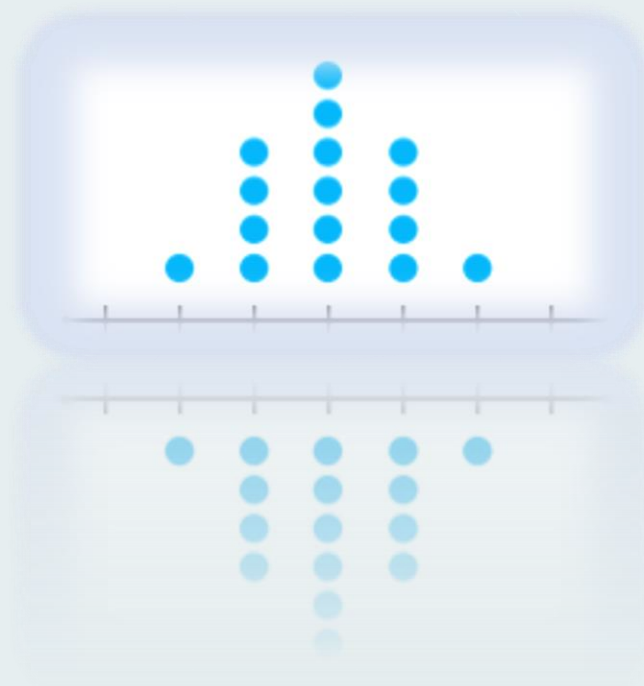
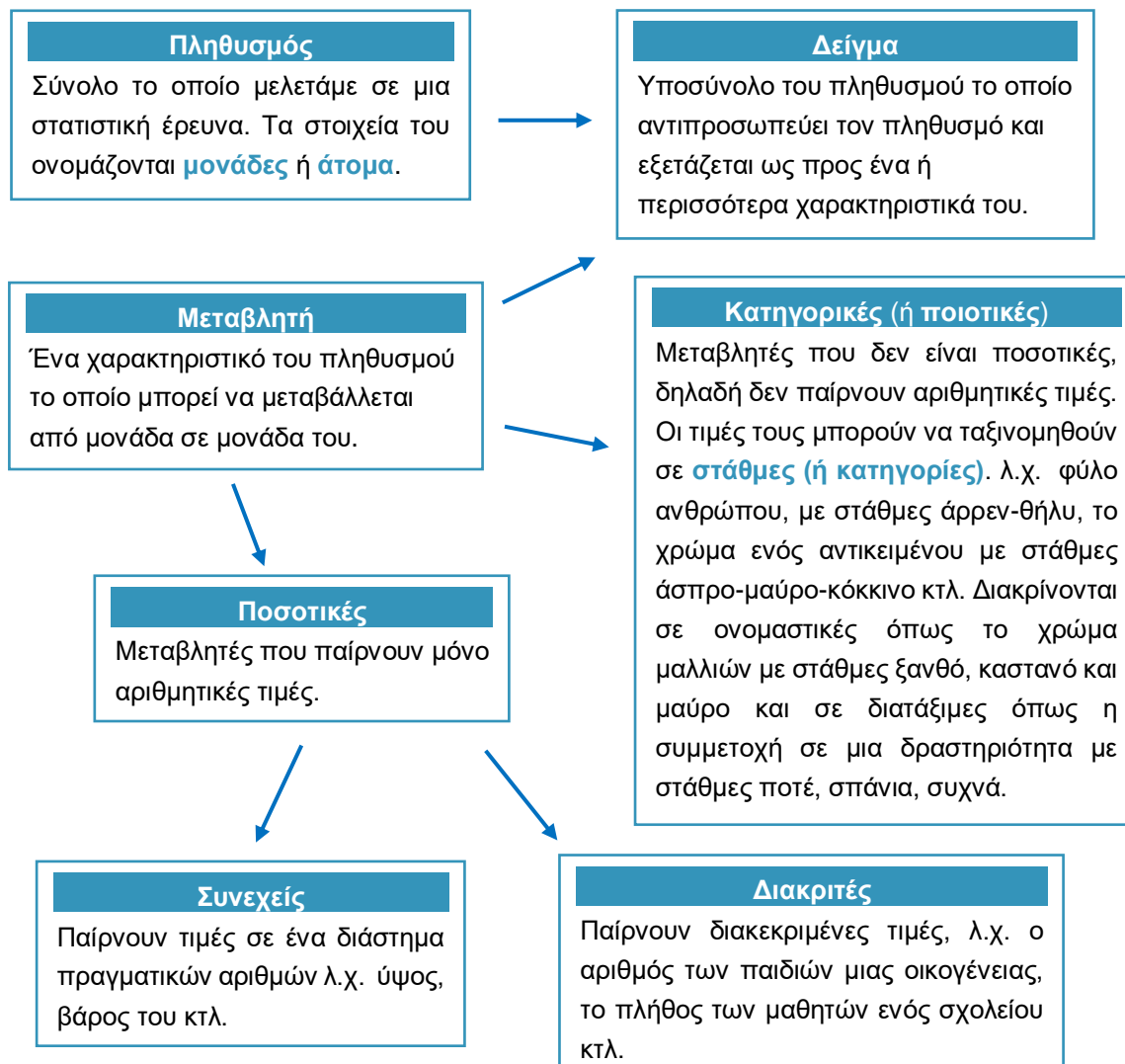


ΟΔΗΓΟΣ ΕΠΑΝΑΛΗΨΗΣ

ΣΤΑΤΙΣΤΙΚΗ



Βασικές έννοιες της Στατιστικής



Σχέσεις εξάρτησης μεταξύ κατηγορικής και ποσοτικής μεταβλητών

Σχέση εξάρτησης: Έχουμε όταν οι αλλαγές στη μία μεταβλητή οδηγούν σε συστηματικές και όχι τυχαίες αλλαγές στην άλλη μεταβλητή.

Μας ενδιαφέρει να μελετήσουμε: Αν οι τιμές της ποσοτικής μεταβλητής διαφέρουν μεταξύ των κατηγοριών της κατηγορικής μεταβλητής.

Παραδείγματα:

1. Υπάρχει διαφορά στους μισθούς (ποσοτική) των δύο φύλων (κατηγορική) που εργάζονται σε μια επιχείρηση;
2. Σχετίζεται ο βαθμός (ποσοτική) που παίρνει ένας μαθητής/τρια σε μια εξέταση μαθήματος στο σχολείο με τον χρόνο προετοιμασίας του (κατηγορική με στάθμες καθόλου, λίγο, πολύ, πάρα πολύ);

3. Υπάρχει διαφορά στον αριθμό προπονήσεων ανά μήνα (ποσοτική) των μαθητών/τριών ενός σχολείου και στο άθλημα (κατηγορική με στάθμες μπάσκετ, ποδόσφαιρο, βόλει) που κάνουν;

Θηκόγραμμα

Το θηκόγραμμα με οριακές τιμές: Η κατασκευή του βασίζεται στους παρακάτω αριθμούς.

- 1^ο τεταρτημόριο (Q_1)
- Διάμεσος (δ) ή (Q_2)
- 3^ο τεταρτημόριο (Q_3)
- Κάτω οριακή τιμή (L)
- Άνω οριακή τιμή (U)
- Ακραίες (απόμακρες) τιμές

Διάμεσος

Είναι η μεσαία διατεταγμένη παρατήρηση (περιττό πλήθος παρατηρήσεων) ή η μέση τιμή των δύο μεσαίων παρατηρήσεων (άρτιο πλήθος παρατηρήσεων).

