ:

Κουνιάς Στρατής
Καθηγητής Παν/μίου Αθηνών

Μακρής Κωνσταντίνος
Σχολικός Σύμβουλος

Τσικαλουδάκης Γεώργιος
Καθηγητής Β/θμιας Εκπαίδευσης

Γλωσσική Επιμέλεια:
Μπουσούνη Λία
Καθηγήτρια Β/θμιας Εκπαίδευσης

Δακτυλογράφηση:
Μπολιώτη Πόπη

Σχήματα:
Μπούτσικας Μιχάλης

ΣΤΟΙΧΕΙΑ ΕΠΑΝΕΚΔΟΣΗΣ

Η επανέκδοση του παρόντος βιβλίου πραγματοποιήθηκε από το Ινστιτούτο Τεχνολογίας Υπολογιστών & Εκδόσεων «Διόφαντος» μέσω ψηφιακής μακέτας, η οποία δημιουργήθηκε με χρηματοδότηση από το ΕΣΠΑ / ΕΠ «Εκπαίδευση & Διά Βίου Μάθηση» / Πράξη «ΣΤΗΡΙΖΩ».

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ & ΘΡΗΣΚΕΥΜΑΤΩΝ, ΠΟΛΙΤΙΣΜΟΥ & ΑΘΛΗΤΙΣΜΟΥ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης

Οι διορθώσεις πραγματοποιήθηκαν κατόπιν έγκρισης του Δ.Σ. του Ινστιτούτου Εκπαιδευτικής Πολιτικής

Η αξιολόγηση, η κρίση των προσαρμογών και η επιστημονική επιμέλεια του προσαρμοσμένου βιβλίου πραγματοποιείται από τη Μονάδα Ειδικής Αγωγής του Ινστιτούτου Εκπαιδευτικής Πολιτικής.

Η προσαρμογή του βιβλίου για μαθητές με μειωμένη όραση από το ΙΤΥΕ – ΔΙΟΦΑΝΤΟΣ πραγματοποιείται με βάση τις προδιαγραφές που έχουν αναπτυχθεί από ειδικούς εμπειρογνώμονες για το ΙΕΠ.

ΠΡΟΣΑΡΜΟΓΗ ΤΟΥ ΒΙΒΛΙΟΥ ΓΙΑ ΜΑΘΗΤΕΣ ΜΕ ΜΕΙΩΜΕΝΗ ΟΡΑΣΗ

ΙΤΥΕ - ΔΙΟΦΑΝΤΟΣ

ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ, ΕΡΕΥΝΑΣ
ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ

ΙΝΣΤΙΤΟΥΤΟ ΕΚΠΑΙΔΕΥΤΙΚΗΣ
ΠΟΛΙΤΙΚΗΣ

ΑΔΑΜΟΠΟΥΛΟΣ ΛΕΩΝΙΔΑΣ
ΔΑΜΙΑΝΟΥ ΧΑΡΑΛΑΜΠΟΣ
ΣΒΕΡΚΟΣ ΑΝΔΡΕΑΣ

Η συγγραφή και η επιστημονική
επιμέλεια του βιβλίου
πραγματοποιήθηκε υπό την αιγίδα
του Παιδαγωγικού Ινστιτούτου

ΜΑΘΗΜΑΤΙΚΑ ΚΑΙ ΣΤΟΙΧΕΙΑ ΣΤΑΤΙΣΤΙΚΗΣ

Γ΄ ΓΕΝΙΚΟΥ ΛΥΚΕΙΟΥ

Τόμος 3ος

Ι.Τ.Υ.Ε. «ΔΙΟΦΑΝΤΟΣ»

2.3 ΜΕΤΡΑ ΘΕΣΗΣ ΚΑΙ ΔΙΑΣΠΟΡΑΣ

Εισαγωγή

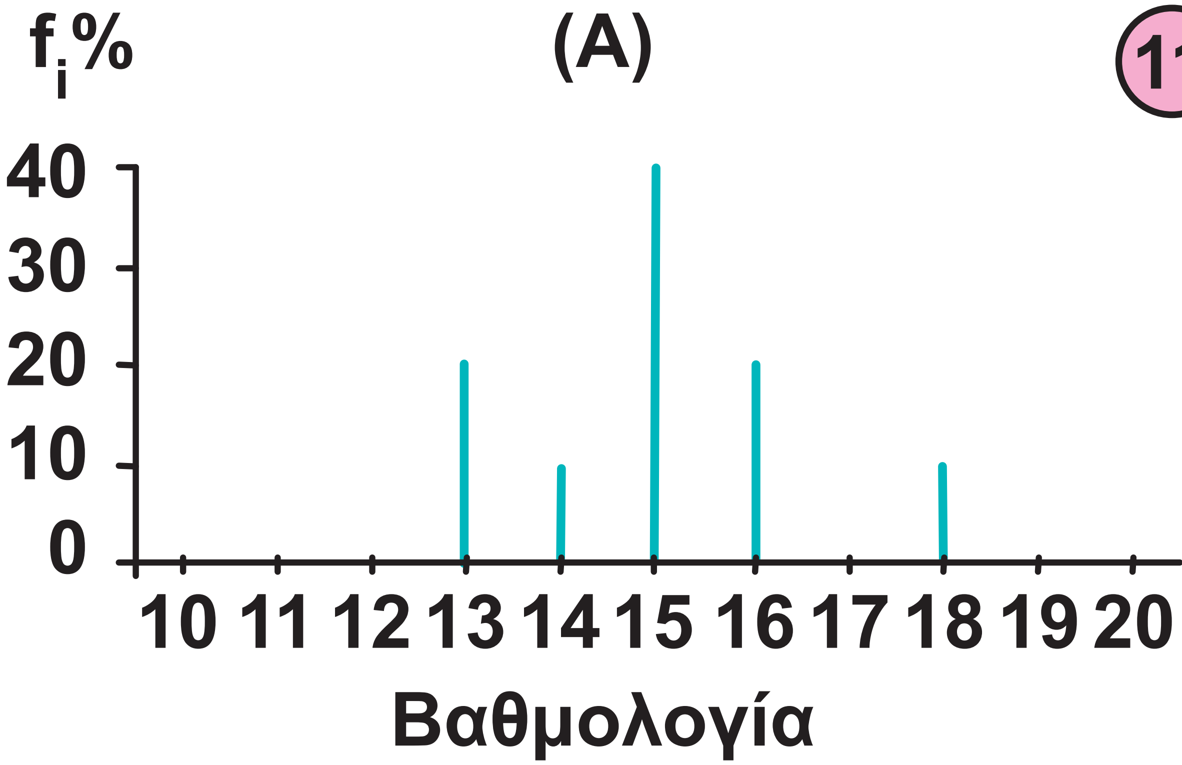
Εκτός από τους στατιστικούς πίνακες και τα διαγράμματα υπάρχουν και αριθμητικά μέτρα με τα οποία μπορούμε να περιγράψουμε με συντομία μια κατανομή συχνοτήτων. Η γνώση των μέτρων αυτών διευκολύνει και την παραπέρα στατιστική επεξεργασία των δεδομένων. Έστω, για παράδειγμα, ένας καθηγητής ο οποίος, για να συγκρίνει δύο διαφορετικά τμήματα Α και Β της ίδιας τάξης ως προς την επίδοσή τους σε ένα μάθημα, πήρε τυχαία 10 μαθητές από κάθε τμήμα. Η βαθμολογία τους στο μάθημα αυτό ήταν:

Τμήμα Α: 13 13 14 15 15 15 15 16 16 18

Τμήμα Β: 10 13 14 14 15 15 15 16 18 20.

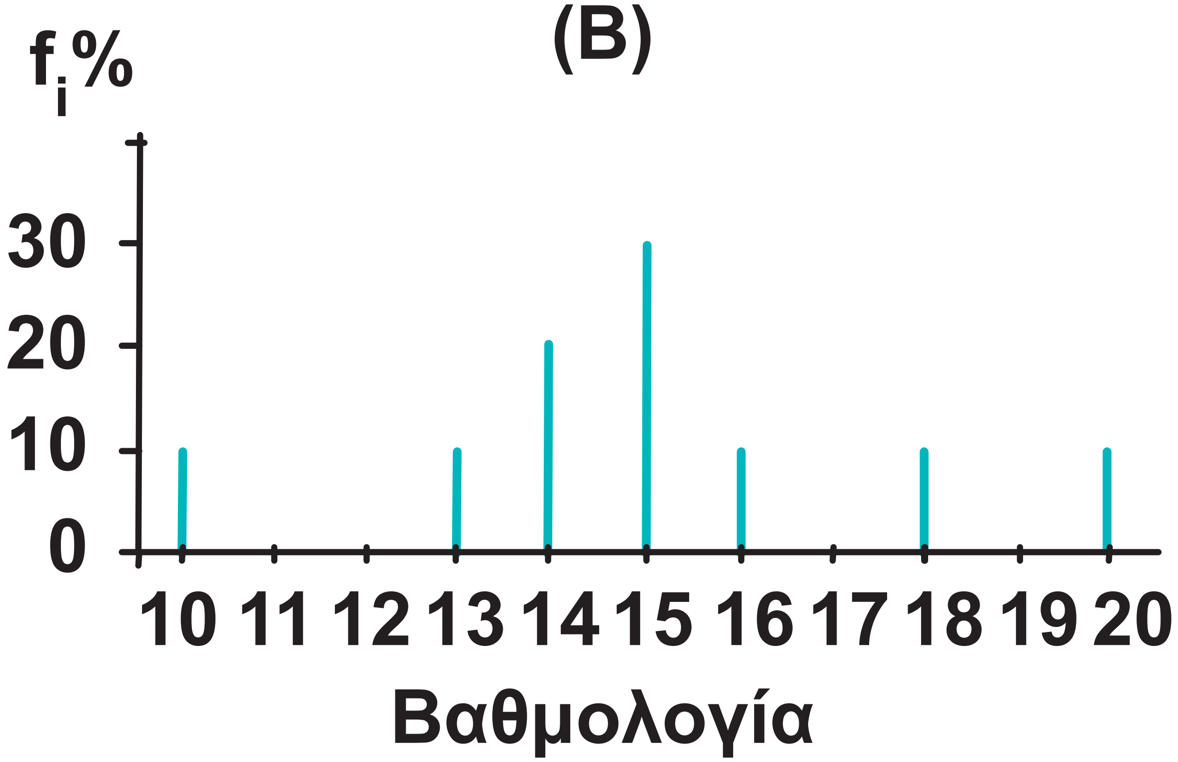
Τα διαγράμματα σχετικών συχνοτήτων δίνονται στα σχήματα 11(α), (β).

(A)


fi%
40
30
20
10
0
10 11 12 13 14 15 16 17 18 19 20
Βαθμολογία

(α)

(B)


fi%
30
20
10
0
10 11 12 13 14 15 16 17 18 19 20
Βαθμολογία

(β)

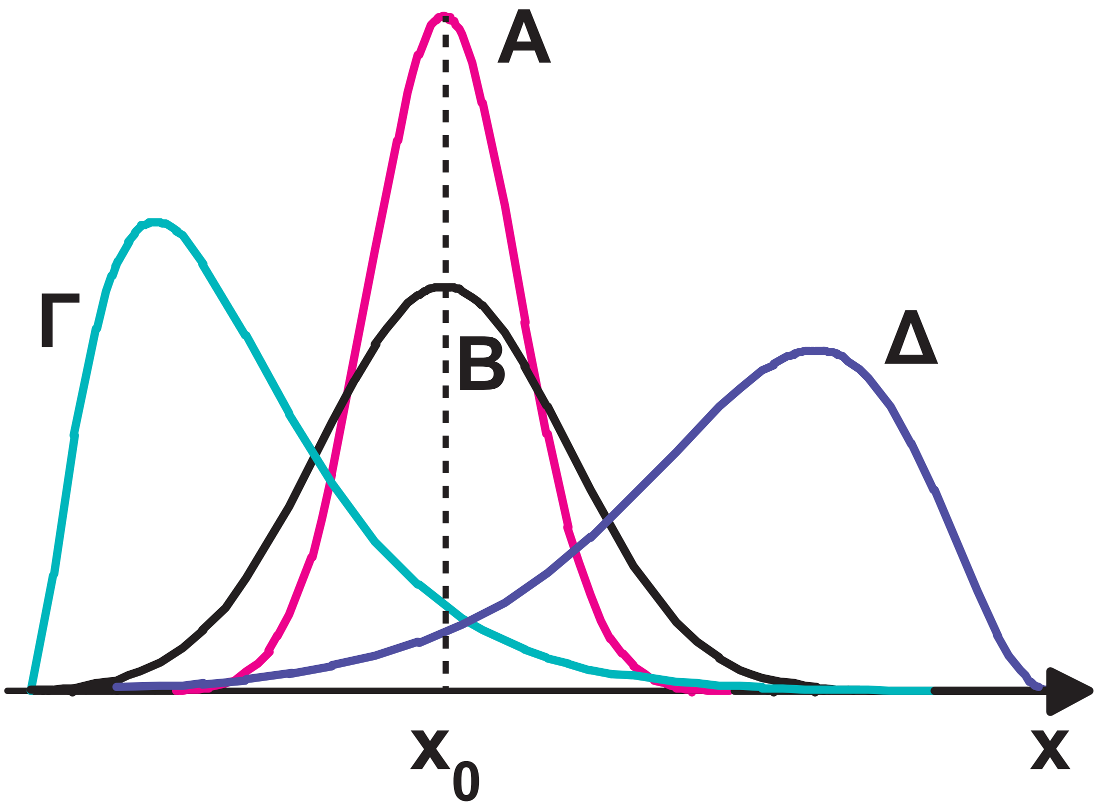
Παρατηρούμε ότι η βαθμολογία και των δύο τμημάτων είναι συγκεντρωμένη γύρω στο 15, αλλά το δεύτερο τμήμα παρουσιάζει μεγαλύτερη διασπορά βαθμών από το πρώτο. Δηλαδή, οι βαθμοί του Β΄ τμήματος είναι περισσότερο διασκορπισμένοι γύρω από μια “κεντρική” τιμή. Οι έννοιες “κεντρική τιμή” και “διασπορά των παρατηρήσεων” μας δίνουν το ερέθισμα για έναν ακόμα πιο σύντομο τρόπο περιγραφής της κατανομής ενός συνόλου δεδομένων. Για να ορίσουμε δηλαδή κάποια **μέτρα** (αριθμητικά μεγέθη), που να μας δίνουν α) τη θέση του “κέντρου” των παρατηρήσεων στον οριζόντιο άξονα και β) τη διασπορά των παρατηρήσεων, δηλαδή πόσο αυτές εκτείνονται γύρω

22-0088_I_c_math_bm_83-117_28b.indd 8 18/09/2017 09:58

από το “κέντρο” τους. Τα πρώτα τα καλούμε **μέτρα θέσης** της κατανομής (location measures), ενώ τα δεύτερα **μέτρα διασποράς** ή **μέτρα μεταβλητότητας** (measures of variability).

Εκτός από τα μέτρα θέσης και διασποράς μιας κατανομής πολλές φορές είναι απαραίτητος και ο προσδιορισμός κάποιων άλλων μέτρων, που καθορίζουν τη **μορφή** της κατανομής. Κατά πόσο δηλαδή η αντίστοιχη καμπύλη συχνοτήτων είναι συμμετρική ή όχι ως προς την ευθεία $x = x_0$, για δεδομένο σημείο x_0 του άξονα 0x. Τα μέτρα αυτά, που συνήθως εκφράζονται σε συνάρτηση με τα μέτρα θέσης και διασποράς, καλούνται **μέτρα ασυμμετρίας** (measures of skewness).

Υπολογίζοντας από ένα σύνολο δεδομένων κάποια από τα ανωτέρω μέτρα, μπορούμε να έχουμε μια σύντομη περιγραφή της μορφής της καμπύλης συχνοτήτων. Στο σχήμα 12 οι καμπύλες συχνοτήτων Α και Β είναι συμμετρικές με το ίδιο “κέντρο” x_0, αλλά η Β έχει μεγαλύτερη μεταβλητότητα από την Α. Οι καμπύλες Γ και Δ είναι ασύμμετρες, με τη Γ όπως λέμε να παρουσιάζει θετική ασυμμετρία και τη Δ αρνητική ασυμμετρία. Το “κέντρο” της Γ είναι αριστερότερα του x_0, ενώ της Δ είναι δεξιότερα του x_0. Η Δ παρουσιάζει μεγαλύτερη μεταβλητότητα από τη Γ.



Μέτρα Θέσης

Τα πιο συνηθισμένα μέτρα που χρησιμοποιούνται για την περιγραφή της θέσης ενός συνόλου δεδομένων πάνω στον οριζόντιο άξονα 0x, εκφράζοντας την "κατά μέσο όρο" απόστασή τους από την αρχή των αξόνων, είναι ο αριθμητικός μέσος ή μέση τιμή (arithmetic mean or average), η διάμεσος (median) και η κορυφή ή επικρατούσα τιμή (mode).

α) Μέση Τιμή ($\bar{x}$)

Η μέση τιμή ενός συνόλου ν παρατηρήσεων αποτελεί το σπουδαιότερο και χρησιμότερο μέτρο της Στατιστικής και ορίζεται ως το άθροισμα των παρατηρήσεων διά του πλήθους των παρατηρήσεων. Όταν σε ένα δείγμα μεγέθους ν οι παρατηρήσεις μιας μεταβλητής Χ είναι $t_1, t_2, \ldots, t_\nu$ τότε η μέση τιμή συμβολίζεται με $\bar{x}$ και δίνεται από τη σχέση:

$$\bar{x} = \frac{t_1 + t_2 + \ldots + t_\nu}{\nu} = \frac{\sum_{i=1}^{\nu} t_i}{\nu} = \frac{1}{\nu}\sum_{i=1}^{\nu} t_i \qquad (1)$$

όπου το σύμβολο $\sum_{i=1}^{\nu} t_i$ παριστάνει μια συντομογραφία του αθροίσματος $t_1 + t_2 + \ldots + t_\nu$ και διαβάζεται "άθροισμα των t_i από $i = 1$ έως ν". Συχνά, όταν δεν υπάρχει πρόβλημα σύγχυσης, συμβολίζεται και ως $\sum t_i$ ή ακόμα πιο απλά με $\sum t$.

Σε μια κατανομή συχνοτήτων, αν $x_1, x_2, \ldots, x_\kappa$ είναι οι τιμές της μεταβλητής Χ με συχνότητες $\nu_1, \nu_2, \ldots, \nu_\kappa$ αντίστοιχα, η μέση τιμή ορίζεται ισοδύναμα από τη σχέση:

$$\bar{x} = \frac{x_1 v_1 + x_2 v_2 + \ldots + x_\kappa v_\kappa}{v_1 + v_2 + \ldots + v_\kappa} = \qquad (2)$$

$$= \frac{\sum_{i=1}^{\kappa} x_i v_i}{\sum_{i=1}^{\kappa} v_i} = \frac{1}{v} \sum_{i=1}^{\kappa} x_i v_i$$

Η παραπάνω σχέση ισοδύναμα γράφεται:

$$\bar{x} = \sum_{i=1}^{\kappa} x_i \frac{v_i}{v} = \sum_{i=1}^{\kappa} x_i f_i$$

όπου f_i οι σχετικές συχνότητες. Για παράδειγμα, η μέση επίδοση των μαθητών στο τμήμα Α θα είναι σύμφωνα με την (1)

$$\bar{x}_A = \frac{13+13+14+...+18}{10} = \frac{150}{10} = 15$$

ή ισοδύναμα από τον αντίστοιχο πίνακα συχνοτήτων σύμφωνα με την (2).

