

Θηκόγραμμα - Boxplot

Στην περιγραφική στατιστική, το θηκόγραμμα (boxplot) είναι ένας βολικός τρόπος γραφικής απεικόνισης πέντε αριθμητικών δεδομένων μιας σειράς παρατηρήσεων: της μικρότερης παρατήρησης του πρώτου τεταρτημρίου (Q1), της διαμέσου (δ) του τρίτου τεταρτημρίου (Q3), και της μεγαλύτερης παρατήρησης. Το θηκόγραμμα δείχνει διαφορές μεταξύ των πληθυσμών. Οι αποστάσεις μεταξύ των διαφόρων τμημάτων του θηκογράμματος βοηθούν να φανεί το μέγεθος της διασποράς και η ασυμμετρία των δεδομένων.

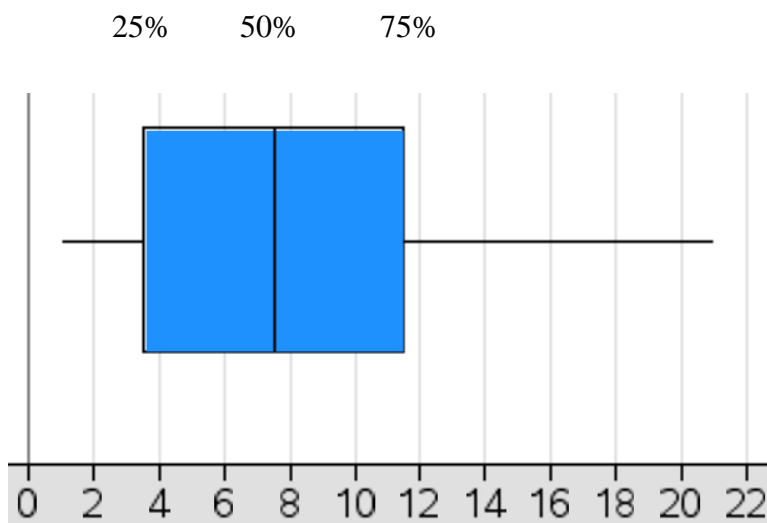
Παράδειγμα 1:

Δίνονται οι παρακάτω 20 παρατηρήσεις μιας μεταβλητής x :

x	1	1	2	3	3	4	5	5	6	7	8	8	8	9	10	13	14	16	19	21
-----	---	---	---	---	---	---	---	---	---	---	---	---	---	---	----	----	----	----	----	----

Οι 5 αριθμοί είναι οι εξής: $\min=1$, $Q1=3.5$, $\delta=7.5$, $Q3=11.5$ $\max=21$

Ακολουθεί το θηκόγραμμα:

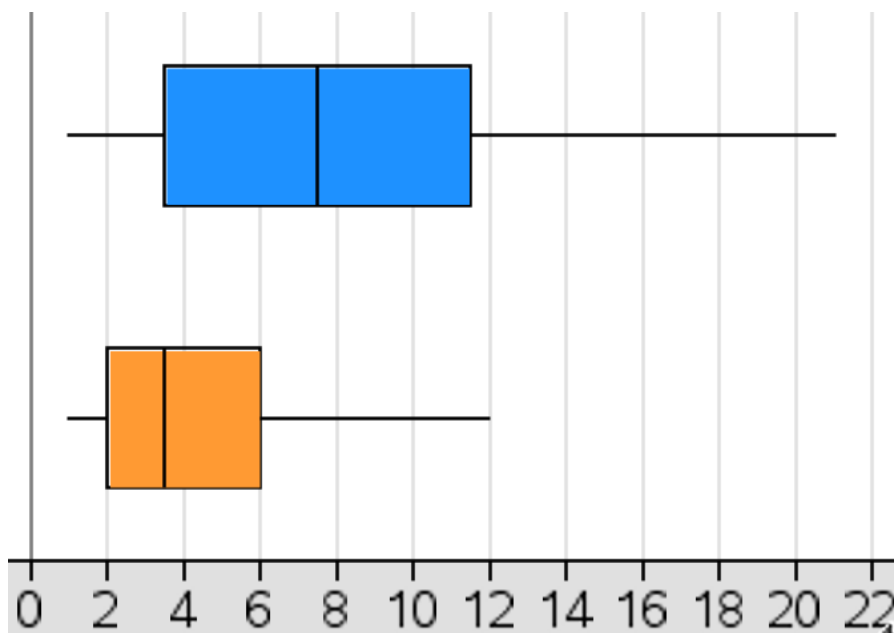


Έστω ότι έχουμε άλλες 14 τιμές από ένα άλλο δείγμα της ίδιας μεταβλητής (π.χ χειμωνιάτικες θερμοκρασίες) και θέλουμε να τις συγκρίνουμε:

y	1	1	1	2	2.5	3	3	4	4	5	6	8	12	12
-----	---	---	---	---	-----	---	---	---	---	---	---	---	----	----

