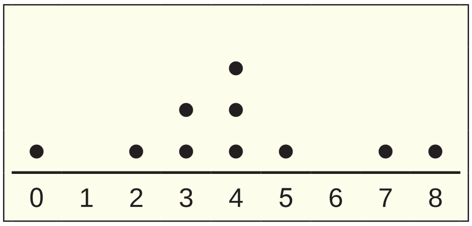
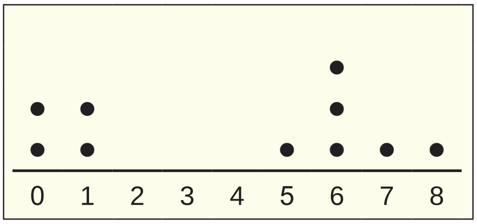
ΕΝΟΤΗΤΑ 2.3: ΜΕΤΡΑ ΘΕΣΗΣ ΚΑΙ ΜΕΤΑΒΛΗΤΟΤΗΤΑΣ, ΘΗΚΟΓΡΑΜΜΑ, ΣΥΝΤΕΛΕΣΤΗΣ ΜΕΤΑΒΛΗΤΟΤΗΤΑΣ

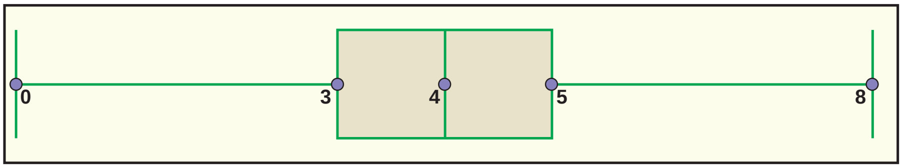
Διερεύνηση

1) Στα επόμενα σημειογράμματα φαίνονται οι πόντοι που σημείωσαν οι 10 καλαθοσφαιριστές κάθε μιας από τις δυο ομάδες μπάσκετ, στο ημίχρονο ενός αγώνα.





1) Να υπολογίσετε την μέση τιμή των πόντων που σημείωσαν οι 10 καλαθοσφαιριστές κάθε ομάδας στο ημίχρονο.
2) Αφού γράψετε τις 10 παρατηρήσεις κάθε ομάδας σε αύξουσα σειρά, να βρείτε για κάθε ομάδα:
   1) Τη μικρότερη παρατήρηση.
   2) Τη μεσαία παρατήρηση από τις πέντε πρώτες, δηλαδή το πρώτο τεταρτημόριο.
   3) Το ημιάθροισμα των δυο μεσαίων παρατηρήσεων, δηλαδή το δεύτερο τεταρτημόριο ή αλλιώς τη διάμεσο.
   4) Τη μεσαία παρατήρηση από τις πέντε τελευταίες, δηλαδή το τρίτο τεταρτημόριο.
   5) Τη μεγαλύτερη παρατήρηση
3) Με βάση τα παραπάνω, το θηκόγραμμα για την πρώτη ομάδα είναι αυτό που βλέπετε στο επόμενο σχήμα. Να σχεδιάσετε το θηκόγραμμα για τη δεύτερη ομάδα.



4) Ποια είναι η διαφορά της μικρότερης από τη μεγαλύτερη παρατήρηση, δηλαδή το εύρος των παρατηρήσεων, για κάθε ομάδα; Είναι το εύρος ένα αξιόπιστο μέτρο μεταβλητότητας;

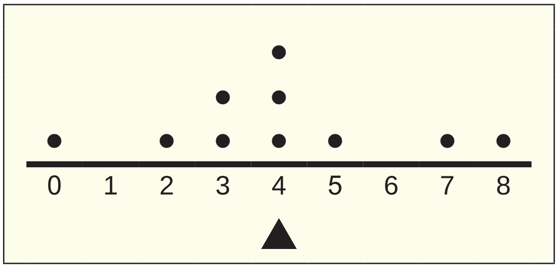
5) Ποια είναι η διαφορά του πρώτου από το τρίτο τεταρτημόριο, δηλαδή το ενδοτεταρτημοριακό εύρος, για κάθε ομάδα;

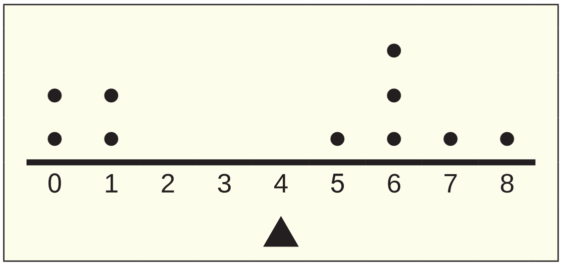
6) Ποια τιμή εμφανίζεται τις περισσότερες φορές, δηλαδή ποια είναι η επικρατούσα τιμή για κάθε ομάδα;

7) Πώς θα υπολογίσουμε τη μέση τιμή των τετραγώνων των διαφορών της μέσης τιμής των παρατηρήσεων από τις παρατηρήσεις, δηλαδή τη διακύμανση των παρατηρήσεων; Περιγράψτε τη διαδικασία και τις αριθμητικές πράξεις που απαιτούνται για κάθε ομάδα ξεχωριστά.

8) Να υπολογίσετε την τετραγωνική ρίζα της διακύμανσης, δηλαδή την τυπική απόκλιση για κάθε ομάδα.

2) Στα επόμενα δύο σχήματα οι κουκίδες συμβολίζουν δυο ομάδες, από 10 βαρίδια ίδιου βάρους η κάθε μια, τοποθετημένα πάνω σε δυο αβαρείς ράβδους αντίστοιχα, αριθμημένες έτσι ώστε οι αποστάσεις των διαδοχικών αριθμών να είναι ίσες.





