

## Μάθημα: Στατιστική (Κωδ. 105)

Διδάσκων: Γιώργος Κ. Παπαδόπουλος

### 5. Κατανόηση της Κατανομής των Δεδομένων

#### Σύντομη ανασκόπηση βασικών εννοιών, προτάσεων και τύπων

|   |   |
|---|---|
| <b>Πληθυσμός (ή στατιστικός πληθυσμός)</b>                | Στατιστικό πληθυσμό (ή πληθυσμό) ονομάζουμε την κατανομή των τιμών μιας τυχαίας μεταβλητής, δηλαδή την κατανομή των τιμών που παίρνει ένα κοινό χαρακτηριστικό μιας ομάδας υποκειμένων. Κάθε υποκείμενο επί του οποίου μετράται/ παρατηρείται η τιμή ενός κοινού χαρακτηριστικού ονομάζεται <b>δειγματοληπτική/πειραματική μονάδα</b> .   |
| <b>Τυχαίο δείγμα και πραγματοποίηση τυχαίου δείγματος</b> | Τυχαίο δείγμα μεγέθους $n$ από έναν πληθυσμό ονομάζουμε $n$ ανεξάρτητες τυχαίες μεταβλητές $X_1, X_2, \dots, X_n$ που παίρνουν τιμές από τον πληθυσμό αυτό (και έχουν επομένως την ίδια κατανομή). Οι συγκεκριμένες τιμές $x_1, x_2, \dots, x_n$ , που έχουμε διαθέσιμες για επεξεργασία μετά τη λήψη του δείγματος αποτελούν μια πραγματοποίηση των $X_1, X_2, \dots, X_n$ και ονομάζονται <b>δεδομένα ή παρατηρήσεις</b> .  |
| <b>Πίνακας κατανομής συχνότητας</b>                       | <b>α) Ποσοτικές μεταβλητές</b><br>Στην πρώτη στήλη του πίνακα κατανομής συχνότητας καταγράφονται σε αύξουσα σειρά οι διαφορετικές τιμές $y_1, y_2, \dots, y_k$ από τις $x_1, x_2, \dots, x_n$ που εμφανίστηκαν στο δείγμα. Στις επόμενες στήλες, για κάθε τιμή $y_i$ , $i = 1, 2, \dots, k$ , καταγράφεται <ul style="list-style-type: none"><li>• η <b>συχνότητά της</b>, <math>v_i</math> (πόσες φορές εμφανίστηκε)</li><li>• η <b>σχετική συχνότητά της</b>, <math>f_i = v_i/n</math></li><li>• η <b>αθροιστική συχνότητά της</b>, <math>N_i</math> (το άθροισμα των συχνοτήτων των τιμών που είναι <math>\leq y_i</math>)</li><li>• η <b>αθροιστική σχετική συχνότητά της</b>, <math>F_i</math> (το άθροισμα των σχετικών συχνοτήτων των τιμών που είναι <math>\leq y_i</math>)</li></ul> Αν (έχει) γίνει ομαδοποίηση των τιμών, στην πρώτη στήλη αντί των διαφορετικών τιμών καταγράφονται οι διαφορετικές κλάσεις τιμών. Στις επόμενες στήλες καταγράφεται η συχνότητα, η σχετική συχνότητα, η αθροιστική συχνότητα και η αθροιστική σχετική συχνότητα κάθε κλάσης τιμών.<br><b>β) Ποιοτικές μεταβλητές</b> <ul style="list-style-type: none"><li>• Στις ποιοτικές μεταβλητές <b>κατηγορίας</b> δεν έχει νόημα η διάταξη των διαφορετικών τιμών <math>y_1, y_2, \dots, y_k</math> και επομένως δεν έχουν νόημα ούτε οι αθροιστικές ούτε οι αθροιστικές σχετικές συχνοτήτες αλλά μόνο οι συχνοτήτες και οι σχετικές συχνοτήτες.</li><li>• Στις ποιοτικές μεταβλητές <b>διάταξης</b> η διάταξη των διαφορετικών τιμών <math>y_1, y_2, \dots, y_k</math> έχει νόημα και επομένως έχουν νόημα τόσο οι συχνοτήτες και οι σχετικές συχνοτήτες όσο και οι αθροιστικές και οι αθροιστικές σχετικές.</li></ul> |
| <b>Γραφική παρουσίαση κατανομής συχνότητας</b>            | <b>α) Ποσοτικές μεταβλητές</b> <ul style="list-style-type: none"><li>• Σημειόγραμμα</li><li>• Ραβδόγραμμα συχνοτήτων και σχετικών συχνοτήτων</li><li>• Διάγραμμα συχνοτήτων και σχετικών συχνοτήτων</li><li>• Κυκλικό διάγραμμα συχνοτήτων και σχετικών συχνοτήτων</li><li>• Ιστόγραμμα συχνοτήτων/σχετικών συχνοτήτων/ αθροιστικών συχνοτήτων/αθροιστικών σχετικών συχνοτήτων</li><li>• Πολύγωνο συχνοτήτων/σχετικών συχνοτήτων/αθροιστικών συχνοτήτων/αθροιστικών σχετικών συχνοτήτων</li><li>• Φυλλογράφημα</li><li>• Θηκόγραμμα</li></ul> <b>β) Ποιοτικές μεταβλητές</b> <ul style="list-style-type: none"><li>• Ραβδόγραμμα συχνοτήτων και σχετικών συχνοτήτων</li><li>• Κυκλικό διάγραμμα συχνοτήτων και σχετικών συχνοτήτων</li></ul>  |