Οι 5 νέοι αριθμοί είναι τώρα οι εξής: $\min=1$, $Q1=2$, $\delta=3.5$, $Q3=6$ $\max=12$.

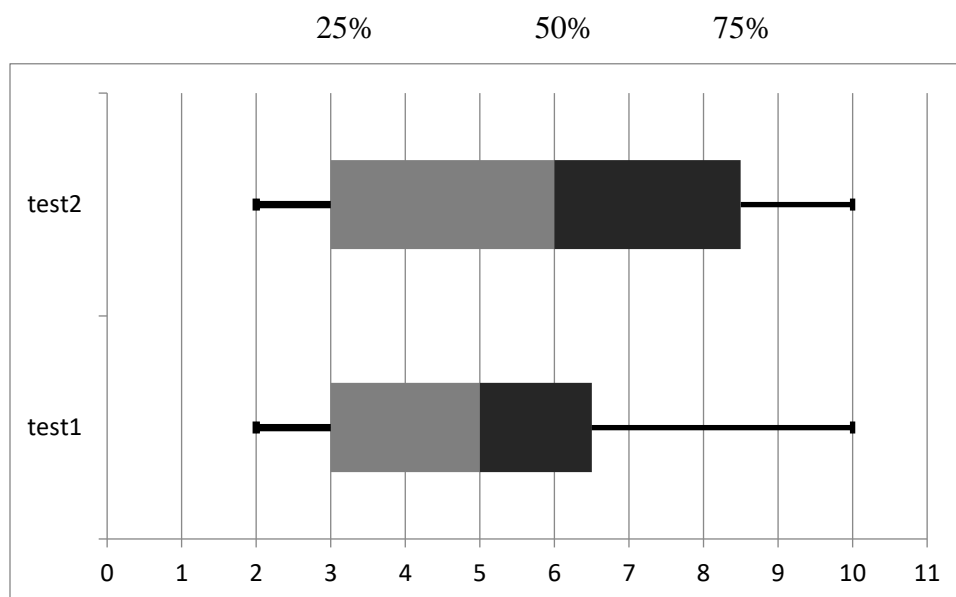
Η παρουσία και των δύο θηκογραμμμάτων μαζί, μας βοηθά να συγκρίνουμε τις δύο ομάδες τιμών.



Παράδειγμα 2:

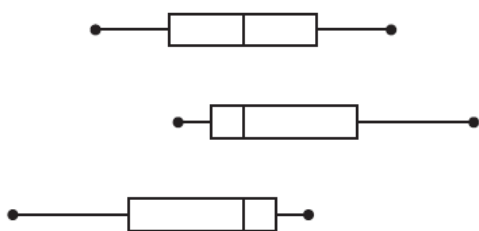
ΤΕΣΤ 1		ΤΕΣΤ 2	
Βαθμοί	Συχνότητα	Βαθμοί	Συχνότητα
2	3	2	3
3	5	3	3
4	2	4	1
5	5	5	2
6	1	6	3
7	2	7	1
8	1	8	3
10	2	9	2
-	-	10	3
Σύνολο	21	-	21

Στον παραπάνω πίνακα είναι οι βαθμοί μιας τάξης από 21 μαθητές σε δύο τεστ. Τα παρακάτω θηκογράμματα κατασκευάστηκαν βάσει αυτού του πίνακα.



Οι αριθμοί της 1^{ης} γραμμής αφορούν τις **τιμές της μεταβλητής**.

Οι αριθμοί πάνω από το 1^ο θηκόγραμμα αφορούν **το τμήμα του πληθυσμού (το ποσοστό)** που οι τιμές του είναι μέχρι την αντίστοιχη τιμή της 1^{ης} γραμμής.



Στα διπλανά θηκογράμματα παρατηρούμε ότι:

α) Στο 1^ο τα δεδομένα κατανέμονται συμμετρικά γύρω από τη διάμεσο.

