ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

Μερικές χρήσιμες έννοιες είναι οι εξής:

- **Πληθυσμός**: Κάθε σύνολο, τα στοιχεία του οποίου εξετάζουμε ως προς ένα ή περισσότερα χαρακτηριστικά.

- **Μεταβλητές**: Τα χαρακτηριστικά του υπό εξέταση πληθυσμού.

- **Δείγμα**: Το υποσύνολο του πληθυσμού το οποίο θα μελετηθεί προκειμένου να εξαχθούν τα γενικότερα συμπεράσματα για τον πληθυσμό.

Συχνότητες

Το άθροισμα των **Συχνοτήτων $\boldsymbol{n_i}$** όλων των τιμών μιας μεταβλητής X με τιμές $x_1, x_2, \dots, x_k$ είναι ίσο με το **Μέγεθος του Δείγματος $\boldsymbol{n}$**, δηλαδή:

$$\sum_{i=1}^{k} n_i = n$$

Η **Σχετική Συχνότητα** $(\boldsymbol{f_i})$ ισούται με:

$$f_i = \frac{n_i}{n}$$

και ισχύει ότι:

$$\sum_{i=1}^{k} f_i = 1$$

<u>Σημείωση</u>: Η επί της εκατό σχετική συχνότητα $(f_i\%)$ ορίζεται ως $f_i \cdot 100$.

Ως **Αθροιστική Συχνότητα** $(\boldsymbol{N_i})$ της τιμής x_i ορίζεται το άθροισμα των συχνοτήτων n_i όλων των τιμών που είναι μικρότερες ή ίσες της x_i.

Επίσης ισχύει ότι:

$$n_1 = N_1, n_2 = N_2 - N_1, \dots, n_k = N_k - N_{k-1}$$

Ως **Αθροιστική Σχετική Συχνότητα** $(\boldsymbol{F_i})$ της τιμής x_i ορίζεται το άθροισμα των σχετικών συχνοτήτων όλων των τιμών που είναι μικρότερες ή ίσες της x_i.

Επίσης ισχύει ότι:

$$f_1 = F_1, f_2 = F_2 - F_1, \dots, f_k = F_k - F_{k-1}$$

ΦΟΙΤΗΤΙΚΟ ΠΡΟΣΗΜΟ

<u>Σημείωση</u>: Η αθροιστική σχετική συχνότητα (F_i) είναι το πηλίκο της αθροιστικής συχνότητας προς το μέγεθος του δείγματος:

$$\boldsymbol{F_i = \frac{N_i}{n}}$$

<u>Σημείωση</u>: Η επί της εκατό αθροιστική σχετική συχνότητα $(F_i\%)$ ορίζεται ως $F_i \cdot 100$.

Ομαδοποίηση Παρατηρήσεων

Κανόνας Sturges: Ο ιδανικός **Αριθμός των Κλάσεων** $(\boldsymbol{c})$ μπορεί να βρεθεί βάσει του παρακάτω τύπου:

$$\boldsymbol{c = 1 + 3.322 \cdot \log n}$$

Βήματα Ομαδοποίησης Παρατηρήσεων

1. Εύρεση του ιδανικού αριθμού κλάσεων (c).

2. Προσδιορισμός του **Πλάτους των Κλάσεων** $(\boldsymbol{k})$. Πλάτος ονομάζεται η διαφορά του κατώτερου από το ανώτερο όριο της κλάσης. Αν θεωρηθεί ως **R** το **Εύρος** του δείγματος $(R = x_{max} - x_{min})$, τότε το πλάτος των κλάσεων ισούται με:
$$\boldsymbol{k = \frac{R}{c}}$$

3. Κατασκευή των κλάσεων. Ξεκινάμε από τη x_{min} και προσθέτουμε κάθε φορά το πλάτος (k), δημιουργώντας τις c κλάσεις.

Περιπτώσεις Δεδομένων

Υπάρχουν οι εξής τρεις περιπτώσεις δεδομένων:

A. Πρωτογενείς παρατηρήσεις
B. Παρατηρήσεις τοποθετημένες σε πίνακα συχνοτήτων χωρίς κλάσεις
C. Παρατηρήσεις τοποθετημένες σε πίνακα συχνοτήτων με κλάσεις

<u>Σημείωση</u>: Ανάλογα με την εκάστοτε περίπτωση χρησιμοποιείται ο αντίστοιχος τύπος για το κάθε μέτρο θέσης ή διασποράς.