Βαθμός x_i	Συχνότητα v_i	$x_i v_i$
13	2	26
14	1	14
15	4	60
16	2	32
18	1	18
Σύνολο	$v_A = 10$	$\sum x_i v_i = 150$

$$\bar{x}_A = \frac{\sum x_i v_i}{v_A} = \frac{150}{10} = 15.$$

Ομοίως, υπολογίζεται και η μέση επίδοση για το τμήμα Β, η οποία είναι πάλι

$$\bar{x}_B = 15$$

Επίσης, το μέσο ύψος των 40 μαθητών της Γ΄ Λυκείου του πίνακα 8, σύμφωνα με τη σχέση (1) είναι

$$\bar{x} = \frac{6918}{40} = 172{,}95 \text{ cm}.$$

Για ευκολότερο όμως υπολογισμό χρησιμοποιούμε τον πίνακα συχνοτήτων, όπως αυτός δίνεται παρακάτω, ομαδοποιώντας τα δεδομένα σε κ = 6 κλάσεις.

Αν x_i είναι το κέντρο της i κλάσης και v_i η αντίστοιχη συχνότητα, τότε σύμφωνα με τη σχέση (2) η μέση τιμή θα είναι:

$$\bar{x} = \frac{\sum x_i v_i}{v} = \frac{6930}{40} = 173{,}25\text{cm}.$$

Παρατηρούμε ότι οι δύο μέσες τιμές του ίδιου συνόλου δεδομένων δεν είναι ακριβώς οι ίδιες. Πού οφείλεται αυτή η, έστω και μικρή, διαφορά;
Η διαφορά αυτή οφείλεται στο γεγονός ότι κατά την ομαδοποίηση υποθέσαμε ότι οι παρατηρήσεις κάθε κλάσης είναι ομοιόμορφα κατανεμημένες και ότι οι τιμές της μεταβλητής σε κάθε κλάση εκπροσωπούνται από την αντίστοιχη κεντρική τιμή x_i. Η υπόθεση αυτή σημαίνει απώλεια πληροφοριών για τις αρχικές τιμές. Χάνουμε λοιπόν λίγο ως προς την ακρίβεια κερδίζουμε όμως χρόνο!

Ύψος σε cm	Κεντρικές τιμές x_i	
156-162	159	
162-168	165	
168-174	171	
174-180	177	
180-186	183	
186-192	189	
	Σύνολο	

	Συχνότητα ν_i	$x_i\,\nu_i$
	2	318
	8	1320
	12	2025
	11	1947
	5	915
	2	378
	$\sum \nu_i = 40$	$\sum x_i \nu_i = 6930$

β) Σταθμικός Μέσος

Στις περιπτώσεις που δίνεται διαφορετική βαρύτητα (έμφαση) στις τιμές $x_1, x_2, ..., x_ν$ ενός συνόλου δεδομένων, τότε αντί του αριθμητικού μέσου χρησιμοποιούμε τον σταθμισμένο αριθμητικό μέσο ή σταθμικό μέσο (weighted mean). Εάν σε κάθε τιμή $x_1, x_2, ..., x_ν$ δώσουμε διαφορετική βαρύτητα, που εκφράζεται με τους λεγόμενους συντελεστές στάθμισης (βαρύτητας) $w_1, w_2, ..., w_ν$, τότε ο σταθμικός μέσος βρίσκεται από τον τύπο:

$$\bar{x} = \frac{x_1 w_1 + x_2 w_2 + ... + x_ν w_ν}{w_1 + w_2 + ... + w_ν} = \frac{\sum_{i=1}^{ν} x_i w_i}{\sum_{i=1}^{ν} w_i}.$$

22-0088_I_c_math_bm_83-117_28b.indd 20 18/09/2017 09:58

Για παράδειγμα, με το νέο σύστημα, για την εισαγωγή ενός μαθητή στην τριτοβάθμια εκπαίδευση θα συνυπολογίζονται ο βαθμός x_1 του απολυτηρίου του Ενιαίου Λυκείου με συντελεστή (βάρος) $w_1 = 7{,}5$, ο βαθμός x_2 στο τεστ δεξιοτήτων με συντελεστή $w_2 = 1$, ο βαθμός x_3 στο 1ο βασικό μάθημα με συντελεστή $w_3 = 1$ και ο βαθμός x_4 στο 2ο βασικό μάθημα με συντελεστή $w_4 = 0{,}5$. Εάν ένας μαθητής πάρει τους βαθμούς $x_1 = 16{,}5$, $x_2 = 18$, $x_3 = 17$ και $x_4 = 16{,}6$, τότε ο σταθμικός μέσος της επίδοσης του θα είναι:

$$\bar{x} = \frac{16,5 \times 7,5 + 18 \times 1 + 17 \times 1 + 16,6 \times 0,5}{7,5 + 1 + 1 + 0,5} = \frac{167}{10} = 16,7.$$

γ) Διάμεσος (δ)

Οι χρόνοι (σε λεπτά) που χρειάστηκαν 9 μαθητές, για να λύσουν ένα πρόβλημα είναι: 3, 5, 5, 36, 6, 7, 4, 7, 8 με μέση τιμή $\bar{x} = 9$. Παρατηρούμε όμως ότι οι οκτώ από τις εννέα παρατηρήσεις είναι μικρότερες του 9 και μία (ακραία τιμή), η οποία επηρεάζει και τη μέση τιμή είναι, αρκετά μεγαλύτερη του 9. Αυτό σημαίνει ότι η μέση τιμή δεν ενδείκνυται ως μέτρο θέσης ("κέντρο") των παρατηρήσεων αυτών. Αντίθετα, ένα άλλο μέτρο θέσης που δεν επηρεάζεται από ακραίες παρατηρήσεις είναι η **διάμεσος** (median), η οποία ορίζεται ως εξής:

Διάμεσος (δ) ενός δείγματος ν παρατηρήσεων οι οποίες έχουν διαταχθεί σε αύξουσα σειρά ορίζεται ως η μεσαία παρατήρηση, όταν το ν είναι περιττός αριθμός, ή ο μέσος όρος (ημιάθροισμα) των δύο μεσαίων παρατηρήσεων όταν το ν είναι άρτιος αριθμός.

Για παράδειγμα, για να βρούμε τη διάμεσο των δεδομένων:

α) 3, 4, 0, 6, 5, 8, 1, 1, 6, 1, 2, 8, 9

β) 3, 4, 0, 6, 5, 8, 1, 1, 6, 1, 2, 8, 9, 9

εργαζόμαστε ως εξής:

α) Έχουμε ν = 13 παρατηρήσεις, οι οποίες σε αύξουσα σειρά είναι:

0 1 1 1 2 3 4 5 6 6 8 8 9 .

Άρα, η διάμεσος είναι η μεσαία παρατήρηση (έβδομη στη σειρά), δ = 4.

β) Έχουμε ν = 14 παρατηρήσεις οι οποίες σε αύξουσα σειρά είναι:

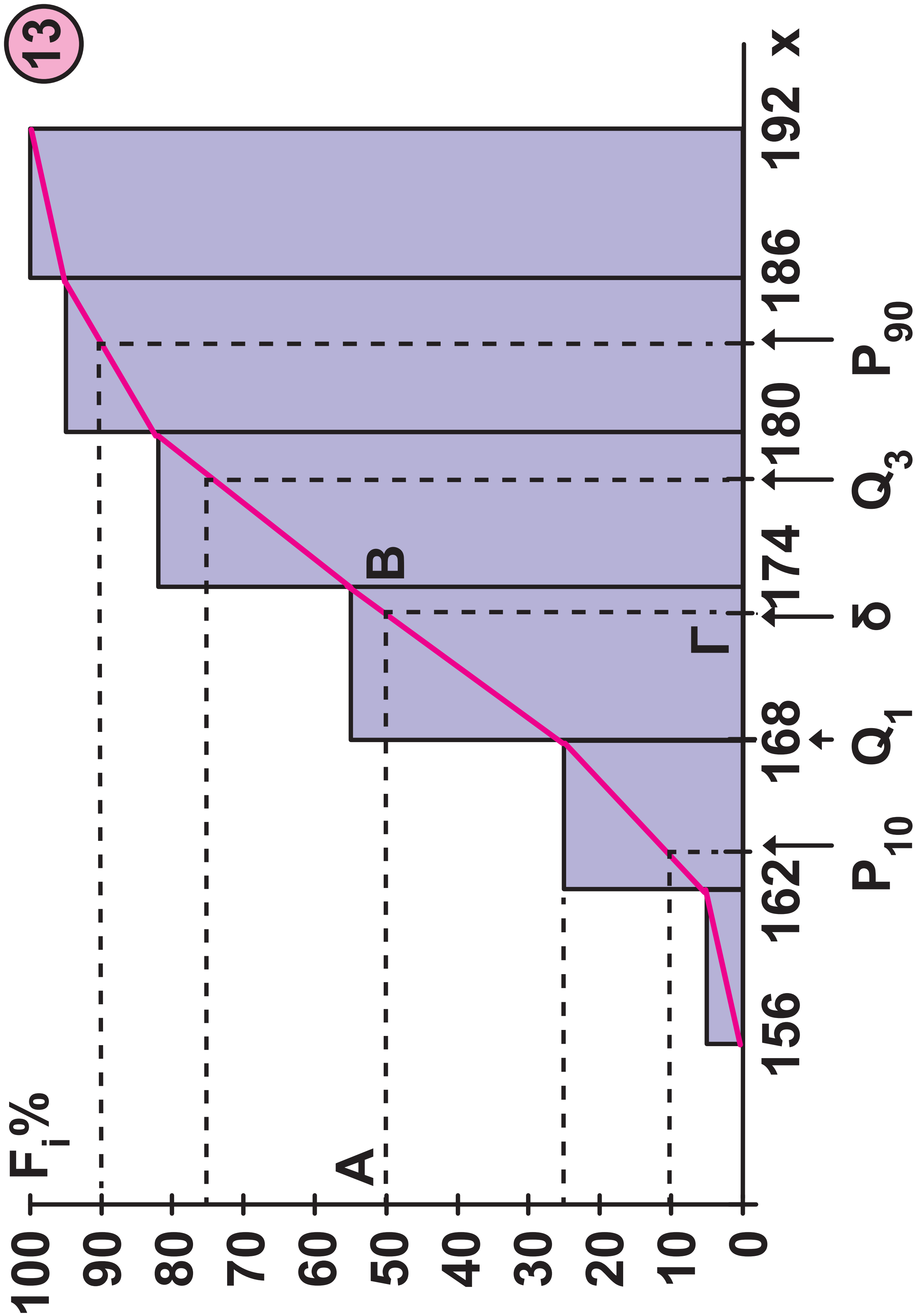
0 1 1 1 2 3 4 5 6 6 8 8 9 9.

Άρα, η διάμεσος είναι το ημιάθροισμα των δύο μεσαίων παρατηρήσεων (της έβδομης και όγδοης στη σειρά), δηλαδή $\delta = \frac{4+5}{2} = 4{,}5$.

Παρατηρούμε ότι, η διάμεσος είναι η τιμή που χωρίζει ένα σύνολο παρατηρήσεων σε δύο ίσα μέρη όταν οι παρατηρήσεις αυτές τοποθετηθούν με σειρά τάξης μεγέθους. Ακριβέστερα, η διάμεσος είναι η τιμή για την οποία το πολύ 50% των παρατηρήσεων είναι μικρότερες από αυτήν και το πολύ 50% των παρατηρήσεων είναι μεγαλύτερες από την τιμή αυτήν.

Διάμεσος σε Ομαδοποιημένα Δεδομένα

Θεωρούμε τα δεδομένα του ύψους των μαθητών στον πίνακα 9 και το αντίστοιχο ιστόγραμμα αθροιστικών σχετικών συχνοτήτων με την πολυγωνική γραμμή, σχήμα 13. Η διάμεσος, όπως ορίστηκε, αντιστοιχεί στην τιμή x = δ της μεταβλητής X (στον οριζόντιο άξονα), έτσι ώστε το 50% των παρατηρήσεων να είναι μικρότερες ή ίσες του δ. Δηλαδή, η διάμεσος θα έχει αθροιστική σχετική συχνότητα F_i = 50%. Εφόσον στον κάθετο άξονα έχουμε τις αθροιστικές σχετικές συχνότητες, από το σημείο Α (50% των παρατηρήσεων) φέρουμε την ΑΒ//0x και στη συνέχεια τη ΒΓ⊥0x. Τότε, στο σημείο Γ αντιστοιχεί η διάμεσος δ των παρατηρήσεων. Δηλαδή, δ ≈ 173.


13
F_i %
100
90
80
70
60
50
40
30
20
10
0
A
B
Γ
δ
156
162
168
174
180
186
192
x
P_{10}
Q_1
Q_3
P_{90}

δ) Εκατοστημόρια ($P_κ$)

Όπως ορίσαμε τη διάμεσο δ, έτσι ώστε το πολύ 50% των παρατηρήσεων να είναι μικρότερες του δ και το πολύ 50% των παρατηρήσεων να είναι μεγαλύτερες του δ, μπορούμε ανάλογα να ορίσουμε και τα **εκατοστημόρια** (percentiles) $P_κ$, $κ = 1, 2, ..., 99$. Οι τιμές $P_1, P_2, ..., P_{99}$ χωρίζουν τη συνολική συχνότητα σε 100 ίσα μέρη. Επομένως, αναλόγως και με τον ορισμό της διαμέσου, ορίζουμε ως κ-εκατοστιαίο σημείο ή $P_κ$ εκατοστημόριο ενός συνόλου παρατηρήσεων την τιμή εκείνη για την οποία το πολύ κ% των παρατηρήσεων είναι μικρότερες του $P_κ$ και το πολύ $(100 – κ)$% των παρατηρήσεων είναι μεγαλύτερες από την τιμή αυτήν.

Ειδική περίπτωση εκατοστημορίων είναι τα P_{25}, P_{50}, P_{75}, τα οποία καλούνται **τεταρτημόρια** (quartiles) και συμβολίζονται με Q_1, Q_2 και Q_3, αντίστοιχα.

Για το Q_1 έχουμε αριστερά το πολύ 25% των παρατηρήσεων και δεξιά το πολύ 75%. Όμοια για το Q_3 έχουμε αριστερά το πολύ 75% των παρατηρήσεων και δεξιά το πολύ 25% των παρατηρήσεων. Προφανώς το $Q_2 = P_{50}$ συμπίπτει και με τη διάμεσο, δηλαδή $Q_2 = δ$. Τα μέτρα αυτά χρησιμοποιούνται αρκετά συχνά για τη μελέτη ενός συνόλου δεδομένων.

Συχνά για ευκολία ο υπολογισμός των τεταρτημορίων Q_1 και Q_3 ενός συνόλου δεδομένων γίνεται **κατά**

προσέγγιση υπολογίζοντας τις διαμέσους του πρώτου και του δεύτερου μισού των διατεταγμένων παρατηρήσεων, αντίστοιχα. Για παράδειγμα, προκειμένου να υπολογίσουμε τα τεταρτημόρια των δεδομένων 3, 4, 0, 6, 5, 8, 1, 1, 6, 1, 2, 8, 9, εργαζόμαστε ως εξής:

- Διατάσσουμε τις παρατηρήσεις σε αύξουσα σειρά μεγέθους:

Έχουμε ν = 13 παρατηρήσεις, οι οποίες σε αύξουσα σειρά είναι:
0 1 1 1 2 3 4 5 6 6 8 8 9 .

- Υπολογίζουμε τη διάμεσο, όπως προαναφέραμε:

Η διάμεσος είναι η έβδομη στη σειρά παρατήρηση, δηλαδή δ = 4.

- Υπολογίζουμε τη διάμεσο του πρώτου μισού των διατεταγμένων

παρατηρήσεων, δηλαδή των παρατηρήσεων που είναι αριστερά του δ. Η τιμή αυτή είναι το Q_1:
Η διάμεσος των παρατηρήσεων που είναι αριστερά του δ, δηλαδή των 0 1 1 1 2 3, είναι το $Q_1 = \frac{1+1}{2} = 1$.

• Υπολογίζουμε τη διάμεσο του δεύτερου μισού των διατεταγμένων παρατηρήσεων, δηλαδή των παρατηρήσεων που είναι δεξιά του δ. Η τιμή αυτή είναι το Q_3.

Η διάμεσος των παρατηρήσεων που είναι δεξιά του δ, δηλαδή των 5 6 6 8 8 9, είναι το $Q_3 = \frac{6+8}{2} = 7$.
(Όμως το ακριβές, σύμφωνα με τον ορισμό είναι $Q_3 = 6$).

Εκατοστημόρια σε Ομαδοποιημένα Δεδομένα

Ο υπολογισμός των εκατοστημορίων (ή τεταρτημορίων) σε ομαδοποιημένα δεδομένα γίνεται όπως και στη διάμεσο από το πολύγωνο αθροιστικών σχετικών συχνοτήτων. Στο σχήμα 13 δίνονται τα Q_1, $Q_2 = δ$, Q_3 και P_{10}, P_{90} για τα δεδομένα του πίνακα 9, από το οποίο βρίσκουμε κατά προσέγγιση:

$P_{10} = 162{,}5$, $Q_1 = 168$, $δ = 173$, $Q_3 = 178$ και $P_{90} = 184$.

ε) Επικρατούσα Τιμή (M_0)

Στην περίπτωση μη ομαδοποιημένων δεδομένων επικρατούσα τιμή ή κορυφή (mode) M_0 ορίζεται ως η

παρατήρηση με τη μεγαλύτερη συχνότητα. Είναι προφανές ότι η επικρατούσα τιμή μπορεί να οριστεί και στην περίπτωση ποιοτικών δεδομένων, ενώ τα άλλα μέτρα που είδαμε ορίζονται μόνο για ποσοτικά δεδομένα. Για παράδειγμα:

α) Η επικρατούσα τιμή (επικρατούσα απασχόληση) για την απασχόληση των μαθητών του πίνακα 7 στον ελεύθερο χρόνο τους είναι M_0 = “Μουσική”.

β) Η επικρατούσα τιμή του αριθμού των αδελφών των μαθητών στον πίνακα 6 είναι $M_0 = 1$, δηλαδή οι περισσότερες οικογένειες (55%) έχουν δύο παιδιά.