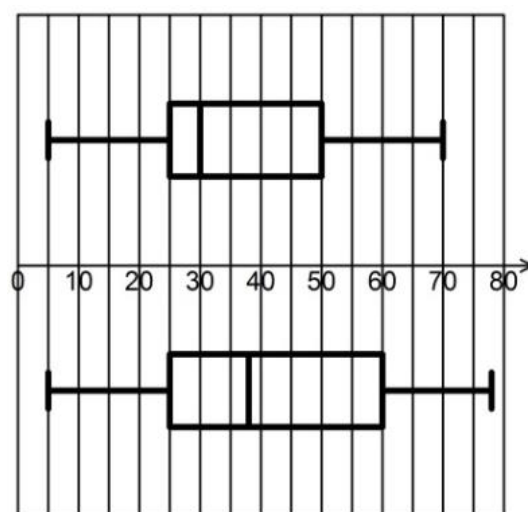
β) Στο δεύτερο έχουμε ασυμμετρία. Συγκεκριμένα τα δεδομένα είναι απλωμένα όσον αφορά τη διάμεσο προς τα δεξιά. Τα δεδομένα που είναι μεγαλύτερα

από την διάμεσο έχουν μεγαλύτερη διασπορά από εκείνα που είναι μικρότερα.

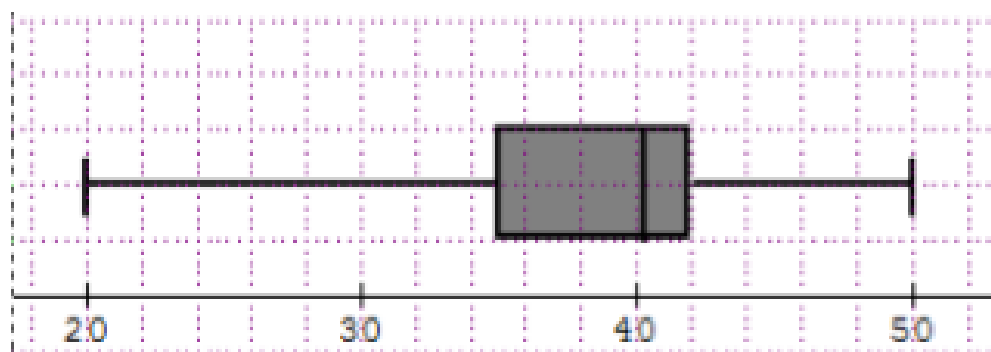
γ) Στο 3^ο έχουμε επίσης ασυμμετρία. Εδώ τα δεδομένα είναι απλωμένα όσον αφορά τη διάμεσο προς τα αριστερά.

Ασκήσεις (Στο τέλος υπάρχουν κάποιες υποδείξεις)

1) Τα διπλανά θηκογράμματα δίνουν τα αποτελέσματα κάποιων μετρήσεων δύο ομάδων A και B. Όλοι οι αριθμοί είναι ακέραιοι και τα αποτελέσματα κάλυψαν όλους τους αριθμούς από την μικρότερη τιμή έως την μεγαλύτερη. Ένα άτομο της ομάδας A είχε τιμή μεγαλύτερη από το 75% των τιμών της B ομάδας. Ποιες είναι οι πιθανές τιμές που αφορούν το άτομο αυτό; (61,62,...,70)

