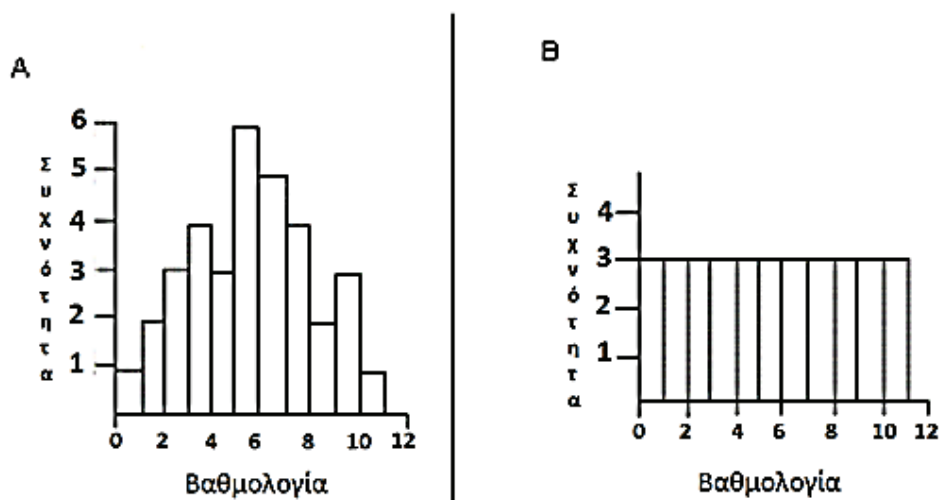


1. Στο παρακάτω ζευγάρι γραφημάτων, να χαρακτηρίσετε ποια μεταβλητή από αυτές που απεικονίζονται έχει την υψηλότερη τυπική απόκλιση (δεν είναι απαραίτητο να κάνετε υπολογισμούς)



- Οι δύο μεταβλητές έχουν την ίδια τυπική απόκλιση
- Στο Β η τυπική απόκλιση είναι μεγαλύτερη απ' ό,τι στο Α
- Στο Α η τυπική απόκλιση είναι μεγαλύτερη απ' ό,τι στο Β
- Δεν μπορούμε να αποφανθούμε μόνο από το γράφημα

Εξηγήστε την απάντησή σας.

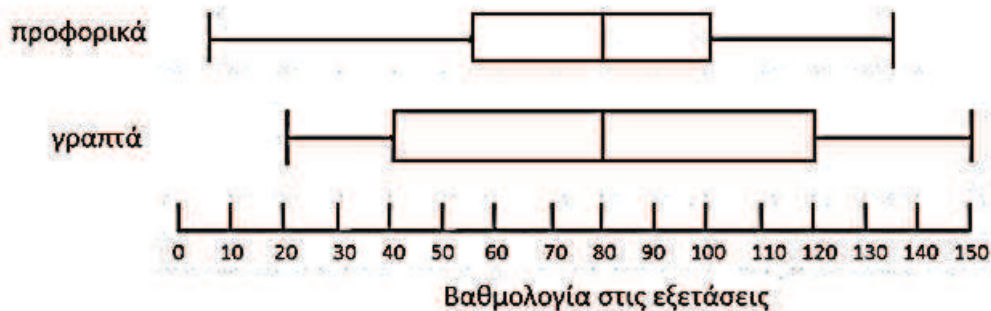
Απάντηση:

Στο γράφημα Α, διαπιστώνουμε ότι ο μέσος όρος θα βρίσκεται κοντά στην τιμή 6 και ότι οι τιμές γύρω από αυτόν εμφανίζονται με μεγαλύτερη συχνότητα, ενώ όσο απομακρυνόμαστε από το μέσο όρο οι τιμές εμφανίζονται με μικρότερη συχνότητα. Στο γράφημα Β, ο μέσος όρος αναμένεται να βρίσκεται επίσης κοντά στο 6, αλλά στην περίπτωση αυτή όλες οι τιμές εμφανίζονται με την ίδια συχνότητα, είτε βρίσκονται κοντά είτε μακριά από το μέσο όρο. Άρα στο γράφημα Β, η διασπορά των τιμών γύρω από το μέσο όρο αναμένεται να είναι μεγαλύτερη απ' ό,τι στο γράφημα Α. Σωστή απάντηση είναι το β.