1) Να εξηγήσετε τον λόγο για τον οποίο, αν τοποθετήσουμε μια τριγωνική σφήνα στη θέση 4 σε κάθε ράβδο, τότε αυτές ισορροπούν. Ποια είναι η φυσική ερμηνεία της μέσης τιμής των βαρών σε σχέση με το σημείο ισορροπίας αυτών;

2) Είναι το εύρος, στην παραπάνω περίπτωση, ένα αξιόπιστο μέτρο μεταβλητότητας των βαρών των δυο ομάδων;

3) Ποια είναι η μέση τιμή των τετραγώνων των διαφορών των παρατηρήσεων από τη μέση τιμή τους, δηλαδή η διακύμανση των βαρών κάθε ομάδας;

4) Εκφράζονται η διακύμανση και οι παρατηρήσεις με την ίδια μονάδα μέτρησης;

5) Ποια είναι η τετραγωνική ρίζα της διακύμανσης, δηλαδή η τυπική απόκλιση των βαρών κάθε ομάδας;

6) Εκφράζονται η τυπική απόκλιση και οι παρατηρήσεις με την ίδια μονάδα μέτρησης;

7) Να εκφράσετε το πηλίκο $\dfrac{\text{τυπική απόκλιση}}{\text{μέση τιμή}}$ ως ποσοστό για κάθε ομάδα βαρών. Ποια ομάδα έχει μικρότερο πηλίκο οπότε είναι περισσότερο ομοιογενής;

Βασικές μαθηματικές έννοιες - Ιδέες - Διεργασίες

Ορισμοί για τα μέτρα θέσης ποσοτικών δεδομένων

- Η **μέση τιμή $\bar{x}$** ενός συνόλου ν παρατηρήσεων ορίζεται, ως το άθροισμα των παρατηρήσεων διά του πλήθους τους.

$$\bar{x}=\frac{t_1+t_2+t_3+\ldots+t_\nu}{\nu}$$

Αν για παράδειγμα οι βαθμοί ενός μαθητή σε δέκα μαθήματα είναι: 18, 17, 19, 17, 18, 16, 18, 17, 16, 18, τότε:

$$\bar{x}=\frac{18+17+19+17+18+16+18+17+16+18}{10}=\frac{174}{10}=17{,}4$$

Αν οι τιμές των παρατηρήσεων είναι x_i με αντίστοιχες συχνότητες ν_i, τότε:

$$\bar{x}=\frac{x_1\cdot\nu_1+x_2\cdot\nu_2+\ldots+x_\kappa\cdot\nu_\kappa}{\nu}$$

x_i	ν_i	$x_i\cdot\nu_i$
16	2	32
17	3	51
18	4	72
19	1	19
Σύνολο	10	174

Για το παράδειγμα με τους βαθμούς του μαθητή στα δέκα μαθήματα έχουμε:

$$\bar{x}=\frac{16\cdot 2+17\cdot 3+18\cdot 4+19\cdot 1}{10}=\frac{174}{10}=17{,}4$$

- Προκειμένου να υπολογίσουμε τη διάμεσο ενός δείγματος ν παρατηρήσεων, αρχικά διατάσσουμε τις παρατηρήσεις σε αύξουσα σειρά. Η **διάμεσος δ** ορίζεται ως:
  - η μεσαία παρατήρηση του διατεταγμένου δείγματος, όταν το ν είναι περιττός αριθμός,
  - το ημιάθροισμα των δύο μεσαίων παρατηρήσεων του διατεταγμένου δείγματος, όταν το ν είναι άρτιος αριθμός.

Για το παράδειγμα με τους βαθμούς του μαθητή στα δέκα μαθήματα, αν τους διατάξουμε σε αύξουσα σειρά, έχουμε: 16, 16, 17, 17, 17, 18, 18, 18, 18, 19. Επομένως:

$$\delta=\frac{17+18}{2}=17{,}5$$

Αν έχουμε τις παρατηρήσεις 4, 4, 5, 5, 6, 7, 39 σε αύξουσα σειρά, τότε η διάμεσος είναι δ = 5 ενώ η μέση τιμή είναι:

$$\bar{x} = \frac{4+4+5+5+6+7+39}{7} = \frac{70}{7} = 10$$

Παρατηρούμε ότι, στην παραπάνω περίπτωση, η διάμεσος επηρεάζεται λιγότερο από την ακραία παρατήρηση «39».

Είναι φανερό ότι το πολύ 50% των παρατηρήσεων είναι μικρότερες και το πολύ 50% είναι μεγαλύτερες από τη διάμεσο δ.

- Τα **τεταρτημόρια** συμβολίζονται με Q_1, Q_2 και Q_3.
  - Για το Q_1 έχουμε αριστερά το πολύ 25% και δεξιά το πολύ 75% των παρατηρήσεων.
  - Για το Q_2 έχουμε Q_2 = δ, δηλαδή συμπίπτει με τη διάμεσο.
  - Για το Q_3 έχουμε αριστερά το πολύ 75% και δεξιά το πολύ 25% των παρατηρήσεων.