**Αριθμητικά περιγραφικά μέτρα**  
(για συγκεκριμένη πραγματοποίηση  $x_1, x_2, \dots, x_n$  του δείγματος με  $y_1, y_2, \dots, y_k$  διαφορετικές τιμές)

**α) Ποσοτικές μεταβλητές**

**Μέτρα θέσης**

- **Δειγματικός μέσος,  $\bar{x}$**

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k x_i = \frac{1}{n} \sum_{i=1}^k v_i y_i = \sum_{i=1}^k f_i y_i$$

- **Κορυφή του δείγματος,  $M_0$**

Η τιμή με τη μεγαλύτερη συχνότητα

- **Διάμεσος του δείγματος,  $\delta$  ή  $Q_2$**

Το πολύ 50% των τιμών του δείγματος είναι μικρότερες από τη διάμεσο και επίσης το πολύ 50% των τιμών του δείγματος είναι μεγαλύτερες από τη διάμεσο.

Σε αύξουσα διάταξη των  $x_1, x_2, \dots, x_n$ , τη θέση της διαμέσου δίνει ο αριθμός  $0.5(n+1)$  εφόσον είναι ακέραιος, ενώ αν δεν είναι ακέραιος, τότε η διάμεσος είναι ίση με το ημίαθροισμα των δύο τιμών που οι θέσεις τους είναι οι πλησιέστερες στον αριθμό  $0.5(n+1)$ .

- **$p$ -ποσοστιαία σημεία του δείγματος,  $x_p, 0 < p < 1$**

Το πολύ  $100p\%$  των τιμών του δείγματος είναι μικρότερες από το  $p$ -ποσοστιαίο σημείο και το πολύ  $100(1-p)\%$  των τιμών του δείγματος είναι μεγαλύτερες από το  $p$ -ποσοστιαίο σημείο.

Σε αύξουσα διάταξη των  $x_1, x_2, \dots, x_n$ , τη θέση του

$p$ -ποσοστιαίου σημείου δίνει ο αριθμός  $p(n+1)$  εφόσον είναι ακέραιος, ενώ αν δεν είναι ακέραιος, τότε το  $p$ -ποσοστιαίο σημείο εκτιμάται με παρεμβολή μεταξύ των δύο τιμών που οι θέσεις τους είναι οι πλησιέστερες στον αριθμό  $p(n+1)$ .

- **Τεταρτημόρια,  $Q_1 = x_{0.25}, Q_2 = x_{0.5} = \delta, Q_3 = x_{0.75}$**

Αν (έχει) γίνει ομαδοποίηση των τιμών του δείγματος σε  $k$  κλάσεις:

- Η **κορυφή** υπολογίζεται από τον τύπο

$$M_0 = L_i + \frac{\Delta_1}{\Delta_1 + \Delta_2} c_i$$

όπου,  $L_i$  το κάτω άκρο της επικρατούσας κλάσης, δηλαδή της κλάσης με τη μεγαλύτερη συχνότητα,  $c_i$  το πλάτος της επικρατούσας κλάσης,  $\Delta_1 = v_i - v_{i-1}$  και  $\Delta_2 = v_i - v_{i+1}$  όπου  $v_i$  η συχνότητα της επικρατούσας κλάσης.

- Στον τύπο υπολογισμού του **δειγματικού μέσου**

$$\bar{x} = \frac{1}{n} \sum_{i=1}^k v_i y_i = \sum_{i=1}^k f_i y_i$$

τα  $y_i, i = 1, 2, \dots, k$  είναι οι κεντρικές τιμές των κλάσεων.

- Η **διάμεσος** υπολογίζεται από τον τύπο

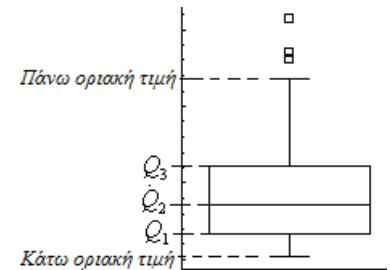
$$\delta = L_i + \frac{0.5n - N_{i-1}}{v_i} c_i$$

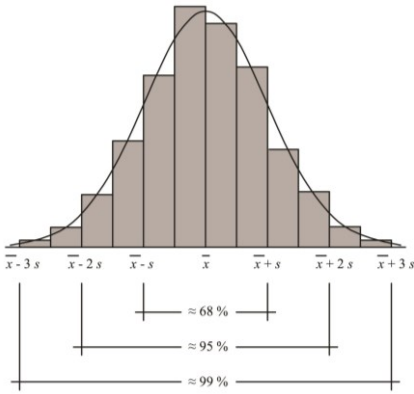
όπου,  $L_i$  το κάτω άκρο της μεσαίας κλάσης, δηλαδή της κλάσης στην οποία ανήκει η διάμεσος,  $c_i$  το πλάτος της μεσαίας κλάσης,  $v_i$  η συχνότητα της μεσαίας κλάσης και  $N_{i-1}$  η αθροιστική συχνότητα της προηγούμενης κλάσης από τη μεσαία.