1^ο τεταρτημόριο

Η διάμεσος των παρατηρήσεων αριστερά της διαμέσου δ .

3^ο τεταρτημόριο

Η διάμεσος των παρατηρήσεων δεξιά της διαμέσου δ .

Κάτω οριακή τιμή

Η μικρότερη παρατήρηση που ανήκει στο διάστημα $[Q_1 - 1,5Q, Q_3 + 1,5Q]$.

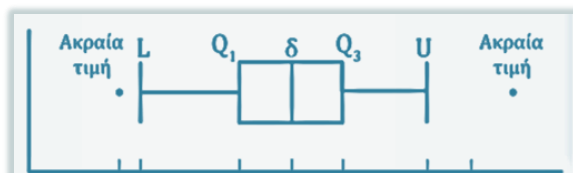
Άνω οριακή τιμή

Η μεγαλύτερη παρατήρηση που ανήκει στο διάστημα $[Q_1 - 1,5Q, Q_3 + 1,5Q]$.

Ακραίες (απόμακρες) τιμές

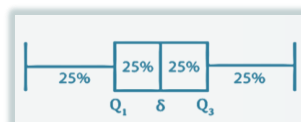
Οι παρατηρήσεις που βρίσκονται εκτός του διαστήματος $[Q_1 - 1,5Q, Q_3 + 1,5Q]$.

Σημείωση: Με Q συμβολίζουμε τη διαφορά $Q_3 - Q_1$ η οποία ονομάζεται **ενδοτεταρτημοριακό εύρος**. Οι αριθμοί $Q_1 - 1,5Q$ και $Q_3 + 1,5Q$ ονομάζονται **κάτω** και **άνω φράγμα**, αντίστοιχα.



Κατανομή παρατηρήσεων στο θηκόγραμμα

Οι αριθμοί Q_1 , δ και Q_3 χωρίζουν το θηκόγραμμα σε τέσσερα μέρη. Σε κάθε μέρος ανήκει, περίπου, το 25% των τιμών των δεδομένων.

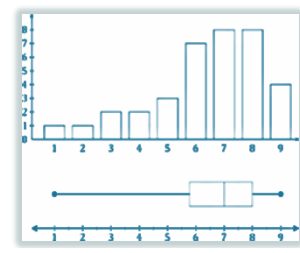
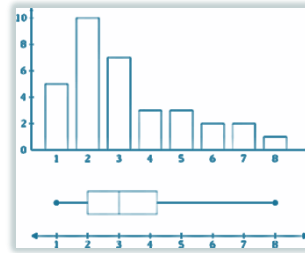
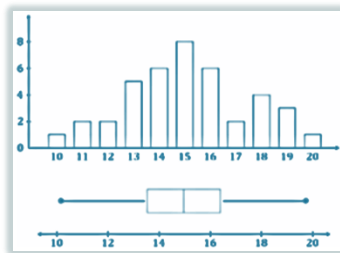


Σχήμα της κατανομής των παρατηρήσεων

Μια **συμμετρική κατανομή** αποτυπώνεται σε θηκόγραμμα με τη διάμεσο να βρίσκεται στο κέντρο του ορθογωνίου, και τις απολήξεις να εκτείνονται σε ίσες περίπου αποστάσεις προς τις δύο κατευθύνσεις (Σχήμα 1).

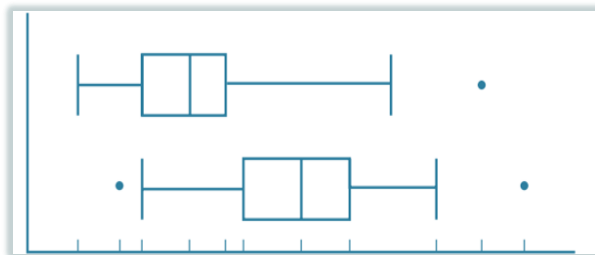
Όταν οι παρατηρήσεις συγκεντρώνονται κοντά στην ελάχιστη τιμή και έχουμε λίγες μεγάλες τιμές, λέμε ότι η κατανομή των παρατηρήσεων παρουσιάζει **θετική ασυμμετρία** (Σχήμα 2).

Όταν οι παρατηρήσεις συγκεντρώνονται κοντά στη μέγιστη τιμή και έχουμε λίγες μικρές τιμές, λέμε ότι η κατανομή των παρατηρήσεων παρουσιάζει **αρνητική ασυμμετρία** (Σχήμα 3).



Πολλαπλά Θηκογράμματα

- ✓ Χρήσιμη μέθοδος για να περιγράψουμε τις τιμές ενός ποσοτικού χαρακτηριστικού σε κάθε στάθμη ενός κατηγορικού χαρακτηριστικού και να συγκρίνουμε τις κατανομές τους.
- ✓ **Κατασκευή:** Σχεδιάζουμε τα θηκογράμματα για τις στάθμες που μας ενδιαφέρουν στο ίδιο σύστημα.



Σύγκριση θηκογραμμάτων

ΠΑΡΑΤΗΡΗΣΕΙΣ

1. Για να εντοπίσουμε τυχόν διαφορές στις θέσεις των ομάδων δεδομένων, συγκρίνουμε τις διαμέσους των θηκογραμμάτων.
2. Για να εντοπίσουμε τυχόν διαφορές ως προς τη μεταβλητότητα των ομάδων δεδομένων, συγκρίνουμε τα ενδοτεταρτημοριακά εύρη καθώς και τις διαφορές μεταξύ άνω και κάτω οριακών τιμών.
3. Συγκρίνουμε τα θηκογράμματα ως προς το σχήμα (συμμετρία-ασυμμετρία)
4. Ελέγχουμε αν υπάρχουν ακραίες τιμές.

Μέση απόλυτη απόκλιση - Διασπορά - Τυπική απόκλιση

Μέση απόλυτη απόκλιση

Είναι η μέση τιμή των απόλυτων αποκλίσεων:

$$MAD = \frac{|x_1 - \bar{x}| + \dots + |x_v - \bar{x}|}{v}$$

Διασπορά

Είναι η μέση τετραγωνική απόκλιση:

$$s^2 = \frac{(x_1 - \bar{x})^2 + \dots + (x_v - \bar{x})^2}{v}$$

Τυπική απόκλιση

Είναι η τετραγωνική ρίζα της διασποράς:

$$s = \sqrt{s^2} = \sqrt{\frac{(x_1 - \bar{x})^2 + \dots + (x_v - \bar{x})^2}{v}}$$

ΠΑΡΑΤΗΡΗΣΕΙΣ

1. Η διαφορά $x_i - \bar{x}$ ονομάζεται **απόκλιση** της τιμής x_i από την μέση τιμή \bar{x} . Η απόλυτη τιμή $|x_i - \bar{x}|$ ονομάζεται **απόλυτη απόκλιση** της τιμής x_i από την μέση τιμή \bar{x} . Το τετράγωνο $(x_i - \bar{x})^2$ ονομάζεται **τετραγωνική απόκλιση** της τιμής x_i από την μέση τιμή \bar{x} .
2. Η MAD και η s περιγράφουν το πόσο μακριά, κατά μέσο όρο, βρίσκονται οι παρατηρήσεις από τη μέση τιμή τους.
3. Η MAD και η s εκφράζονται στην ίδια μονάδα μέτρησης με τις παρατηρήσεις. Η s^2 εκφράζεται στη μονάδα μέτρησης των παρατηρήσεων υψωμένη στο τετράγωνο.
4. Η MAD, η s^2 και η s παίρνουν μη αρνητικές τιμές και είναι ίσες με μηδέν όταν οι παρατηρήσεις είναι ίσες μεταξύ τους και αντιστρόφως.
5. Η MAD, η s^2 και η s, παραμένουν αμετάβλητες αν σε κάθε παρατήρηση προστεθεί ο ίδιος αριθμός α . Αντίθετα, η μέση τιμή \bar{x} γίνεται $\bar{x} + \alpha$.
6. Αν πολλαπλασιάσουμε κάθε παρατήρηση με τον ίδιο αριθμό α , τότε οι παρατηρήσεις που προκύπτουν έχουν μέση τιμή $\alpha\bar{x}$, μέση απόλυτη απόκλιση $|\alpha|MAD$ και τυπική απόκλιση $|\alpha|s$.