  Για τον ευκολότερο υπολογισμό των Q_1 και Q_3 έχουμε:
  - Όταν το πλήθος των διατεταγμένων παρατηρήσεων είναι άρτιος αριθμός, διακρίνουμε το πρώτο μισό, του οποίου η διάμεσος είναι το Q_1 και το δεύτερο μισό, του οποίου η διάμεσος είναι το Q_3.
  - Όταν το πλήθος των διατεταγμένων παρατηρήσεων είναι περιττός αριθμός, αφαιρούμε από το δείγμα τη διάμεσο και διακρίνουμε το πρώτο μισό, του οποίου η διάμεσος είναι το Q_1 και το δεύτερο μισό, του οποίου η διάμεσος είναι το Q_3.

  Για το παράδειγμα με τους βαθμούς του μαθητή στα δέκα μαθήματα είναι:

  $$Q_1 = 17,\ \ Q_2 = δ = 17{,}5,\ \ Q_3 = 18$$

- **Επικρατούσα τιμή M_0** είναι η παρατήρηση με τη μεγαλύτερη συχνότητα. Στο παράδειγμα με τους βαθμούς του μαθητή στα δέκα μαθήματα, επικρατούσα τιμή είναι M_0 = 18. Η επικρατούσα τιμή μπορεί να μην είναι μοναδική. Όταν όλες οι παρατηρήσεις είναι διαφορετικές, τότε λέμε ότι δεν υπάρχει επικρατούσα τιμή.

Ορισμοί για τα μέτρα διασποράς ποσοτικών δεδομένων

- Το **εύρος R** είναι το απλούστερο από τα μέτρα διασποράς και ορίζεται ως η διαφορά της ελάχιστης x_{min} παρατήρησης από τη μέγιστη x_{max} παρατήρηση. Δηλαδή:

  $$R = \text{μέγιστη παρατήρηση} - \text{ελάχιστη παρατήρηση} = x_{max} - x_{min}$$

- Το **ενδοτεταρτημοριακό εύρος Q** είναι η διαφορά του πρώτου τεταρτημορίου Q_1 από το τρίτο τεταρτημόριο Q_3. Δηλαδή:

  $$Q = Q_3 - Q_1$$

Όσες παρατηρήσεις βρίσκονται έξω από το διάστημα $[Q_1 - 1{,}5\cdot Q\ ,\ Q_3 + 1{,}5\cdot Q]$ ονομάζονται **ακραίες τιμές**. Η διάμεσος είναι ένα μέτρο θέσης το οποίο δεν επηρεάζεται από το μέγεθος των ακραίων τιμών.

Το εύρος και το ενδοτεταρτημοριακό εύρος είναι πολύ απλά μέτρα διασποράς και δεν εκφράζουν την απόκλιση που έχει κάθε παρατήρηση από το «κέντρο» των παρατηρήσεων. Για τον λόγο αυτό χρησιμοποιούμε και τα επόμενα μέτρα διασποράς.

- Η **διακύμανση s^2** ορίζεται ως η μέση τιμή των τετραγώνων των διαφορών της μέσης τιμής των παρατηρήσεων από τις παρατηρήσεις. Δηλαδή:

$$s^2 = \frac{(t_1 - \bar{x})^2 + (t_2 - \bar{x})^2 + (t_3 - \bar{x})^2 + \ldots + (t_\nu - \bar{x})^2}{\nu}$$

Για παράδειγμα, αν οι βαθμοί έξι μαθητών στα Μαθηματικά είναι: 18, 15, 18, 18, 15, 18, τότε:

$$\bar{x} = \frac{18+15+18+18+15+18}{6} = \frac{102}{6} = 17 \text{ και}$$

$$s^2 = \frac{(18-17)^2 + (15-17)^2 + (18-17)^2 + (18-17)^2 + (15-17)^2 + (18-17)^2}{6} = \frac{12}{6} = 2$$

Αν οι τιμές των παρατηρήσεων είναι x_i με αντίστοιχες συχνότητες ν_i, τότε:

$$s^2 = \frac{(x_1 - \bar{x})^2 \cdot \nu_1 + (x_2 - \bar{x})^2 \cdot \nu_2 + \ldots + (x_\kappa - \bar{x})^2 \cdot \nu_\kappa}{\nu}$$

Για το παράδειγμα των βαθμών των έξι μαθητών στα Μαθηματικά, σχετικός είναι ο επόμενος πίνακας:

x_i	ν_i	$x_i \cdot \nu_i$	$x_i - \bar{x}$	$(x_i - \bar{x})^2$	$(x_i - \bar{x})^2 \cdot \nu_i$
15	2	30	-2	4	8
18	4	72	1	1	4
Σύνολο	6	102			12

$$\bar{x} = \frac{102}{6} = 17 \quad \text{και} \quad s^2 = \frac{12}{6} = 2$$

Η διακύμανση δεν εκφράζεται με τις μονάδες που εκφράζονται οι παρατηρήσεις.

- Η **τυπική απόκλιση s** είναι: $s = \sqrt{s^2}$. Στο παραπάνω παράδειγμα είναι: $s = \sqrt{2} \approx 1{,}41$. Η τυπική απόκλιση έχει το πλεονέκτημα ότι εκφράζεται με τις μονάδες, που εκφράζονται οι παρατηρήσεις.

Θηκόγραμμα

- Έστω ότι έχουμε τον αριθμό των εργαζομένων σε 20 βιοτεχνίες.