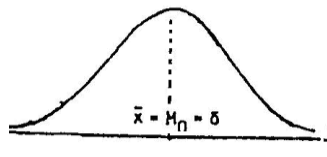
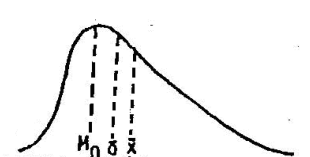
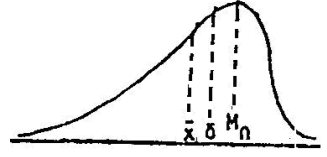
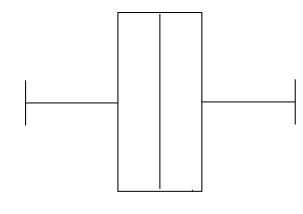
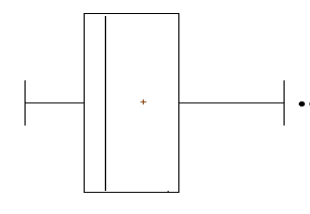
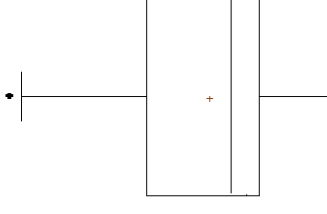
- Τα  **$p$ -ποσοστιαία σημεία** υπολογίζονται από τον τύπο

$$x_p = L_i + \frac{pn - N_{i-1}}{v_i} c_i$$

όπου,  $L_i$  το κάτω άκρο της κλάσης στην οποία βρίσκεται το  $x_p$ ,  $c_i$  το πλάτος της,  $v_i$  η συχνότητά της και  $N_{i-1}$  η αθροιστική συχνότητα της προηγούμενης κλάσης.

|   |   |
|---|---|
|   | <p><b>Μέτρα μεταβλητότητας/διασποράς</b></p> <ul style="list-style-type: none"> <li>• <b>Εύρος</b><br/> <math>R = x_{\max} - x_{\min}</math></li> <li>• <b>Ενδοτεταρτημοριακό εύρος</b><br/> <math>Q_3 - Q_1</math></li> <li>• <b>Δειγματική διακύμανση</b><br/> <math display="block">s^2 = \frac{1}{v-1} \sum_{i=1}^v (x_i - \bar{x})^2 = \frac{1}{v-1} \left( \sum_{i=1}^v x_i^2 - v\bar{x}^2 \right) =</math> <math display="block">= \frac{1}{v-1} \sum_{i=1}^k (y_i - \bar{x})^2 v_i = \frac{1}{v-1} \left( \sum_{i=1}^k v_i y_i^2 - v\bar{x}^2 \right)</math></li> <li>• <b>Δειγματική τυπική απόκλιση</b><br/> <math display="block">s = \sqrt{\frac{1}{v-1} \sum_{i=1}^v (x_i - \bar{x})^2} = \sqrt{\frac{1}{v-1} \left( \sum_{i=1}^v x_i^2 - v\bar{x}^2 \right)} =</math> <math display="block">= \sqrt{\frac{1}{v-1} \sum_{i=1}^k (y_i - \bar{x})^2 v_i} = \sqrt{\frac{1}{v-1} \left( \sum_{i=1}^k v_i y_i^2 - v\bar{x}^2 \right)}</math></li> <li>• <b>Συντελεστής μεταβλητότητας</b><br/> <math display="block">CV = \frac{s}{ \bar{x} } 100\%</math></li> </ul> <p>Αν (έχει) γίνει ομαδοποίηση των τιμών του δείγματος σε <math>k</math> κλάσεις τα <math>y_i, i = 1, 2, \dots, k</math> είναι οι κεντρικές τιμές των κλάσεων.</p> <p><b>β) Ποιοτικές μεταβλητές</b><br/> Ορίζεται (και έχει νόημα) μόνο η <b>κορυφή</b> της κατανομής.</p> |
| <p><b>Μέτρα θέσης και μεταβλητότητας γραμμικού μετασχηματισμού των παρατηρήσεων/δεδομένων</b></p> | <p>Αν <math>t_i = \alpha x_i + \beta</math>, τότε</p> <ul style="list-style-type: none"> <li>• <math>\bar{t} = \alpha \bar{x} + \beta</math></li> <li>• <math>s_t^2 = \alpha^2 s_x^2</math></li> <li>• <math>s_t =  \alpha  s_x</math></li> <li>• <math>\delta_t = \alpha \delta_x + \beta</math></li> <li>• <math>M_{0t} = \alpha M_{0x} + \beta</math></li> </ul>   |
| <p><b>Θηκόγραμμα</b></p>  |  <p>Πάνω οριακή τιμή: η μεγαλύτερη τιμή του δείγματος που είναι <math>\leq Q_3 + 1.5(Q_3 - Q_1)</math> ή <math>\leq Q_3 + 3(Q_3 - Q_1)</math><br/> Κάτω οριακή τιμή: η μικρότερη τιμή του δείγματος που είναι <math>\geq Q_1 - 1.5(Q_3 - Q_1)</math> ή <math>\geq Q_1 - 3(Q_3 - Q_1)</math></p>   |