2. Ένα τυχαίο δείγμα μαθητών βαθμολογήθηκε σε ένα μάθημα με τη χρήση 20βάθμιας κλίμακας. Στη συνέχεια υπολογίστηκαν τα παρακάτω στατιστικά: μέσος όρος = 12,8, διάμεσος = 12,7, τυπική απόκλιση = 1,4 και ενδοτεταρτημοριακό εύρος = 2. Η κατανομή της βαθμολογίας προσεγγίζει την κανονική. Με βάση τις παραπάνω πληροφορίες να εκτιμήσετε τη μικρότερη (min) και τη μεγαλύτερη (max) τιμή της βαθμολογίας.

Απάντηση: Γνωρίζουμε ότι στην κανονική κατανομή περίπου το 99,7% των παρατηρήσεων βρίσκεται μεταξύ ± 3 τυπικών αποκλίσεων από το μέσο όρο. Άρα με δεδομένο ότι το ελάχιστο (min), θα βρίσκεται περίπου 3 τυπικές αποκλίσεις κάτω από το μέσο όρο και ότι το μέγιστο (max), θα βρίσκεται περίπου 3 τυπικές αποκλίσεις πάνω από το μέσο όρο, μια εκτίμηση του min είναι $12,8 - 3 \times 1,4 = 8,6$ και του max $12,8 + 3 \times 1,4 = 17$.

3. Τα παρακάτω boxplot αναπαριστούν τη βαθμολογία στην εξέταση ενός μαθήματος (προφορική και γραπτή). Με βάση το γράφημα να απαντήσετε στις τρεις ερωτήσεις που ακολουθούν.



- i) Ποια κατανομή βαθμολογίας έχει το μεγαλύτερο ποσοστό μαθητών που έλαβαν βαθμό κάτω από 30;
- Η βαθμολογία στα γραπτά.
 - Η βαθμολογία στα προφορικά.
 - Το ποσοστό είναι περίπου το ίδιο στις δύο κατανομές.
 - Δεν έχουμε αρκετές πληροφορίες για να απαντήσουμε σε αυτή την ερώτηση.
- Εξηγήστε την απάντησή σας

Απάντηση: Αρχικά παρατηρούμε, ότι και οι δύο κατανομές βαθμολογίας έχουν τιμές κάτω από το 30. Με βάση, όμως, τις πληροφορίες που μας δίνουν τα δύο boxplot, δεν μπορούμε να γνωρίζουμε πως ακριβώς κατανέμονται οι τιμές της βαθμολογίας στις δύο κατανομές κάτω από την τιμή 30, άρα και το ποσοστό αυτών βρίσκεται κάτω από 30. Άρα, σωστή απάντηση είναι το δ.

- ii) Ποιο είδος εξέτασης αναμένεται να έχει μεγαλύτερη τυπική απόκλιση βαθμολογίας;
- Η γραπτή.
 - Η προφορική.
 - Θα έχουν περίπου την ίδια τυπική απόκλιση.
 - Δεν έχουμε αρκετές πληροφορίες για να απαντήσουμε σε αυτή την ερώτηση.
- Εξηγήστε την απάντησή σας

Απάντηση: Αρχικά παρατηρούμε ότι το εύρος των δύο κατανομών ($\max - \min$) είναι σχεδόν το ίδιο, η διάμεσος είναι η ίδια και δεν υπάρχουν ακραίες τιμές. Αυτό που διαφοροποιεί, κυρίως, τα δύο boxplot, είναι ότι το ενδοτεταρτημοριακό εύρος της βαθμολογίας στα γραπτά (περίπου από 40 έως 120) είναι σημαντικά μεγαλύτερο απ' ότι στα προφορικά (περίπου από 55 έως 100). Από αυτό συμπεραίνουμε ότι η τυπική απόκλιση της βαθμολογίας των γραπτών αναμένεται να είναι μεγαλύτερη από αυτή των προφορικών (άρα σωστή απάντηση το α.)

- iii) Ποια κατανομή βαθμολογίας έχει το μεγαλύτερο ποσοστό μαθητών που έλαβαν βαθμό τουλάχιστον 80;
- Η βαθμολογία στα γραπτά.
 - Η βαθμολογία στα προφορικά.
 - Το ποσοστό είναι περίπου το ίδιο στις δύο κατανομές.
 - Δεν έχουμε αρκετές πληροφορίες για να απαντήσουμε σε αυτή την ερώτηση.