γ) Για να βρούμε την επικρατούσα τιμή των παρατηρήσεων 0 1 1 2 2 2

22-0088_I_c_math_bm_83-117_28b.indd 33 18/09/2017 09:58

3 4 4 4 5 5 7 8, κατασκευάζουμε τον παρακάτω πίνακα συχνοτήτων. Οι τιμές 2 και 4 είναι και οι δύο επικρατούσες τιμές, γιατί καθεμιά έχει συχνότητα 3. Βλέπουμε εδώ ότι η επικρατούσα τιμή μπορεί να μην είναι μοναδική. Όταν έχουμε δύο κορυφές, η αντίστοιχη κατανομή συχνοτήτων λέγεται δικόρυφη (bimodal), ενώ όταν έχουμε πολλές κορυφές λέγεται πολυκόρυφη (multimodal).

x_i	v_i
0	1
1	2
2	3
3	1
4	3
5	2
7	1
8	1

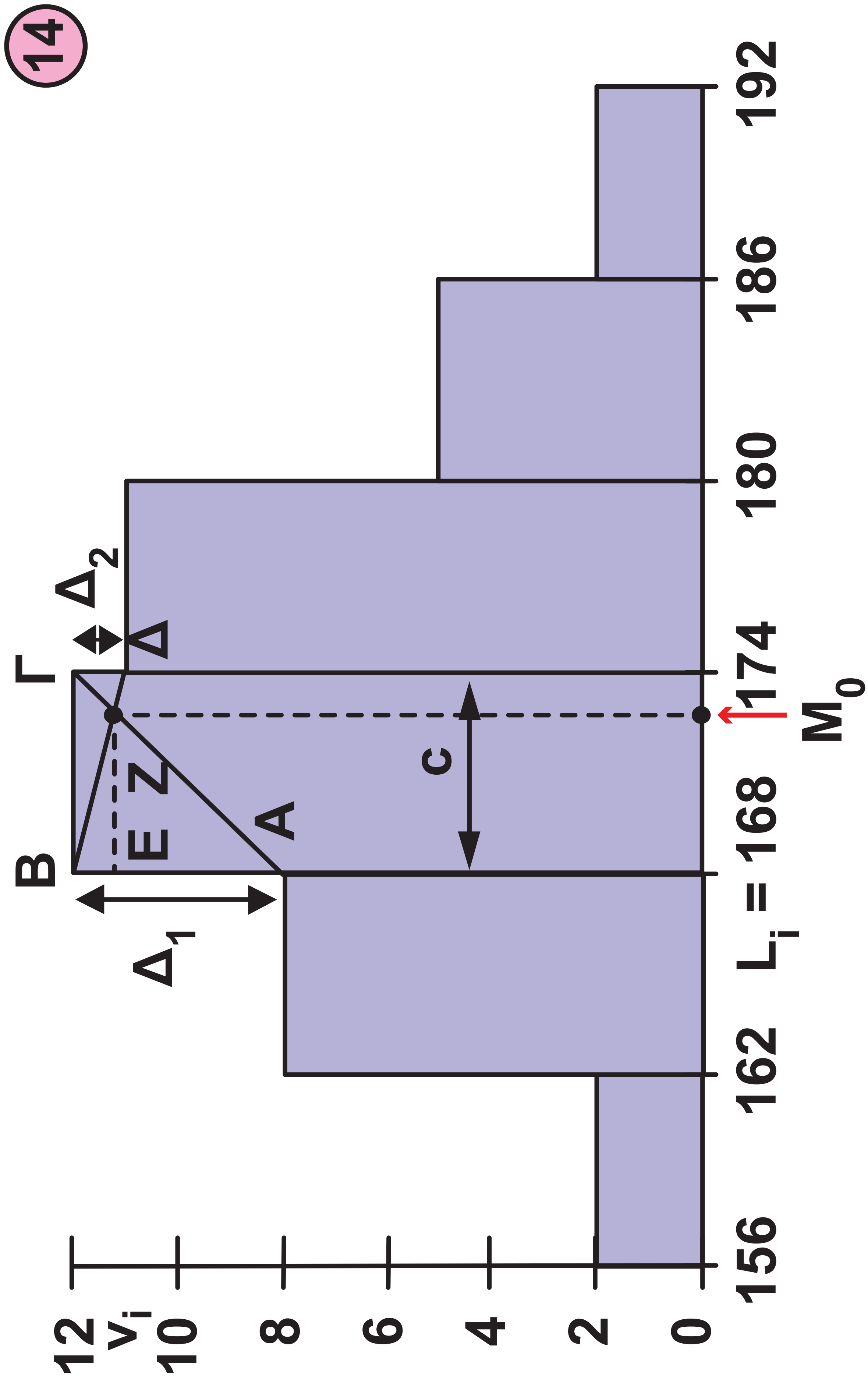
δ) Όταν όλες οι παρατηρήσεις είναι διαφορετικές, τότε λέμε ότι δεν υπάρχει επικρατούσα τιμή. Έτσι, για τις παρατηρήσεις 0, 1, 2, 7, 8, 9 δεν έχουμε επικρατούσα τιμή.

22-0088_I_c_math_bm_83-117_28b.indd 35 18/09/2017 09:58

Επικρατούσα Τιμή σε Ομαδοποιημένα Δεδομένα

Όταν έχουμε ομαδοποιημένα (ποσοτικά) δεδομένα σε **ισοπλατείς** κλάσεις, τότε βρίσκουμε πρώτα την **επικρατούσα κλάση** i, δηλαδή την κλάση με τη μεγαλύτερη συχνότητα. Υποθέτοντας, όπως έχουμε ήδη αναφέρει και προηγουμένως, ότι οι παρατηρήσεις στις κλάσεις κατανέμονται ομοιόμορφα, η επικρατούσα τιμή προσδιορίζεται, όπως φαίνεται στο παρακάτω σχήμα 14, ως η τετμημένη του σημείου τομής Ζ των ευθύγραμμων τμημάτων ΑΓ και ΒΔ. Στο σχήμα αυτό δίνεται η κορυφή για το ύψος των μαθητών του πίνακα 9. Κατά προσέγγιση η κορυφή (επικρατέστερο ύψος) είναι
$M_0 \approx 173$ cm.

22-0088_I_c_math_bm_83-117_28b.indd 36 18/09/2017 09:58


14
B
Γ
A
E
Z
Δ
Δ_1
Δ_2
c
v_i
12
10
8
6
4
2
0
156
162
L_i = 168
174
180
186
192
M_0

37 / 91

Μέτρα Διασποράς

Στα προηγούμενα είδαμε ότι τα μέτρα θέσης παρέχουν κάποια πληροφορία για την κατανομή ενός πληθυσμού. Αυτά όμως δεν επαρκούν, για να περιγράψουν πλήρως την κατανομή, όπως διαπιστώσαμε στην αρχή της § 2.3 συγκρίνοντας τις βαθμολογίες των μαθητών δύο τμημάτων Α και Β στα σχήματα 11(α), (β).

Ενώ οι βαθμολογίες των δύο τμημάτων Α και Β έχουν ίσες μέσες τιμές $\bar{x}_A = \bar{x}_B = 15$ και ίσες διαμέσους $\delta_A = \delta_B = 15$, είναι φανερό ότι οι κατανομές τους διαφέρουν σημαντικά ως προς τη μεταβλητότητά τους. Οι βαθμοί του τμήματος Α είναι περισσότερο "συγκεντρωμένοι" γύρω από τη μέση τιμή, ενώ, αντίθετα, οι

βαθμοί του τμήματος Β διασπείρονται περισσότερο, έχουν δηλαδή μεγάλες αποκλίσεις γύρω από τη μέση τιμή τους.
Παράλληλα λοιπόν με τα μέτρα θέσης κρίνεται απαραίτητη και η εξέταση κάποιων μέτρων διασποράς ή μεταβλητότητας, δηλαδή μέτρων που εκφράζουν τις αποκλίσεις των τιμών μιας μεταβλητής γύρω από τα μέτρα κεντρικής τάσης.
Τέτοια μέτρα λέγονται **μέτρα διασποράς** (measures of variation, dispersion measures). Τα σπουδαιότερα μέτρα διασποράς είναι το εύρος, η ενδοτεταρτημοριακή απόκλιση, η διακύμανση και η τυπική απόκλιση.

α) Εύρος (R)

Το απλούστερο από τα μέτρα διασποράς είναι το εύρος ή κύμανση (range) (R), που ορίζεται ως η διαφορά της ελάχιστης παρατήρησης από τη μέγιστη παρατήρηση, δηλαδή:

Εύρος R = Μεγαλύτερη παρατήρηση − Μικρότερη παρατήρηση

Έτσι, για τη βαθμολογία του τμήματος Α το εύρος είναι $R_A = 18 - 13 = 5$, ενώ για το τμήμα $R_B = 20 - 10 = 10$, τιμές που επιβεβαιώνουν ότι πράγματι στο τμήμα Β έχουμε μεγαλύτερη διασπορά βαθμολογίας παρά στο τμήμα Α.

Όταν έχουμε ομαδοποιημένα δεδομένα, το εύρος δίνεται από τη

διαφορά του κατώτερου ορίου της πρώτης κλάσης από το ανώτερο όριο της τελευταίας κλάσης. Το εύρος των υψών των μαθητών του δείγματος στον πίνακα 9 είναι $R = 192 - 156 = 36$. Προφανώς, το εύρος σε ομαδοποιημένα δεδομένα μπορεί να διαφέρει ελαφρώς από τα αντίστοιχα δεδομένα πριν αυτά ομαδοποιηθούν. Για παράδειγμα, το εύρος των υψών στον πίνακα 8, πριν αυτά ομαδοποιηθούν, βρήκαμε ότι είναι $R = 191 - 156 = 35$.
Το εύρος είναι ένα αρκετά απλό μέτρο, που υπολογίζεται εύκολα δε θεωρείται όμως αξιόπιστο μέτρο διασποράς, γιατί βασίζεται μόνο στις δυο ακραίες παρατηρήσεις.

β) Ενδοτεταρτημοριακό Εύρος (Q)

Το **ενδοτεταρτημοριακό εύρος** (interquartile range) είναι η διαφορά του πρώτου τεταρτημορίου Q_1 από το τρίτο τεταρτημόριο Q_3, δηλαδή:

$$Q = Q_3 - Q_1$$

Στο μεταξύ τους διάστημα περιλαμβάνεται το 50% των παρατηρήσεων. Επομένως όσο μικρότερο είναι αυτό το διάστημα, τόσο μεγαλύτερη θα είναι η συγκέντρωση των τιμών και άρα μικρότερη η διασπορά των τιμών της μεταβλητής.

Από τα δεδομένα του σχήματος 13 βρήκαμε κατά προσέγγιση $Q_1 = 168$, $Q_3 = 178$ επομένως το ενδοτεταρτημοριακό εύρος είναι $Q = 10$. Δηλαδή το 50% των μαθητών έχουν ύψος μεταξύ 168 και 178 cm.

γ) Διακύμανση (s^2)

Ένας άλλος τρόπος για να υπολογίσουμε τη διασπορά των παρατηρήσεων $t_1, t_2, ..., t_\nu$ μιας μεταβλητής Χ θα ήταν να αφαιρέσουμε τη μέση τιμή $\bar{x}$ από κάθε παρατήρηση και να βρούμε τον αριθμητικό μέσο των διαφορών αυτών, δηλαδή τον αριθμό:

$$\frac{(t_1-\bar{x})+(t_2-\bar{x})+\dots+(t_\nu-\bar{x})}{\nu}=\frac{\sum_{i=1}^{\nu}(t_i-\bar{x})}{\nu}.$$

Ο αριθμός όμως αυτός είναι ίσος με μηδέν, αφού

$$\frac{(t_1-\bar{x})+(t_2-\bar{x})+\ldots+(t_\nu-\bar{x})}{\nu}=$$

$$=\frac{t_1+t_2+\ldots+t_\nu}{\nu}-\frac{\nu\bar{x}}{\nu}=\bar{x}-\bar{x}=0.$$

Γι’ αυτό, ως ένα μέτρο διασποράς παίρνουμε τον μέσο όρο των τετραγώνων των αποκλίσεων των t_i από τη μέση τιμή τους $\bar{x}$. Το μέτρο αυτό καλείται **διακύμανση** ή **διασπορά** (variance) και ορίζεται από τη σχέση

$$s^2=\frac{1}{\nu}\sum_{i=1}^{\nu}(t_i-\bar{x})^2 \qquad (1)$$

Ο τύπος αυτός αποδεικνύεται ότι μπορεί να πάρει την ισοδύναμη μορφή:

$$s^2 = \frac{1}{v}\left\{\sum_{i=1}^{v} t_i^2 - \frac{\left(\sum_{i=1}^{v} t_i\right)^2}{v}\right\} \qquad (2)$$

η οποία διευκολύνει σημαντικά τους υπολογισμούς κυρίως όταν η μέση τιμή $\bar{x}$ δεν είναι ακέραιος αριθμός. Όταν έχουμε πίνακα συχνοτήτων ή ομαδοποιημένα δεδομένα, η διακύμανση ορίζεται από τη σχέση:

$$s^2 = \frac{1}{v}\sum_{i=1}^{\kappa}(x_i - \bar{x})^2 v_i \qquad (3)$$

ή την ισοδύναμη μορφή:

$$s^2 = \frac{1}{\nu}\left\{\sum_{i=1}^{\kappa} x_i^2 \nu_i - \frac{\left(\sum_{i=1}^{\kappa} x_i \nu_i\right)^2}{\nu}\right\}. \qquad (4)$$

όπου $x_1, x_2, ..., x_\kappa$ οι τιμές της μεταβλητής (ή τα κέντρα των κλάσεων) με αντίστοιχες συχνότητες $\nu_1, \nu_2, ..., \nu_\kappa$.

Για παράδειγμα, η διακύμανση της βαθμολογίας των μαθητών του τμήματος Α είναι σύμφωνα με την (1)

$$s_A^2 = \frac{(13-15)^2 + (13-15)^2 + (14-15)^2 + \ldots + (18-15)^2}{10} =$$

$$= \frac{20}{10} = 2,$$

ενώ για τους μαθητές του τμήματος Β βρίσκουμε $s_B^2 = 6,6$, που επιβεβαιώνει τη διαπίστωσή μας ότι η βαθμολογία των μαθητών του τμήματος Β παρουσιάζει μεγαλύτερη μεταβλητότητα από τη βαθμολογία των μαθητών του τμήματος Α.

Ομοίως, η διακύμανση του ύψους των μαθητών για τα ομαδοποιημένα δεδομένα του πίνακα 9, υπολογίζεται σύμφωνα με τον τύπο (3), όπως φαίνεται στον επόμενο πίνακα:

Κλάσεις [-)	Κεντρικές τιμές x_i	Συχνότητα v_i	
156-162	159	2	
162-168	165	8	
168-174	171	12	
174-180	177	11	
180-186	183	5	
186-192	189	2	
Σύνολο:		$v = 40$	

	x_i^2	$x_i v_i$	$x_i^2 v_i$
	25281	318	50562
	27225	1320	217800
	29241	2052	350892
	31329	1942	344619
	33489	915	167445
	35721	378	71442
	—	$\Sigma x_i v_i = 6930$	$\Sigma x_i^2 v_i = 1202760$

Επομένως:

$$s^2 = \frac{1}{v}\left\{ \sum_{i=1}^{\kappa} x_i^2 v_i - \frac{\left(\sum_{i=1}^{\kappa} x_i v_i \right)^2}{v} \right\} =$$

$$= \frac{1}{40}\left\{ 1202776 - \frac{6930^2}{40} \right\} = 53{,}4$$

Εάν υπολογίσουμε τη διακύμανση από τα μη ομαδοποιημένα δεδομένα του πίνακα 8, βρίσκουμε $s^2 = 50{,}9$. Η διαφορά αυτή οφείλεται στην απώλεια πληροφορίας λόγω ομαδοποίησης των παρατηρήσεων.

δ) Τυπική Απόκλιση (s)

Η διακύμανση είναι μια αξιόπιστη παράμετρος διασποράς, αλλά έχει ένα μειονέκτημα. Δεν εκφράζεται με τις μονάδες με τις οποίες εκφράζονται οι παρατηρήσεις. Για παράδειγμα, αν οι παρατηρήσεις εκφράζονται σε cm, η διακύμανση εκφράζεται σε cm^2. Αν όμως πάρουμε τη θετική τετραγωνική ρίζα της διακύμανσης, θα έχουμε ένα μέτρο διασποράς που θα εκφράζεται με την ίδια μονάδα μέτρησης του χαρακτηριστικού, όπως ακριβώς είναι και όλα τα άλλα μέτρα θέσης, που εξετάσαμε έως τώρα. Η ποσότητα αυτή λέγεται τυπική απόκλιση (standard deviation), συμβολίζεται με s και δίνεται από τη σχέση:

22-0088_I_c_math_bm_83-117_28b.indd 51 18/09/2017 09:58

$$s = \sqrt{s^2}$$

Η τυπική απόκλιση για το ύψος των μαθητών του πίνακα 4 είναι από το προηγούμενο παράδειγμα

$s = \sqrt{53,4} = 7,3$ cm, αν αυτή υπολογιστεί από τα ομαδοποιημένα δεδομένα του πίνακα 9, ή

$s = \sqrt{50,9} = 7,13$ cm, αν υπολογιστεί από τα μη ομαδοποιημένα δεδομένα του πίνακα 8.