Μέτρα Θέσης ή Κεντρικής Τάσης

Τα **Μέτρα Θέσης** ή **Κεντρικής Τάσης** περιγράφουν τη θέση της κατανομής από την οποία προέρχονται τα δεδομένα του δείγματος.

- **Αριθμητικός Μέσος** $(\boldsymbol{\bar{x}})$: Για την περίπτωση Α, ο αριθμητικός μέσος ισούται με:

$$\boldsymbol{\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i}$$

ΦΟΙΤΗΤΙΚΟ ΠΡΟΣΗΜΟ

Για περιπτώσεις δεδομένων Β και C, ο τύπος μετατρέπεται σε:

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} n_i \cdot x_i \quad \text{ή} \quad \bar{x} = \sum_{i=1}^{n} f_i x_i$$

- **Γεωμετρικός Μέσος** $(\bar{x}_G)$:

$$\bar{x}_G = \sqrt[n]{x_1 \cdot x_2 \cdot \ldots \cdot x_n}$$

- **Αρμονικός Μέσος** $(\bar{x}_H)$**:**

$$\bar{x}_H = \frac{1}{\sum_{i=1}^{n} \frac{1}{x_i}}$$

<u>Σημείωση</u>: Γενικά ισχύει:

$$\bar{x}_H < \bar{x}_G < \bar{x}$$

- **Σταθμικός Μέσος** $(\bar{x}_w)$**:** Αν οι τιμές του δείγματος $x_1, x_2, \ldots, x_n$ έχουν διαφορετική βαρύτητα $w_1, w_2, \ldots, w_n$, τότε ο σταθμικός μέσος υπολογίζεται ως:

$$\bar{x}_w = \frac{\sum_{i=1}^{n} x_i w_i}{\sum_{i=1}^{n} w_i}$$

- **Επικρατούσα Τιμή** (M_0) ή **Κορυφή** (T): Για τις περίπτωση δεδομένων Α και Β αποτελεί την τιμή της μεταβλητής που εμφανίζει την υψηλότερη συχνότητα.

  Για την περίπτωση δεδομένων C, ο τύπος μετατρέπεται σε:

$$M_0 = L.B. + k\,\frac{D_1}{D_1 + D_2}$$

  όπου $L.B.$: το κάτω όριο της κλάσης, k: το πλάτος της κλάσης, $D_1 = n_i - n_{i-1}$ και $D_2 = n_i - n_{i+1}$

- **Διάμεσος** (M): Για τις περίπτωση δεδομένων Α και Β, ισχύει ότι:

  Αν n είναι περιττός αριθμός, η Μ ισούται με τη μεσαία διατεταγμένη παρατήρηση. Αν n είναι άρτιος αριθμός, η Μ ισούται με το μέσο όρο των δυο μεσαίων διατεταγμένων παρατηρήσεων.

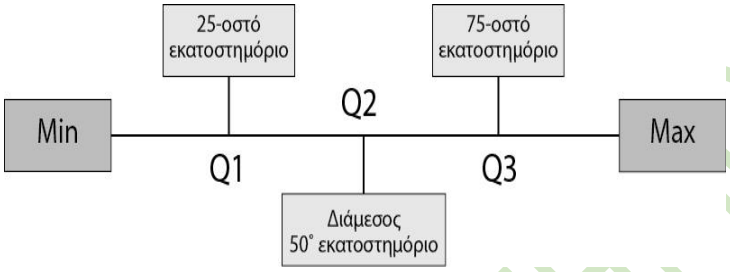
  Για την περίπτωση δεδομένων C, ο τύπος μετατρέπεται σε:

ΦΟΙΤΗΤΙΚΟ ΠΡΟΣΗΜΟ

$$M = L.B. + k\,\frac{\frac{1}{2}\,n - N_{i-1}}{n_i}$$

όπου N_{i-1}: η αθροιστική συχνότητα της προηγούμενης κλάσης

- **Ποσοστημόρια**: Τα συγκεκριμένα μέτρα θέσης χωρίζονται στις εξής δυο κατηγορίες:
  I. **Τεταρτημόρια** (Q_i)
  II. **Εκατοστημόρια** (P_i)



<u>Σημείωση</u>: Προκειμένου να υπολογίσουμε οποιοδήποτε ποσοστημόριο θα πρέπει πρωτίστως να διατάξουμε τις παρατηρήσεις σε αύξουσα σειρά.