Μέτρα σχετικής μεταβλητότητας

Συντελεστής μεταβλητότητας

Αν ένα σύνολο αριθμητικών δεδομένων έχει μέση τιμή $\bar{x} \neq 0$ και τυπική απόκλιση s τότε ο συντελεστής μεταβλητότητας ορίζεται ως εξής:

$$CV = \frac{s}{|\bar{x}|} \quad \text{ή} \quad CV = \frac{s}{|\bar{x}|} 100\%$$

Ομοιογένεια

Ένα σύνολο δεδομένων λέγεται ομοιογενές, όταν έχει συντελεστή μεταβλητότητας $CV \leq 10\%$.

Σχετική ομοιογένεια

Αν δύο σύνολα δεδομένων A και B έχουν συντελεστές μεταβλητότητας CV_A και CV_B αντίστοιχα, με $CV_A < CV_B$, τότε λέμε ότι το σύνολο A είναι **πιο ομοιογενές** από το σύνολο B.

ΠΑΡΑΤΗΡΗΣΕΙΣ

Χρησιμοποιούμε τον συντελεστή μεταβλητότητας όταν:

1. Θέλουμε να συγκρίνουμε τη μεταβλητότητα δύο ή περισσότερων ομάδων ποσοτικών δεδομένων με διαφορετικές μέσες τιμές.
2. Θέλουμε να συγκρίνουμε τη μεταβλητότητα δύο ή περισσότερων ομάδων ποσοτικών δεδομένων με διαφορετικές μονάδες μέτρησης.
3. Θέλουμε να εξετάσουμε αν μια ομάδα ποσοτικών δεδομένων είναι ομοιογενής.

Επίδραση των ακραίων τιμών στα μέτρα θέσης και μεταβλητότητας

Μέση τιμή – Διάμεσος

Οι ακραίες τιμές επηρεάζουν σε μεγαλύτερο βαθμό τη μέση τιμή απ'ότι τη διάμεσο.

Γενικά:

- Όταν απομακρύνουμε τη μικρότερη ακραία τιμή, η μέση τιμή αυξάνεται.
- Όταν απομακρύνουμε τη μεγαλύτερη ακραία τιμή, η μέση τιμή μειώνεται.

Ενδ/κο εύρος – Εύρος – Διασπορά – Τυπική απόκλιση

Οι ακραίες τιμές επηρεάζουν σε μεγαλύτερο βαθμό το εύρος, τη διασπορά και την τυπική απόκλιση απ'ότι το ενδοτεταρτημοριακό εύρος.

Επιλογή μέτρων θέσης και μεταβλητότητας ανάλογα με την ύπαρξη ακραίων τιμών

Ύπαρξη ακραίων τιμών

Όταν έχουμε ένα σύνολο δεδομένων με ακραίες τιμές χρησιμοποιήσουμε τη διάμεσο και το ενδοτεταρτημοριακό εύρος για να περιγράψουμε τα δεδομένα αφού επηρεάζονται λιγότερο από αυτές συγκριτικά με τα άλλα μέτρα θέσης και μεταβλητότητας.

Μη ύπαρξη ακραίων τιμών

Όταν δεν περιέχονται ακραίες τιμές στα δεδομένα, χρησιμοποιούμε τη μέση τιμή και την τυπική απόκλιση για να περιγράψουμε τα δεδομένα αφού για τον υπολογισμό τους λαμβάνουν υπόψη το σύνολο των δεδομένων.

Αιτιότητα

- ✓ Μια σχέση εξάρτησης μεταξύ μιας ποσοτικής και μιας κατηγορικής μεταβλητής δεν συνεπάγεται απαραίτητα ότι οι δύο μεταβλητές συνδέονται με σχέση αιτίας και αποτελέσματος.

Χαρακτηριστικό παράδειγμα:

Παρατηρήσαμε από δεδομένα πολλών ετών ότι οι πωλήσεις παγωτού (κατηγορική με στάθμες: χαμηλές, υψηλές) σχετίζονται με τον αριθμό ηλιακών εγκαυμάτων (ποσοτική).





Οι πωλήσεις παγωτού δεν επηρεάζουν τον αριθμό των ηλιακών εγκαυμάτων αλλά ούτε το αντίστροφο. Άλλοι παράγοντες, όπως οι καιρικές συνθήκες, μπορεί να επηρεάσουν τη σχέση ανάμεσα στις δύο μεταβλητές.

ΚΕΦΑΛΑΙΟ 1^ο

ΔΙΑΧΕΙΡΙΣΗ ΔΕΔΟΜΕΝΩΝ

ΚΡΙΤΗΡΙΟ ΑΞΙΟΛΟΓΗΣΗΣ

ΘΕΜΑ 1. Μια εταιρεία δημοσκοπήσεων θέλει να ερευνήσει την πρόθεση ψήφου των κατοίκων της Θεσσαλονίκης στις ερχόμενες βουλευτικές εκλογές. Για το σκοπό αυτό ρώτησε 1000 μόνιμους κατοίκους της Θεσσαλονίκης ποιο πολιτικό κόμμα θα ψηφίσουν. Να επιλέξετε τη σωστή απάντηση στα παρακάτω ερωτήματα.

α. Ο πληθυσμός της έρευνας είναι:

- i. Οι Έλληνες πολίτες.
- ii. Οι 1000 δημότες που ρωτήθηκαν.
- iii. Οι δημότες της Θεσσαλονίκης.

β. Το δείγμα αποτελούν:

- i. Οι 1000 δημότες που ρωτήθηκαν.
- ii. Οι Έλληνες πολίτες.
- iii. Οι δημότες της πόλης.

γ. Η μεταβλητή της έρευνας είναι:

- i. Η πρόσθεση ψήφου των 1000 δημοτών της Θεσσαλονίκης που ρωτήθηκαν.
- ii. Η πρόσθεση ψήφου των δημοτών της Θεσσαλονίκης.
- iii. Η πρόσθεση ψήφου των Ελλήνων πολιτών.

ΘΕΜΑ 2. Να βρείτε το είδος (κατηγορική ονομαστική (ΚΟ) ή κατηγορική διατάξιμη (ΚΔ), ποσοτική συνεχής (ΠΣ) ή ποσοτική διακριτή (ΠΔ)) καθεμιάς από τις παρακάτω μεταβλητές.

