10ο ΜΑΘΗΜΑ

ΕΝΟΤΗΤΑ 2.3.

ΜΕΤΡΑ ΘΕΣΗΣ

ΜΕΤΡΑ ΔΙΑΣΠΟΡΑΣ

Σκοπός:

Στην ενότητα αυτή παρουσιάζονται τα μέτρα θέσης και τα μέτρα διασποράς. Ο ορισμός τους και διάφοροι μέθοδοι υπολογισμού. Γίνεται επίσης αναφορά στα μέτρα ασυμμετρίας.

Προσδοκώμενα αποτελέσματα:

Όταν μελετήσετε αυτή την ενότητα, θα πρέπει:

- Να διατυπώνετε τους ορισμούς των μέτρων θέσης και διασποράς που μας ενδιαφέρουν στη συγκεκριμένη μελέτη.
- Να βρίσκετε (υπολογίζετε) τα μέτρα θέσης και διασποράς που σας ζητούνται.
- Να βρίσκετε το συντελεστή μεταβολής και να εξετάζετε αν ένα δείγμα είναι ομοιογενές.
- Να συγκρίνετε δυο διαφορετικά δείγματα ως προς την ομοιογένεια.
- Να βρίσκετε τα μέτρα θέσης και διασποράς μιας μεταβλητής μετά από γραμμικό μετασχηματισμό.

Συμβολισμοί:

Το άθροισμα:

$$x_1 + x_2 + ... + x_\nu$$

μπορεί να γραφεί εν' συντομία:

$$\boxed{\sum_{i=1}^{\nu} x_i}$$

Παράδειγμα:

$$x_1 + x_2 + \ldots + x_{10} = \sum_{i=1}^{10} x_i$$

Έτσι μπορούμε να γράφουμε συνοπτικά και άλλες μορφές αθροισμάτων, όπως:

- $x_1\nu_1 + x_2\nu_2 + \ldots + x_\kappa\nu_\kappa = \sum_{i=1}^{\kappa} x_i\nu_i$
- $\nu_1(x_1 - x)^2 + \nu_2(x_2 - x)^2 + \ldots + \nu_\kappa(x_\kappa - x)^2 = \sum_{i=1}^{\kappa} \nu_i(x_i - x)^2$
- $\lambda x_1 + \lambda x_2 + \ldots + \lambda x_\kappa = \sum_{i=1}^{\kappa} \lambda x_i = \lambda \sum_{i=1}^{\kappa} x_i$
- $(x_1 + y_1) + (x_2 + y_2) + \ldots + (x_\kappa + y_\kappa) = \sum_{i=1}^{\kappa}(x_i + y_i) = \sum_{i=1}^{\kappa} x_i + \sum_{i=1}^{\kappa} y_i$
- $(\lambda x_1 + \mu y_1) + (\lambda x_2 + \mu y_2) + \ldots + (\lambda x_\kappa + \mu y_\kappa) = \sum_{i=1}^{\kappa}(\lambda x_i + \mu y_i) =$

$$= \lambda \sum_{i=1}^{\kappa} x_i + \mu \sum_{i=1}^{\kappa} y_i$$

Πως ορίζεται η μέση τιμή ν παρατηρήσεων;

Απάντηση:

Η μέση τιμή των παρατηρήσεων $x_1, x_2, \ldots, x_\nu$ είναι ο αριθμός:

$$\overline{x} = \frac{x_1 + x_2 + \ldots + x_\nu}{\nu}$$

ή

$$\overline{x} = \frac{\sum_{i=1}^{\nu} x_i}{\nu} = \frac{1}{\nu} \cdot \sum_{i=1}^{\nu} x_i$$

Πως ορίζεται ο σταθμικός μέσος των παρατηρήσεων $x_1,\ x_2,\ldots,x_\kappa$ **με συντελεστές βαρύτητας** $w_1,\ w_2,\ldots,w_\kappa$ **αντίστοιχα;**

Απάντηση:

Ο σταθμικός μέσος των παρατηρήσεων είναι ο αριθμός:

$$\boxed{\bar{x} = \frac{x_1 \cdot w_1 + x_2 \cdot w_2 + \ldots + x_\nu \cdot w_\nu}{w_1 + w_2 + \ldots + w_\nu}}$$

ή

$$\boxed{\bar{x} = \frac{\sum_{i=1}^{\kappa} x_i \cdot w_i}{\sum_{i=1}^{\nu} w_i}}$$

Παρατηρήσεις:

1. Όταν υπάρχει κατανομή συχνοτήτων:

$$(x_i, \nu_i), \quad i = 1, 2, \ldots, \kappa$$

ή κατανομή σχετικών συχνοτήτων:

$$(x_i, f_i), \quad i = 1, 2, \ldots, \kappa$$

η μέση τιμή βρίσκεται με τη χρήση των τύπων:

$$\boxed{\bar{x} = \frac{\sum_{i=1}^{\kappa} x_i \cdot \nu_i}{\nu} = \frac{1}{\nu}\sum_{i=1}^{\kappa} x_i \cdot \nu_i} \qquad (1)$$

και

$$\boxed{\bar{x} = \sum_{i=1}^{\kappa} x_i \cdot f_i} \qquad (2)$$

2. Στην περίπτωση που η μεταβλητή είναι διακριτή, οι τύποι (1) και (2) δίνουν τη μέση τιμή όπως αυτή θα προέκυπτε με βάση τον ορισμό της. Στην περίπτωση όμως που τα δεδομένα έχουν ομαδοποιηθεί η μέση τιμή όπως υπολογίζεται από τους τύπους (1) και (2) συνήθως δεν είναι ίση με την τιμή που θα παίρναμε από την εφαρμογή του τύπου του ορισμού. Αυτό οφείλεται στο ότι το x_i στους τύπους (1) και (2) είναι η κεντρική τιμή της αντίστοιχης κλάσης η οποία μπορεί και να μην είναι τιμή του δείγματος.

3. Ο σταθμικός μέσος μπορεί να χρησιμοποιηθεί για τον προσδιορισμό της μέσης τιμής περισσοτέρων ομάδων ομοειδών δεδομένων με διαφορετικό μέγεθος των οποίων γνωρίζουμε τις μέσες τιμές.

Παράδειγμα: Σε ένα λύκειο η Α΄ λυκείου έχει τρία τμήματα με 25, 20 και 30 παιδιά. Αν η μέση βαθμολογία στο μάθημα της Ιστορίας είναι 17, 16 και 15 ανά τμήμα αντίστοιχα, τότε η μέση βαθμολογία στο μάθημα της ιστορίας για όλη την Α΄ λυκείου, είναι:

$$\overline{x}=\frac{25\cdot 17+20\cdot 16+30\cdot 15}{25+20+30}=\frac{425+320+450}{75}$$

άρα:

$$\boxed{\overline{x}=\frac{1195}{75}\simeq 15,93}$$

◈ **Να δικαιολογήσετε τον τύπο:**

$$\overline{x}=\sum_{i=1}^{\kappa} x_i\cdot f_i$$

Είναι ο τύπος (2) της προηγούμενης παρατήρησης.

Απάντηση:

Στην κατανομή συχνοτήτων:

$$(x_i,\nu_i),\qquad i=1,2,\ldots\kappa$$

ισχύει:

$$\overline{x}=\frac{1}{\nu}\cdot\sum_{i=1}^{\kappa} x_i\cdot \nu_i$$

Άρα:

$$\overline{x}=\frac{x_1\cdot\nu_1+x_2\cdot\nu_2+\ldots+x_\kappa\cdot\nu_\kappa}{\nu}=\frac{x_1\cdot\nu_1}{\nu}+\frac{x_2\cdot\nu_2}{\nu}+\ldots+\frac{x_\kappa\cdot\nu_\kappa}{\nu}=$$

$$=x_1\cdot\frac{\nu_1}{\nu}+x_2\cdot\frac{\nu_2}{\nu}+\ldots+x_\kappa\cdot\frac{\nu_\kappa}{\nu}=x_1\cdot f_1+x_2\cdot f_2+\ldots+x_\kappa\cdot f_\kappa$$

Επειδή είναι:

$$f_i=\frac{\nu_i}{\nu},\qquad i=1,2,\ldots,\kappa$$

έχουμε:

$$\boxed{\overline{x}=\sum_{i=1}^{\kappa} x_i\cdot f_i}$$

◈ **Πως ορίζεται η διάμεσος ενός δείγματος ν παρατηρήσεων;**

Απάντηση:

Κατ' αρχάς πρέπει να διατάξουμε τις παρατηρήσεις σε αύξουσα σειρά. Από τις διατεταγμένες παρατηρήσεις, **διάμεσος** (δ) είναι η μεσαία παρατήρηση αν το πλήθος ν των παρατηρήσεων είναι περιττός αριθμός ή το ημιάθροισμα των δυο μεσαίων παρατηρήσεων αν το πλήθος ν είναι άρτιος αριθμός.

Παρατηρήσεις:

1. Η διάμεσος χωρίζει ένα διατεταγμένο δείγμα παρατηρήσεων σε δυο ίσες ομάδες. Η πρώτη ομάδα περιέχει τις παρατηρήσεις που είναι μικρότερες από την διάμεσο και το πλήθος δεν υπερβαίνει το 50% του συνόλου των παρατηρήσεων. Το ίδιο συμβαίνει και με την άλλη ομάδα, που περιέχει τις παρατηρήσεις που είναι μεγαλύτερες από τη διάμεσο.