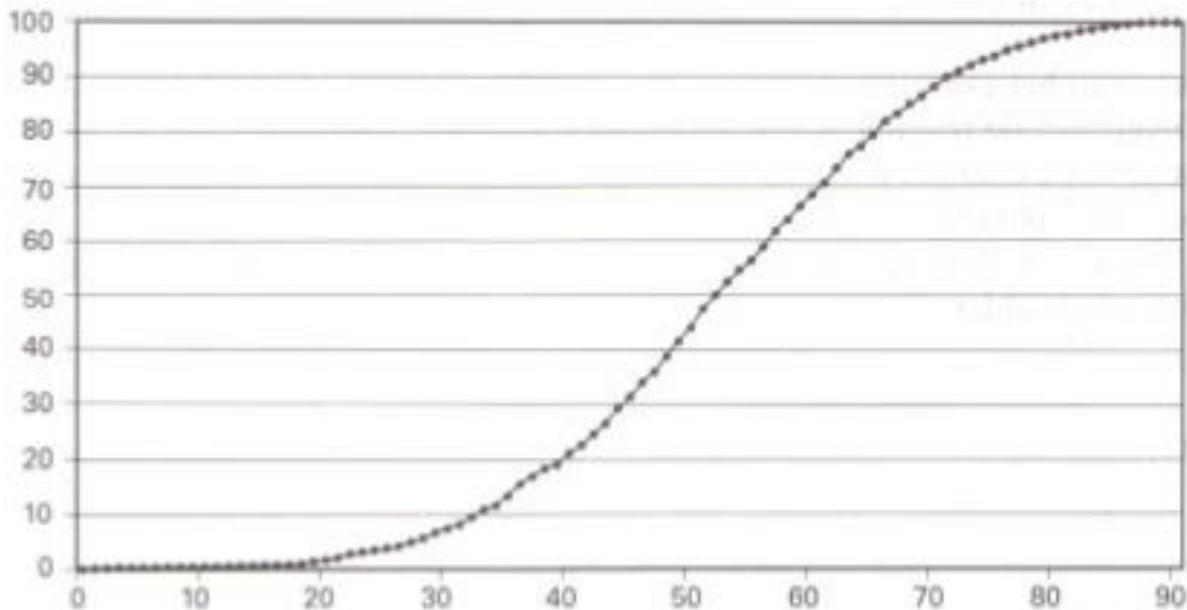


2) Το παρακάτω θηκόγραμμα παρουσιάζει τα αποτελέσματα κάποιων μετρήσεων στα μέλη ενός τυχαίου δείγματος:

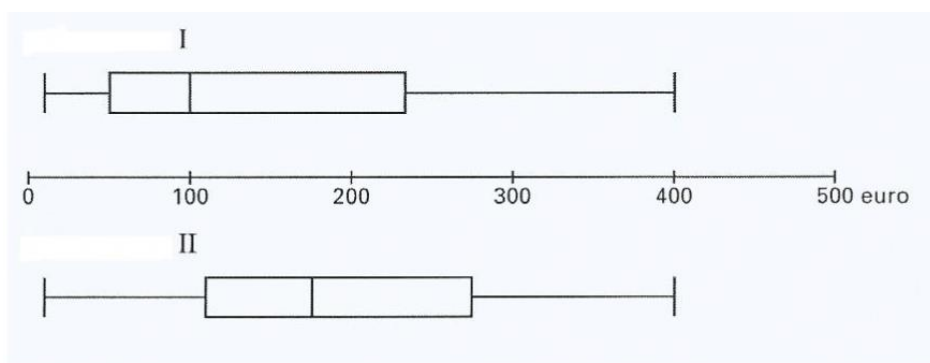


Είναι αληθές ότι στο διάστημα $[35, 42]$ βρίσκονται τουλάχιστον το 40% των παρατηρήσεων;

3) Τα αποτελέσματα του διαγωνίσματος των μαθηματικών κυμάνθηκαν από 0 έως 90 μόρια και δίνονται από το παρακάτω διάγραμμα αθροιστικών σχετικών συχνοτήτων. Να γίνει το θηκόγραμμα.



4) Τα παρακάτω θηκογράμματα σχετίζονται με τις αναλήψεις χρημάτων μια συγκεκριμένη μέρα από δύο διαφορετικά μηχανήματα αυτόματης διανομής χρημάτων σε μία πόλη:



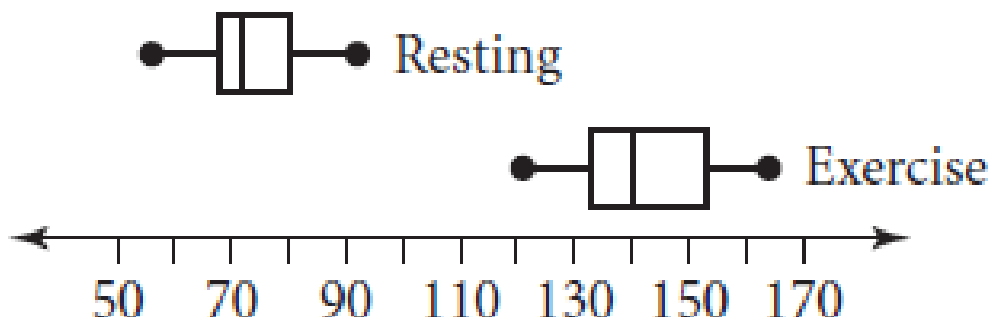
Οι παρακάτω προτάσεις συνδέονται με τα δύο αυτά θηκογράμματα. Εξετάστε αν είναι αληθείς ή όχι και δικαιολογείστε την απάντησή σας:

A. Το μικρότερο και το μεγαλύτερο ποσό που σηκώθηκε εκείνη την ημέρα και στα δύο μηχανήματα ήταν το ίδιο.

B. Το συνολικό ποσό των χρημάτων που σηκώθηκαν εκείνη την ημέρα στο μηχανήμα II ήταν μεγαλύτερο από εκείνο του μηχανήματος I.