10	14	25	7	31	8	12	19	10	24	9	13	5	28	24	19	26	51	68	14

Κατατάσσουμε τις παρατηρήσεις σε αύξουσα σειρά:

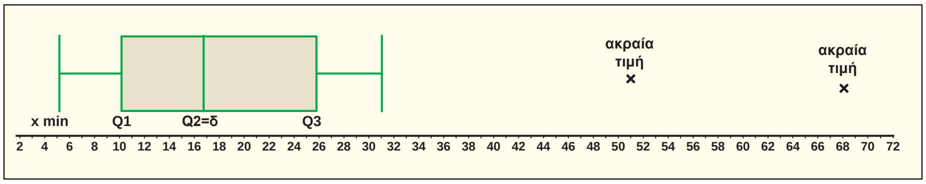
5, 7, 8, 9, 10, 10, 12, 13, 14, 14, 19, 19, 24, 24, 25, 26, 28, 31, 51, 68

Είναι: $Q_1 = \dfrac{10+10}{2} = 10$, $Q_2 = \delta = \dfrac{14+19}{2} = 16{,}5$, $Q_3 = \dfrac{25+26}{2} = 25{,}5$

Το ενδοτεταρτημοριακό εύρος είναι: Q = 25,5 - 10 = 15,5 και 1,5 · Q = 23,25

Επομένως $\left[Q_1 - 1{,}5 \cdot Q\, ,\, Q_3 + 1{,}5 \cdot Q\right] = \left[-13{,}25\, ,\, 48{,}75\right]$. Οπότε ακραίες παρατηρήσεις είναι το 51 και το 68.

Τα μέτρα αυτά απεικονίζονται στο επόμενο διάγραμμα που λέγεται **θηκόγραμμα**.



Από τα μέσα των πλευρών, που παριστάνουν το πρώτο και το τρίτο τεταρτημόριο (δηλαδή τα Q_1 και Q_3 αντιστοίχως) φέρουμε ευθύγραμμα τμήματα με μήκος που προσδιορίζεται ως εξής:

Η μικρότερη παρατήρηση που είναι μεγαλύτερη του Q_1-1,5·Q είναι το 5, επομένως το αριστερό άκρο του τμήματος που θα φέρουμε από το Q_1 είναι το 5 και δεν υπάρχουν παρατηρήσεις μικρότερες του Q_1-1,5·Q για να σημειωθούν χωριστά.

Η μεγαλύτερη παρατήρηση που είναι μικρότερη του Q_3+1,5·Q είναι η 31, επομένως το δεξί άκρο του τμήματος που θα φέρουμε από το Q_3 είναι το 31 και θα σημειώσουμε χωριστά τις ακραίες τιμές 51 και 68 οι οποίες είναι μεγαλύτερες του Q_3+1,5·Q.

Συντελεστής μεταβλητότητας

- Δυο όμιλοι επιχειρήσεων, που ο καθένας αποτελείται από 5 επιχειρήσεις, είχαν ετήσιες δαπάνες για το οικονομικό έτος 2017 τα ποσά, σε χιλιάδες ευρώ, που αναγράφονται στον παρακάτω πίνακα.

Όμιλος Α	200	250	300	300	350
Όμιλος Β	5.000	5.050	5.100	5.100	5.150

Οι μέσες τιμές των δαπανών για κάθε όμιλο είναι:

$$\overline{x}_A = \frac{200+250+300+300+350}{5} = 280 \text{ χιλιάδες ευρώ.}$$

$$\overline{x}_B = \frac{5.000+5.050+5.100+5.100+5.100+5.150}{5} = 5.080 \text{ χιλιάδες ευρώ.}$$

Οι διακυμάνσεις των δαπανών για κάθε όμιλο είναι:

$$s_A^2 = \frac{(200-280)^2+(250-280)^2+(300-280)^2+(300-280)^2+(350-280)^2}{5}$$

$$s_A^2 = 2.600$$

$$s_B^2 = \frac{(5.000-5.080)^2+(5.050-5.080)^2+(5.100-5.080)^2+(5.100-5.080)^2+(5.150-5.080)^2}{5}$$

$$s_B^2 = 2.600$$

Οι τυπικές αποκλίσεις των δαπανών για κάθε όμιλο είναι ίσες με:

$$s_A = s_B = \sqrt{2.600} \approx 51 \text{ χιλιάδες ευρώ.}$$

Μπορούμε άραγε να ισχυριστούμε ότι οι 51 χιλιάδες ευρώ των ίσων τυπικών αποκλίσεων έχουν την ίδια βαρύτητα για τους ομίλους Α και Β, που έχουν διαφορετικές μέσες τιμές $\overline{x}_A = 280$ χιλιάδες ευρώ και $\overline{x}_B = 5.080$ χιλιάδες ευρώ; Ένας τέτοιος ισχυρισμός θα ήταν λανθασμένος, γιατί η σχέση τιμών της μεταβλητής «ετήσιες δαπάνες» με την τιμή της τυπικής απόκλισης, είναι διαφορετικές για τις επιχειρήσεις των δυο ομίλων.

Ένα μέτρο, που μας βοηθάει να αντιμετωπίσουμε τέτοια προβλήματα και μας δίνει τη δυνατότητα συγκρίσεων, είναι ο **συντελεστής μεταβλητότητας CV**, ο οποίος είναι καθαρός αριθμός και εκφράζεται συνήθως ως ποσοστό. Ορίζεται ως ακολούθως:

$$CV = \frac{s}{|\overline{x}|}$$

Στο παράδειγμά μας ο συντελεστής μεταβλητότητας για τις δυο επιχειρήσεις είναι:

$$CV_A = \frac{51}{280} \approx 0,1821 = 18,21\% \text{ και } CV_B = \frac{51}{5.080} \approx 0,01004 \approx 1,00\%$$

Όσο μικρότερο είναι το ποσοστό αυτό, τόσο περισσότερη ομοιογένεια υπάρχει στις τιμές της μεταβλητής που μας ενδιαφέρει. Στο παράδειγμά μας είναι $CV_A > CV_B$, επομένως η διασπορά των τιμών στον όμιλο Α σε σχέση με τη μέση τιμή είναι μεγαλύτερη από την αντίστοιχη διασπορά στον όμιλο Β.