- **1ο Τεταρτημόριο** (Q_1) ή **25ο Εκατοστημόριο** (P_{25}): ορίζουμε την τιμή κάτω από την οποία βρίσκεται το 25% των παρατηρήσεων και πάνω από την οποία βρίσκεται το υπόλοιπο 75% των παρατηρήσεων.

  Για τις περίπτωση δεδομένων Α και Β, ισχύει ότι:

  $$Q_1 \text{ ή } P_{25} = x_{0.25(n+1)}$$

  Για την περίπτωση δεδομένων C, ο τύπος μετατρέπεται σε:

  $$Q_1 \text{ ή } P_{25} = L.B. + k\,\frac{0.25n - N_{i-1}}{n_i}$$

- **2ο Τεταρτημόριο** (Q_2) ή **50ο Εκατοστημόριο** (P_{50}) ή **Διάμεσος** (M): ορίζουμε την τιμή κάτω από την οποία βρίσκεται το 50% των παρατηρήσεων και πάνω από την οποία βρίσκεται το υπόλοιπο 50% των παρατηρήσεων.

  Για τις περίπτωση δεδομένων Α και Β, ισχύει ότι:

  $$Q_2 \text{ ή } P_{50} \text{ ή } M = x_{0.5(n+1)}$$

  Για την περίπτωση δεδομένων C, ο τύπος μετατρέπεται σε:

  $$Q_1 \text{ ή } P_{25} \text{ ή } M = L.B. + k\,\frac{0.5n - N_{i-1}}{n_i}$$

- **3ο Τεταρτημόριο** (Q_3) ή **75ο Εκατοστημόριο** (P_{75}): ορίζουμε την τιμή κάτω από την οποία βρίσκεται το 75% των παρατηρήσεων και πάνω από την οποία βρίσκεται το υπόλοιπο 25% των παρατηρήσεων.

ΦΟΙΤΗΤΙΚΟ ΠΡΟΣΗΛΟ

Για τις περίπτωση δεδομένων Α και Β, ισχύει ότι:

$$\boldsymbol{Q_3} \text{ ή } \boldsymbol{P_{75}} = \boldsymbol{x_{0.75(n+1)}}$$

Για την περίπτωση δεδομένων C, ο τύπος μετατρέπεται σε:

$$\boldsymbol{Q_3} \text{ ή } \boldsymbol{P_{75}} = \boldsymbol{L.B.} + \boldsymbol{k}\,\frac{\boldsymbol{0.75n - N_{i-1}}}{\boldsymbol{n_i}}$$

- **kᵒ Εκατοστημόριο** $(\boldsymbol{P_k})$: ορίζουμε την τιμή κάτω από την οποία βρίσκεται το $k\%$ των παρατηρήσεων και πάνω από την οποία βρίσκεται το υπόλοιπο $(1-k)\%$ των παρατηρήσεων.

  Για τις περίπτωση δεδομένων Α και Β, ισχύει ότι:

$$\boldsymbol{P_k} = \boldsymbol{x_{k\%(n+1)}}$$

  Για την περίπτωση δεδομένων C, ο τύπος μετατρέπεται σε:

$$\boldsymbol{P_k} = \boldsymbol{L.B.} + \boldsymbol{k}\,\frac{\boldsymbol{k\% \cdot n - N_{i-1}}}{\boldsymbol{n_i}}$$

Μέτρα Διασποράς

Τα **Μέτρα Διασποράς** εκφράζουν τις αποκλίσεις των τιμών μιας μεταβλητής γύρω από τα μέτρα κεντρικής τάσης.