Γ. Οι αναλήψεις στο μηχανήμα I είναι σχετικά μικρότερες από το μηχανήμα II.

5) Μετρήθηκαν οι παλμοί της καρδιάς 30 φοιτητών σε κατάσταση ανάπαυσης και σε κατάσταση σωματικής καταπόνησης (άσκηση-exercise). Τα αποτελέσματα φαίνονται στα παρακάτω θηκογράμματα.



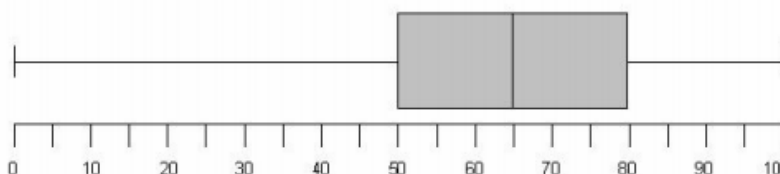
Απαντήστε στα παρακάτω:

Από τα δύο θηκογράμματα συμπεραίνουμε ότι:

- Το ελάχιστο των παλμών κατά την άσκηση είναι κατά 30 περίπου μεγαλύτερο από το μέγιστο των παλμών κατά την ανάπαυση. ΣΩΣΤΟ - ΛΑΘΟΣ
- Και τα δύο σύνολα δεδομένων έχουν το ίδιο είδος ασυμμετρίας. Οι τιμές οι μεγαλύτερες της διαμέσου είναι περισσότερο διεσπαρμένες από εκείνες που είναι μικρότερες. ΣΩΣΤΟ - ΛΑΘΟΣ
- Η διάμεσος των παλμών κατά τη σωματική καταπόνηση είναι σχεδόν διπλάσια από εκείνη των παλμών κατά την ανάπαυση. ΣΩΣΤΟ - ΛΑΘΟΣ
- Και τα δύο σύνολα δεδομένων έχουν το ίδιο ενδοτεταρτομοριακό εύρος. ΣΩΣΤΟ - ΛΑΘΟΣ

6) Το παρακάτω θηκογράμματα δείχνει τη βαθμολογία των μαθητών στα σχολεία μιας περιοχής σε ένα διαγώνισμα με κλίμακα βαθμολογίας από 0-100.

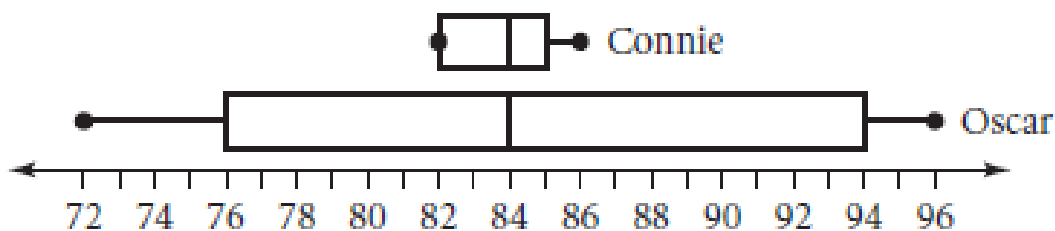
- α) Να συμπληρωθεί ο παρακάτω πίνακας:



Μέγιστη τιμή	Ελάχιστη τιμή	1 ^ο τετ	Διάμεσος	3 ^ο τεταρτ

- Τι κλάσμα των μαθητών είχε πάνω από 50;
- Πόσο είναι το ενδοτεταρτημοριακό εύρος.

7) Οι δύο συμμαθητές η Connie και ο Oscar σε 7 τεστ πήραν βαθμούς και οι δύο με μέση τιμή και διάμεσο 84. Η επίδοσή τους φαίνεται στα παρακάτω θηκογράμματα. Σχολιάστε την.