Εφαρμογές

Εφαρμογή 1

Μία ομάδα δέκα μαθητών μέτρησε το μήκος ενός θρανίου με ακρίβεια δέκατου του εκατοστού του μέτρου (χιλιοστού του μέτρου). Οι μαθητές χρησιμοποίησαν την ίδια μετροταινία και κατέγραψαν τις τιμές των δέκα μετρήσεων, όπως φαίνεται στον παρακάτω πίνακα.

Μήκος σε cm	120,2	120,1	120,1	119,8	119,7	120,3	120,2	119,9	120,1	119,8

1) Να αναφέρετε δύο λόγους που, κατά τη γνώμη σας, δικαιολογούν τις διαφορές στις μετρήσεις του θρανίου.
2) Πώς υπολογίζουμε τη μέση τιμή των μετρήσεων;
3) Ποια είναι η τιμή του μήκους του θρανίου που θα χρησιμοποιήσουμε;
4) Για ποιον λόγο είναι χρήσιμος ο υπολογισμός της μέσης τιμής πολλών μετρήσεων;
5) Δικαιολογήστε με παράδειγμα -που θα δημιουργήσετε από τις τιμές του πίνακα- γιατί ζητήθηκε από τους μαθητές να κάνουν δέκα μετρήσεις και όχι λιγότερες (π.χ. τρεις μετρήσεις).

<u>Λύση</u>

1) Δύο λόγοι που δικαιολογούν τις διαφορές μπορεί να είναι:
   1) Η αρχή της μετροταινίας δε συμπίπτει με την αρχή της μετρούμενης απόστασης
   2) Η μετροταινία δεν ακολουθεί ευθεία και παράλληλη γραμμή προς τη μετρούμενη απόσταση.
2) Για τη μέση τιμή των μετρήσεων έχουμε:

$$\bar{x}=\frac{120,2+120,1+120,1+119,8+119,7+120,3+120,2+119,9+120,1+119,8}{10}$$

$$=\frac{1200,2}{10}=120,02$$

3) Για την άγνωστη τιμή του μήκους του θρανίου, μπορούμε να χρησιμοποιήσουμε τη μέση τιμή $\bar{x}=120,02\,\text{cm}$ του δείγματος.
4) Θέλουμε τη μέση τιμή πολλών μετρήσεων, δηλαδή θέλουμε μεγάλο μέγεθος δείγματος, ώστε το αποτέλεσμα να έχει μεγαλύτερη πιθανότητα να προσεγγίσει ικανοποιητικά το πραγματικό μήκος του θρανίου.
5) Αν για παράδειγμα πάρουμε τη μέση τιμή των τριών πρώτων μετρήσεων έχουμε:

$$\bar{x}'=\frac{120,2+120,1+120,1}{3}=\frac{360,4}{3}=120,133...\,\text{cm}$$

Αυτή είναι η μέση τιμή ενός άλλου δείγματος. Το μήκος όμως του θρανίου είναι μοναδικό. Το μεγαλύτερο μέγεθος του δείγματος μας δίνει τη δυνατότητα ακριβέστερης εκτίμησης του πραγματικού μήκους.

Εφαρμογή 2

Τα αποτελέσματα του SAT τεστ στις ΗΠΑ, για μαθητές και μαθήτριες που τελειώνοντας το σχολείο προτίθενται να συνεχίσουν τις σπουδές τους, έδειξαν το 2019 τα εξής :

Πήραν μέρος 2.220.087 μαθητές και μαθήτριες και η μεγαλύτερη δυνατή επίδοση (χωρίς κανένα λάθος/καμία απώλεια βαθμών) ήταν 1.600. Η μέση επίδοση του παραπάνω πληθυσμού στο τεστ ήταν 1.059.

Στο πλαίσιο μίας διερευνητικής εργασίας που έγινε σε εννέα διαφορετικές τάξεις, δέκα ομάδες μαθητών/τριών από κάθε τάξη συγκέντρωσαν η καθεμία τις επιδόσεις ενός διαφορετικού δείγματος μαθητών και μαθητριών πλήθους 100 (προσπαθώντας κάθε δείγμα να είναι, κατά το δυνατόν, αντιπροσωπευτικό του πληθυσμού) και υπολόγισαν τη μέση επίδοση του δείγματος.

Τα αποτελέσματα της δουλειάς των ομάδων φαίνονται στον παρακάτω πίνακα:

		Ομάδες									
		1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Τάξεις	**1**	1063	1059	1047	1056	1057	1040	1038	1076	1053	1038
	2	1071	1039	1062	1056	1071	1054	1043	1067	1078	1048
	3	1061	1074	1068	1060	1076	1047	1056	1061	1061	1057
	4	1040	1072	1039	1076	1043	1063	1072	1067	1061	1056
	5	1051	1071	1051	1072	1049	1049	1080	1055	1049	1044
	6	1072	1056	1070	1049	1045	1051	1061	1049	1076	1071
	7	1065	1058	1080	1056	1079	1040	1068	1075	1053	1072
	8	1077	1038	1069	1065	1054	1055	1054	1065	1060	1072
	9	1069	1076	1069	1079	1041	1051	1048	1064	1048	1063

Ο καθηγητής της τάξης 1 αποφάσισε να υπολογίσει μία νέα μέση επίδοση συνολικά, με χρήση και των δέκα που υπολόγισαν οι ομάδες των μαθητών/τριών του.