Αξίζει να σημειωθεί ότι αν η καμπύλη συχνοτήτων για το χαρακτηριστικό που εξετάζουμε είναι κανονική ή περίπου κανονική, τότε η τυπική απόκλιση s έχει τις παρακάτω ιδιότητες:

i) το 68% περίπου των παρατηρήσεων βρίσκεται στο διάστημα

$$(\bar{x} - s, \bar{x} + s)$$

ii) το 95% περίπου των παρατηρήσεων βρίσκεται στο διάστημα

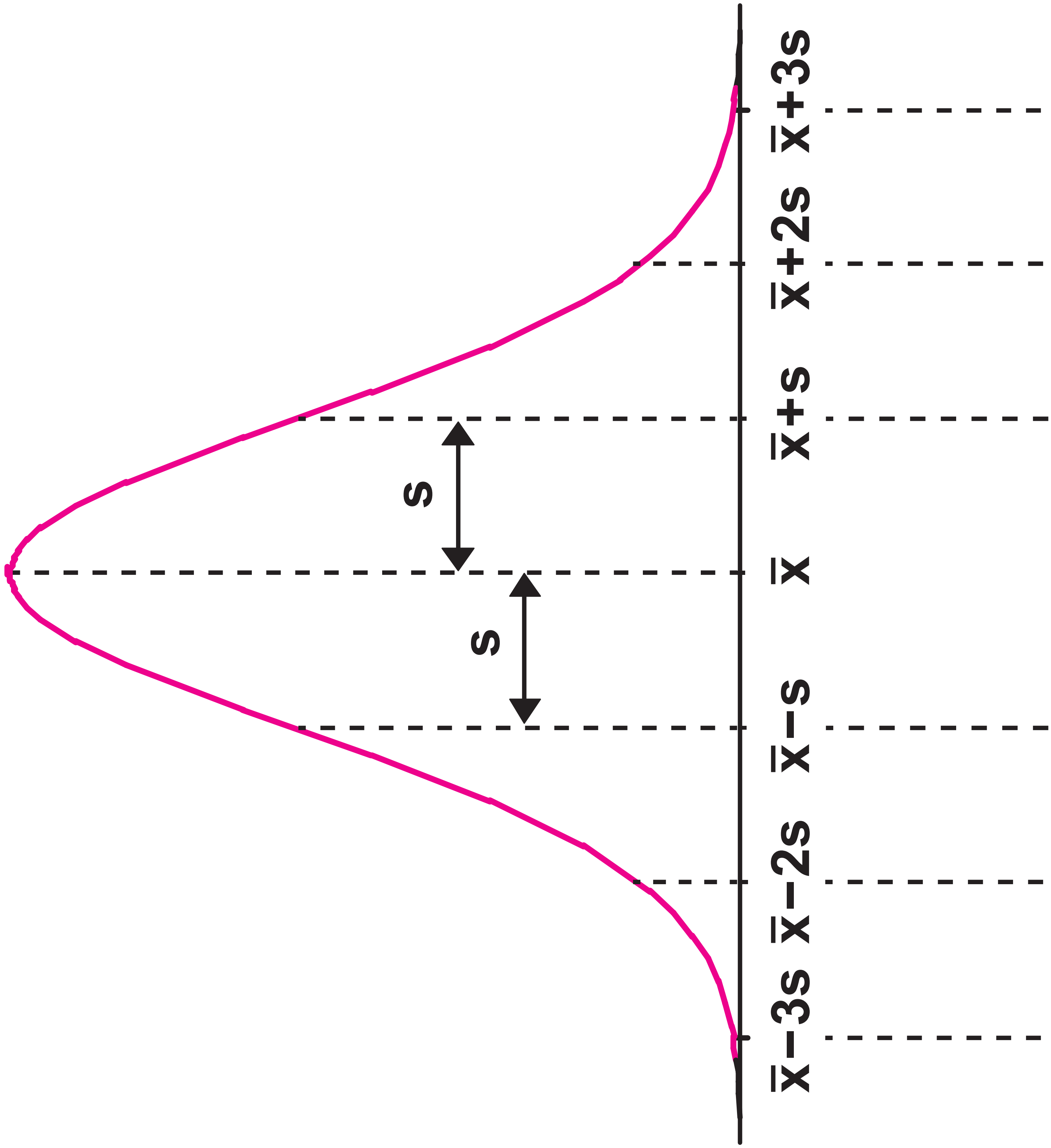
$$(\bar{x} - 2s, \bar{x} + 2s)$$

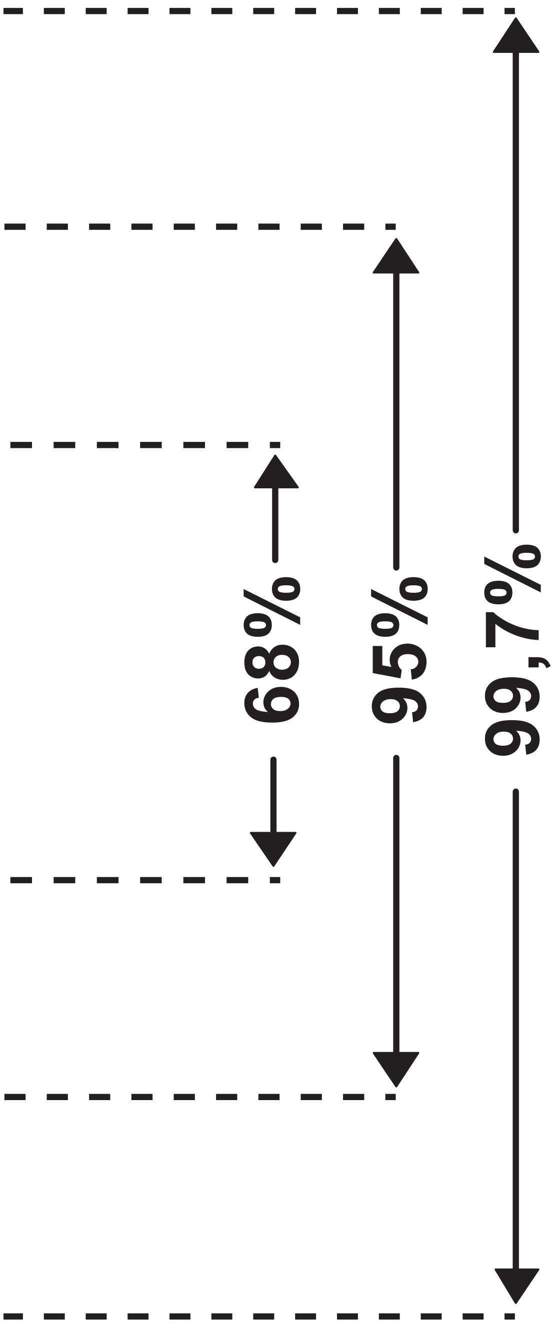
iii) το 99,7% περίπου των παρατηρήσεων βρίσκεται στο διάστημα

$$(\bar{x} - 3s, \bar{x} + 3s)$$

iv) το εύρος ισούται περίπου με έξι τυπικές αποκλίσεις, δηλαδή $R \approx 6s$.

15




68%
95%
99,7%

Συντελεστής Μεταβολής (CV)

Έστω ότι από ένα δείγμα είκοσι μαθητών της Α΄ Γυμνασίου βρήκαμε μέσο βάρος $\bar{x}_A$ = 40 kgr και τυπική απόκλιση s_A = 6 kgr, ενώ από ένα δεύτερο δείγμα τριάντα μαθητών της Γ΄ Λυκείου βρήκαμε μέσο βάρος $\bar{x}_B$ = 75 kgr και τυπική απόκλιση s_B = 6 kgr. Όπως αντιλαμβανόμαστε, είναι λάθος να πούμε ότι το βάρος των μαθητών του Λυκείου έχει τον ίδιο βαθμό μεταβλητότητας με το βάρος των μαθητών του Γυμνασίου, καθόσον η βαρύτητα που έχουν τα 6 kgr στο μέσο βάρος των 40 kgr είναι διαφορετική από αυτήν που έχουν στο μέσο βάρος των 75 kgr. Ακόμη, ας υποθέσουμε ότι ο μέσος

μισθός των υψηλόβαθμων υπαλλήλων μιας εταιρείας Α είναι $\bar{x}_A = 2.500$ € με τυπική απόκλιση $s_A = 420$ €, ενώ για τους υπαλλήλους μιας εταιρείας Β είναι $\bar{x}_B = 1.400$ \$ με τυπική απόκλιση $s_B = 350$ \$. Στην περίπτωση αυτή έχουμε διαφορετικές μονάδες μέτρησης του μισθού, επομένως οι διασπορές των παρατηρήσεων δεν είναι άμεσα συγκρίσιμες.

Ένα μέτρο με το οποίο μπορούμε να ξεπεράσουμε τις παραπάνω δυσκολίες και το οποίο μας βοηθά στη σύγκριση ομάδων τιμών, που είτε εκφράζονται σε διαφορετικές μονάδες μέτρησης είτε εκφράζονται στην ίδια μονάδα μέτρησης, αλλά έχουν σημαντικά διαφορετικές μέσες τιμές, είναι ο **συντελεστής μεταβολής** ή **συντελεστής μεταβλητότητας** (coefficient of variation), ο οποίος για $\bar{x} \neq 0$ ορίζεται από το λόγο:

$$CV = \frac{\text{τυπική απόκλιση}}{\text{μέση τιμή}} = \frac{s}{\bar{x}}$$

Αν $\bar{x} < 0$, τότε αντί της $\bar{x}$ χρησιμοποιούμε την $|\bar{x}|$.
Ο συντελεστής μεταβολής είναι ανεξάρτητος από τις μονάδες μέτρησης, εκφράζεται επί τοις εκατό και παριστάνει ένα μέτρο **σχετικής διασποράς** των τιμών και όχι της απόλυτης διασποράς, όπως έχουμε δει έως τώρα.
Για το πρώτο παράδειγμα του βάρους έχουμε συντελεστή μεταβολής για τις δύο ομάδες μαθητών:

$$CV_A = \frac{s_A}{\bar{x}_A} = \frac{6}{40} = 0{,}15 = 15\% \text{ και}$$

$$CV_B = \frac{s_B}{\bar{x}_B} = \frac{6}{75} = 0{,}08 = 8\%$$

δηλαδή, ο βαθμός διασποράς του βάρους των μαθητών Γυμνασίου είναι μεγαλύτερος από το βαθμό διασποράς του βάρους των μαθητών Λυκείου (για τα συγκεκριμένα δείγματα).
Ανάλογα συμπεράσματα βγάζουμε και για το δεύτερο παράδειγμα, όπου βρίσκουμε $CV_A = 16{,}8\%$ και $CV_B = 25\%$. Παρ' όλο που η τυπική απόκλιση των μισθών στην εταιρεία Α είναι μεγαλύτερη από την τυπική απόκλιση στην εταιρεία Β,

ο συντελεστής μεταβολής δίνει μεγαλύτερη σχετική διασπορά στην εταιρεία Β. Αυτό μεταφράζεται στο να λέμε ότι έχουμε μεγαλύτερη **ομοιογένεια** μισθών στην εταιρεία Α παρά στη Β.

Γενικά δεχόμαστε ότι ένα δείγμα τιμών μιας μεταβλητής θα είναι ομοιογενές, εάν ο συντελεστής μεταβολής δεν ξεπερνά το 10%.

ΕΦΑΡΜΟΓΕΣ

1. Ο παρακάτω πίνακας συχνοτήτων δίνει την κατανομή του χρόνου X (σε sec) 60 μαθητών που χρειάστηκαν, για να τρέξουν μια δεδομένη απόσταση. Να υπολογιστούν:

α) ο μέσος, ο διάμεσος και ο επικρατέστερος χρόνος για την κάλυψη της συγκεκριμένης απόστασης,

β) η τυπική απόκλιση,

γ) σε πόσο χρόνο από της στιγμή της εκκίνησης κάλυψε την απόσταση το 25% των μαθητών.

x_i	v_i
50	4
55	6
60	8
65	12
70	14
75	10
80	6

ΛΥΣΗ

α) • Για τον υπολογισμό της μέσης τιμής συμπληρώνουμε τις τρεις πρώτες στήλες του παρακάτω πίνακα:

x_i	v_i	$x_i v_i$	$x_i^2 v_i$
50	4	200	10000
55	6	330	18150
60	8	480	28800
65	12	780	50700
70	14	980	68600
75	10	750	56250
80	6	480	38400
Σύνολο	$v = 60$	$\sum x_i v_i = 4000$	$\sum x_i^2 v_i = 270900$

Επομένως, ο μέσος χρόνος για την κάλυψη της συγκεκριμένης απόστασης είναι:

$$\bar{x} = \frac{\sum x_i v_i}{v} = \frac{4000}{60} \approx 66,67 \text{ sec.}$$

- Έχουμε $v = 60$ παρατηρήσεις σε αύξουσα σειρά, άρα η διάμεσος είναι ο μέσος όρος της 30ής και 31ης παρατήρησης, δηλαδή ο μέσος όρος των παρατηρήσεων 65 και 70, άρα $\delta = \frac{65+70}{2} = 67,5$ sec.

- Η επικρατούσα τιμή είναι η τιμή με τη μεγαλύτερη συχνότητα, άρα $M_0 = 70$ sec.

β) Για τον υπολογισμό της τυπικής απόκλισης είναι προτιμότερο να εφαρμόσουμε τη σχέση (4), μιας και

η μέση τιμή δεν είναι ακέραιος αριθμός.
Με βάση τον παραπάνω πίνακα η διακύμανση της μεταβλητής Χ είναι:

$$s^2 = \frac{1}{v}\left\{ \sum x_i^2 v_i - \frac{\left(\sum x_i v_i\right)^2}{v} \right\} =$$

$$= \frac{1}{60}\left\{ 270900 - \frac{4000^2}{60} \right\} = 70,56 \text{ sec}^2$$

και η τυπική απόκλιση
$s = \sqrt{70,56} = 8,4$ sec.

γ) Θέλουμε να υπολογίσουμε το πρώτο τεταρτημόριο, Q_1. Αριστερά της διαμέσου δ = 67,5 έχουμε

30 παρατηρήσεις. Η διάμεσος αυτών των 30 πρώτων παρατηρήσεων είναι το ημιάθροισμα της 15ης και 16ης παρατήρησης, δηλαδή

$$Q_1 = \frac{(60+60)}{2} = 60\,\text{sec}.$$ Δηλαδή, ύστερα από μία ώρα από τη στιγμή της εκκίνησης το 25% των μαθητών κάλυψαν τη συγκεκριμένη απόσταση.

2. Να αποδειχτεί ότι η συνάρτηση

$$f(\lambda) = \sum_{i=1}^{\nu}(x_i - \lambda)^2 =$$

$$= (x_1 - \lambda)^2 + (x_2 - \lambda)^2 + \ldots + (x_\nu - \lambda)^2$$

γίνεται ελάχιστη, όταν $\lambda = \bar{x}$.

ΛΥΣΗ

Λαμβάνοντας την πρώτη παράγωγο της f(λ), βρίσκουμε $f'(\lambda) = -2(x_1 - \lambda) - 2(x_2 - \lambda) - \ldots -2(x_\nu - \lambda)$.

Έχουμε διαδοχικά:

$$f'(\lambda) = 0$$

$$x_1 - \lambda + x_2 - \lambda + \ldots + x_\nu - \lambda = 0$$

$$x_1 + x_2 + \ldots + x_\nu - \nu\lambda = 0$$

$$\lambda = \frac{x_1 + x_2 + \ldots + x_\nu}{\nu} = \bar{x}.$$

Η δεύτερη παράγωγος της f(λ) είναι:

$$f''(\lambda) = \underbrace{2 + 2 + \ldots + 2}_{\nu} = 2\nu$$

και επειδή $f''(\bar{x}) = 2ν > 0$, συνεπάγεται ότι για $λ = \bar{x}$ η $f(λ)$ γίνεται ελάχιστη.

3. Έστω $x_1, x_2, ..., x_ν$ ν παρατηρήσεις με μέση τιμή $\bar{x}$ και τυπική απόκλιση s_x.

α) Αν $y_1, y_2, ..., y_ν$ είναι οι παρατηρήσεις που προκύπτουν αν προσθέσουμε σε καθεμιά από τις $x_1, x_2, ..., x_ν$ μια σταθερά c, να δειχτεί ότι:

i) $\bar{y} = \bar{x} + c$ ii) $s_y = s_x$

β) Αν $y_1, y_2, ..., y_ν$ είναι οι παρατηρήσεις που προκύπτουν αν πολλαπλασιάσουμε τις $x_1, x_2, ..., x_ν$ επί μια σταθερά c, να αποδειχτεί ότι:

i) $\bar{y} = c\,\bar{x}$, ii) $s_y = |c|s_x$

ΑΠΟΔΕΙΞΗ

α) Έχουμε $y_i = x_i + c$, $i = 1, 2, \ldots, \nu$ επομένως:

i) $$\bar{y} = \frac{y_1 + y_2 + \ldots + y_\nu}{\nu} = \frac{x_1 + c + x_2 + c + \ldots + x_\nu + c}{\nu} =$$

$$= \frac{x_1 + x_2 + \ldots + x_\nu}{\nu} + \frac{\nu c}{\nu} = \bar{x} + c.$$

ii) $$s_y^2 = \frac{(y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \ldots + (y_\nu - \bar{y})^2}{\nu} =$$

$$= \frac{(x_1 + c - \bar{x} - c)^2 + (x_2 + c - \bar{x} - c)^2 + \ldots + (x_\nu + c - \bar{x} - c)^2}{\nu} =$$

$$= \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \ldots + (x_\nu - \bar{x})^2}{\nu} = s_x^2.$$

Άρα και $s_y = s_x$.

β) Έχουμε $y_i = cx_i$, $i = 1, 2, \ldots, \nu$, επομένως:

i) $$\bar{y} = \frac{y_1 + y_2 + \ldots + y_\nu}{\nu} = \frac{cx_1 + cx_2 + \ldots + cx_\nu}{\nu} =$$

$$= c\frac{x_1 + x_2 + \ldots + x_\nu}{\nu} = c\bar{x}$$

22-0088_I_c_math_bm_83-117_28b.indd 69
18/09/2017 09:58

ii) $s_y^2 = \dfrac{(y_1-\bar{y})^2+(y_2-\bar{y})^2+\ldots+(y_\nu-\bar{y})^2}{\nu} =$

$$= \frac{(cx_1-c\bar{x})^2+(cx_2-c\bar{x})^2+\ldots+(cx_\nu-c\bar{x})^2}{\nu} =$$

$$= \frac{c^2[(x_1-\bar{x})^2+(x_2-\bar{x})^2+\ldots+(x_\nu-\bar{x})^2]}{\nu} = c^2 s_x^2$$

Άρα και $s_y = |c|s_x$.

ΑΣΚΗΣΕΙΣ

Α΄ ΟΜΑΔΑΣ

1. Έξι διαδοχικοί άρτιοι αριθμοί έχουν μέση τιμή 15. Να βρείτε τους αριθμούς και τη διάμεσό τους.

2. Έχουμε ένα δείγμα ν = 10 παρατηρήσεων, όπου κάθε παρατήρηση μπορεί να είναι 1, 2 ή 3. Είναι δυνατό η μέση τιμή να είναι α) 1 β) 4 γ) 1,8;

3. Ένας επενδυτής επένδυσε το ίδιο ποσό χρημάτων σε 8 διαφορετικές μετοχές στο χρηματιστήριο. Κατά τη διάρκεια του περασμένου έτους οι μετοχές

είχαν τις παρακάτω εκατοστιαίες μεταβολές στην αξία τους: 5, 16, −10, 0, 27, 14, −20, 34. Να βρεθεί η μέση εκατοστιαία απόδοση της επένδυσης.

4. Το μέσο ύψος 9 καλαθοσφαιριστών μιας ομάδας είναι 205 cm.
   α) Για να "ψηλώσει" την ομάδα ο προπονητής πήρε έναν ακόμη παίκτη με ύψος 216 cm. Ποιο είναι το μέσο ύψος της ομάδας τώρα;

   β) Εάν ήθελε να "ψηλώσει" την ομάδα στα 208 cm, πόσο ύψος έπρεπε να έχει ο καλαθοσφαιριστής που πήρε;

5. Η μέση ηλικία 18 αγοριών και 12 κοριτσιών μιας τάξης είναι 15,4 χρόνια. Εάν η μέση ηλικία των αγοριών είναι 15,8 χρόνια, να βρείτε τη μέση ηλικία των κοριτσιών.