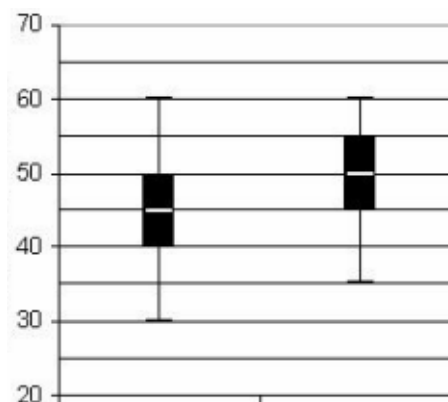
8) Το διπλανό θηκογράμμα αφορά τη βαθμολογία στα μαθηματικά δύο τάξεων (A η 1^η από αριστερά, B η 2^η από αριστερά).

α) Σε ποια τάξη το εύρος της βαθμολογίας είναι μεγαλύτερο;

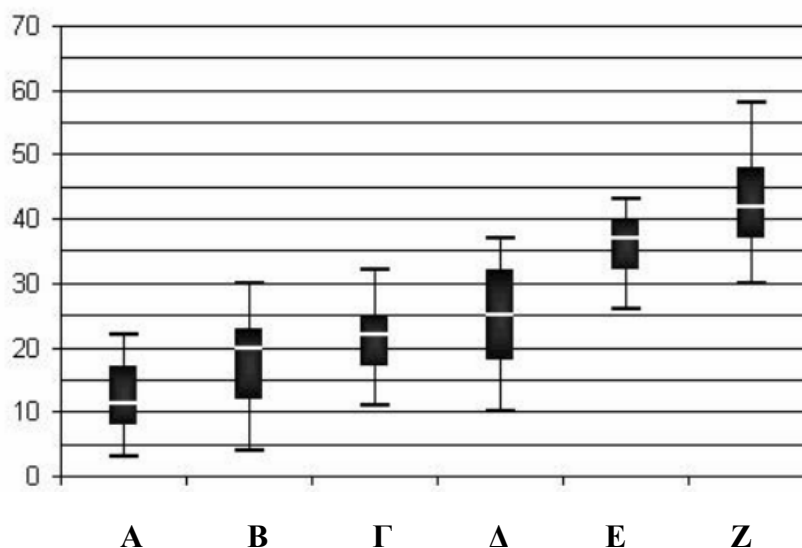
β) Τι διαφορά έχουν οι δύο διάμεσοι στις δύο τάξεις;

γ) Πόσο είναι το ενδοτεταρτομοριακό εύρος σε κάθε τάξη;

δ) Σε κάθε τάξη τι κλάσμα των μαθητών έχει πάνω από 45;



9)



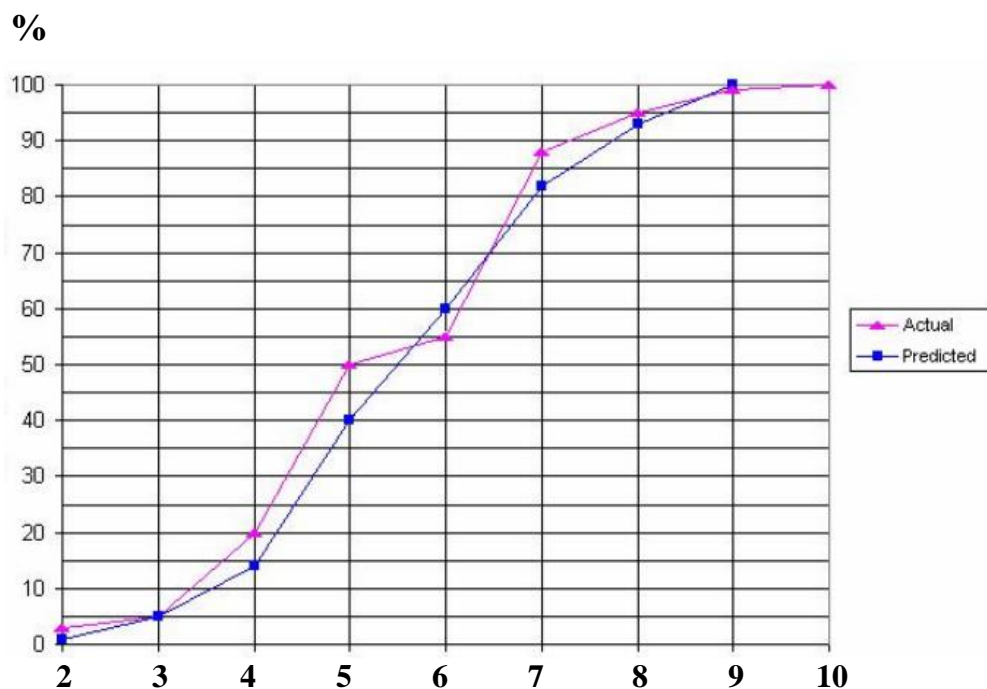
Οι μαθητές 6 τάξεων Α, Β κλπ (οριζόντιος άξονας) πέτυχαν σε ένα διαγωνισμό βαθμούς που στοιχεία τους φαίνονται στα παραπάνω θηκογράμματα.