α. Το ύψος των μαθητών/-τριών ενός σχολείου.	ΚΟ	ΚΔ	ΠΣ	ΠΔ
β. Ο χρόνος αναμονής στην ουρά ενός σουπερμάρκετ.	ΚΟ	ΚΔ	ΠΣ	ΠΔ
γ. Η σειρά κατάταξης των μαθητών σε έναν αγώνα δρόμου.	ΚΟ	ΚΔ	ΠΣ	ΠΔ
δ. Ο βαθμός τετραμήνου στα Στοχαστικά Μαθηματικά των μαθητών/-τριών της Α' Λυκείου ενός σχολείου.	ΚΟ	ΚΔ	ΠΣ	ΠΔ
ε. Η εθνικότητα των αθλητών/-τριών μιας ποδοσφαιρικής ομάδας.	ΚΟ	ΚΔ	ΠΣ	ΠΔ
στ. Η ειδικότητα των γιατρών ενός νοσοκομείου.	ΚΟ	ΚΔ	ΠΣ	ΠΔ
ζ. Το θρήσκευμα των πολιτών μιας χώρας.	ΚΟ	ΚΔ	ΠΣ	ΠΔ
η. Ο αριθμός των κατοίκων μιας χώρας.	ΚΟ	ΚΔ	ΠΣ	ΠΔ

ΘΕΜΑ 3. Θέλετε να ερευνήσετε αν υπάρχει σχέση εξάρτησης μεταξύ της κατηγορικής μεταβλητής «**Τύπος κινητού τηλεφώνου**» και της ποσοτικής μεταβλητής «**Αριθμός εγκατεστημένων εφαρμογών**». Ποια από τα παρακάτω ερωτήματα έχουν

συνάφεια με την έρευνα που θέλετε να πραγματοποιήσετε; Να σημειώσετε την ένδειξη X στα κουτάκια των ερωτημάτων που θεωρείτε συναφή.

α. Ποιες είναι οι πιο δημοφιλείς εφαρμογές στην αγορά και πώς σχετίζονται με τον τύπο του κινητού τηλεφώνου;

β. Εξαρτάται ο αριθμός των εγκατεστημένων εφαρμογών από τον τύπο του κινητού τηλεφώνου;

γ. Υπάρχει σχέση εξάρτησης μεταξύ του αριθμού των εγκατεστημένων εφαρμογών και του τύπου του κινητού τηλεφώνου;

ΘΕΜΑ 4. Δίνεται το παρακάτω σύνολο παρατηρήσεων:

1, 6, 7, 7, 8, 8, 8, 9, 10, 13, 17

Να επιλέξετε τη σωστή απάντηση στις παρακάτω ερωτήσεις.

α. Το εύρος των παρατηρήσεων είναι:

i. 17 **ii.** 3 **iii.** 7 **iv.** 16

β. Η διάμεσος των παρατηρήσεων είναι:

i. 9 **ii.** 8 **iii.** 7 **iv.** 6

γ. Το ενδοτεταρτημοριακό εύρος είναι:

i. 16 **ii.** 8 **iii.** 7 **iv.** 3

δ. Το διάστημα $[Q_1 - 1,5Q, Q_3 + 1,5Q]$ είναι το:

i. [2,5, 14,5] **ii.** [1, 17] **iii.** [7, 10] **iv.** [6, 13]

ε. Ακραίες τιμές είναι οι παρατηρήσεις:

i. 1, 6, 13 και 17 **ii.** 1 και 17 **iii.** 17 και 13 **iv.** 1 και 6

ΘΕΜΑ 5. Ο παρακάτω πίνακας περιλαμβάνει την περίληψη των πέντε αριθμών για τη διάρκεια (σε ώρες) καθημερινού ύπνου τον προηγούμενο μήνα του Δημήτρη και του Λάμπρου.

α. Να εξετάσετε αν υπάρχουν ακραίες τιμές στα δεδομένα;

β. Αν, επιπλέον, γνωρίζετε ότι ο δεύτερος μεγαλύτερος σε διάρκεια καθημερινός ύπνος

του Δημήτρη είναι 9 ώρες να κατασκευάσετε στον ίδιο άξονα τα θηκογράμματα των παραπάνω δεδομένων και έπειτα να τα σχολιάσετε ως προς το σχήμα (συμμετρία-ασυμμετρία).

γ. Αν ο προηγούμενος μήνας είχε 30 ημέρες να βρείτε:

γ₁. Πόσες ημέρες, τουλάχιστον, ο Δημήτρης κοιμήθηκε κατ'ελάχιστο 6 ώρες;

γ₂. Πόσες ημέρες, τουλάχιστον, ο Λάμπρος κοιμήθηκε το πολύ 5 ώρες;

	min	Q_1	δ	Q_3	max
Δημήτρης	4	5	6	7	10,5
Λάμπρος	4,5	5	5,5	6	7,5

ΛΥΣΕΙΣ

ΘΕΜΑ 1 α. iii β. i γ. ii	ΘΕΜΑ 2 α. ΠΣ β. ΠΣ γ. ΚΔ δ. ΠΣ ε. ΚΟ στ. ΚΟ ζ. ΚΟ η. ΠΔ	ΘΕΜΑ 3 β και γ	ΘΕΜΑ 4 α. iv β. ii γ. iv δ. i ε. ii
------------------------------------	---	--------------------------	--

ΘΕΜΑ 5

α. Από την 1^η γραμμή του πίνακα, που αφορά τις ώρες ύπνου του Δημήτρη, βρίσκουμε:

$$Q = Q_3 - Q_1 = 7 - 5 = 2 \quad \text{και} \quad [Q_1 - 1,5Q, Q_3 + 1,5Q] = [2, 10].$$

Παρατηρούμε ότι η ελάχιστη (min) παρατήρηση είναι το 4 το οποίο ανήκει στο διάστημα [2, 10].

Επομένως δεν υπάρχει ακραία τιμή στο αριστερό άκρο της κατανομής των παρατηρήσεων. Επίσης, η μέγιστη (max) παρατήρηση είναι το 10,5 που δεν ανήκει στο διάστημα [2, 10].

Επομένως υπάρχει τουλάχιστον μία ακραία τιμή (η τιμή 10,5) στο δεξί άκρο της κατανομής των παρατηρήσεων.

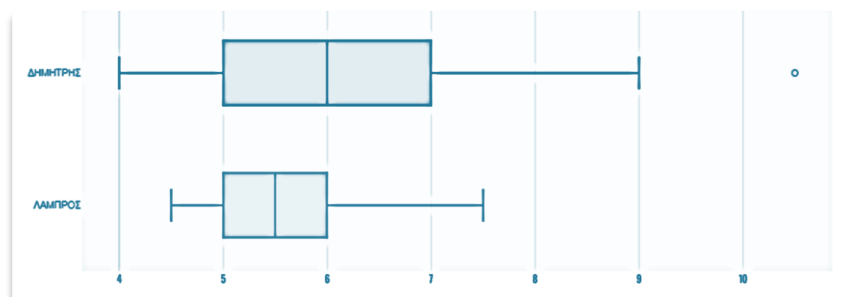
Από την 2η γραμμή του πίνακα, που αφορά τις ώρες ύπνου του Λάμπρου, βρίσκουμε:

$$Q = Q_3 - Q_1 = 6 - 5 = 1 \quad \text{και} \quad [Q_1 - 1,5Q, Q_3 + 1,5Q] = [3,5, 7,5].$$

Παρατηρούμε ότι η ελάχιστη (min) παρατήρηση είναι το 4,5 και η μέγιστη (max) το 7,5. Οι δύο αυτές τιμές ανήκουν στο διάστημα [3,5, 7,5]. Επομένως όλες οι παρατηρήσεις ανήκουν σε αυτό το διάστημα και δεν υπάρχουν ακραίες τιμές στα δεδομένα του Λάμπρου.

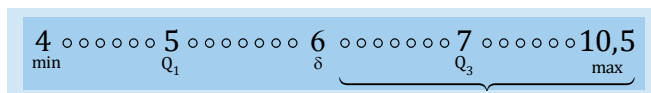
β. Επειδή ο δεύτερος μεγαλύτερος σε διάρκεια καθημερινός ύπνος του Δημήτρη είναι 9 και ο αριθμός 9 ανήκει στο διάστημα [2, 10], στα δεδομένα του Δημήτρη υπάρχει μόνο μία ακραία τιμή, η τιμή 10,5.