Πώς σχολιάζετε την ενέργεια αυτή;

Τι παρατηρείτε σχετικά με το αποτέλεσμα που προκύπτει;

(Πηγή: SAT Suite of Assessments Annual Report 2019, ανακτήθηκε από https://research.collegeboard.org/programs/sat/data/2019-sat-suite-annual-report, τελευταία προσπέλαση, 28/11/2019)

Λύση

Η ενέργεια του καθηγητή έχει νόημα καθώς το δείγμα συνεχίζει να είναι αντιπροσωπευτικό, με τη διαφορά ότι έχει μέγεθος 1000 (δεκαπλάσιο).
Το αποτέλεσμα προκύπτει ως εξής:

$$\bar{x}_1 = \frac{1063+1059+1047+1056+1057+1040+1038+1076+1053+1038}{10} = 1052,7$$

Με όμοιο τρόπο έχουμε (για τις υπόλοιπες τάξεις):

$\bar{x}_2$	$\bar{x}_3$	$\bar{x}_4$	$\bar{x}_5$	$\bar{x}_6$	$\bar{x}_7$	$\bar{x}_8$	$\bar{x}_9$
1058,9	1062,1	1058,9	1057,1	1060	1064,6	1060,9	1060,8

Όπως παρατηρούμε, οι νέες μέσες επιδόσεις των νέων δειγμάτων με πλήθος 1000 αντί για 100 είναι «συνολικά πιο κοντά» στη μέση επίδοση του πληθυσμού (που είναι 1059), από τις μέσες επιδόσεις των δειγμάτων με πλήθος 100.
Επίσης, αν θεωρήσουμε όλα τα δείγματα σαν ένα δείγμα πλήθους 9.000, η μέση επίδοση αυτού του ακόμα μεγαλύτερου δείγματος θα είναι 1059,556.

Στατιστικά μέτρα δείγματος και πληθυσμού

Γενικότερα

Σε ένα δείγμα παρατηρήσεων το μέσο $\bar{x}$ τον λέμε δειγματικό μέσο και την τυπική απόκλιση s τη λέμε δειγματική τυπική απόκλιση και μπορούμε να τα υπολογίσουμε. Εξαρτώνται βέβαια κάθε φορά από την επιλογή του δείγματος. Όμως, μέσω του δείγματος, τον μέσο μ του πληθυσμού και την τυπική του απόκλιση σ μπορούμε μόνο να τα εκτιμήσουμε και όχι να τα υπολογίσουμε.

Εφαρμογή 3

Η μέση τιμή των μηνιαίων μισθών των υπαλλήλων μιας εταιρείας είναι 1.200 ευρώ και η τυπική απόκλιση 100 ευρώ.

1) Αν γίνει μια κράτηση σε κάθε μηνιαίο μισθό κατά 50 ευρώ, τότε ποια μεταβολή επέρχεται στη μέση τιμή, στην τυπική απόκλιση και στον συντελεστή μεταβλητότητας των μισθών;

2) Αν γίνει μια κράτηση σε κάθε μηνιαίο μισθό κατά 5%, τότε ποια μεταβολή επέρχεται στη μέση τιμή, στην τυπική απόκλιση και στον συντελεστή μεταβλητότητας των μισθών;

Λύση

Έστω ότι $t_1, t_2, t_3, \dots, t_\nu$ είναι οι μηνιαίοι μισθοί των υπαλλήλων της εταιρείας. Τότε:

✓ $\bar{x} = \dfrac{t_1 + t_2 + t_3 + \dots + t_\nu}{\nu} = 1.200$

✓ $s = \sqrt{\dfrac{(t_1 - 1.200)^2 + (t_2 - 1.200)^2 + \dots + (t_\nu - 1.200)^2}{\nu}} = 100$

✓ $CV = \frac{s}{|\bar{x}|} = \frac{100}{1.200} \approx 8,33\%$

1) Αν γίνει μια κράτηση σε κάθε μηνιαίο μισθό κατά 50 ευρώ, τότε:

$$\bar{y} = \frac{(t_1 - 50) + (t_2 - 50) + ... + (t_\nu - 50)}{\nu} = \frac{t_1 + t_2 + ... + t_\nu}{\nu} - \frac{50 \cdot \nu}{\nu} = 1.200 - 50 = 1.150$$

$$s_y = \sqrt{\frac{(t_1 - 50 - 1.150)^2 + ... + (t_\nu - 50 - 1.150)^2}{\nu}} = \sqrt{\frac{(t_1 - 1.200)^2 + ... + (t_\nu - 1.200)^2}{\nu}} = s = 100$$

$$CV_y = \frac{s_y}{|\bar{y}|} = \frac{100}{1.150} \approx 8,7\%$$

Επομένως:

✓ Η μέση τιμή $\bar{x} = 1.200$€ μειώνεται κατά 50€ και γίνεται $\bar{y} = 1.150$ €

✓ Η τυπική απόκλιση δε μεταβάλλεται.