6. Σε μια κάλπη υπάρχουν άσπρες, μαύρες, κόκκινες και πράσινες μπάλες σε αναλογία 10%, 20%, 30% και 40% αντίστοιχα. Μια άσπρη μπάλα έχει βάρος 10 gr, μια μαύρη 11 gr, μια κόκκινη 12 gr και μια πράσινη 13 gr. Να βρείτε τη μέση τιμή, τη διάμεσο και την επικρατούσα τιμή του βάρους για όλες τις μπάλες, αν ξέρουμε ότι στην κάλπη υπάρχουν

α) 10 μπάλες, β) 20 μπάλες,
γ) δε γνωρίζουμε πόσες μπάλες υπάρχουν στην κάλπη.

7. Η επίδοση ενός μαθητή σε πέντε μαθήματα είναι 12, 10, 16, 18, 14.
α) Να βρείτε τη μέση επίδοση.
β) Αν τα μαθήματα είχαν συντελεστές στάθμισης 2, 3, 1, 1 και 3, ποια θα ήταν η μέση επίδοση; Σε ποια μαθήματα έπρεπε να δώσει ιδιαίτερη προσοχή ο μαθητής;

8. Μία εταιρεία απασχολεί 5 υπαλλήλους στο Τμήμα Α με μέσο (μηνιαίο) μισθό 1249 ευρώ, 6 υπαλλήλους στο Τμήμα Β με μέσο μισθό 1280

ευρώ, και 4 υπαλλήλους στο Τμήμα Γ με μέσο μισθό 1360 ευρώ. Ποιος είναι ο μέσος μισθός όλων των υπαλλήλων;

9. Η μέση τιμή και η διάμεσος πέντε αριθμών είναι 6. Οι τρεις από αυτούς είναι οι 5, 8, 9. Να βρείτε τους άλλους δύο.

10. Στον παρακάτω πίνακα δίνονται οι τιμές μιας μεταβλητής Χ με τις αντίστοιχες συχνότητές τους. Η πέμπτη συχνότητα χάθηκε! Μπορείτε να την "ανακαλύψετε", εάν γνωρίζετε ότι

α) η μέση τιμή είναι 4,4,

β) η διάμεσος είναι το 4,5,

γ) υπάρχουν δύο επικρατούσες τιμές;

x_i	v_i
2	1
3	3
4	1
5	2
6	;
7	1

11. Για την κατανομή του βαθμού των Μαθηματικών της Β΄ τάξης των 40 μαθητών και μαθητριών της Γ΄ Λυκείου του πίνακα 4 να βρείτε:
α) τη μέση τιμή,
β) τη διάμεσο,
γ) την επικρατούσα τιμή,

δ) το πρώτο και το τρίτο τεταρτημόριο και να ερμηνεύσετε τα αποτελέσματα.

12. Ο παρακάτω πίνακας δίνει τον αριθμό των επισκέψεων 40 μαθητών σε διάφορα μουσεία της χώρας κατά τη διάρκεια ενός έτους.

Επισκέψεις	Συχνότητα
[0-2)	8
2-4	12
4-6	10
6-8	6
8-10	4

Να υπολογιστούν:

α) η μέση τιμή,
β) η επικρατούσα τιμή,
γ) η διάμεσος,
δ) το πρώτο και τρίτο τεταρτημόριο.

13. Το μέσο ύψος των 30 μαθητών και μαθητριών μιας τάξης είναι 170 cm. Ποιο θα είναι το μέσο ύψος της τάξης:
α) αν φύγει ένας μαθητής με ύψος 180 cm,
β) αν έρθει μια νέα μαθήτρια με ύψος 170 cm,
γ) αν φύγει ένας μαθητής με ύψος 180 cm και έλθει μια μαθήτρια με ύψος 170 cm;

14. Καθεμία από τις παρακάτω λίστες δεδομένων έχουν μέση τιμή 50.

α) Σε ποια λίστα υπάρχει (i) μεγαλύτερη (ii) μικρότερη διασπορά παρατηρήσεων; (Να μη γίνουν πράξεις).

0, 20, 40, 50, 60, 80, 100

0, 48, 49, 50, 51, 52, 100

0, 1, 2, 50, 98, 99, 100.

β) Μπορεί να χρησιμοποιηθεί για σύγκριση των δεδομένων αυτών το εύρος;

15. Η βαθμολογία δέκα μαθητών σε ένα διαγώνισμα ήταν: 7, 11, 10, 13, 15, 3, 12, 11, 4, 14. Να υπολογίσετε:

α) τη μέση τιμή, την επικρατούσα τιμή και τη διάμεσο,

β) τα Q_1 και Q_3,

γ) το εύρος, την τυπική από-κλιση και το συντελεστή μεταβολής.

16. **Να υπολογιστεί η τυπική απόκλιση των δεδομένων της άσκησης 12.**

17. **Ο μέσος χρόνος που χρειάζο-νται οι μαθητές ενός σχολείου να πάνε το πρωί από το σπί-τι τους μέχρι το σχολείο είναι 10 λεπτά με τυπική απόκλι-ση 2 λεπτά. Υποθέτοντας ότι έχουμε περίπου κανονική κα-τανομή, να βρείτε κατά προ-σέγγιση το ποσοστό των μα-θητών που χρειάζονται:**

α) κάτω από 8 λεπτά
β) πάνω από 14 λεπτά
γ) το πολύ 10 λεπτά
δ) από 6 έως 12 λεπτά
για να πάνε στο σχολείο τους.

18. Να υπολογίσετε τη μέση τιμή και τη διάμεσο για τα παρακάτω δείγματα δεδομένων και να σχολιάσετε τα αποτελέσματα:
α) 1 2 6
β) 2 4 12
γ) 11 12 16
δ) 12 14 22.

19. Να υπολογίσετε τη διακύμανση και την τυπική απόκλιση για καθεμιά από τις παρακάτω λίστες δεδομένων. Συγκρίνοντας τα δεδομένα και τα αποτελέσματα τι συμπέρασμα βγάζετε;

α) 1, 3, 4, 5, 7

β) 3, 9, 12, 15, 21

γ) 6, 8, 9, 10, 12

δ) −1, −3, −4, −5, −7.

20. Οι μαθητές του Γ2 ξόδεψαν ετησίως κατά μέσο όρο 625 ευρώ αγοράζοντας διάφορα τρόφιμα από το κυλικείο του σχολείου. Εάν ο συντελεστής μεταβολής είναι 27,2%, να βρείτε την τυπική απόκλιση.

22-0088_I_c_math_bm_83-117_28b.indd 82 18/09/2017 09:58

Εάν επιπλέον γνωρίζετε ότι το $\sum x_i^2 = 11.746.700$, πόσοι είναι οι μαθητές του Γ2;

Β΄ ΟΜΑΔΑΣ

1. Η βαθμολογία 50 μαθητών στην Ιστορία κυμαίνεται από 10 μέχρι 20 (κανένας δεν είναι κάτω από τη βάση). Γνωρίζουμε επίσης ότι πέντε μαθητές έχουν βαθμό κάτω από 12, δεκαπέντε κάτω από 14, πέντε μεγαλύτερο ή ίσο του 18 και δεκαπέντε μεγαλύτερο ή ίσο του 16.

α) Να παρασταθούν τα δεδομένα σε έναν πίνακα συχνοτήτων.

β) Να υπολογίσετε: i) τη μέση τιμή, ii) τη διάμεσο.

γ) Εάν στο 5% των μαθητών με την καλύτερη επίδοση δοθεί έπαινος, πόσο βαθμό πρέπει να έχει κάποιος μαθητής για να πάρει έπαινο;

2. Η μέση τιμή και η διακύμανση των 5 τιμών ενός δείγματος είναι $\bar{x} = 4$ και $s^2 = 10$, αντίστοιχα. Εάν, για τις τέσσερις τιμές ισχύει $\sum_{i=1}^{4}(x_i - \bar{x})^2 = 14$, να βρεθεί η πέμπτη τιμή.

3. **Ένας μαθητής αγόρασε 10 βιβλία που κόστιζαν χωρίς Φ.Π.Α. 15, 9, 6, 18, 21, 6, 18, 27, 9, 12 ευρώ αντίστοιχα.**

α) Ποια είναι η μέση, η διάμεση και η επικρατούσα αξία (τιμή) των βιβλίων;

β) Πώς μεταβάλλονται οι απαντήσεις του ερωτήματος (α), αν προσθέσουμε και το Φ.Π.Α., που είναι 18%;

γ) Αν ο μαθητής πληρώσει επί πλέον 0,3 ευρώ (χωρίς Φ.Π.Α.) για το ντύσιμο κάθε βιβλίου, πώς διαμορφώνονται τώρα οι απαντήσεις στο ερώτημα (β);

4. Να δείξετε ότι εάν από όλες τις τιμές 0, 2, 4, 6, 8, 10 και 12 ενός δείγματος αφαιρέσουμε τη μέση τιμή τους και διαιρέσουμε με την τυπική τους απόκλιση, τότε οι τιμές που θα προκύψουν θα έχουν μέση τιμή 0 και διασπορά 1.

5. Στο παρακάτω σχήμα φαίνονται τα ύψη των πωλήσεων σε χιλιάδες ευρώ που έγιναν από τους πωλητές μιας εταιρείας κατά τη διάρκεια ενός έτους.

22-0088_I_c_math_bm_83-117_28b.indd 86 18/09/2017 09:58

α) Πόσοι είναι οι πωλητές;
β) Πόσοι πωλητές έκαναν πωλήσεις πάνω από 5 χιλιάδες ευρώ;
γ) Να βρεθεί η επικρατούσα τιμή των πωλήσεων.
δ) Να κατασκευάσετε τον πίνακα συχνοτήτων και να

υπολογίσετε τη μέση τιμή και τη διακύμανση.

ε) Από το πολύγωνο αθροιστικών σχετικών συχνοτήτων να εκτιμήσετε τα τρία τεταρτημόρια Q_1, Q_2, Q_3.

6. Στον παρακάτω πίνακα δίνεται η κατανομή της ηλικίας των ατόμων μιας πόλης. Να υπολογίσετε:

α) Την τυπική απόκλιση και το συντελεστή μεταβολής.

β) Το ενδοτεταρτημοριακό εύρος.

22-0088_I_c_math_bm_83-117_28b.indd 88 18/09/2017 09:58

Ηλικία (σε έτη)	Συχνότητα (σε χιλιάδες)
0-20	12
20-40	14
40-60	20
60-80	10
80-100	4

7. Από τη Στατιστική της Φυσικής Κίνησης Πληθυσμού της Ελλάδας οι θάνατοι λόγω υπερτασικής νόσου το 1995 δίνονται στον παρακάτω πίνακα (ΕΣΥΕ):
Να κατασκευάσετε στο ίδιο σχήμα τα πολύγωνα αθροιστικών σχετικών συχνοτήτων για την ηλικία ανδρών και

γυναικών, αντίστοιχα, που πέθαναν από υπερτασική νόσο το 1995, και στη συνέχεια να τα συγκρίνετε.

	Θάνατοι	
Ηλικία	Άνδρες	Γυναίκες
50-54	10	7
55-59	10	4
60-64	17	21
65-69	36	57
70-74	44	61
75-79	73	109
80-84	117	162
85-89	123	195

2.4 ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

Εισαγωγή

Στα διάφορα προβλήματα που εξετάσαμε έως τώρα στη Στατιστική ασχοληθήκαμε κάθε φορά με μία μόνο μεταβλητή (ξεχωριστά), π.χ. με το βάρος, με το ύψος, με την επίδοση μαθητών κτλ. Για καθεμιά μεμονωμένη μεταβλητή αρκεστήκαμε στη μελέτη της κατανομής συχνοτήτων, στον υπολογισμό διάφορων μέτρων όπως μέση τιμή, διάμεσος, διακύμανση κτλ. Σε αρκετές όμως περιπτώσεις εξίσου ενδιαφέρουσα είναι και η ταυτόχρονη μελέτη δύο ή περισσότερων μεταβλητών, για να προσδιορίσουμε με ποιο τρόπο οι μεταβλητές αυτές σχετίζονται μεταξύ τους. Για παράδειγμα:

α) Η ηλικία και το βάρος ενός παιδιού έχουν κάποια θετική εξάρτηση (συσχέτιση) μεταξύ τους με την έννοια ότι όσο πιο μεγάλο είναι το παιδί τόσο μεγαλύτερο βάρος θα έχει.

β) Η διάρκεια ζωής των ζώντων οργανισμών σε μια περιοχή και το ποσοστό μόλυνσης της περιοχής έχουν αρνητική εξάρτηση μεταξύ τους, με την έννοια ότι όσο πιο μεγάλο είναι το ποσοστό μόλυνσης της περιοχής τόσο μικρότερη είναι η διάρκεια ζωής των οργανισμών που ζουν στην περιοχή.

γ) Όσο μεγαλύτερη (μέχρι ένα ανώτερο όριο) είναι η περιεκτικότητα σε φθόριο στο πόσιμο νερό, τόσο μικρότερες είναι οι περιπτώσεις στη φθορά των δοντιών των μικρών παιδιών.

δ) Η συνολική παραγωγή καλαμποκιού εξαρτάται από τη θέση του χωραφιού, από την ποσότητα λιπάσματος, από την επίδραση της θερμοκρασίας, της υγρασίας κτλ.

ε) Το ύψος των αποδοχών των υπαλλήλων μιας εταιρείας δεν εξαρτάται από το βάρος τους.

Έτσι λοιπόν είναι ενδιαφέρον να εξεταστούν οι επιδράσεις που κάποιες μεταβλητές ασκούν σε κάποιες άλλες μεταβλητές. Η ύπαρξη

μιας συναρτησιακής σχέσης (εξίσωσης) μεταξύ των μεταβλητών μπορεί να είναι εξαιρετικά πολύτιμη για την πρόβλεψη των τιμών μιας μεταβλητής από τις γνώσεις που διαθέτουμε για τις άλλες μεταβλητές, όταν ισχύουν κάποιες συγκεκριμένες συνθήκες.
Ο κλάδος της Στατιστικής που εξετάζει τη σχέση μεταξύ δύο ή περισσότερων μεταβλητών με απώτερο σκοπό την πρόβλεψη μιας απ' αυτές μέσω των άλλων χαρακτηρίζεται με την ονομασία **ανάλυση παλινδρόμησης** (regression analysis). Ιστορικά, ο όρος “regression” χρησιμοποιήθηκε για πρώτη φορά από τον Άγγλο ανθρωπολόγο Galton (1822-1911) το 1885. Με τη μελέτη του ύψους των παιδιών σε σχέση

22-0088_I_c_math_bm_83-117_28b.indd 94 18/09/2017 09:58

με το ύψος των γονέων διαπιστώθηκε ότι παιδιά ψηλών γονέων τείνουν, κατά μέσο όρο, να είναι κοντύτερα των γονιών τους, ενώ παιδιά κοντών γονέων τείνουν, κατά μέσο όρο, να γίνονται ψηλότερα των γονιών τους.

Απλή Γραμμική Παλινδρόμηση

Η απλούστερη περίπτωση παλινδρόμησης είναι η απλή γραμμική παλινδρόμηση (simple linear regression), κατά την οποία υπάρχει μόνο μια **ανεξάρτητη μεταβλητή** X (independent or input variable), και η **εξαρτημένη μεταβλητή** Y (dependent or response variable), η οποία μπορεί να προσεγγιστεί ικανοποιητικά από μία γραμμική συνάρτηση του X. Η περίπτωση αυτή εμφανίζεται τόσο σε πειραματικές όσο και σε μη πειραματικές μελέτες. Στις πειραματικές μελέτες ο ερευνητής καθορίζει, για παράδειγμα, από πριν τις δόσεις ενός φαρμάκου (ανεξάρτητη μεταβλητή) που δίνει στα πειραματόζωα και μετρά τις αντιδράσεις τους (εξαρτημένη μεταβλητή). Με την παλινδρόμηση ενδιαφέρεται να προσδιορίσει μία σχέση δόσης-αντίδρασης για το συγκεκριμένο φάρμακο. Στις μη πειραματικές μελέτες ή δειγματοληψίες, γίνονται μετρήσεις σε δύο χαρακτηριστικά (μεταβλητές) για κάθε άτομο (μονάδα) του δείγματος. Σε ένα δείγμα 10 μαθητών μετράμε, για παράδειγμα, το βάρος και το ύψος τους. Η διάκριση εδώ μεταξύ ανεξάρτητης και εξαρτημένης μεταβλητής είναι δύσκολη. Αν αυτό που μας ενδιαφέρει είναι το “τι συμβαίνει με το βάρος των παιδιών όταν αλλάζει το ύψος τους”, τότε θεωρούμε ως ανεξάρτητη μεταβλητή Χ το ύψος και ως εξαρτημένη μεταβλητή Υ το βάρος. Οπότε, ενδιαφερόμαστε για την **παλινδρόμηση του βάρους (Υ) πάνω στο ύψος (Χ).** Αντίθετα, αν μας ενδιαφέρει το “τι συμβαίνει με το ύψος των παιδιών όταν αλλάζει το βάρος τους”, τότε θεωρούμε ως ανεξάρτητη μεταβλητή Χ το βάρος και ως εξαρτημένη μεταβλητή Υ το ύψος. Τότε έχουμε παλινδρόμηση του ύψους (Υ) πάνω στο βάρος (Χ).

Διάγραμμα Διασποράς

Ο παρακάτω πίνακας 10 δίνει τα ύψη X (σε cm) και τα βάρη Y (σε kg) των 18 αγοριών της Γ΄ Λυκείου του πίνακα 4. Οι τιμές του ύψους δίνονται σε αύξουσα σειρά.