α) Κατατάξτε τις τάξεις κατά αύξουσα τάξη μεγέθους ανάλογα με

- 1) Την ελάχιστη τιμή βαθμολογίας.
- 2) Την μέγιστη τιμή βαθμολογίας.
- 3) Την τιμή q_1 του πρώτου τεταρτημορίου.
- 4) Την τιμή της διαμέσου.
- 5) Την τιμή q_3 του τρίτου τεταρτημορίου.

β) Ποιά τάξη έχει το μεγαλύτερο ενδοτεταρτομοριακό εύρος;

10) Το παρακάτω διάγραμμα είναι το πολύγωνο των σχετικών αθροιστικών συχνοτήτων επί τοις εκατό που παριστάνει την βαθμολογία (από 2 έως 10) σε ένα διαγώνισμα (actual) καθώς και την αναμενόμενη (predicted) από τον καθηγητή του τμήματος.

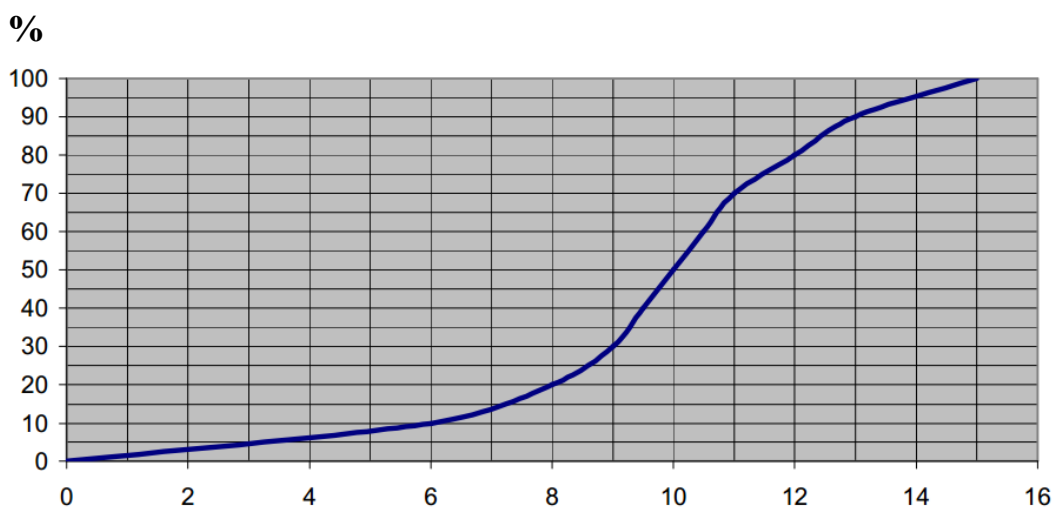
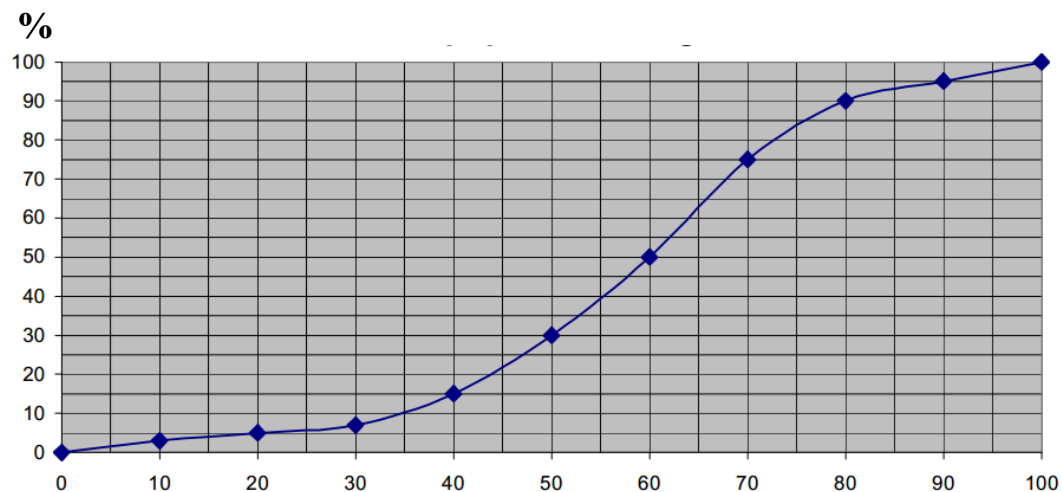