✓ Ο συντελεστής μεταβλητότητας CV = 8,33% αυξάνεται και γίνεται CV_y = 8,7%

2) Αν γίνει μια κράτηση σε κάθε μηνιαίο μισθό κατά 5%, τότε όλοι οι μηνιαίοι μισθοί πολλαπλασιάζονται με 0,95. Έτσι έχουμε:

$$\bar{z} = \frac{0,95 \cdot t_1 + 0,95 \cdot t_2 + ... + 0,95 \cdot t_\nu}{\nu} = 0,95 \cdot \frac{t_1 + t_2 + ... + t_\nu}{\nu} = 0,95 \cdot 1200$$

$$\bar{z} = 1.140$$

$$s_z = \sqrt{\frac{(0,95 \cdot t_1 - 0,95 \cdot \bar{x})^2 + (0,95 \cdot t_2 - 0,95 \cdot \bar{x})^2 + ... + (0,95 \cdot t_\nu - 0,95 \cdot \bar{x})^2}{\nu}}$$

$$= \sqrt{0,95^2 \cdot s^2} = |0,95| \cdot s = 0,95 \cdot 100 = 95$$

$$CV_z = \frac{s_z}{|\bar{z}|} = \frac{0,95 \cdot s}{0,95 \cdot \bar{x}} = CV = 8,33\%$$

Επομένως:

✓ Η μέση τιμή $\bar{x} = 1.200$€ μειώνεται κατά 5% και γίνεται $\bar{z} = 1.140$€

✓ Η τυπική απόκλιση s = 100 € μειώνεται κατά 5% και γίνεται s_z = 95 €

✓ Ο συντελεστής μεταβλητότητας δε μεταβάλλεται.

Έστω ότι έχουμε τις τιμές x_i, i = 1,2,...,ν, μιας ποσοτικής μεταβλητής X με μέση τιμή $\bar{x}$ και τυπική απόκλιση s_x. Τότε για τη μέση τιμή $\bar{y}$ και την τυπική απόκλιση s_y των τιμών της μεταβλητής Y για τις οποίες ισχύει $y_i = \alpha x_i + \beta$, i=1,2,...,ν και $\alpha, \beta \in R$ είναι: $\bar{y} = \alpha\bar{x} + \beta$ και $s_y = |\alpha| \cdot s_x$

Εφαρμογή 4

Στην εθνική οδό Αθηνών - Λαμίας τα αυτοκίνητα τρέχουν με σταθερή ταχύτητα. Εμείς, με το δικό μας αυτοκίνητο, τρέχουμε με 100 Km/h και προσπερνάμε σε πλήθος όσα ακριβώς μας προσπερνούν. Να βρείτε τη διάμεσο των ταχυτήτων του δικού μας αυτοκινήτου και όλων όσων μας προσπέρασαν ή προσπεράσαμε στο ταξίδι μας αυτό.

Λύση

Η διάμεσος είναι δ = 100 Km/h διότι το πολύ 50% των ταχυτήτων είναι μικρότερες και το πολύ 50% είναι μεγαλύτερες από 100 Km/h.

Εφαρμογή 5

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 του βάρους τους, αν γνωρίζουμε ότι στην κάλπη υπάρχουν:

1) 10 μπάλες.
2) 20 μπάλες.
3) Δε γνωρίζουμε πόσες μπάλες υπάρχουν στην κάλπη.

Λύση

x_i	v_i	f_i
10	1	0,10
11	2	0,20
12	3	0,30
13	4	0,40
Σύνολο	10	1,00

x_i	v_i	f_i
10	2	0,10
11	4	0,20
12	6	0,30
13	8	0,40
Σύνολο	20	1,00

Έστω ότι έχουμε τις τιμές x_i, i = 1, 2,…,ν *με σχετικές συχνότητες* f_i, i = 1, 2,…,ν. *Τότε για τη μέση τιμή* $\bar{x}$ *είναι:*

$\bar{x} = f_1 \cdot x_1 + f_2 \cdot x_2 + \ldots + f_\nu \cdot x_\nu$

1) Με βάση την αναλογία της εκφώνησης, στις 10 μπάλες θα έχουμε 1 άσπρη, 2 μαύρες, 3 κόκκινες και 4 πράσινες. Αν οι τιμές των βαρών είναι οι παρατηρήσεις x_i με αντίστοιχες συχνότητες v_i και σχετικές συχνότητες f_i, τότε:

$$\bar{x} = \frac{1 \cdot 10 + 2 \cdot 11 + 3 \cdot 12 + 4 \cdot 13}{10}$$
$$= 0{,}10 \cdot 10 + 0{,}20 \cdot 11 + 0{,}30 \cdot 12 + 0{,}40 \cdot 13 = 12 \text{ gr}$$

2) Με βάση την αναλογία της εκφώνησης, στις 20 μπάλες θα έχουμε 2 άσπρες, 4 μαύρες, 6 κόκκινες και 8 πράσινες. Αν οι τιμές των βαρών είναι οι παρατηρήσεις x_i με αντίστοιχες συχνότητες v_i και σχετικές συχνότητες f_i, τότε:

$$\bar{x} = \frac{2 \cdot 10 + 4 \cdot 11 + 6 \cdot 12 + 8 \cdot 13}{20}$$
$$= 0{,}10 \cdot 10 + 0{,}20 \cdot 11 + 0{,}30 \cdot 12 + 0{,}40 \cdot 13 = 12 \text{ gr}$$

3) Παρατηρώντας προσεκτικά την παραπάνω διαδικασία καταλήγουμε ότι ανεξαρτήτως του πλήθους από τις μπάλες έχουμε:

$$\bar{x} = f_1 \cdot x_1 + f_2 \cdot x_2 + f_3 \cdot x_3 + f_4 \cdot x_4 = 0{,}10 \cdot 10 + 0{,}20 \cdot 11 + 0{,}30 \cdot 12 + 0{,}40 \cdot 13 =$$
$$= 12 \text{ gr}$$

Ασκήσεις – Προβλήματα – Δραστηριότητες

1) Οι βαθμοί του Αντρέα σε 4 διαγωνίσματα στα Μαθηματικά ήταν 15, 18, 18, 17. Για τα ίδια διαγωνίσματα, ο Βασίλης είχε πάρει 2 μονάδες περισσότερες σε κάθε διαγώνισμα από τον Αντρέα, ενώ ο Γιάννης είχε πάρει 4 μονάδες λιγότερες από τον Αντρέα σε κάθε διαγώνισμα. Να βρείτε τη μέση τιμή των βαθμών του κάθε παιδιού.