- **Εύρος** $(\boldsymbol{R})$

$$\boldsymbol{R} = \boldsymbol{x_{max}} - \boldsymbol{x_{min}}$$

- **Ενδοτεταρτημοριακό Εύρος** $(\boldsymbol{IQR})$

$$\boldsymbol{IQR} = \boldsymbol{Q_3} - \boldsymbol{Q_1}$$

  <u>Σημείωση</u>: Το ημι – ενδοτεταρτημοριακό εύρος ισούται με:

$$\frac{IQR}{2}$$

- **Διακύμανση** ή **Διασπορά** $(\boldsymbol{s^2})$: Αν έχουμε δείγμα μεγέθους n και η περίπτωση δεδομένων είναι η Α, τότε έχουμε:

$$\boldsymbol{s^2} = \frac{\sum_{i=1}^{n}(\boldsymbol{x_i} - \overline{\boldsymbol{x}})^2}{\boldsymbol{n-1}} \quad \text{ή} \quad \boldsymbol{s^2} = \frac{\sum_{i=1}^{n}\boldsymbol{x_i^2} - \frac{(\sum_{i=1}^{n}\boldsymbol{x_i})^2}{\boldsymbol{n}}}{\boldsymbol{n-1}} \quad \text{ή} \quad \boldsymbol{s^2} = \frac{\sum_{i=1}^{n}\boldsymbol{x_i^2} - \boldsymbol{n}\overline{\boldsymbol{x}}^2}{\boldsymbol{n-1}}$$

  Για περιπτώσεις δεδομένων Β και C, οι παραπάνω τύποι μετατρέπονται ως εξής:

ΦΟΙΤΗΤΙΚΟ ΠΡΟΣΗΜΟ

$$s^2 = \frac{\sum_{i=1}^{n} n_i (x_i - \overline{x})^2}{n-1} \quad \text{ή} \quad s^2 = \frac{\sum_{i=1}^{n} n_i {x_i}^2 - \frac{(\sum_{i=1}^{n} n_i x_i)^2}{n}}{n-1} \quad \text{ή} \quad s^2 = \frac{\sum_{i=1}^{n} n_i {x_i}^2 - n\overline{x}^2}{n-1}$$

<u>Σημείωση</u>: Στην περίπτωση που έχουμε πεπερασμένο πληθυσμό μεγέθους N, στον παρανομαστή όλων των παραπάνω κλασμάτων έχουμε N αντίο για $n-1$.

- **Τυπική Απόκλιση** (s): Σε κάθε μια από τις παραπάνω περιπτώσεις των τύπων υπολογισμού της διακύμανσης, η τυπική απόκλιση ισούται με:

$$s = \sqrt{s^2}$$

- **Μέση Απόκλιση** (MD): Για την περίπτωση Α, η μέση απόκλιση ισούται με:

$$MD = \frac{1}{n}\sum_{i=1}^{n} |x_i - \overline{x}|$$

Για περιπτώσεις δεδομένων Β και C, έχουμε:

$$MD = \frac{1}{n}\sum_{i=1}^{n} n_i \cdot |x_i - \overline{x}|$$

- **Συντελεστής Μεταβλητότητας** (CV)

$$CV = \frac{s}{|\overline{x}|} 100\%$$

Ο CV μας δείχνει πόσο ομοιογενές είναι το δείγμα. Εμπειρικά, αν $CV \leq 10\%$, δεχόμαστε ότι το δείγμα είναι ομοιογενές.

Εμπειρικός Κανόνας Συμμετρίας

Αν το Ιστόγραμμα των παρατηρήσεων έχει σχήμα καμπάνας (περίπου συμμετρικό) ισχύουν τα εξής:

- Στο διάστημα $\overline{x} \pm s$ βρίσκεται περίπου το 68% των παρατηρήσεων.
- Στο διάστημα $\overline{x} \pm 2s$ βρίσκεται περίπου το 95% των παρατηρήσεων.
- Στο διάστημα $\overline{x} \pm 3s$ βρίσκεται περίπου το 99.7% των παρατηρήσεων.

Θεώρημα Chebyshev

Για οποιοδήποτε σύνολο παρατηρήσεων, ανεξάρτητα από τη μορφή της κατανομής τους, το ποσοστό των τιμών που βρίσκεται σε απόσταση $\pm k$ τυπικών αποκλίσεων από το μέσο είναι τουλάχιστον, $1 - \frac{1}{k^2}$, όπου $k > 1$.