Τα θηκογράμματα δίνονται στο παρακάτω σχήμα.



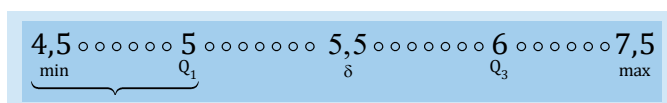
Στο παραπάνω σχήμα παρατηρούμε ότι οι κατανομές των παρατηρήσεων στα δεδομένα του Δημήτρη και του Λάμπρου εμφανίζουν θετική ασυμμετρία. Σε κάθε θηκόγραμμα, οι παρατηρήσεις συγκεντρώνονται κυρίως κοντά στην ελάχιστη τιμή, ενώ οι μεγαλύτερες τιμές εκτείνονται προς τα δεξιά. Ιδιαίτερα, το μήκος των δεξιών απολήξεων είναι μεγαλύτερο από αυτό των αριστερών, υποδεικνύοντας ότι οι υψηλότερες τιμές σε κάθε θηκόγραμμα απέχουν σημαντικά από το Q_3 .

Υ1. Η διάμεσος στα δεδομένα του Δημήτρη είναι $\delta=6$ και είναι ο μέσος όρος της 15^{ης} και 16^{ης} παρατήρησης, όταν οι 30 παρατηρήσεις έχουν διαταχθεί σε αύξουσα σειρά. Δεξιά της διαμέσου υπάρχουν 15 παρατηρήσεις.



Επομένως ο Δημήτρης είχε κατ'ελάχιστο 6 ώρες καθημερινού ύπνου τουλάχιστον 15 ημέρες τον προηγούμενο μήνα.

Υ2. Το πρώτο τεταρτημόριο στα δεδομένα του Λάμπρου είναι $Q_1=5$ και είναι η 8^η παρατήρηση, όταν οι 30 παρατηρήσεις έχουν διαταχθεί σε αύξουσα σειρά.



Επομένως ο Λάμπρος είχε το πολύ 5 ώρες καθημερινού ύπνου τουλάχιστον 8 ημέρες τον προηγούμενο μήνα.

ΚΕΦΑΛΑΙΟ 2^ο

ΜΕΤΡΑ ΘΕΣΗΣ ΚΑΙ ΜΕΤΑΒΛΗΤΟΤΗΤΑΣ

ΚΡΙΤΗΡΙΟ ΑΞΙΟΛΟΓΗΣΗΣ

ΘΕΜΑ 1. Έστω ένα σύνολο n παρατηρήσεων x_1, x_2, \dots, x_n με μέση τιμή \bar{x} .

- α. Πως ορίζεται η μέση απόλυτη απόκλιση (MAD) των παρατηρήσεων;
- β. Πως ορίζεται η διασπορά (s^2) και η τυπική απόκλιση (s) των παρατηρήσεων;
- γ. Πως ορίζεται ο συντελεστής μεταβλητότητας (CV) των παρατηρήσεων;

ΘΕΜΑ 2. Έστω ένα σύνολο n παρατηρήσεων x_1, x_2, \dots, x_n με μέση τιμή \bar{x} .

Να χαρακτηρίσετε ως σωστή (Σ) ή λανθασμένη (Λ) καθεμία από τις επόμενες προτάσεις:

- α. Απόκλιση της τιμής x_1 από την μέση τιμή \bar{x} ονομάζουμε το άθροισμα $x_1 + \bar{x}$. Σ Λ
- β. Η μέση απόλυτη απόκλιση των παρατηρήσεων μπορεί να είναι αρνητικός αριθμός. Σ Λ
- γ. Αν στις παρατηρήσεις υπάρχουν ακραίες τιμές και τις απομακρύνουμε, το εύρος των παρατηρήσεων θα αυξηθεί. Σ Λ
- δ. Για τη σύγκριση της μεταβλητότητας δύο δειγμάτων με διαφορετικές μονάδες μέτρησης ένα κατάλληλο μέτρο είναι ο συντελεστής μεταβλητότητας CV. Σ Λ
- ε. Αν στις παρατηρήσεις δεν υπάρχουν ακραίες τιμές, τότε ως μέτρο θέσης των παρατηρήσεων προτιμούμε τη μέση τιμή από τη διάμεσο. Σ Λ
- στ. Η διασπορά και η τυπική απόκλιση επηρεάζονται σημαντικά από τις ακραίες τιμές. Σ Λ
- ζ. Αν πολλαπλασιάσουμε τις παρατηρήσεις με τον αριθμό 3, τότε το ενδοτεταρτημοριακό εύρος θα τριπλασιαστεί. Σ Λ
- η. Αν από κάθε παρατήρηση αφαιρέσουμε τον αριθμό 1, τότε τυπική απόκλιση των παρατηρήσεων θα ελαττωθεί κατά μία μονάδα. Σ Λ

ΘΕΜΑ 3. Παρακάτω δίνονται οι χρόνοι (σε λεπτά) που χρειάστηκαν 5 άνθρωποι για να πάνε περπατώντας από το σπίτι τους στη δουλειά τους.

4, 5, 4, 6, 12, 5, 6

- α. Να υπολογίσετε τη μέση τιμή, τη διάμεσο, το εύρος, το ενδοτεταρτημοριακό εύρος και την τυπική απόκλιση των παραπάνω χρόνων.
- β.
 - β1. Να εξετάσετε αν η τιμή 12 είναι ακραία.

β₂. Να απομακρύνεται την τιμή 12 και έπειτα να υπολογίσετε εκ νέου τα μέτρα του (α) ερωτήματος. Ποια μέτρα επηρεάστηκαν περισσότερο από την απομάκρυνση αυτής της τιμής;

β₃. Ποια μέτρα θέσης και μεταβλητότητας θεωρείτε κατάλληλα για να περιγράψουν τα δεδομένα πριν την απομάκρυνση της τιμής 12;

γ. Να συγκρίνετε ως προς την ομοιογένεια το αρχικό σύνολο παρατηρήσεων με το σύνολο που προκύπτει αν αφαιρέσουμε την τιμή 12.

ΘΕΜΑ 4. Σε ένα παιχνίδι μπόουλινγκ η ομάδα με τη μεγαλύτερη μέση τιμή πόντων κερδίζει. Οι πόντοι που πέτυχαν τα μέλη δύο ομάδων μετά τη λήξη του παιχνιδιού δίνονται παρακάτω.

Ομάδα A: 172, 170, 173, 177

Ομάδα B: 174, 174, 172, 164

α. Ποια ομάδα κέρδισε το παιχνίδι;

β. Αν η νικήτρια ομάδα ήταν αυτή με τη μεγαλύτερη διάμεσο πόντων, θα ήταν ίδιο το αποτέλεσμα;

γ. Τα μέλη ποιας ομάδας παρουσίασαν μεγαλύτερη ομοιογένεια πόντων;

δ. Σε ένα άλλο παιχνίδι μεταξύ των δύο ομάδων το οποίο έληξε ισόπαλο, όλα τα μέλη της Ομάδας A αύξησαν τους πόντους τους κατά 55 μονάδες και όλα τα μέλη της Ομάδας B αύξησαν τους πόντους τους κατά α%.

δ₁. Να βρείτε το α.