α) Να συμπληρωθεί ο παρακάτω πίνακας.

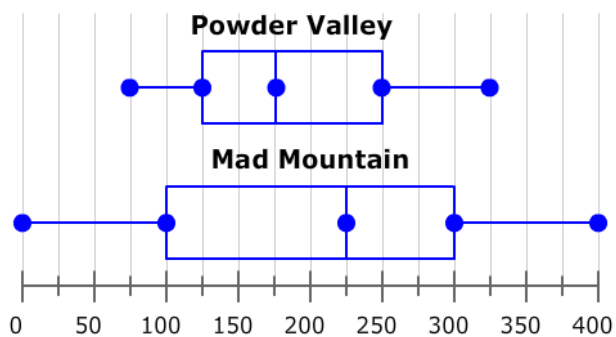
Βαθμολογία	Μέγιστη τιμή	Ελάχιστη τιμή	1 ^ο τεταρτ.	Διάμεσος	3 ^ο τεταρτ.	Ενδότ. εύρος
Πραγματική						
Αναμενόμενη						

β) Να γίνουν τα δύο θηκογράμματα που αφορούν τις βαθμολογίες.

11) Με τη βοήθεια του κάθε ενός από τα παρακάτω διαγράμματα (πολύγωνο αθροιστικών σχετικών συχνοτήτων) να γίνει το θηκόγραμμα.



12) Η ετήσια χιονόπτωση σε ίντσες σε δύο χειμερινά καταλύματα στο Powder Valley και στο Mad Mountain για τα τελευταία 50 χρόνια φαίνεται στα παρακάτω δύο θηκογράμματα:



α) Να συμπληρωθεί ο παρακάτω πίνακας:

Χιονόπτωση	Μέγιστη τιμή	Ελάχιστη τιμή	1 ^ο τεταρτ.	Διάμεσος	3 ^ο τεταρτ.	Ενδότ. εύρος
Powder Valley						
Mad Mountain						

β) Σε ποιο από τα δύο καταλύματα είναι μεγαλύτερη η πιθανότητα η χιονόπτωση να ξεπεράσει τις 300 ίντσες την επόμενη χρονιά;

Απαντήσεις – Υποδείξεις

5) Σε περίπτωση που θέλετε να σχεδιάσετε τα θηκογράμματα τα δεδομένα είναι τα παρακάτω: οι παλμοί της καρδιάς 30 φοιτητών σε κατάσταση ανάπαυσης (resting): 68, 76, 84, 80, 76, 72, 60, 68, 68, 80, 68, 80, 64, 64, 72, 76, 72, 68, 56, 88, 80, 76, 68, 56, 64, 60, 92, 72, 84, 72. Η ελάχιστη τιμή 56, η μέγιστη 92, η διάμεσος 72, το 1^ο τεταρτημόριο είναι 68 και το 3^ο 80. Τα αντίστοιχα νούμερα σε κατάσταση σωματικής καταπόνησης (άσκηση-exercise) ήταν τα εξής: 148, 136, 157, 151, 121, 139, 137, 129, 127, 129, 155, 141, 133, 153, 161, 153, 127, 135, 144, 146, 136, 131, 133, 159, 127, 142, 133, 150, 164, 161. Στα δεύτερα η ελάχιστη τιμή τους είναι 121, η μέγιστη 164, η διάμεσος 140, το 1^ο τεταρτημόριο είναι 133 και το 3^ο 153).

6) α) 0-100-50-70-80-30 β) 3/4 γ) 30.

7) Σε περίπτωση που θέλετε να σχεδιάσετε τα θηκογράμματα:

Βαθμοί του Connie: 82, 82, 84, 84, 85, 85, 86

Βαθμοί του Oscar: 72, 76, 76, 84, 90, 94, 96

8) A: 50%, B: 75%