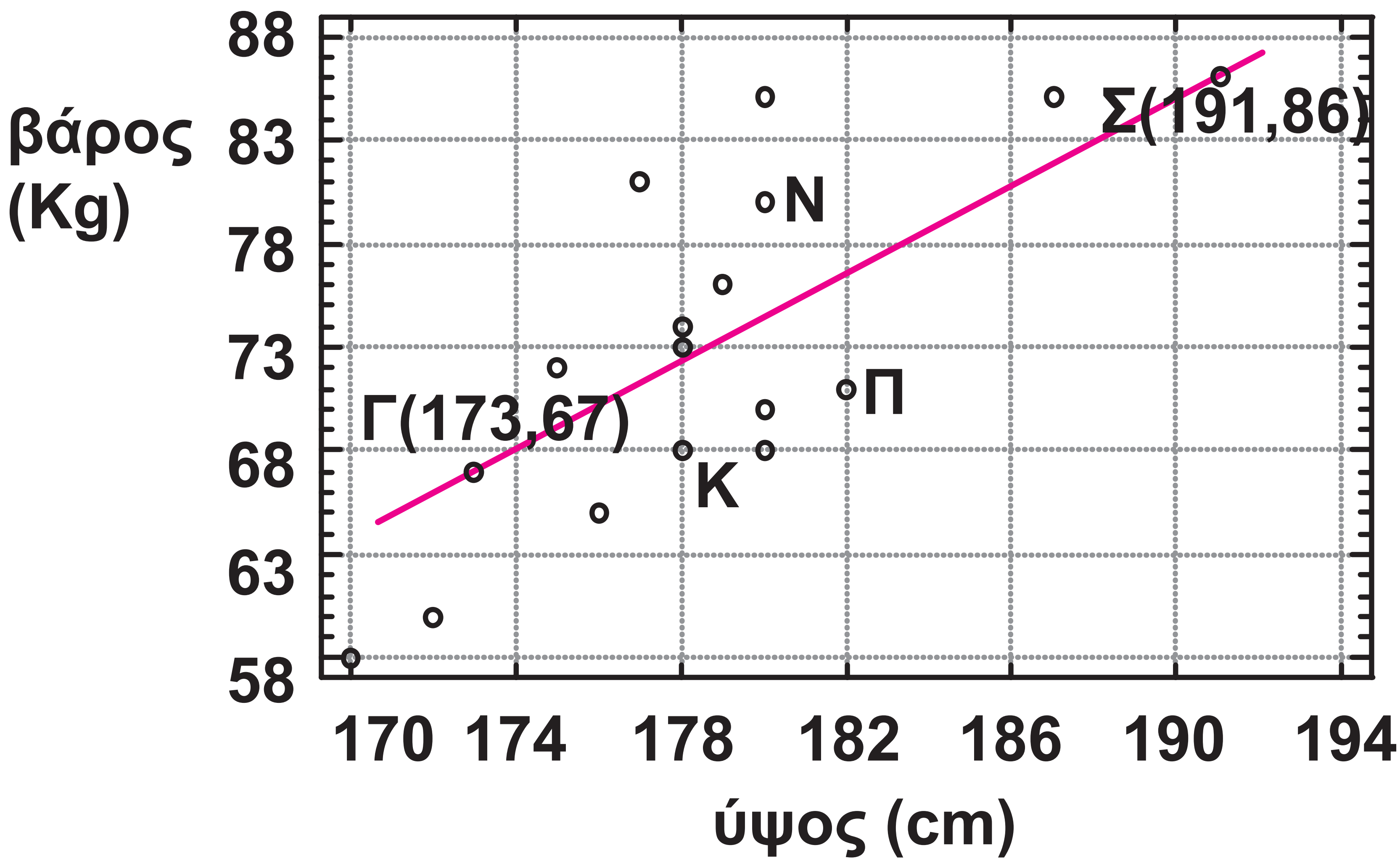
Πίνακας 10

Λίστα υψών (σε cm) και βαρών (σε kg) των 18 αγοριών του πίνακα 4.

Μαθη-τής	Ύψος X	Βά-ρος Y	Μαθη-τής	Ύψος X	Βά-ρος Y
Α	170	58	Κ	178	68
Β	172	60	Λ	179	76
Γ	173	67	Μ	180	68
Δ	175	72	Ν	180	80
Ε	176	65	Ξ	180	70
Ζ	177	81	Ο	180	85
Η	178	73	Π	182	71
Θ	178	74	Ρ	187	85
Ι	178	73	Σ	191	86

Στο παράδειγμα αυτό έχουμε την περίπτωση όπου σε κάθε άτομο (μαθητή) γίνονται δύο μετρήσεις. Δηλαδή το δείγμα αποτελείται από τα ζεύγη τιμών των συνεχών μεταβλητών X (ύψος) και Y (βάρος).

Αν παραστήσουμε τα ζεύγη (x, y) των παρατηρήσεων σε ένα σύστημα ορθογώνιων αξόνων, παρατηρούμε ότι προκύπτει μία “διασπορά” των σημείων που αντιστοιχούν στους μαθητές που εξετάζουμε.
Η παράσταση αυτή των σημείων καλείται διάγραμμα διασποράς (scatter diagram), βλέπε σχήμα 16.



Διάγραμμα διασποράς και ευθεία προσαρμοσμένη “με το μάτι” για τα δεδομένα του πίνακα 10.

Η προσεκτική παρατήρηση ενός διαγράμματος διασποράς μπορεί να μας δώσει σημαντικές πληροφορίες για τη σχέση εξάρτησης που ενδεχομένως υπάρχει μεταξύ των μεταβλητών τις οποίες εξετάζουμε. Η πείρα μας λέει ότι υπάρχει κάποια σχέση μεταξύ του ύψους και του βάρους κάθε μαθητή. Στο παράδειγμα αυτό το διάγραμμα διασποράς δείχνει, γενικά, ότι οι ψηλοί μαθητές είναι συνήθως και πιο βαρείς. Για παράδειγμα, ο Ν είναι ψηλότερος και βαρύτερος από τον Κ, ο Π είναι ψηλότερος και βαρύτερος από τον Κ, αλλά υπάρχουν και εξαιρέσεις, όπως ο Π είναι ψηλότερος από τον Ν αλλά ο Ν είναι βαρύτερος από τον Π.

Ευθεία Παλινδρόμησης

Από το διάγραμμα διασποράς του προηγούμενου παραδείγματος φαίνεται καθαρά ότι υπάρχει μία σχέση ανάμεσα στο ύψος X και το βάρος Y των 18 αγοριών της Γ΄ Λυκείου. Τα σημεία (x, y) είναι συγκεντρωμένα περίπου γύρω από μία ευθεία, δηλαδή η σχέση μεταξύ των X και Y είναι κατά προσέγγιση γραμμική. Όπως έχουμε ήδη αναφέρει, μπορούμε να θεωρήσουμε τη μία μεταβλητή ως ανεξάρτητη μεταβλητή και την άλλη ως εξαρτημένη. Εδώ θεωρούμε ως ανεξάρτητη μεταβλητή το ύψος X και ως εξαρτημένη μεταβλητή το βάρος Y, οπότε η ευθεία που θα προσαρμόζεται καλύτερα στα σημεία αυτά καλείται **ευθεία παλινδρόμησης της Y πάνω στη X.**

Όπως γνωρίζουμε, η εξίσωση μιας ευθείας δίνεται από τη σχέση:

$$y = α + βx \qquad (1)$$

όπου α και β είναι παράμετροι τις οποίες θέλουμε να υπολογίσουμε ή, όπως λέμε, να “εκτιμήσουμε”, έτσι ώστε η ευθεία που θα προκύψει να μας δίνει όσο το δυνατόν την καλύτερη περιγραφή της σχέσης (εξάρτησης) που υπάρχει μεταξύ των μεταβλητών X και Y. Η παράμετρος α μας δίνει το σημείο, (0, α) όπου η ευθεία αυτή τέμνει τον άξονα y΄y, ενώ η παράμετρος β παριστάνει το συντελεστή διεύθυνσης της ευθείας. Ο πιο εύκολος τρόπος χάραξης της ευθείας είναι αυτός που γίνεται “με το μάτι”. Μια τέτοια ευθεία έχουμε φέρει και στο διάγραμμα διασποράς του σχήματος 16. Για να βρούμε τα α και β, εργαζόμαστε ως εξής:

- **Επιλέγουμε δύο σημεία, έστω τα Γ(173,67) και Σ(191,86) πάνω στην ευθεία που φέραμε “με το μάτι”.**
- **Αντικαθιστούμε τις συντεταγμένες (x, y) των σημείων αυτών στην (1), οπότε προκύπτει το σύστημα:**

$$\begin{cases} y_1 = \alpha + \beta x_1 \\ y_2 = \alpha + \beta x_2 \end{cases} \Leftrightarrow \begin{cases} 67 = \alpha + 173\beta \\ 86 = \alpha + 191\beta \end{cases}$$

- **Επιλύοντας το σύστημα αυτό βρίσκουμε α = −115,6 και β = 1,06 οπότε η εξίσωση της ευθείας (1) γίνεται:**

$$y = -115{,}6 + 1{,}06x. \qquad (2)$$

Επομένως, η ευθεία που κατά τη γνώμη μας προσαρμόζεται καλύτερα στα σημεία του διαγράμματος διασποράς διέρχεται από το σημείο (0, −115,6) και έχει συντελεστή διεύθυνσης 1,06.

Μέθοδος των Ελαχίστων Τετραγώνων

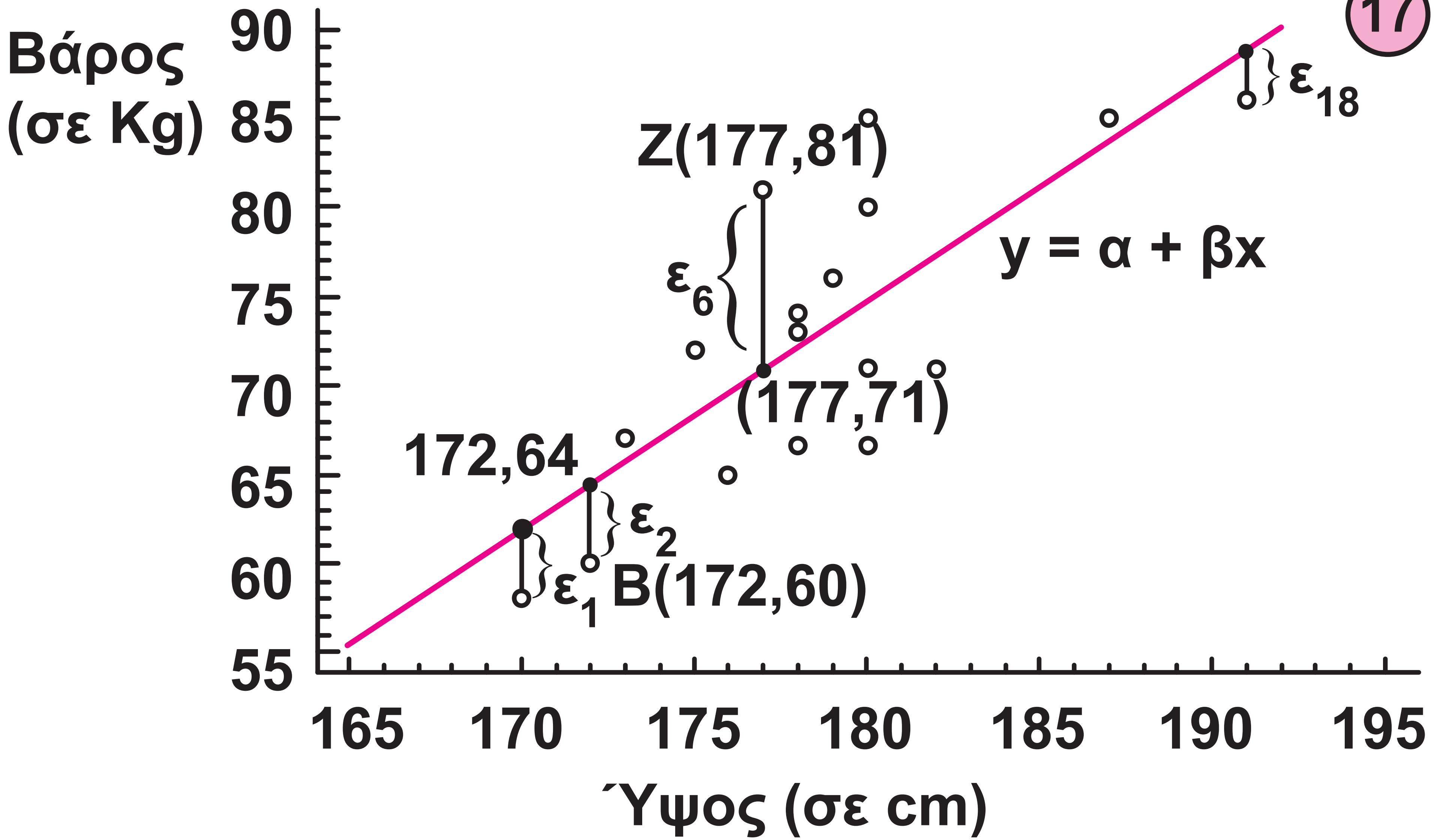
Είδαμε ότι η πιο απλή διαδικασία προσαρμογής μιας ευθείας γραμμής σε ένα διάγραμμα διασποράς είναι “με το μάτι”. Αυτή όμως έχει πολλά μειονεκτήματα παρά την απλότητά της. Το κυριότερο είναι η έλλειψη αντικειμενικότητας, αφού διάφορα άτομα μπορούν να χαράξουν διαφορετικές μεταξύ τους ευθείες. Ακόμα και το ίδιο άτομο μπορεί να χαράζει διαφορετικές ευθείες κάθε φορά. Χρειαζόμαστε λοιπόν μια ακριβέστερη μέθοδο για την προσαρμογή μιας ευθείας γραμμής σε τέτοιου είδους δεδομένα.

Μια μέθοδος που χρησιμοποιείται για την εκτίμηση των παραμέτρων α και β, άρα και για την εύρεση της

εξίσωσης της καλύτερης ευθείας που προσαρμόζεται στα δεδομένα, είναι η "μέθοδος ελαχίστων τετραγώνων".
Η πρώτη αναφορά με ολοκληρωμένη ανάπτυξη της μεθόδου των ελαχίστων τετραγώνων εμφανίζεται το 1805 σε μια εργασία του Γάλλου μαθηματικού Legendre, (1752-1833) και αμέσως μετά από το Γερμανό μαθηματικό Gauss, (1777-1855) στην αστρονομική του πραγματεία "Theoria Motus" για τον προσδιορισμό της τροχιάς του μικρού πλανήτη Δήμητρα. Μάλιστα εδώ ο Gauss αναφέρει ότι χρησιμοποίησε την αρχή των ελαχίστων τετραγώνων πριν από το 1794 (σε ηλικία μόλις 17 ετών), έτσι ώστε να προηγείται του Legendre ως προς την ανακάλυψη αυτής της μεθόδου.

Ας δούμε ξανά το διάγραμμα διασποράς στο σχήμα 17 του προηγούμενου παραδείγματος για τα ύψη X και τα βάρη Y των 18 μαθητών του πίνακα 10. Στο διάγραμμα αυτό έχουμε φέρει και μία ευθεία $y = α + βx$, που πιστεύουμε ότι προσαρμόζεται καλύτερα στα σημεία (x_i, y_i) για τις $ν = 18$ συνολικά μετρήσεις των μεταβλητών X και Y.

17



Προσαρμογή ευθείας ελαχίστων τετραγώνων στο διάγραμμα διασποράς του δεδομένων του πίνακα 10.

Έτσι, για παράδειγμα, για το μαθητή Β, σημείο Β(172, 60), με ύψος x_2 = 172 cm έχουμε βρει, όπως φαίνεται στον πίνακα 10, βάρος y_2 = 60 kg, ενώ, σύμφωνα με την ευθεία που φέραμε, το βάρος του αναμένεται να είναι (περίπου) 64 kg, έχουμε δηλαδή ένα σφάλμα $\varepsilon_2 = 60 - 64 = -4$, δηλαδή βάρος 4 kg λιγότερο από το αναμενόμενο. Ομοίως για το μαθητή Ζ, σημείο Ζ(177, 81), το βάρος του που μετρήθηκε ήταν y_6 = 81 kg, ενώ το αναμενόμενο βάρος του σύμφωνα με την ευθεία που φέραμε είναι 71 kg, έχουμε δηλαδή ένα σφάλμα $\varepsilon_6 = 81 - 71 = 10$, δηλαδή βάρος 10 kg περισσότερο από το αναμενόμενο.

Ανάλογα σφάλματα υπολογίζονται

και για τους άλλους μαθητές. Θα θέλαμε λοιπόν να βρούμε με κάποια μέθοδο εκείνη την ευθεία $y = α + βx$, έτσι ώστε τα σφάλματα που προκύπτουν να είναι όσο το δυνατόν μικρότερα.

Η μέθοδος των ελαχίστων τετραγώνων συνίσταται στον προσδιορισμό των παραμέτρων α, β, έτσι ώστε να ελαχιστοποιείται το άθροισμα των τετραγώνων των κατακόρυφων αποστάσεων των σημείων (x_i, y_i) από την ευθεία $y = α + βx$, δηλαδή το

$$\sum_{i=1}^{ν} ε_i^2 = \sum_{i=1}^{ν} (y_i - α - βx_i)^2 \qquad (4)$$

να γίνεται ελάχιστο.

Οι τιμές των παραμέτρων α και β, που ελαχιστοποιούν την (4), καλούνται **εκτιμήτριες ελαχίστων τετραγώνων** (least square estimators), συμβολίζονται με $\hat{\alpha}$ (“α καπέλο”) και $\hat{\beta}$ (“β καπέλο”), αντιστοίχως, και αποδεικνύεται (η απόδειξη εδώ παραλείπεται) ότι δίνονται από τις σχέσεις:

$$\hat{\beta} = \frac{\nu \sum_{i=1}^{\nu} x_i y_i - \left(\sum_{i=1}^{\nu} x_i\right)\left(\sum_{i=1}^{\nu} y_i\right)}{\nu \sum_{i=1}^{\nu} x_i^2 - \left(\sum_{i=1}^{\nu} x_i\right)^2} \qquad (5)$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$$

όπου $\bar{y} = \frac{1}{\nu}\sum_{i=1}^{\nu} y_i$, $\bar{x} = \frac{1}{\nu}\sum_{i=1}^{\nu} x_i$.

Η ευθεία

$$\hat{y} = \hat{\alpha} + \hat{\beta}x \qquad (6)$$

καλείται **ευθεία ελαχίστων τετραγώνων ή ευθεία παλινδρόμησης** της Υ (πάνω) στη Χ. Αντικαθιστώντας το $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x}$ στη σχέση (6) βρίσκουμε την

$$\hat{y} - \bar{y} = \hat{\beta}(x - \bar{x}),$$

η οποία φανερώνει ότι η ευθεία ελαχίστων τετραγώνων $\hat{y} = \hat{\alpha} + \hat{\beta}x$ διέρχεται από το σημείο με συντεταγμένες $(\bar{x}, \bar{y})$ και έχει συντελεστή διεύθυνσης το $\hat{\beta}$.

Αντικαθιστώντας τις τιμές x_i και y_i από τον πίνακα 10 στις σχέσεις (5) βρίσκουμε:

$$\hat{\beta} = 1{,}28 \text{ και } \hat{\alpha} = -156{,}1$$

οπότε η ευθεία ελαχίστων τετραγώνων που προσαρμόζεται καλύτερα στα δεδομένα είναι από τη σχέση (6), η

$$\hat{y} = -156{,}1 + 1{,}28x.$$

Παρατηρούμε ότι υπάρχει σημαντική διαφορά από την ευθεία $y = -115{,}6 + 1{,}06x$ που προσαρμόσαμε “με το μάτι” στο σχήμα 16.

Ερμηνεία των $\hat{\alpha}$ και $\hat{\beta}$

Στην εξίσωση ελαχίστων τετραγώνων $\hat{y} = \hat{\alpha} + \hat{\beta}x$ η τιμή της εκτιμήτριας $\hat{\alpha}$ της παραμέτρου α παριστάνει την τεταγμένη του σημείου στο οποίο η ευθεία τέμνει τον άξονα $y'y$, δηλαδή την τιμή της εξαρτημένης μεταβλητής Y όταν $x = 0$. Όταν το $\hat{\alpha} = 0$ τότε η ευθεία διέρχεται από την αρχή των αξόνων.