Εξηγήστε την απάντησή σας

Απάντηση: Η διάμεσος και στις δύο κατανομές βαθμολογίας είναι περίπου 80, άρα το ποσοστό των μαθητών που έλαβαν βαθμό τουλάχιστον 80 αναμένεται να είναι περίπου το ίδιο στις δύο κατανομές (σωστή απάντηση το γ.)

4. Ο παρακάτω υποθετικός πίνακας συνοψίζει τις μέσες εβδομαδιαίες δαπάνες σε ευρώ ανά νοικοκυριό για αλκοόλ και καπνό σε 11 πόλεις της Ελλάδας:

Περιοχή	Αλκοόλ	Καπνός
Ρέθυμνο	6,47	4,03
Βόλος	6,13	3,76
Λάρισα	6,19	3,77
Καβάλα	4,89	3,34
Ορεστιάδα	5,63	3,47
Γρεβενά	4,52	2,92
Αλεξανδρούπολη	5,89	3,20
Τρίπολη	4,79	2,71

Άρτα	5,27	3,53
Κοζάνη	6,08	4,51
Δράμα	4,02	4,56

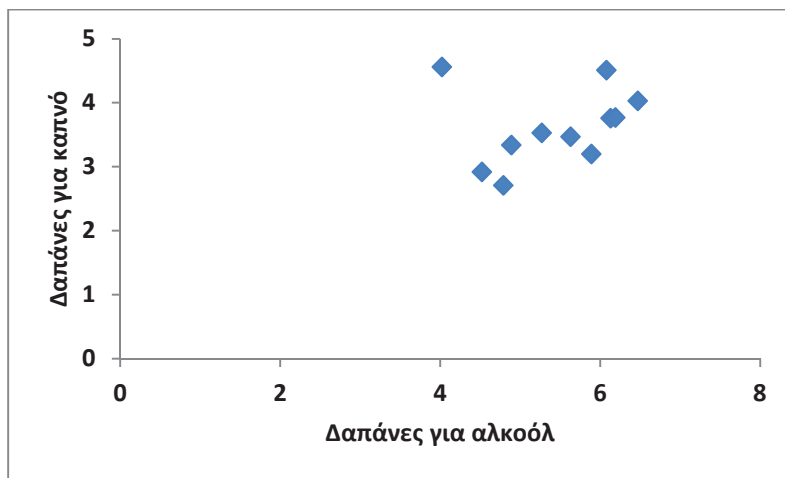
Ο δείκτης γραμμικής συσχέτισης r του Pearson ανάμεσα στις δύο μεταβλητές (δαπάνες για αλκοόλ και καπνό) βρέθηκε ίσος με +0,22.

i) Με βάση την τιμή του δείκτη r , να χαρακτηρίσετε τη συσχέτιση ανάμεσα στις δαπάνες για αλκοόλ και τις δαπάνες για καπνό.

Απάντηση: Η συσχέτιση ανάμεσα στις δαπάνες για καπνό και τις δαπάνες για αλκοόλ βρέθηκε θετική, χαμηλή.

ii) Είναι ο δείκτης r κατάλληλος για να περιγράψει τη συσχέτιση αυτή; Να δικαιολογήσετε και διαγραμματικά την απάντησή σας.

Απάντηση: Για να απαντήσουμε στο ερώτημα κατασκευάζουμε αρχικά το διάγραμμα διασποράς των δύο μεταβλητών:



Παρατηρούμε ότι, με εξαίρεση το ζευγάρι τιμών για τη Δράμα (4,02, 4,56), η συσχέτιση ανάμεσα στις δύο μεταβλητές είναι θετική και τα σημεία βρίσκονται πολύ κοντά σε μια ευθεία γραμμή. Αυτό, όμως, δεν αντικατοπτρίζεται στην τιμή του δείκτη r , η τιμή του οποίου έχει επηρεαστεί σημαντικά από την παρουσία του ακραίου σημείου. Αν αφαιρεθεί το σημείο αυτό, η τιμή του r αναμένεται να αυξηθεί. Άρα ο δείκτης r δεν είναι κατάλληλος για να περιγράψει τη συσχέτιση αυτή, λόγω της παρουσίας του ακραίου σημείου.