δ₂. Πως μεταβλήθηκε η τυπική απόκλιση και ο συντελεστής μεταβλητότητας των πόντων κάθε ομάδας από το πρώτο στο δεύτερο παιχνίδι;

ΛΥΣΕΙΣ

ΘΕΜΑ 1

βλέπε σελ. 5 .

ΘΕΜΑ 2

α. Λ β. Λ γ. Λ δ. Σ ε. Σ στ. Σ ζ. Σ η. Λ

ΘΕΜΑ 3

α. **Μέση τιμή:** $\bar{x} = \frac{4+5+4+6+12+5+6}{7} = \frac{42}{7} = 6$ λεπτά.

Διάμεσος: Τα διατεταγμένα δεδομένα είναι

$$4, 4, 5, 5, 6, 6, 12$$

Η διάμεσος είναι η τέταρτη παρατήρηση, δηλαδή $\delta = 5$ λεπτά.

Εύρος: Βρίσκουμε πρώτα την ελάχιστη (min) και τη μέγιστη (max) παρατήρηση. Είναι $\min = 4$ και $\max = 12$. Επομένως το εύρος των παρατηρήσεων είναι

$$R = \max - \min = 12 - 4 = 8 \text{ λεπτά}$$

Ενδοτεταρτημοριακό εύρος: Το πρώτο τεταρτημόριο Q_1 είναι η δεύτερη παρατήρηση, δηλαδή $Q_1 = 4$. Το τρίτο τεταρτημόριο Q_3 είναι η έκτη παρατήρηση, δηλαδή $Q_3 = 6$. Επομένως το ενδοτεταρτημοριακό εύρος είναι

$$Q = Q_3 - Q_1 = 6 - 4 = 2 \text{ λεπτά}$$

Τυπική απόκλιση: $s = \sqrt{\frac{2 \cdot (4-6)^2 + 2(5-6)^2 + 2(6-6)^2 + (12-6)^2}{7}} = \sqrt{\frac{46}{7}} \approx 2,56$ λεπτά.

β₁. Στο προηγούμενο ερώτημα βρήκαμε: $Q_1 = 4$, $Q_3 = 6$ και $Q = 2$.

Επομένως $[Q_1 - 1,5Q, Q_3 + 1,5Q] = [1, 9]$.

Η τιμή 12 είναι ακραία αφού δεν ανήκει στο διάστημα $[1, 9]$.

β₂. Τα δεδομένα μετά την απομάκρυνση της τιμής 12 είναι: 4, 4, 5, 5, 6, 6

Μέση τιμή: $\bar{x}' = \frac{4+4+5+5+6+6}{6} = \frac{30}{6} = 5$ λεπτά.

Διάμεσος: Η διάμεσος είναι η μέση τιμή της τρίτης και της τέταρτης παρατήρησης, δηλαδή

$$\delta' = \frac{5+5}{2} = 5 \text{ λεπτά.}$$

Εύρος: Η ελάχιστη και η μέγιστη παρατήρηση είναι αντίστοιχα 4 και 6. Επομένως το εύρος των παρατηρήσεων είναι

$$R' = 6 - 4 = 2 \text{ λεπτά}$$

Ενδοτεταρτημοριακό εύρος: Το πρώτο τεταρτημόριο Q_1' είναι η δεύτερη παρατήρηση, δηλαδή $Q_1' = 4$. Το τρίτο τεταρτημόριο Q_3' είναι η πέμπτη παρατήρηση, δηλαδή $Q_3' = 6$. Επομένως το ενδοτεταρτημοριακό εύρος είναι

$$Q' = Q_3' - Q_1' = 6 - 4 = 2 \text{ λεπτά}$$

Τυπική απόκλιση: $s' = \sqrt{\frac{2 \cdot (4-5)^2 + 2 \cdot (5-5)^2 + 2 \cdot (6-5)^2}{6}} = \sqrt{\frac{4}{6}} \approx 0,82 \text{ λεπτά.}$

Παρατηρούμε ότι τα μέτρα που επηρεάστηκαν περισσότερο από την απομάκρυνση της τιμής 12 είναι η μέση τιμή, το εύρος και η τυπική απόκλιση.

β₃. Λόγω ύπαρξης της ακραίας τιμής 12 στα δεδομένα κατάλληλο μέτρο θέσης για την περιγραφή των δεδομένων είναι η διάμεσος και κατάλληλο μέτρο μεταβλητότητας το ενδοτεταρτημοριακό εύρος.

γ. Από τους υπολογισμούς που κάναμε στα ερωτήματα (β_2) και (β_3) έχουμε:

Αρχικό σύνολο δεδομένων: $CV = \frac{s}{\bar{x}} \approx \frac{2,56}{6} \approx 0,43$

Σύνολο δεδομένων χωρίς την τιμή 12: $CV' = \frac{s'}{\bar{x}'} \approx \frac{0,82}{5} \approx 0,16$

Παρατηρούμε ότι $CV' < CV$ και επομένως το σύνολο δεδομένων χωρίς την τιμή 12 είναι πιο ομοιογενές από το αρχικό σύνολο.

ΘΕΜΑ 4

α. Υπολογίζουμε τη μέση τιμή των πόντων των μελών κάθε ομάδας.

$$\text{Ομάδα A: } \bar{x}_A = \frac{170 + 172 + 173 + 177}{4} = \frac{692}{4} = 173 \text{ πόντοι}$$

$$\text{Ομάδα B: } \bar{x}_B = \frac{174 + 174 + 172 + 164}{4} = \frac{684}{4} = 171 \text{ πόντοι}$$

Επειδή $\bar{x}_A > \bar{x}_B$ η Ομάδα A κέρδισε το παιχνίδι.

β. Διατάσσουμε του πόντους των μελών κάθε ομάδας σε αύξουσα σειρά:

Ομάδα A: 170, 172, 173, 177

Ομάδα B: 164, 172, 174, 174

Η διάμεσος των πόντων στην Ομάδα A είναι η μέση τιμή της δεύτερης και της τρίτης παρατήρησης, δηλαδή $\delta_A = \frac{172 + 173}{2} = 172,5$ πόντοι.

Η διάμεσος των πόντων στην Ομάδα B είναι η μέση τιμή της δεύτερης και της τρίτης παρατήρησης, δηλαδή $\delta_B = \frac{172 + 174}{2} = 173$ πόντοι.

Παρατηρούμε ότι $\delta_B > \delta_A$ και επομένως, σε αυτή την περίπτωση, νικήτρια ομάδα είναι η Ομάδα Β. Το αποτέλεσμα διαφέρει από αυτό του ερωτήματος (α).

γ. Για να συγκρίνουμε ως προς την ομοιογένεια του πόντους των μελών κάθε ομάδας θα υπολογίσουμε τον συντελεστή μεταβλητότητας κάθε ομάδας. Στο (α) ερώτημα βρήκαμε:

$$\bar{x}_A = 173 \quad \text{και} \quad \bar{x}_B = 171$$

Επίσης

$$s_A = \sqrt{\frac{(170-173)^2 + (172-173)^2 + (173-173)^2 + (177-173)^2}{4}} = \sqrt{\frac{9+1+0+16}{4}} = \sqrt{\frac{26}{4}} \approx 2,55$$

$$s_B = \sqrt{\frac{2 \cdot (174-171)^2 + (172-171)^2 + (164-171)^2}{4}} = \sqrt{\frac{18+1+0+49}{4}} = \sqrt{\frac{68}{4}} \approx 4,12$$

Άρα

$$CV_A = \frac{s_A}{\bar{x}_A} = \frac{2,55}{173} \approx 0,01 \quad \text{και} \quad CV_B = \frac{s_B}{\bar{x}_B} = \frac{4,12}{171} \approx 0,02$$

Επειδή $CV_A < CV_B$ οι πόντοι των μελών της Ομάδας Α παρουσίασαν μεγαλύτερη ομοιογένεια.