|                                     |   |
|-------------------------------------|---|
| <p><b>Ο εμπειρικός κανόνας</b></p>  | <p>Αν η κατανομή του δείγματος προσομοιάζει με μια κανονική κατανομή (έχει κωδωνοειδή μορφή), τότε</p> <ul style="list-style-type: none"> <li>• στο διάστημα <math>(\bar{x} - s, \bar{x} + s)</math> βρίσκεται περίπου το 68% των παρατηρήσεων</li> <li>• στο διάστημα <math>(\bar{x} - 2s, \bar{x} + 2s)</math> βρίσκεται περίπου το 95% των παρατηρήσεων</li> <li>• στο διάστημα <math>(\bar{x} - 3s, \bar{x} + 3s)</math> βρίσκονται περίπου όλες οι παρατηρήσεις (πάνω από το 99%).</li> </ul>  |
| <p><b>Η ανισότητα Chebyshev</b></p> | <p>Το ποσοστό των τιμών του δείγματος που βρίσκονται στο διάστημα <math>(\bar{x} - ks, \bar{x} + ks)</math> είναι <b>τουλάχιστον</b> <math>1 - (1/k^2)</math>,</p>  |

|  |   |   |
|--|---|---|
|    |   |   |
|  <p style="text-align: center;"><i>Συμμετρική</i></p> |  <p style="text-align: center;"><i>Λοξή με θετική ασυμμετρία</i></p> |  <p style="text-align: center;"><i>Λοξή με αρνητική ασυμμετρία</i></p> |

## Προβλήματα και Ασκήσεις

1. Η πτυχιακή μελέτη ενός φοιτητή αφορούσε, μεταξύ άλλων, στον αριθμό των πετάλων, έστω  $X$ , των ανθέων μιας συγκεκριμένης ποικιλίας ενός φυτού που καλλιεργείται στο νομό Κοζάνης. Στο πλαίσιο αυτής της μελέτης, ο φοιτητής μέτρησε τον αριθμό των πετάλων σε 115 άνθη της συγκεκριμένης ποικιλίας που επέλεξε τυχαία από καλλιέργειες του νομού Κοζάνης. Τα αποτελέσματα αυτών των μετρήσεων φαίνονται στον πίνακα που ακολουθεί.

|   |   |   |   |   |   |    |   |   |   |   |    |   |   |   |   |   |
|---|---|---|---|---|---|----|---|---|---|---|----|---|---|---|---|---|
| 7 | 5 | 8 | 7 | 5 | 5 | 6  | 6 | 5 | 7 | 5 | 5  | 5 | 9 | 6 | 8 | 5 |
| 5 | 5 | 6 | 6 | 5 | 5 | 6  | 5 | 9 | 6 | 5 | 5  | 7 | 6 | 6 | 7 | 5 |
| 7 | 5 | 5 | 6 | 6 | 5 | 6  | 5 | 6 | 5 | 5 | 5  | 5 | 6 | 6 | 5 | 5 |
| 8 | 5 | 5 | 5 | 5 | 6 | 5  | 5 | 5 | 6 | 5 | 5  | 6 | 5 | 5 | 5 | 6 |
| 7 | 5 | 7 | 5 | 5 | 8 | 5  | 5 | 5 | 6 | 5 | 10 | 5 | 6 | 5 | 5 | 6 |
| 5 | 7 | 5 | 5 | 5 | 9 | 5  | 5 | 7 | 5 | 5 | 5  | 5 | 6 | 7 | 5 | 5 |
| 6 | 5 | 6 | 5 | 7 | 5 | 10 | 5 | 6 | 5 | 5 | 5  | 8 |   |   |   |   |