ΦΟΙΤΗΤΙΚΟ ΠΡΟΣΗΜΟ

Μετασχηματισμός Δεδομένων

Αν $y_i = ax_i + b$, τότε:

- $\overline{y} = a\overline{x} + b$
- $s_y^2 = a^2 s_x^2$
- $s_y = |a| s_x$
- $M_y = aM_x + b$
- $M_{0y} = aM_{0x} + b$
- Αν $a > 0$, τότε $Q_{y_i} = aQ_{x_i} + b, \quad i = 1,3$

Ροπές

Η **ροπή τάξης r** γύρω από το **σημείο x_0** μιας κατανομής δίνεται από τον εξής τύπο (σε παρενθέσεις αναφέρονται οι πιθανές περιπτώσεις δεδομένων):

$$\frac{\sum_{i=1}^{n}(x_i - x_0)^r}{n} \quad (A) \qquad\qquad \frac{\sum_{i=1}^{n} n_i(x_i - x_0)^r}{n} \quad (B, C)$$

- Αν $x_0 = 0$, τότε έχουμε τη **Ροπή γύρω από την Αρχή τάξης r**:

$$\nu_r = \frac{\sum_{i=1}^{n} x_i^r}{n} \quad (A) \qquad\qquad \nu_r = \frac{\sum_{i=1}^{n} n_i x_i^r}{n} \quad (B, C)$$

- Αν $x_0 = \overline{x}$, τότε έχουμε την **Κεντρική Ροπή τάξης r**:

$$\mu_r = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^r}{n} \quad (A) \qquad\qquad \mu_r = \frac{\sum_{i=1}^{n} n_i(x_i - \overline{x})^r}{n} \quad (B, C)$$

Μέτρα Ασυμμετρίας

Εμπειρικός Κανόνας

- Αν $\bar{x} = M = \mathrm{M}_0 \Rightarrow$ **Συμμετρία**
- Αν $\bar{x} > M > \mathrm{M}_0 \Rightarrow$ **Θετική Ασυμμετρία**
- Αν $\bar{x} < M < \mathrm{M}_0 \Rightarrow$ **Αρνητική Ασυμμετρία**

Συντελεστής Ασυμμετρίας Βάσει Κεντρικών Ροπών

$$\alpha_3 = \frac{\mu_3}{(\sqrt{\mu_2})^3} = \frac{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^3}{n}}{\left[\sqrt{\frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n}}\right]^3}$$

- Αν $\alpha_3 = 0 \Rightarrow$ **Συμμετρία**
- Αν $\alpha_3 > 0 \Rightarrow$ **Θετική Ασυμμετρία**
- Αν $\alpha_3 < 0 \Rightarrow$ **Αρνητική Ασυμμετρία**

ΦΟΙΤΗΤΙΚΟ ΠΡΟΣΗΜΟ

Συντελεστής Ασυμμετρίας του Bowley

$$S_k = \frac{(Q_3 - M) - (M - Q_1)}{Q_3 - Q_1}$$

- Αν $S_k = 0 \Rightarrow$ **Συμμετρία**
- Αν $0 < S_k < 1 \Rightarrow$ **Θετική Ασυμμετρία**
- Αν $-1 < S_k < 0 \Rightarrow$ **Αρνητική Ασυμμετρία**

Συντελεστής Ασυμμετρίας του Pearson

$$\gamma_2 = \frac{3(\bar{x} - M)}{s}$$

- Αν $\gamma_2 = 0 \Rightarrow$ **Συμμετρία**
- Αν $\gamma_2 > 0 \Rightarrow$ **Θετική Ασυμμετρία**
- Αν $\gamma_2 < 0 \Rightarrow$ **Αρνητική Ασυμμετρία**

Μέτρα Κύρτωσης

Συντελεστής Κύρτωσης Βάσει Κεντρικών Ροπών

$$\alpha_4 = \frac{\mu_4}{(\mu_2)^2} = \frac{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{n}}{\left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}\right]^2}$$

- Αν $\alpha_4 = 3 \Rightarrow$ **Μεσόκυρτη Κατανομή**
- Αν $\alpha_4 > 3 \Rightarrow$ **Λεπτόκυρτη Κατανομή**
- Αν $\alpha_4 < 3 \Rightarrow$ **Πλατύκυρτη Κατανομή**

ΦΟΙΤΗΤΙΚΟ ΠΡΟΣΗΜΟ