δ₁. Η μέση τιμή των πόντων της Ομάδας Α θα αυξηθεί κατά 55 μονάδες:

$$\bar{x}'_A = \bar{x}_A + 55 = 173 + 55 = 228$$

Η μέση τιμή των πόντων της Ομάδας Β θα αυξηθεί κατά α%:

$$\bar{x}'_B = \bar{x}_B(1 + \alpha\%) = 171(1 + \alpha\%)$$

Επειδή το παιχνίδι έληξε ισοπαλία θα είναι:

$$171(1 + \alpha\%) = 228 \quad \text{ή} \quad 1 + \alpha\% = \frac{228}{171} \quad \text{ή} \quad 1 + \alpha\% = \frac{4}{3} \quad \text{ή} \quad \alpha\% = \frac{1}{3} = 0,3\bar{3}$$

Επομένως $\alpha = 33.\bar{3}$.

δ₂. Η αύξηση των πόντων της Ομάδας Α κατά 55 μονάδες στο δεύτερο παιχνίδι δεν θα μεταβάλει την τυπική απόκλιση των πόντων της ομάδας στο πρώτο παιχνίδι. Επομένως η τυπική απόκλιση s'_A στο δεύτερο παιχνίδι θα είναι ίδια με αυτή του πρώτου παιχνιδιού: $s'_A = s_A$.

Η αύξηση των πόντων της Ομάδας Β κατά $33.\bar{3}\%$ στο δεύτερο παιχνίδι θα μεταβάλει την τυπική απόκλιση των πόντων της ομάδας στο πρώτο παιχνίδι.

Συγκεκριμένα:

$$s'_B = (1 + 33.\bar{3}\%)s_B = 133.\bar{3}\%s_B$$

Επομένως η τυπική απόκλιση αυξήθηκε κατά $33.\bar{3}\%$.

Επίσης, για τον συντελεστή μεταβλητότητας των πόντων της Ομάδας Α στο πρώτο και στο δεύτερο παιχνίδι έχουμε:

$$CV'_A = \frac{s'_A}{\bar{x}'_A} = \frac{s_A}{\bar{x}_A + 55} < \frac{s_A}{\bar{x}_A} = CV_A$$

Παρατηρούμε ότι ο συντελεστής μεταβλητότητας μειώθηκε.

Για τον συντελεστή μεταβλητότητας των πόντων της Ομάδας Β στο πρώτο και στο δεύτερο παιχνίδι έχουμε:

$$CV'_B = \frac{s'_B}{\bar{x}'_B} = \frac{133,3 s_B}{133,3 \bar{x}_B} = \frac{s_B}{\bar{x}_B} = CV_B$$

Παρατηρούμε ότι ο συντελεστής μεταβλητότητας παρέμεινε ίδιος.

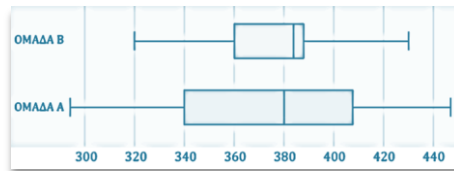
ΚΕΦΑΛΑΙΟ 3^ο

ΣΧΕΣΕΙΣ ΕΞΑΡΤΗΣΗΣ ΜΕΤΑΞΥ ΔΥΟ ΜΕΤΑΒΛΗΤΩΝ ΚΡΙΤΗΡΙΟ ΑΞΙΟΛΟΓΗΣΗΣ

ΘΕΜΑ 1. Να χαρακτηρίσετε ως σωστή (Σ) ή λανθασμένη (Λ) καθεμία από τις επόμενες προτάσεις:

α. Αν υπάρχει σχέση εξάρτησης μεταξύ μιας κατηγορικής και μιας ποσοτικής μεταβλητής τότε η μία μεταβλητή είναι η αιτία και η άλλη το αποτέλεσμα. Σ Λ

β. Από τα θηκογράμματα του παρακάτω σχήματος προκύπτει ότι η μεταβλητότητα των παρατηρήσεων της Ομάδας Α είναι μικρότερη από αυτή της Ομάδας Β.

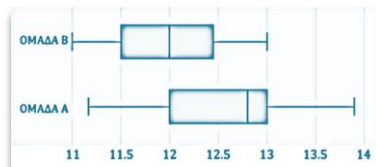


Σ Λ

γ. Παρατηρήσαμε, αναλύοντας δεδομένα πολλών ετών, ότι οι άνθρωποι με σκουρόχρωμα μάτια έχουν καλύτερες επιδόσεις στα αγωνίσματα του στίβου από τους ανθρώπους με ανοικτό χρώμα ματιών. Σ Λ

Δεν μπορούμε να συμπεράνουμε ότι το χρώμα των ματιών επηρεάζει την επίδοση ενός αθλητή σε κάποιο αγώνισμα του στίβου.

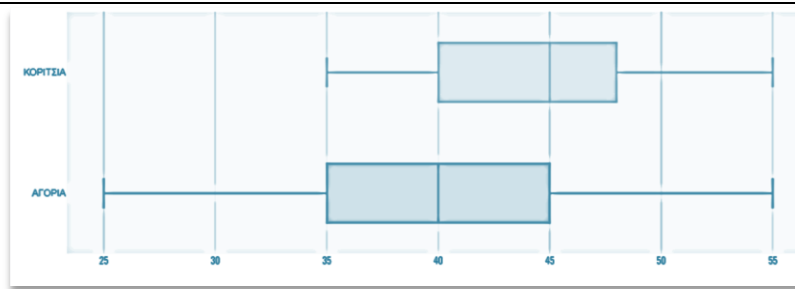
δ. Στα παρακάτω θηκογράμματα παρουσιάζονται οι καλύτερες επιδόσεις δύο ομάδων αθλητριών σε έναν αγώνα δρόμου των 100m.



Σ Λ

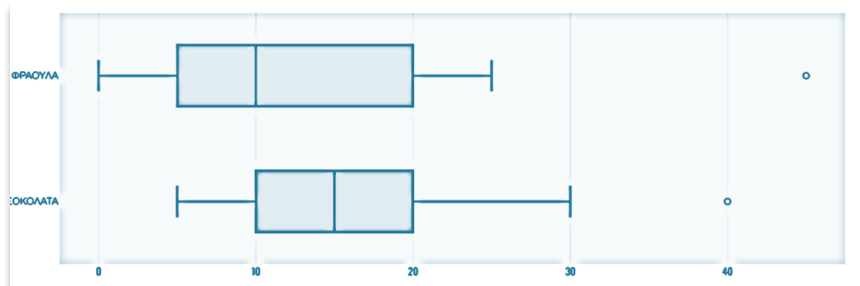
Η Ομάδα Α πέτυχε καλύτερες επιδόσεις από την Ομάδα Β.

ΘΕΜΑ 2. Στα παρακάτω θηκογράμματα παρουσιάζεται ο αριθμός των ακόλουθων (followers) σε ένα δημοφιλές κοινωνικό δίκτυο των αγοριών και των κοριτσιών ενός σχολείου του Δήμου Αθηναίων.



Να γράψετε μια συνοπτική αναφορά για το ποια από τις δύο ομάδες παιδιών έχει γενικά περισσότερους ακόλουθους.