2) Αν ο μέσος όρος του μηνιαίου μισθού των υπαλλήλων ενός εργοστασίου πέρυσι ήταν 850€ και φέτος σε κάθε υπάλληλο δοθεί αύξηση 50€, να βρείτε τον νέο μέσο όρο των μισθών.

3) Καθεμία από τις παρακάτω λίστες δεδομένων έχουν μέση τιμή 50.

(I) 0, 20, 40, 50, 60, 80, 100
(II) 0, 48, 49, 50, 51, 52, 100
(III) 0, 1, 2, 50, 98, 99, 100

1) Μπορεί να χρησιμοποιηθεί το εύρος για σύγκριση της μεταβλητότητας των δεδομένων αυτών;
2) Χωρίς να γίνουν οι πράξεις, να βρείτε σε ποια λίστα υπάρχει μεγαλύτερη και σε ποια μικρότερη διασπορά παρατηρήσεων.

4) 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

5) 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

6) Η βαθμολογία 16 μαθητών σε ένα διαγώνισμα ήταν: 8, 15, 13, 20, 9, 13, 17, 19, 20, 9, 10, 10, 15, 13, 14, 17. Να υπολογίσετε:

1) Τα τρία μέτρα θέσης, μέση τιμή, διάμεσο και επικρατούσα τιμή.
2) Το εύρος, την τυπική απόκλιση και τον συντελεστή μεταβλητότητας.

7) Η μέση τιμή ηλικίας των υπαλλήλων μιας εταιρείας είναι 32 χρόνια. Ποια θα είναι η μέση τιμή ηλικίας των ίδιων υπαλλήλων ύστερα από τρία χρόνια;

8) Οι βαθμοί στα Μαθηματικά 20 μαθητών της Β΄ τάξης ενός Λυκείου είναι:

12	14	15	13	17	15	16	14	18	15	17	13	19	15	16	12	16	18	13	14

1) Να βρείτε τη μέση τιμή και την επικρατούσα τιμή.
2) Να βρείτε τη διάμεσο.
3) Να βρείτε το πρώτο και το τρίτο τεταρτημόριο.
4) Να σχεδιάσετε το θηκόγραμμα.

Πρόσθετο Υλικό

Σύνθετες ασκήσεις

1) Η μέση επίδοση 17 αγοριών και 13 κοριτσιών στο μάθημα των Μαθηματικών μιας τάξης είναι 16,8. Η μέση επίδοση των κοριτσιών είναι 15,6. Να βρείτε τη μέση επίδοση των αγοριών.

2) Σε ένα Λύκειο υπάρχουν 500 μαθητές. Η Α΄ τάξη έχει 200 μαθητές με μέσο όρο ηλικίας 15,7 χρόνια, ενώ η Β΄ τάξη έχει 180 μαθητές με μέσο όρο ηλικίας 16,9 χρόνια. Οι υπόλοιποι μαθητές της Γ΄ τάξης έχουν μέσο όρο ηλικίας 17,7 χρόνια. Να υπολογίσετε τον μέσο όρο ηλικίας όλων των μαθητών του σχολείου.

3) Η μέση τιμή 40 παρατηρήσεων είναι 20. Αν από αυτές οι 7 μειώνονται κατά 2 και οι 9 αυξάνονται κατά 6, να βρεθεί η νέα μέση τιμή.

4) Η τυπική απόκλιση ενός δείγματος με ν παρατηρήσεις είναι ίση με μηδέν. Ποιο συμπέρασμα προκύπτει για τις παρατηρήσεις;

5) Η μέση τιμή 6 διαδοχικών ακέραιων αριθμών είναι 7,5. Να υπολογίσετε την τυπική τους απόκλιση.

6) Αν σε μία τάξη ο μέσος όρος της βαθμολογίας ν_1 αγοριών είναι $\overline{x}$ και ο μέσος όρος της βαθμολογίας ν_2 κοριτσιών είναι $\overline{y}$, να αποδείξετε ότι ο μέσος όρος της βαθμολογίας όλων των παιδιών της τάξης είναι $\overline{z} = \dfrac{\nu_1\overline{x} + \nu_2\overline{y}}{\nu_1 + \nu_2}$.

7) Να αποδείξετε ότι η μέση τιμή είναι το άθροισμα των γινομένων των τιμών της μεταβλητής επί τις αντίστοιχες σχετικές συχνότητες.

8) Σε ένα εργοστάσιο με 100 εργαζόμενους η μέση τιμή των αμοιβών τους είναι 900€. Οι 40 από αυτούς πληρώνονται με μισθό μικρότερο της μέσης τιμής και οι μισθοί τους έχουν μέση τιμή 800€. Αν οι αποδοχές των εργαζομένων με μισθό μικρότερο της μέσης τιμής αυξηθούν και γίνουν όσο η μέση τιμή, τότε ποια θα είναι η νέα μέση τιμή των αμοιβών των 100 εργαζομένων;