- α) Να υπολογίσετε και να ερμηνεύσετε τα μέτρα θέσης και διασποράς της κατανομής του δείγματος. β) Να κατασκευάσετε το θηκόγραμμα του δείγματος. Τι συμπεραίνετε για την κατανομή του δείγματος; γ) Για κάποιο άνθος βρέθηκε  $x = 7$ . Τι μπορούμε να πούμε για τη θέση αυτής της τιμής στην κατανομή του δείγματος; δ) Αν  $x_{0.98} = 9.68$ , τι μπορούμε να πούμε για τη θέση της τιμής  $x = 10$  στην κατανομή του δείγματος; ε) Να κατασκευάσετε το θηκόγραμμα των  $z$ -τιμών,  $z_1, z_2, \dots, z_{115}$ , των τιμών  $x_1, x_2, \dots, x_{115}$  της  $X$ . Τι συμπεραίνετε για την κατανομή των  $z$ -τιμών;

2. Στον πίνακα που ακολουθεί φαίνεται ο αριθμός σταφίδων που περιέχονται σε καθένα από 14, τυχαία επιλεγμένα, μικρά φακελάκια (των 30gr) παραγωγής μιας μεγάλης εταιρείας συσκευασίας τροφίμων και αντίστοιχα, σε 14 τυχαία επιλεγμένα φακελάκια (επίσης των 30gr) οικοτεχνικής παραγωγής.

| Αριθμός σταφίδων σε φακελάκια (των 30 gr) |    |    |    |    |                        |    |    |    |    |
|---|----|----|----|----|------------------------|----|----|----|----|
| Εταιρείας τροφίμων                        |    |    |    |    | Οικοτεχνικής παραγωγής |    |    |    |    |
| 25  | 26 | 25 | 28 | 26 | 25                     | 29 | 24 | 24 | 28 |
| 26  | 28 | 27 | 26 | 27 | 24                     | 28 | 22 | 25 | 28 |
| 24  | 25 | 26 | 26 |    | 30                     | 27 | 28 | 24 |    |

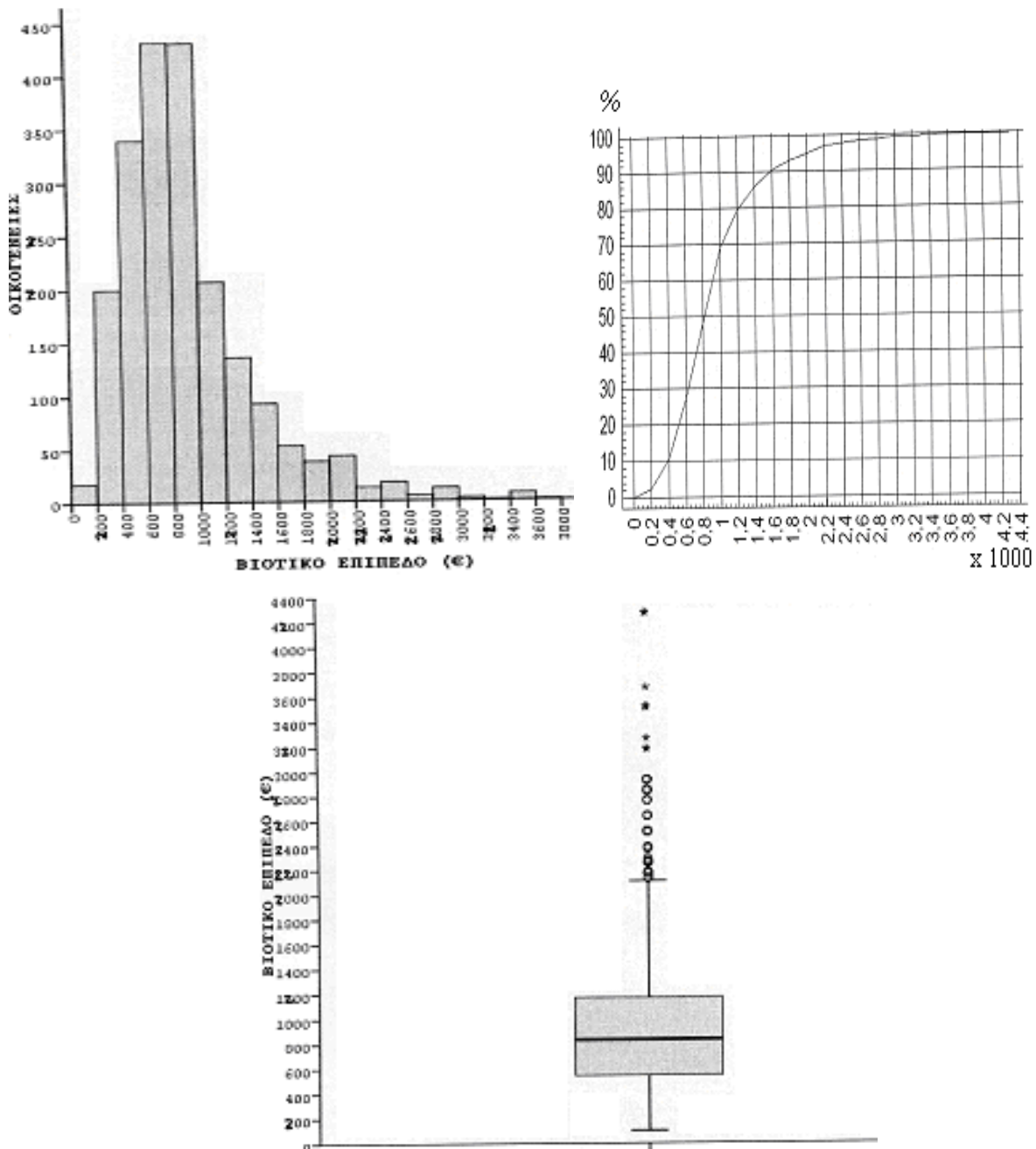
Να υπολογίσετε και να ερμηνεύσετε τα μέτρα θέσης και διασποράς καθώς και το θηκόγραμμα της κατανομής κάθε δείγματος. Πώς συγκρίνετε τις δύο κατανομές. Τι μπορούμε να πούμε για τη θέση της τιμής  $x=28$  στην κατανομή του δείγματος