Έστω τώρα δυο τιμές x_1 και $x_2 = x_1 + 1$ της ανεξάρτητης μεταβλητής. Τότε λαμβάνοντας τη διαφορά των αντίστοιχων προβλεπόμενων τιμών της εξαρτημένης μεταβλητής βρίσκουμε

$$\hat{y}_2 - \hat{y}_1 = (\hat{\alpha} + \hat{\beta}x_2) - (\hat{\alpha} + \hat{\beta}x_1) =$$
$$= \hat{\alpha} + \hat{\beta}(x_1 + 1) - (\hat{\alpha} + \hat{\beta}x_1) = \hat{\beta}$$

δηλαδή $\hat{y}_2 = \hat{y}_1 + \hat{\beta}$. Συνεπώς ο συντελεστής διεύθυνσης $\hat{\beta}$ της ευθείας $\hat{y} = \hat{\alpha} + \hat{\beta}x$ παριστά τη μεταβολή της εξαρτημένης μεταβλητής Y όταν το X μεταβληθεί κατά μια μονάδα. Έτσι, όταν το x αυξηθεί κατά μια μονάδα τότε το $\hat{y}$ αυξάνεται κατά $\hat{\beta}$ μονάδες όταν $\hat{\beta} > 0$ ή ελαττώνεται κατά $\hat{\beta}$ μονάδες όταν $\hat{\beta} < 0$.

ΕΦΑΡΜΟΓΕΣ

1. Ένας ερευνητής, για να εξετάσει την επίδραση ενός αναισθητικού, εμβολίασε 10 ποντίκια με διαφορετική δόση κάθε φορά. Οι χρόνοι που μεσολάβησαν ώσπου τα ποντίκια να χάσουν τις αισθήσεις τους (λιποθυμήσουν) καταγράφονται στον παρακάτω πίνακα.

Δόση (σε mgr)	0,30	0,35	0,40	0,45	0,55	0,60	0,65	0,70	0,75	0,80
Χρόνος λιπο-θυμίας (sec)	12,5	11,5	11	8,5	7	6	5	4	2,5	2

α) Να γίνει το διάγραμμα διασποράς

β) Να επιλεγούν δύο πιθανά σημεία από τα οποία διέρχεται η προσαρμοσμένη ευθεία παλινδρόμησης και με βάση αυτά να βρεθεί η εξίσωση της ευθείας.

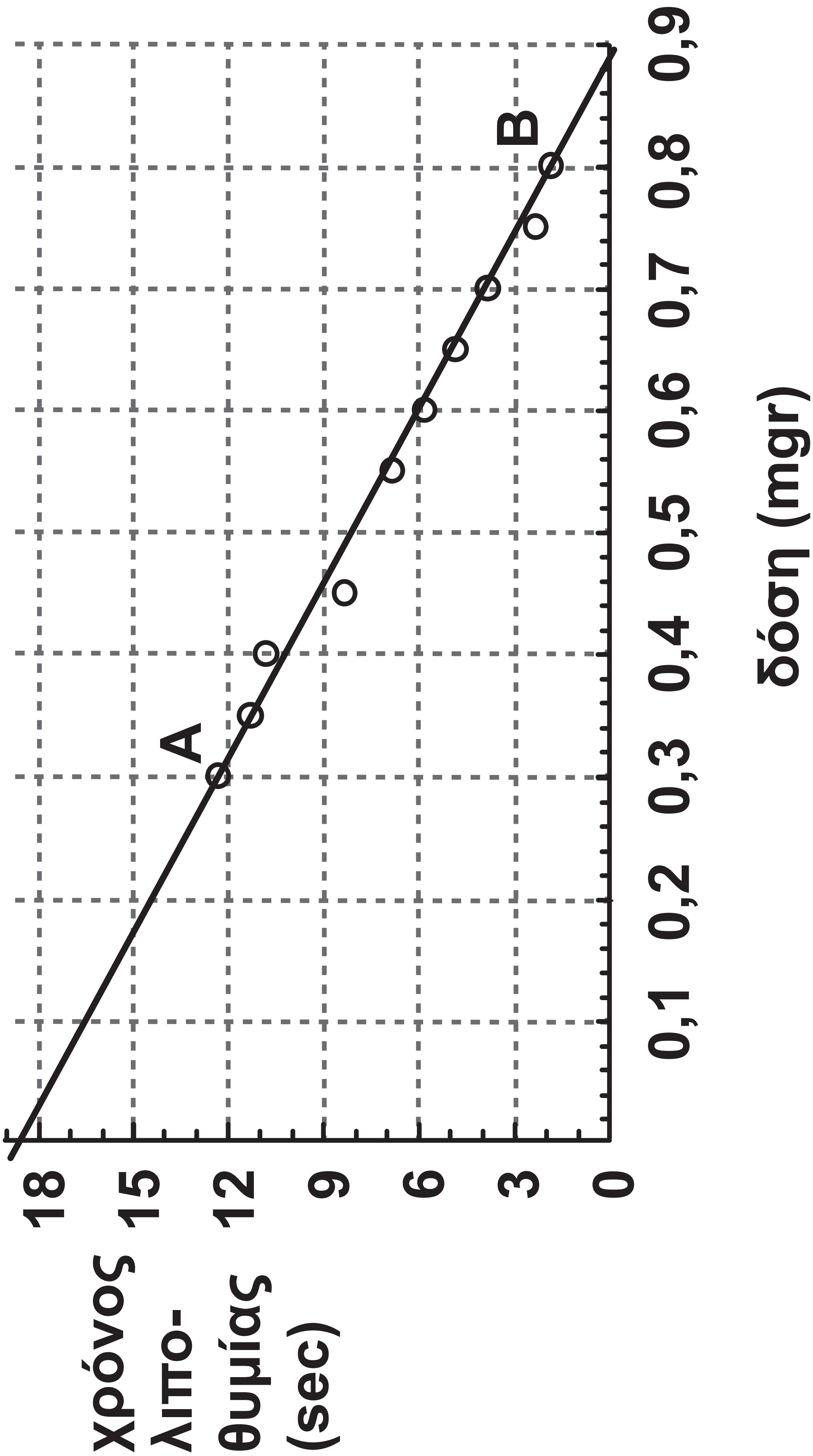
22-0088_I_c_math_bm_83-117_28b.indd 117 18/09/2017 09:58

γ) Ύστερα από πόσο χρόνο αναμένεται να λιποθυμήσει ένα ποντίκι, εάν του γίνει ένεση (εμβολιαστεί) με 0,   0,5,   1 mgr αναισθητικού, αντίστοιχα; Να σχολιαστούν τα αποτελέσματα.

ΛΥΣΗ

α) Η περίπτωση που εξετάζεται εδώ αναφέρεται σε μια πειραματική κατάσταση, κατά την οποία ο ερευνητής καθορίζει εκ των προτέρων τη δόση του αναισθητικού που θα δώσει στα πειραματόζωα και μετρά τις αντιδράσεις τους. Ενδιαφέρεται δηλαδή να προσδιορίσει μια σχέση δόσης αναισθητικού και χρόνου λιποθυμίας. Έτσι, η δόση αναισθητικού παριστάνει την ανεξάρτητη

μεταβλητή (X) και ο χρόνος λιποθυμίας την εξαρτημένη μεταβλητή (Y). Επομένως, το διάγραμμα διασποράς παριστάνεται παρακάτω:


Χρόνος λιποθυμίας (sec)
18
15
12
9
6
3
0
0,1
0,2
0,3
0,4
0,5
0,6
0,7
0,8
0,9
δόση (mgr)
A
B

β) Δύο σημεία από τα οποία εκτιμάται “με το μάτι” ότι θα διέρχεται η καλύτερη ευθεία παλινδρόμησης είναι τα Α(0,3 , 12,3) και Β(0,8 , 1,9). Αντικαθιστώντας στην εξίσωση της ευθείας $y = α + βx$ έχουμε το σύστημα:

$$\begin{cases} 12,2 = α + 0,3β \\ 1,9 \quad = α + 0,8β, \end{cases}$$

από την επίλυση του οποίου βρίσκουμε $α = 18,5$ και $β = -20,8$, οπότε η ευθεία παλινδρόμησης έχει εξίσωση

$$y = 18,5 - 20,8x.$$

Παρατηρούμε, όπως άλλωστε αναμενόταν και από το διάγραμμα διασποράς, ότι ο συντελεστής διεύθυνσης είναι αρνητικός, δηλαδή έχουμε

αρνητική εξάρτηση του χρόνου λιποθυμίας ως προς τη δόση αναισθητικού. Μεγαλύτερη δόση επιφέρει γρηγορότερη (σε μικρότερο χρόνο) λιποθυμία.
γ) Αντικαθιστώντας στην παραπάνω εξίσωση τις τιμές της δόσης $x = 0$, $0{,}5$ και 1 mgr, βρίσκουμε τον προβλεπόμενο χρόνο λιποθυμίας του ποντικιού $y = 18{,}5$, $8{,}1$ και $-2{,}3$ sec, αντίστοιχα.

ΣΧΟΛΙΟ

Παρατηρούμε ότι εμφανίζονται εδώ δυο παράδοξα:

- Μηδενική δόση του αναισθητικού προκαλεί λιποθυμία σε 18,5 sec.

Υποθέτουμε όμως εδώ ότι τα ποντίκια χάνουν τις αισθήσεις τους μόνο από τη δόση του αναισθητικού και όχι από το φόβο τους!

- Δόση αναισθητικού 1 mgr προκαλεί λιποθυμία σε −2,3 sec. Όμως και πάλι εδώ υποθέτουμε ότι η ένεση γίνεται σε ποντίκι που έχει τις αισθήσεις του και όχι να είναι πριν από 2,3 sec λιπόθυμο!

Οι δύο αυτές παρατηρήσεις οδηγούν στο βασικό συμπέρασμα ότι οι προβλέψεις της εξαρτημένης μεταβλητής έχουν νόημα, είναι δηλαδή δυνατές, μόνο για τις τιμές της ανεξάρτητης μεταβλητής οι οποίες βρίσκονται στο διάστημα που έχουμε εξετάσει ή τουλάχιστον πολύ κοντά στα άκρα του διαστήματος αυτού.

2. Ο παρακάτω πίνακας δίνει τις πωλήσεις (ζήτηση) ενός προϊόντος (π.χ. των κερασιών) Υ (σε κιλά) από το ψαράδικο μιας περιοχής και τις αντίστοιχες τιμές Χ του προϊόντος σε ευρώ ανά κιλό για μια ορισμένη χρονική περίοδο

Τιμή ψαριών ανά κιλό (σε ευρώ), Χ	15	13	11	9	9	6	5	4
Πωλήσεις σε κιλά, Υ	5	6	8	10	9	12	15	11

α) Να βρεθεί η ευθεία ελαχίστων τετραγώνων $\hat{y} = \hat{\alpha} + \hat{\beta}x$ (των πωλήσεων σε συνάρτηση με την τιμή) και να χαραχτεί στο αντίστοιχο διάγραμμα διασποράς.

β) Να ερμηνευθεί η έννοια των $\hat{\alpha}$ και $\hat{\beta}$

γ) Ποια είναι η αναμενόμενη ζήτηση (πωλήσεις), όταν η τιμή είναι 8 ευρώ/κιλό;

δ) Με βάση την ευθεία αυτή μπορούμε να προβλέψουμε την τιμή του προϊόντος, όταν η ζήτηση είναι 10 κιλά;

ΛΥΣΗ

α) Για τον προσδιορισμό της εξίσωσης της ευθείας ελάχιστων τετραγώνων διευκολύνει ο παρακάτω

πίνακας με τις απαραίτητες πρά-ξεις.
Επομένως, έχουμε:

x	y	x^2	xy
15	5	225	75
13	6	169	78
11	8	121	88
9	10	81	90
9	9	81	81
6	12	36	72
5	15	25	75
4	11	16	44
$\Sigma x = 72$	$\Sigma y = 76$	$\Sigma x^2 = 754$	$\Sigma xy = 603$

- $\nu = 8$

- $\bar{x} = \dfrac{\sum x}{\nu} = \dfrac{72}{8} = 9$

22-0088_I_c_math_bm_83-117_28b.indd 126 18/09/2017 09:58

- $\bar{y} = \dfrac{\sum y}{\nu} = \dfrac{76}{8} = 9{,}5$

- $\hat{\beta} = \dfrac{\nu \sum xy - (\sum x)(\sum y)}{\nu \sum x^2 - (\sum x)^2} =$

$$= \frac{8(603) - (72)(76)}{8(754) - (72)^2} = \frac{-648}{848} = -0{,}76$$

- $\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} = 9{,}5 + (0{,}76)(9) = 16{,}34.$

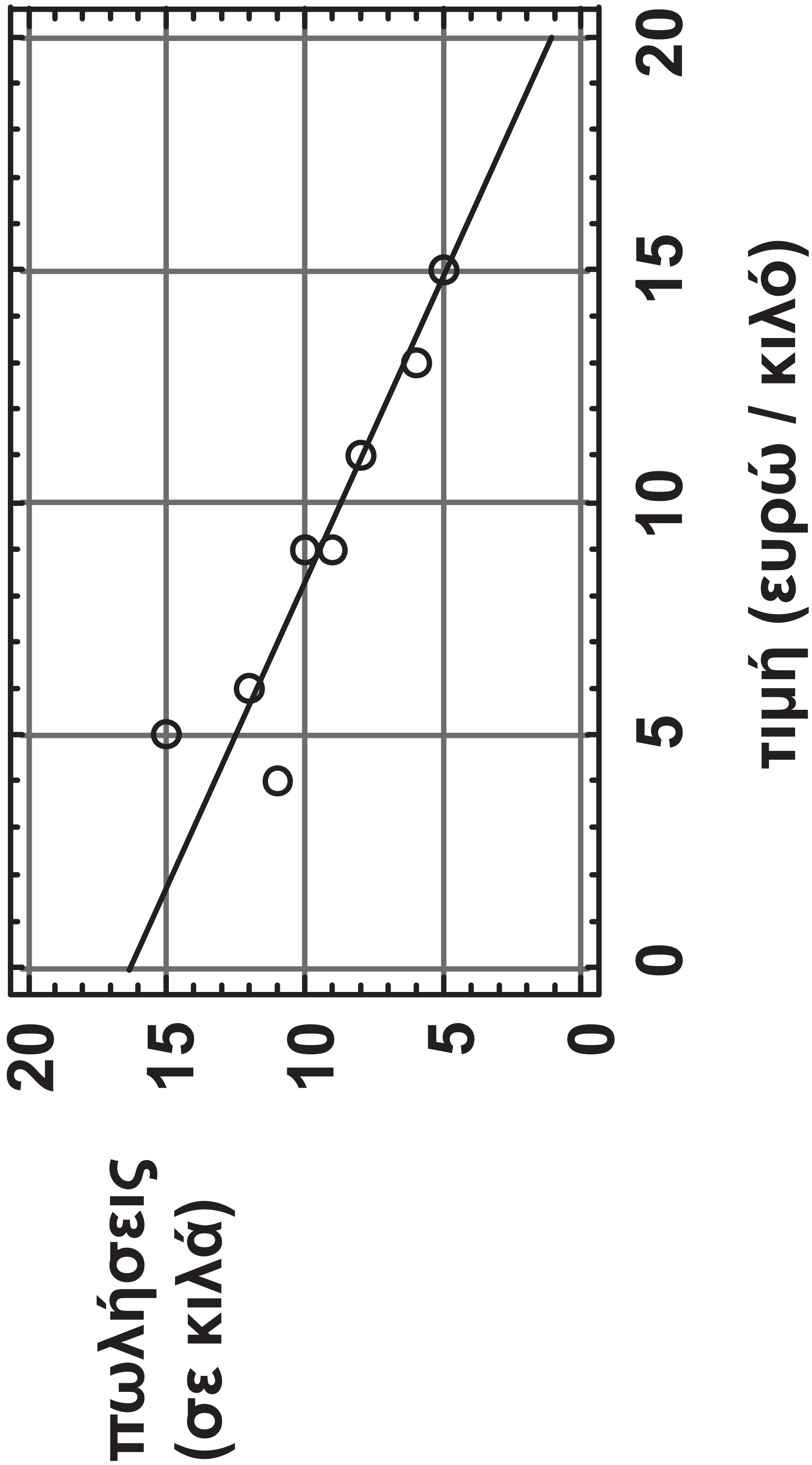
Άρα, η εξίσωση της ζήτησης (των ψαριών) σε συνάρτηση με την τιμή τους είναι $\hat{y} = 16{,}34 - 0{,}76x$.
Επειδή γνωρίζουμε ότι η ευθεία διέρχεται από τα σημεία $(0, \hat{\alpha})$ και $(\bar{x}, \bar{y})$, είναι εύκολο να χαραχτεί στο διάγραμμα διασποράς, όπως φαίνεται παρακάτω.

β) Το $\hat{\alpha} = 16,34$ προσδιορίζει την προβλεπόμενη ζήτηση του προϊόντος, όταν η τιμή του είναι 0 ευρώ/κιλό. Προφανώς εδώ τέτοια περίπτωση δεν είναι ρεαλιστική.
Το $\hat{\beta}$ προσδιορίζει τη μεταβολή που επέρχεται στην εξαρτημένη μεταβλητή Y, όταν η X μεταβληθεί κατά μία μονάδα. Δηλαδή όταν η τιμή των ψαριών αυξηθεί κατά 1 ευρώ/κιλό (μία μονάδα), οι πωλήσεις θα ελαττωθούν κατά 0,76 κιλά.
γ) Όταν η αξία του προϊόντος είναι 8 ευρώ/κιλό, σημαίνει ότι $x = 8$. Συνεπώς, για $x = 8$ η αναμενόμενη ζήτηση $\hat{y}$ είναι

$$\hat{y} = 16,34 - 0,76 \cdot 8 = 16,34 - 6,08 =$$

$$= 10,26 \text{ κιλά.}$$

δ) Προφανώς, στην περίπτωση που χρησιμοποιείται η μέθοδος των ελαχίστων τετραγώνων για την εκτίμηση της ευθείας παλινδρόμησης της εξαρτημένης μεταβλητής Y πάνω στην ανεξάρτητη μεταβλητή X, δεν μπορεί να χρησιμοποιηθεί η προκύπτουσα ευθεία για πρόβλεψη της X, όταν δίνεται το Y. Για να γίνει κάτι τέτοιο, πρέπει εξαρχής να εκτιμηθεί η ευθεία παλινδρόμησης της X πάνω στην Y.


Πωλήσεις
(σε κιλά)
0
5
10
15
20
Τιμή (ευρώ / κιλό)
0
5
10
15
20

ΑΣΚΗΣΕΙΣ

Α΄ ΟΜΑΔΑΣ

1. Να κατασκευάσετε το διάγραμμα διασποράς και να βρείτε την εξίσωση της ευθείας που προσαρμόζεται (με το μάτι) καλύτερα στα δεδομένα.