2. Σε ομαδοποιημένα δεδομένα η διάμεσος βρίσκεται με τη βοήθεια των πολυγώνων αθροιστικών συχνοτήτων ή αθροιστικών σχετικών συχνοτήτων. Πιο συγκεκριμένα στο διάγραμμα αθροιστικών σχετικών συχνοτήτων $(F_i\%)$ η διάμεσος είναι η τετμημένη, σημείο του οριζόντιου άξονα, η οποία αντιστοιχεί στο σημείο του κατακόρυφου άξονα με τιμή 50%.

Αν το σημείο του οριζόντιου άξονα που δίνει τη διάμεσο δεν φαίνεται καθαρά και οι παρατηρήσεις σε κάθε κλάση είναι κατανεμημένες ομοιόμορφα η διάμεσος δίνεται από τον τύπο:

$$\boxed{\delta = L_i + \frac{\frac{\nu}{2} - N_{i-1}}{\nu_i}\cdot c}$$

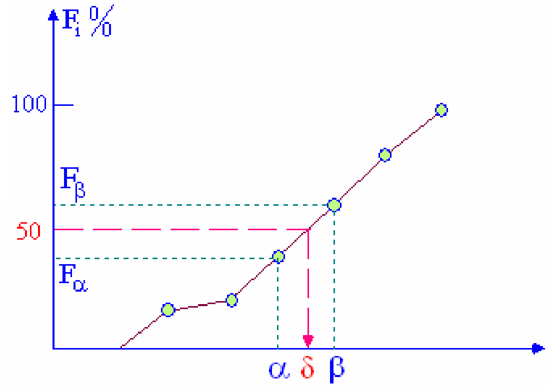
Όπου:

- L_i: Είναι το κάτω όριο της κλάσης που περιέχει τη διάμεσο.
- ν: Είναι το μέγεθος του δείγματος.
- N_{i-1}: Είναι η αθροιστική συχνότητα της προηγούμενης κλάσης (από αυτήν που περιέχει τη διάμεσο)
- ν_i: Είναι η συχνότητα της κλάσης που περιέχει τη διάμεσο.
- c: Είναι το πλάτος των κλάσεων.

Ο πιο πάνω τύπος προκύπτει από τις αναλογίες που περιγράψαμε στην προηγούμενη παράγραφο.

Πιο απλά:

Έστω ότι το πολύγωνο αθροιστικών συχνοτήτων είναι το παρακάτω:



Η διάμεσος δ προκύπτει από την αναλογική σχέση:

$$\boxed{\frac{\delta-\alpha}{\beta-\alpha}=\frac{50-F_\alpha}{F_\beta-F_\alpha}}$$

όπου F_α και F_β οι αθροιστικές σχετικές συχνότητες, στο άξονα $F_i\%$, που αντιστοιχούν στα α, β που είναι τα όρια της κλάσης στην οποία ανήκει η διάμεσος.

3. Σύγκριση μέσης τιμής και διαμέσου

Μέση τιμή

Πλεονεκτήματα

- Για τον υπολογισμό της χρησιμοποιούνται όλες οι τιμές
- Είναι μοναδική για κάθε σύνολο δεδομένων.
- Είναι εύκολα κατανοητή.
- Υπολογίζεται εύκολα.
- Έχει μεγάλη εφαρμογή για περαιτέρω στατιστική ανάλυση.

Μειονεκτήματα

- Επηρεάζεται πολύ από ακραίες τιμές
- Μπορεί να μην αντιστοιχεί σε δυνατή τιμή της μεταβλητής. Όταν η μεταβλητή x παίρνει ακέραιες τιμές η μέση τιμή μπορεί να μην είναι ακέραιος αριθμός.
- Δεν υπολογίζεται σε ποιοτικά δεδομένα
- Υπολογίζεται δύσκολα σε ομαδοποιημένα δεδομένα με ανοικτές τις ακραίες κλάσεις.

Διάμεσος

Πλεονεκτήματα

- Είναι εύκολα κατανοητή.
- Δεν επηρεάζεται από ακραίες τιμές
- Υπολογίζεται και στην περίπτωση που οι ακραίες κλάσεις είναι ανοικτές.
- Ο υπολογισμός της είναι απλός.
- Είναι μοναδική για κάθε σύνολο δεδομένων.

Μειονεκτήματα

- Δεν χρησιμοποιούνται όλες οι τιμές για τον υπολογισμό της.
- Είναι δύσκολη η εφαρμογή της για περαιτέρω στατιστική ανάλυση
- Δεν υπολογίζεται για ποιοτικά δεδομένα.
- Για τον υπολογισμό της μπορεί να χρειαστεί παρεμβολή.