3. Μια ομάδα ερευνητών, στο πλαίσιο ενός πειράματος, ράντισε μια καλλιέργεια σέλινου με παραθείο με σκοπό να εκτιμήσει το υπόλοιπο παραθείου στο σέλινο μετά ορισμένο χρονικό διάστημα από το ράντισμα. Στον πίνακα που ακολουθεί φαίνονται ομαδοποιημένες σε πέντε κλάσεις οι μετρήσεις (σε milligrams ) που έκανε η ερευνητική ομάδα σε 100 τυχαία επιλεγμένα φυτά.

| Ποσότητα παραθείου (σε mgr) | Αριθμός φυτών |
|-----------------------------|---------------|
| [0, 20)                     | 10            |
| [20, 40)                    | 10            |
| [40, 60)                    | 20            |
| [60, 80)                    | 40            |
| [80, 100)                   | 20            |

Να εφαρμόσετε κατάλληλες μεθόδους περιγραφικής στατιστικής για να περιγράψετε την κατανομή του παραπάνω δείγματος. Να συνοψίσετε τα συμπεράσματά σας σε μια σύντομη παράγραφο.

4. Έστω  $X$  τυχαία μεταβλητή που εκφράζει το μηνιαίο βιοτικό επίπεδο των μελών των οικογενειών στο Νομό Αττικής<sup>1</sup> το έτος 2008. Στα σχήματα που ακολουθούν παρουσιάζεται η κατανομή ενός αντιπροσωπευτικού δείγματος τιμών της  $X$  μεγέθους  $n = 2051$ . Για την κατανομή αυτή, δίνεται επίσης, ο δειγματικός μέσος,  $\bar{x} = 923.12 \text{ €}$  και η δειγματική τυπική απόκλιση,  $s = 538.01 \text{ €}$ .



- α) Ποιο πληθυσμό μελετάμε και ποια είναι η δειγματοληπτική μονάδα. β) Να υπολογίσετε (κατά προσέγγιση) και να ερμηνεύσετε τη διάμεσο και το 1<sup>ο</sup> και 3<sup>ο</sup> τεταρτημόριο της κατανομής του δείγματος. γ) Τι ποσοστό (περίπου) των οικογενειών του δείγματος έχει μηνιαίο βιοτικό επίπεδο πάνω από 2000€; δ) Αν

<sup>1</sup> Το μηνιαίο βιοτικό επίπεδο μιας οικογένειας είναι ίδιο για όλα τα μέλη της οικογένειας και προκύπτει από τη διαίρεση του συνολικού καθαρού μηνιαίου εισοδήματος της οικογένειας με ένα σταθμικό άθροισμα των μελών της. Το σταθμικό άθροισμα προκύπτει ως εξής: για τον πρώτο ενήλικα βάρους 1, για κάθε άλλο ενήλικα και κάθε παιδί άνω των 14 ετών βάρους 0.5 και για κάθε παιδί κάτω των 14 ετών βάρους 0.3. Για παράδειγμα, το μηνιαίο βιοτικό επίπεδο μιας οικογένειας με συνολικό καθαρό μηνιαίο εισόδημα 2800€ που αποτελείται από τον πατέρα, τη μητέρα, τη γιαγιά, ένα παιδί 8 ετών και ένα παιδί 16 ετών είναι,  $2800 / (1 + 0.5 + 0.5 + 0.3 + 0.5) = 1000 \text{ €}$ .