ΘΕΜΑ 3. Ένα κατάστημα γλυκών θέλοντας να μάθει ποια από τις δύο γεύσεις παγωτού «Φράουλα» και «Σοκολάτα» είναι πιο δημοφιλής στους πελάτες του, κατέγραψε τις ημερήσιες πωλήσεις κάθε γεύσης ενός καλοκαιρινού μήνα. Από τα δεδομένα των πωλήσεων προέκυψαν τα παρακάτω θηκογράμματα.



- α. Να βρείτε τη γεύση με τον μεγαλύτερο ημερήσιο αριθμό πωλήσεων.
- β. Να συγκρίνετε τις διαμέσους των ημερήσιων πωλήσεων των δύο γεύσεων.
- γ. Συγκρίνοντας τα αντίστοιχα ενδοτεταρτημοριακά εύρη και τις αποστάσεις μεταξύ των κάτω και άνω οριακών τιμών, να πείτε σε ποια από τις δύο γεύσεις ο ημερήσιος αριθμός πωλήσεων παρουσίασε μεγαλύτερη μεταβλητότητα.
- δ. Ποια από τις δύο γεύσεις θεωρείτε ότι είναι πιο δημοφιλής στους πελάτες του συγκεκριμένου καταστήματος;

ΛΥΣΕΙΣ

ΘΕΜΑ 1

α. Λ β. Λ γ. Σ δ. Λ

ΘΕΜΑ 2

Παρατηρώντας προσεκτικά τα δύο θηκογράμματα συμπεραίνουμε ότι τα κορίτσια έχουν, γενικά, περισσότερους ακόλουθους από τα αγόρια. Οι αριθμοί των ακόλουθων των κοριτσιών παρουσιάζουν μεγάλη συγκέντρωση στις τιμές 40 έως 48 σε αντίθεση με αυτούς των αγοριών των οποίων το «κέντρο» εντοπίζεται στις τιμές 35 έως 45. Επίσης, οι τέσσερις αριθμοί: L, Q₁, δ και Q₃, εκτός των άνω οριακών τιμών U που ταυτίζονται, βρίσκονται δεξιότερα στο θηκόγραμμα των κοριτσιών από τους αντίστοιχους αριθμούς στο θηκόγραμμα των αγοριών.

ΘΕΜΑ 3

α. Παρατηρούμε ότι μέγιστος αριθμός ημερήσιων πωλήσεων σοκολάτας είναι 40 (ακραία τιμή στο θηκόγραμμα της σοκολάτας) ενώ ο αντίστοιχος για την φράουλα είναι μεγαλύτερος από 40 (ακραία τιμή στο θηκόγραμμα της φράουλας).

β. Παρατηρούμε ότι $\delta_{\text{φράουλα}} = 10$ και $\delta_{\text{σοκολάτα}} = 15$. Επομένως $\delta_{\text{σοκολάτα}} > \delta_{\text{φράουλα}}$.

γ. **Φράουλα:** $Q = Q_3 - Q_1 = 20 - 5 = 15$ και $L - U = 25 - 0 = 25$.

Σοκολάτα: $Q = Q_3 - Q_1 = 20 - 10 = 10$ και $L - U = 30 - 5 = 25$.

Παρατηρούμε ότι και οι δύο γεύσεις παρουσίασαν την ίδια διαφορά μεταξύ άνω και κάτω οριακής τιμής. Η φράουλα όμως παρουσίασε μεγαλύτερο ενδοτεταρτημοριακό εύρος. Επομένως ο ημερήσιος αριθμός πωλήσεων φράουλας παρουσίασε μεγαλύτερη μεταβλητότητα.

δ. Και οι πέντε αριθμοί: L, Q₁, δ και Q₃ και U, βρίσκονται δεξιότερα στο θηκόγραμμα της σοκολάτας από τους αντίστοιχους αριθμούς στο θηκόγραμμα της φράουλας. Μπορούμε να πούμε λοιπόν ότι η γεύση «Σοκολάτα» είναι δημοφιλέστερη στους πελάτες του συγκεκριμένου καταστήματος.

Το παρόν αναπτύχθηκε στο πλαίσιο της Πράξης «Συγγραφή, Αξιολόγηση και Ένταξη διδακτικών βιβλίων στο Μητρώο Διδακτικών Βιβλίων και στην Ψηφιακή Βιβλιοθήκη Διδακτικών Βιβλίων» με κωδικό ΟΠΣ (MIS) 6010165, του Προγράμματος «Ανθρώπινο Δυναμικό και Κοινωνική Συνοχή 2021-2027» που υλοποιείται από το Ινστιτούτο Εκπαιδευτικής Πολιτικής και συγχρηματοδοτείται από το Ευρωπαϊκό Κοινωνικό Ταμείο.



Τίτλος: Οδηγός επανάληψης-Στατιστική

Έκδοση: 1.0 Ημερομηνία: 26.04.2024

ΣΥΝΤΕΛΕΣΤΕΣ:

ΕΜΠΝΕΥΣΤΕΣ/ ΟΜΑΔΑ ΣΧΕΔΙΑΣΜΟΥ ΚΑΙ ΑΝΑΠΤΥΞΗΣ/ ΤΕΧΝΙΚΗ ΥΛΟΠΟΙΗΣΗ

Κωνσταντίνος Ρεκούμης
Εκπαιδευτικός Ιδιωτικής Εκπαίδευσης ΠΕ03
Λάμπρος Κατσάπας
Εκπαιδευτικός Ιδιωτικής Εκπαίδευσης ΠΕ03

Νικόλαος Κουμάντος
Εκπαιδευτικός Ιδιωτικής Εκπαίδευσης ΠΕ03
Ελένη Ρεκούμη
Εκπαιδευτικός Ιδιωτικής Εκπαίδευσης ΠΕ03



Το παρόν χορηγείται με άδεια Creative Commons Αναφορά Δημιουργού-Μη Εμπορική Χρήση 4.0 Διεθνές (CC BY-NC 4.0).

Με τη συγκεκριμένη άδεια, μπορείτε να:

- Μοιραστείτε — αντιγράψετε και αναδιανείμετε το υλικό με κάθε μέσο και τρόπο
- Προσαρμόσετε — αναμίξετε, τροποποιήσετε και δημιουργήσετε πάνω στο υλικό

Υπό τους ακόλουθους όρους:

- **Αναφορά Δημιουργού** — Θα πρέπει να καταγράψετε αναφορά στον δημιουργό, με σύνδεσμο της άδειας, και με αναφορά αν έχουν γίνει αλλαγές. Μπορείτε να το κάνετε αυτό με οποιοδήποτε εύλογο τρόπο, αλλά όχι με τρόπο που να υπονοεί ότι ο δημιουργός αποδέχεται το έργο σας ή τη χρήση που εσείς κάνετε.
- **Μη Εμπορική Χρήση** — Δεν μπορείτε να χρησιμοποιήσετε το υλικό για εμπορικούς σκοπούς.

- **Παρόμοια Διανομή** — Αν αναμίξετε, τροποποιήσετε, ή δημιουργήσετε πάνω στο υλικό, πρέπει να διανεμήτε τις δικές σας συνεισφορές υπό την ίδια άδεια όπως και το πρωτότυπο.
- **Δεν υπάρχουν πρόσθετοι περιορισμοί** — Δεν μπορείτε να εφαρμόσετε νομικούς όρους ή τεχνολογικά μέσα που να περιορίζουν νομικά τους άλλους από το να κάνουν οτιδήποτε επιτρέπει η άδεια. Ο αδειοδότης δεν μπορεί να ανακαλέσει αυτές τις ελευθερίες όσο εσείς ακολουθείτε τους όρους της άδειας.