α)

x	12	15	16	18	18
y	13	14	18	18	20

β)

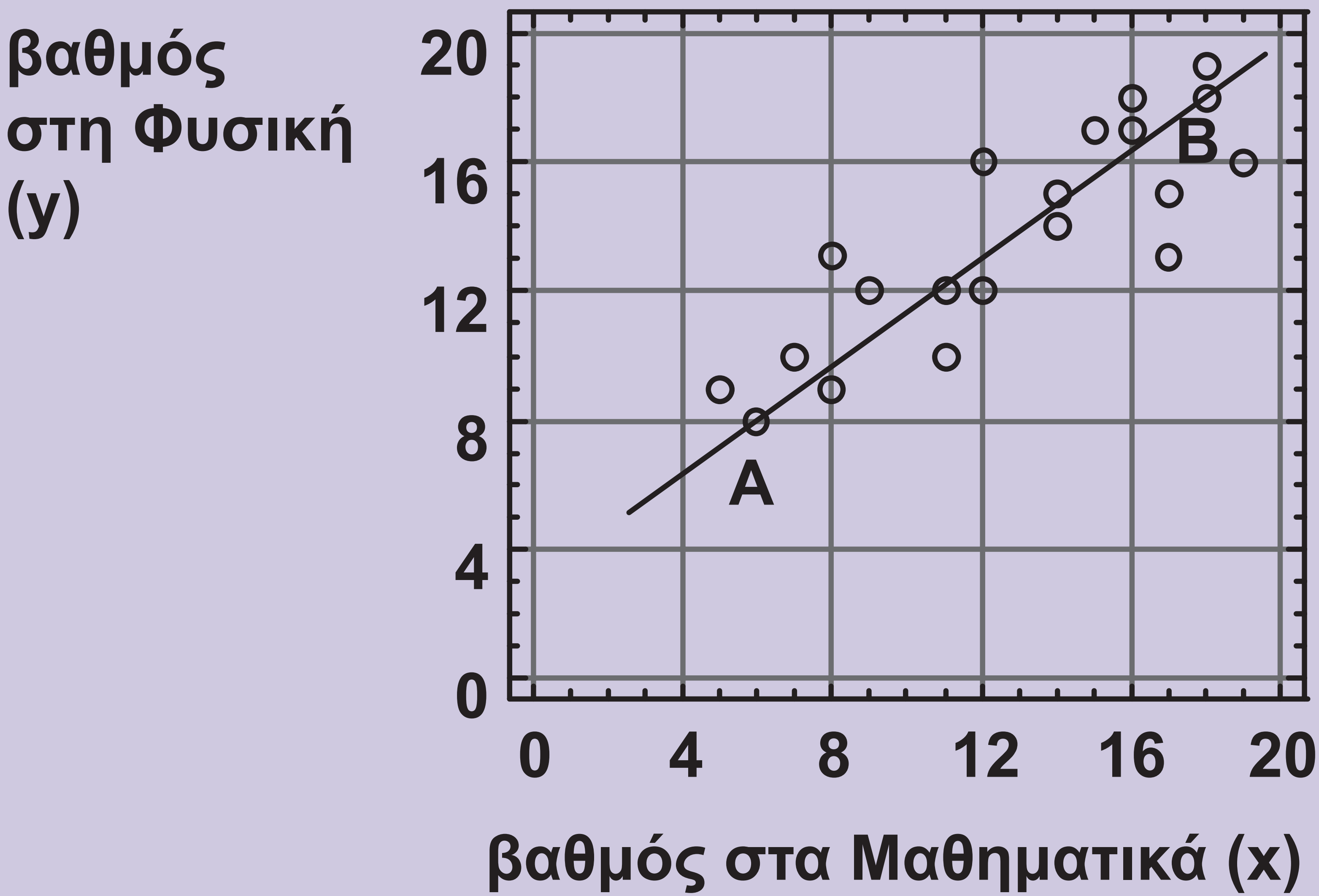
x	1	2	3	4	5	6	7	8	9
y	12	10	10	6	6	3	5	2	4

2. Παρακάτω δίνονται δύο διαγράμματα διασποράς με τις προσαρμοσμένες ευθείες,

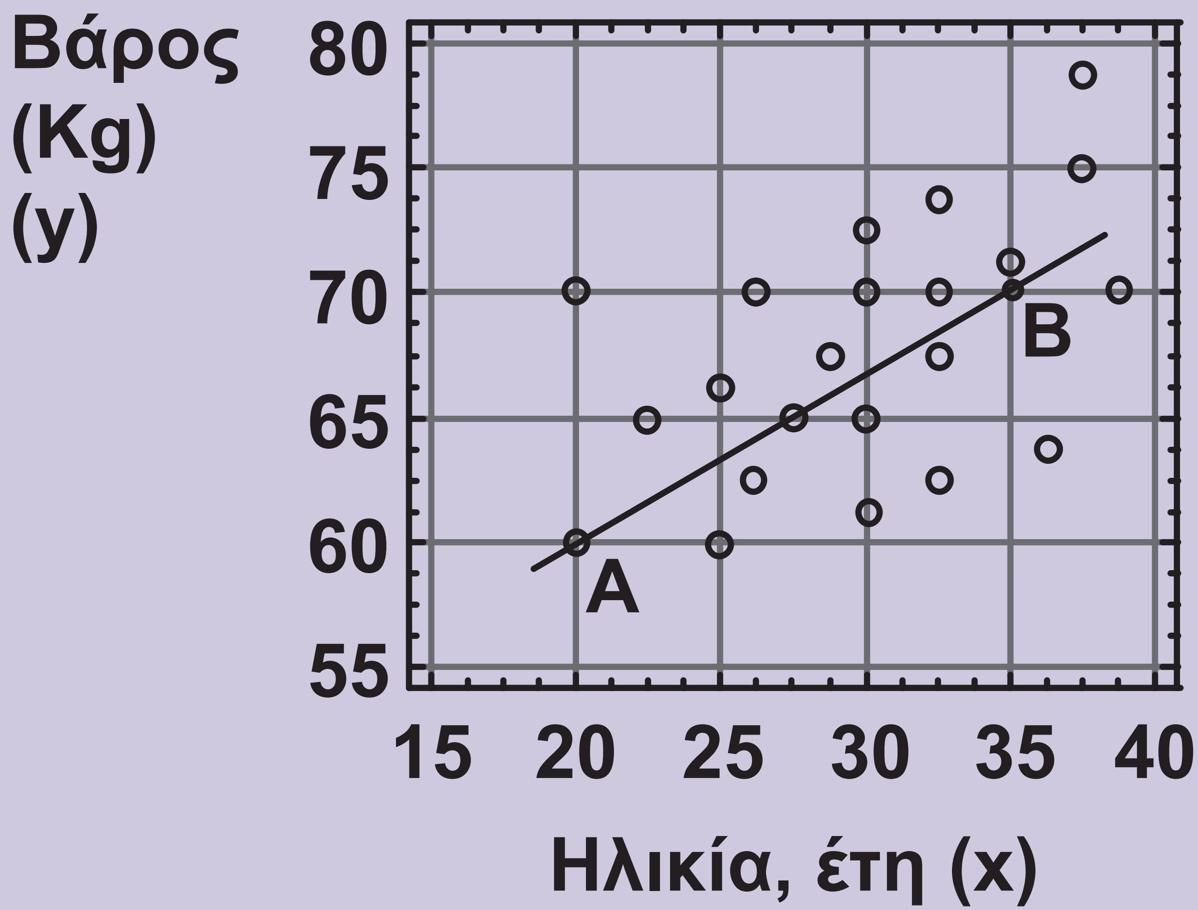
22-0088_l_c_math_bm_83-117_28b.indd 131 18/09/2017 09:58

όπως τις χάραξε “με το μάτι” ένας μαθητής. Χρησιμοποιώντας τα σημεία Α και Β να βρείτε τις εξισώσεις $y = α + βx$ των αντίστοιχων ευθειών.

(i)


βαθμός
στη Φυσική
(y)
20
16
12
8
4
0
A
B
0
4
8
12
16
20
βαθμός στα Μαθηματικά (x)

(ii)



22-0088_I_c_math_bm_83-117_28b.indd 134
18/09/2017 09:58

3. Για καθένα από τα παρακάτω ζεύγη τιμών να βρείτε την εξίσωση ελαχίστων τετραγώνων $\hat{y} = \hat{\alpha} + \hat{\beta}x$ και να τη χαράξετε στο αντίστοιχο διάγραμμα διασποράς. Στη συνέχεια να προβλέψετε την τιμή της y, όταν x = 6.

α)

x	1	2	3	4	5
y	1	2	3	4	5

β)

x	1	2	3	4	5
y	5	4	3	2	1

γ)

x	1	2	3	4	5
y	4	2	5	1	3

δ)

x	1	2	3	4	5
y	3	1	5	2	4

22-0088_I_c_math_bm_83-117_28b.indd 135 18/09/2017 09:58

4. Να εφαρμόσετε τη μέθοδο ελαχίστων τετραγώνων για τα δεδομένα της εφαρμογής 1, για να εκτιμήσετε την ευθεία γραμμικής παλινδρόμησης του χρόνου λιποθυμίας (Υ) των ποντικιών στη δόση αναισθητικού (Χ). Να συγκρίνετε τα αποτελέσματα που προκύπτουν με τη μέθοδο αυτή με τα αντίστοιχα αποτελέσματα που προέκυψαν με την προσαρμογή της ευθείας "με το μάτι".

5. α) Με τη μέθοδο ελαχίστων τετραγώνων να βρείτε την ευθεία γραμμικής παλινδρόμησης της Υ πάνω στη Χ για τα παρακάτω δεδομένα 5 μαθητών.

Επίδοση στα Μαθηματικά, x	12 15 16 18 18
Επίδοση στη Φυσική, y	13 14 18 18 20

β) Αν υποθέσουμε ότι χάθηκε ο βαθμός της Φυσικής για το μαθητή που πήρε 15 στα Μαθηματικά και για να μην υποχρεωθεί ο μαθητής να ξαναδώσει εξετάσεις στη Φυσική, ποιο βαθμό, κατά τη γνώμη σας, πρέπει να πάρει;

Β΄ ΟΜΑΔΑΣ

1. Ο παρακάτω πίνακας δίνει τα αποτελέσματα των μετρήσεων

της συστολικής πίεσης και της ηλικίας 10 γυναικών:

Ηλικία (έτη)	25	30	35	40	45	50	55	60	65	70
Πίεση (mm Hg)	116	117	121	147	111	133	105	153	155	176

α) Ποια από τις δύο μεταβλητές ηλικία και συστολική πίεση μπορεί να θεωρηθεί ως ανεξάρτητη μεταβλητή;

β) Να γίνει το αντίστοιχο διάγραμμα διασποράς.

γ) Να χαράξετε “με το μάτι” την ευθεία που προσαρμόζεται καλύτερα στα δεδομένα.

δ) Τι συστολική πίεση προβλέπετε για μια γυναίκα 75 ετών;

2. Να χρησιμοποιήσετε τα δεδομένα του Πίνακα 4 μόνο για τα αγόρια, για να προβλέψετε το ύψος ενός μαθητή όταν:

α) ο πατέρας του έχει ύψος 180 cm

β) ο μέσος όρος του ύψους των γονιών του είναι 170 cm.

3. Να χρησιμοποιήσετε τα δεδομένα του Πίνακα 4 μόνο για τα κορίτσια, για να προβλέψετε το ύψος μιας μαθήτριας όταν:
   α) η μητέρα της έχει ύψος 167 cm
   β) ο μέσος όρος του ύψους των γονιών της είναι 170 cm.

4. Από 8 γάμους που έγιναν σε μια εκκλησία κατά τη διάρκεια ενός μηνός, οι ηλικίες των ανδρογύνων ήσαν:

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

α) Να υπολογίσετε με τη μέθοδο ελαχίστων τετραγώνων την ευθεία γραμμικής παλινδρόμησης της Y πάνω στη X και να τη χαράξετε στο αντίστοιχο διάγραμμα διασποράς.

β) Να βρείτε την αναμενόμενη ηλικία του γαμπρού για μια υποψήφια νύφη 25 ετών.

γ) Για κάθε έτος που μια γυναίκα καθυστερεί να παντρευτεί πόσο αυξάνεται η ηλικία του υποψήφιου γαμπρού.

5. Για τα ίδια δεδομένα της προηγούμενης άσκησης (4) να βρείτε:

α) την ευθεία γραμμικής παλινδρόμησης της X πάνω στη Y,

β) την αναμενόμενη ηλικία της νύφης για έναν υποψήφιο γαμπρό 28 ετών,

γ) για κάθε έτος που ένας άνδρας καθυστερεί να παντρευτεί πόσο αυξάνεται η ηλικία της υποψήφιας νύφης;

ΥΠΟΔΕΙΞΕΙΣ – ΑΠΑΝΤΗΣΕΙΣ ΑΣΚΗΣΕΩΝ

2 ΣΤΑΤΙΣΤΙΚΗ

§ 2.3 Α΄ Ομάδας

1. 10, 12, 14, 16, 18, 20 με διάμεσο δ = 15.
2. α) ΝΑΙ, β) ΟΧΙ, γ) ΝΑΙ.
3. 8,25%.
4. α) 206,1 cm β) 235 cm.
5. 14,8.
6. Και στις 3 περιπτώσεις έχουμε: $\bar{x}$ = 12gr, δ = 12gr, M_0 = 13gr.

7. α) 14, β) 13.

8. 1291 ευρώ.

9. Οι 2 και 6.

10. α) 2, β) 2, γ) 3.

11. α) $\overline{x} = 15{,}45$, β) $\delta = 15$,
γ) $M_0 = 15$,
δ) $Q_1 = 14$, $Q_3 = 17$.

12. α) $\overline{x} = 4{,}3$ β) $M_0 = 3{,}3$,
γ) $Q_1 \approx 2{,}33$, $\delta = 4$, $Q_3 \approx 6$.

13. α) 169,66 cm, β) 170 cm,
γ) 169,67 cm.

14. α) Μικρότερη διασπορά έχουμε στη δεύτερη λίστα και μεγαλύτερη διασπορά στην τρίτη.
β) ΟΧΙ.

15. α) $\overline{x} = 10$, $M_0 = 11$, $\delta = 11$

β) $Q_1 = 7$, $Q_3 = 13$

γ) $R = 12$, $s = 3{,}87$, $cv = 38{,}7\%$.

16. 2,47.

17. α) 16%, β) 2,5%, γ) 50%,

δ) 81,5%.

18. α) $\overline{x} = 3$, $\delta = 2$

β) $\overline{x} = 6$, $\delta = 4$

γ) $\overline{x} = 13$, $\delta = 12$

δ) $\overline{x} = 16$, $\delta = 14$.

19. α) $s^2 = 4$, β) $s^2 = 36$,

γ) $s^2 = 4$, $s^2 = 4$.

20. $v = 28$.

§ 2.3 Β΄ Ομάδας

1. β) $\bar{x} = 15$, $\delta \approx 15$ γ) $P_{95} \approx 19$.

2. 10 ή −2.

3. α) $\bar{x} = 13{,}20$ ευρώ,
   $\delta = 10{,}50$ ευρώ,
   $M_0 = 9$ ευρώ.

   β) Αυξάνουν κατά 18%.
   γ) Αυξάνουν κατά 0,30 ευρώ.

4. Να γίνουν οι πράξεις.

5. α) 60, β) 33,
   γ) 5,2 χιλιάδες ευρώ,
   δ) $\bar{x} = 5{,}7$ χιλιάδες ευρώ,
   $s^2 = 12{,}24$
   ε) $Q_1 \approx 2{,}8$, $\delta \approx 5{,}4$, $Q_3 \approx 8{,}3$.

6. α) $s = 23{,}29$, $cv = 53{,}75\%$
   β) $Q = Q_3 - Q_1 \approx 59 - 24 = 35$.

7. Να εργαστείτε όπως στο σχήμα 7.

§ 2.4 Α΄ Ομάδας

1. Να εργαστείτε όπως στο σχήμα 16.

2. i) $y = 3 + \frac{5}{6}x$ ii) $y = 46{,}7 + 0{,}67x$.

3. α) $\hat{y} = x$,   β) $\hat{y} = 6 - x$,

   γ) $\hat{y} = 3{,}9 - 0{,}3x$

   δ) $\hat{y} = 2{,}1 + 0{,}3x$.

4. $\hat{y} = 18{,}98 - 21{,}6x$.

5. α) $\hat{y} = -0{,}35 + 1{,}07x$,   β) 16.

§ 2.4 Β΄ Ομάδας

1. α) Η ηλικία.
   γ,δ) Καθένας από τους μαθητές μπορεί να φέρει τη “δική του” ευθεία, οπότε θα έχει και διαφορετική πρόβλεψη συστολικής πίεσης.

2. α) 178,7 cm β) 177,5 cm.

3. α) 168,3 cm β) 168,1 cm.

4. α) $\hat{y} = -1{,}88 + 1{,}18x$.
   β) Περίπου 27 έτη και 7 μήνες.
   γ) Περίπου 1 έτος και 2 μήνες.

5. α) $\bar{x} = 5{,}25 + 0{,}72y$.
   β) Περίπου 25 έτη και 5 μήνες.
   γ) Περίπου 9 μήνες.

Ευρετήριο Όρων

Στο Ευρετήριο όρων τα γράμματα Α, Β, Γ, Δ και Ε δηλώνουν τον 1ο, 2ο, 3ο, 4ο και 5ο τόμο αντίστοιχα, ενώ οι αριθμοί αναφέρονται στην πρώτη από τις δύο ενδείξεις που αναγράφονται σε κάθε σελίδα.

22-0088_I_c_math_bm_83-117_28b.indd 155 18/09/2017 09:58

22-0088_I_c_math_bm_83-117_28b.indd 160 18/09/2017 09:58

ΠΕΡΙΕΧΟΜΕΝΑ 3ου ΤΟΜΟΥ

Βάσει του ν. 3966/2011 τα διδακτικά βιβλία του Δημοτικού, του Γυμνασίου, του Λυκείου, των ΕΠΑ.Λ. και των ΕΠΑ.Σ. τυπώνονται από το ΙΤΥΕ - ΔΙΟΦΑΝΤΟΣ και διανέμονται δωρεάν στα Δημόσια Σχολεία. Τα βιβλία μπορεί να διατίθενται προς πώληση, όταν φέρουν στη δεξιά κάτω γωνία του εμπροσθόφυλλου ένδειξη «ΔΙΑΤΙΘΕΤΑΙ ΜΕ ΤΙΜΗ ΠΩΛΗΣΗΣ». Κάθε αντίτυπο που διατίθεται προς πώληση και δεν φέρει την παραπάνω ένδειξη θεωρείται κλεψίτυπο και ο παραβάτης διώκεται σύμφωνα με τις διατάξεις του άρθρου 7 του νόμου 1129 της 15/21 Μαρτίου 1946 (ΦΕΚ 1946,108, Α').

22-0088_I_c_math_bm_83-117_28b.indd 164 18/09/2017 09:58