είστε εκπρόσωπος των εργαζομένων, ποιες πληροφορίες από την κατανομή του δείγματος θα χρησιμοποιούσατε ως επιχειρήματα σε μια συνάντηση με τον υπουργό οικονομικών; ε) Τι ποσοστό (περίπου) των οικογενειών του δείγματος βρίσκεται κάτω από το όριο της φτώχειας (το όριο της φτώχειας ορίζεται ως το 60% του διάμεσου μηνιαίου βιοτικού επιπέδου). στ) Αν η  $z$ -τιμή μιας τιμής του δείγματος είναι  $-1.3$ , ποια είναι η θέση αυτής της τιμής στην κατανομή του δείγματος; ζ) Αν μια τιμή του δείγματος είναι  $1500€$ , ποια είναι η θέση της στην κατανομή του δείγματος; η) Τι ποσοστό (κατά προσέγγιση) των τιμών του δείγματος βρίσκεται στο διάστημα  $(\bar{x} - 2s, \bar{x} + 2s)$ ; Συμφωνεί αυτό το ποσοστό με αυτό που αναμένουμε από την ανισότητα *Chebyshev*;

5. Ένας ερευνητής σχεδίασε και εκτέλεσε ένα πείραμα για να μελετήσει το χρόνο, έστω  $X$  (σε ημέρες), που απαιτείται για την αποδόμηση μιας συγκεκριμένης χημικής ουσίας από το μέλι (η ουσία αυτή χρησιμοποιείται για την καταπολέμηση των ακάρεων). Στον πίνακα που ακολουθεί φαίνονται 50 σχετικές παρατηρήσεις.

|    |    |    |    |    |    |    |    |    |    |    |    |    |
|----|----|----|----|----|----|----|----|----|----|----|----|----|
| 38 | 47 | 32 | 55 | 42 | 40 | 36 | 35 | 45 | 45 | 40 | 35 | 34 |
| 39 | 50 | 48 | 41 | 40 | 42 | 38 | 30 | 34 | 41 | 33 | 37 | 36 |
| 43 | 30 | 41 | 46 | 35 | 43 | 30 | 32 | 39 | 31 | 48 | 46 | 36 |
| 36 | 39 | 41 | 46 | 32 | 33 | 36 | 40 | 37 | 50 | 31 |    |    |

α) Να υπολογίσετε τον μέσο, την τυπική απόκλιση, την κορυφή και τη διάμεσο του δείγματος. β) Να ομαδοποιήσετε τις παρατηρήσεις σε 6 κλάσεις με πλάτος 5 ημέρες η κάθε μια και αριστερό άκρο της πρώτης κλάσης τις 30 ημέρες. Να υπολογίσετε και πάλι τον μέσο, την τυπική απόκλιση, την κορυφή και τη διάμεσο του δείγματος χρησιμοποιώντας τώρα τις ομαδοποιημένες παρατηρήσεις και να συγκρίνετε τα αποτελέσματα με αυτά του ερωτήματος (α). γ) Να κατασκευάσετε το ιστόγραμμα συχνοτήτων της κατανομής με βάση την ομαδοποίηση που κάνατε στο (β). Τι συμπεραίνετε για τη μορφή της; δ) Να σχολιάσετε τη θέση της κορυφής, της διαμέσου και του μέσου του δείγματος σε σχέση με τη μορφή της κατανομής που προκύπτει από το (γ). ε) Να υπολογίσετε τα ποσοστά των παρατηρήσεων που βρίσκονται εντός των διαστημάτων  $(\bar{x} - s, \bar{x} + s)$ ,  $(\bar{x} - 2s, \bar{x} + 2s)$ ,  $(\bar{x} - 3s, \bar{x} + 3s)$  και να τα συγκρίνετε με τα αντίστοιχα ποσοστά που αναμένονται από την ανισότητα *Chebyshev* και από τον εμπειρικό κανόνα.

6. Σε μια περιοχή του Μαινάλου αιχμαλωτίστηκαν από μια ομάδα ερευνητών, με βάση ένα σχέδιο τυχαίας δειγματοληψίας, 100 αλεπούδες για να ελεγχθούν ως προς το αν έχουν προσβληθεί από παράσιτα (ενός συγκεκριμένου τύπου). Στη συνέχεια οι ερευνητές κατέγραψαν τον αριθμό, έστω  $X$ , των παράσιτων που βρέθηκαν ανά αλεπού. Στον πίνακα που ακολουθεί φαίνονται οι συχνότητες όλων των τιμών  $x$  της μεταβλητής  $X$  που εμφανίστηκαν στο δείγμα (μηδέν παράσιτα σε κάθε μία από 69 αλεπούδες, ένα παράσιτο σε κάθε μια από 17 αλεπούδες, 2 παράσιτα σε κάθε μία από 6 αλεπούδες, κτλ).

|                   |    |    |   |   |   |   |   |   |   |
|-------------------|----|----|---|---|---|---|---|---|---|
| Αριθμός παράσιτων | 0  | 1  | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| Αριθμός αλεπούδων | 69 | 17 | 6 | 3 | 1 | 2 | 1 | 0 | 1 |

α) Να υπολογίσετε και να ερμηνεύσετε τα μέτρα θέσης και διασποράς της κατανομής του δείγματος. β) Να κατασκευάσετε το θηκόγραμμα του δείγματος. Τι συμπεραίνετε για την κατανομή του δείγματος; γ) Να υπολογίσετε τα ποσοστιαία σημεία  $x_{0.95}$  και  $x_{0.98}$ . Τι μπορούμε να πούμε για τη θέση των τιμών,  $x = 4$  και  $x = 6$  στην κατανομή του δείγματος;

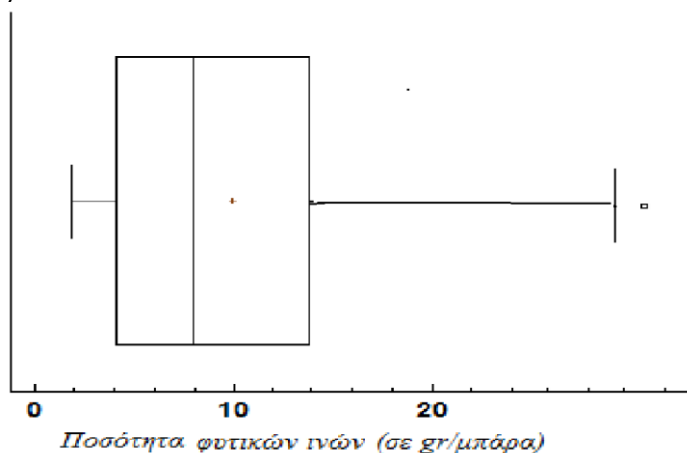
7. Στον πίνακα που ακολουθεί φαίνεται για καθένα από 25 βάζα μελιού που επιλέξαμε τυχαία από τα ράφια του πρατηρίου ενός συνεταιρισμού μελισσοκόμων i) ο παραγωγός ( $A, B, \Gamma$ ) ii) το είδος του μελιού (ανθέων, ελάτης,

θυμαρίσιο, πεύκου) iii) το μέγεθος της συσκευασίας (μικρό, μεσαίο, μεγάλο) iv) η περιεκτικότητα του μελιού σε σάκχαρα και v) η ποσότητα μελιού που περιέχεται σε κάθε βάζο.

| Παραγωγός | Είδος     | Μέγεθος συσκευασίας | Περιεκτικότητα σε σάκχαρα (%) | Ποσότητα (σε gr) |
|-----------|-----------|---------------------|-------------------------------|------------------|
| A         | Ανθέων    | Μικρό               | 75                            | 250              |
| B         | Ανθέων    | Μεσαίο              | 77                            | 500              |
| A         | Ανθέων    | Μεσαίο              | 70                            | 490              |
| Γ         | Ανθέων    | Μικρό               | 78                            | 240              |
| A         | Θυμαρίσιο | Μεγάλο              | 77                            | 1000             |
| A         | Ανθέων    | Μεγάλο              | 75                            | 950              |
| A         | Πεύκου    | Μεγάλο              | 52                            | 1100             |
| Γ         | Πεύκου    | Μεσαίο              | 55                            | 550              |
| Γ         | Ανθέων    | Μεσαίο              | 77                            | 450              |
| Γ         | Ελάτης    | Μεσαίο              | 60                            | 500              |
| A         | Θυμαρίσιο | Μικρό               | 77                            | 250              |
| B         | Θυμαρίσιο | Μικρό               | 75                            | 270              |
| A         | Θυμαρίσιο | Μεγάλο              | 78                            | 1000             |
| B         | Θυμαρίσιο | Μεγάλο              | 78                            | 1050             |
| B         | Πεύκου    | Μεγάλο              | 60                            | 1000             |
| A         | Πεύκου    | Μεσαίο              | 50                            | 500              |
| B         | Ελάτης    | Μεσαίο              | 55                            | 550              |
| Γ         | Ελάτης    | Μικρό               | 59                            | 250              |
| A         | Ελάτης    | Μικρό               | 60                            | 250              |
| B         | Θυμαρίσιο | Μεσαίο              | 75                            | 560              |
| Γ         | Ανθέων    | Μεσαίο              | 77                            | 500              |
| Γ         | Πεύκου    | Μεσαίο              | 55                            | 500              |
| Γ         | Ανθέων    | Μικρό               | 77                            | 240              |
| B         | Ανθέων    | Μεγάλο              | 72                            | 990              |
| A         | Ανθέων    | Μικρό               | 75                            | 250              |

Να εφαρμόσετε κατάλληλες (κατά περίπτωση) μεθόδους περιγραφικής στατιστικής για να περιγράψετε την κατανομή καθενός από τα δείγματα αυτά και να συνοψίσετε τα συμπεράσματά σας σε μια σύντομη παράγραφο.

8. Στο σχήμα που ακολουθεί φαίνεται σε μορφή θηκογραμματος η κατανομή ενός δείγματος 50 τιμών της τυχαίας μεταβλητής  $X$  που εκφράζει την ποσότητα φυτικών ιών σε μπάρες δημητριακών ορισμένου τύπου. Δίνεται επίσης ότι ο μέσος του συγκεκριμένου δείγματος είναι  $\bar{x} = 10 \text{ gr/μπάρα}$  με τυπική απόκλιση  $s = 4.2 \text{ gr/μπάρα}$ .



Τι μπορείτε να συμπεράνετε για τη θέση, τη μεταβλητότητα και τη μορφή της κατανομής του συγκεκριμένου δείγματος τιμών της  $X$ ;