11^{o} ΜΑΘΗΜΑ

ΜΕΤΡΑ ΔΙΑΣΠΟΡΑΣ

Τι ονομάζουμε μέτρα διασποράς μιας κατανομής;
Ποια είναι τα μέτρα διασποράς;

Απάντηση:

Τα μέτρα διασποράς είναι τα αριθμητικά μεγέθη που μας δίνουν την διασπορά των παρατηρήσεων γύρω από τις κεντρικές τιμές της κατανομής.

Τα μέτρα διασποράς είναι το **εύρος** η **διακύμανση** η **τυπική απόκλιση** και το **ενδοτεταρτημοριακό εύρος.**

Πως ορίζεται το εύρος;

Απάντηση:

Το εύρος (ή κύμανση) R ορίζεται ως η διαφορά της μικρότερης παρατήρησης από τη μεγαλύτερη:

$$\boxed{R = \text{μεγαλύτερη παρατήρηση - μικρότερη παρατήρηση}}$$

Παρατήρηση:

Σε ομαδοποιημένα δεδομένα το εύρος της κατανομής δίνεται από τη διαφορά του κατώτερου ορίου της πρώτης κλάσης από το ανώτερο όριο της τελευταίας κλάσης. Σ' αυτή την περίπτωση ίσως υπάρχει διαφορά μεταξύ του εύρους πριν την ομαδοποίηση και του εύρους μετά την ομαδοποίηση.

Πως ορίζεται η διακύμανση των παρατηρήσεων $t_1, t_2, \ldots, t_\nu$ **μιας μεταβλητής** x **;**

Απάντηση:

Η διακύμανση (ή διασπορά) s^2 ορίζεται ως η μέση τιμή των τετραγώνων των αποκλίσεων των τιμών t_i από τη μέση τιμή τους $\bar{x}$. Δηλαδή:

$$\boxed{s^2 = \frac{1}{\nu}\sum(t_i - \bar{x})^2 = \frac{(t_1-\bar{x})^2 + (t_2-\bar{x})^2 + \ldots + (t_\nu - \bar{x})^2}{\nu}}$$

Παρατηρήσεις:

1. Αποδεικνύεται ότι η διακύμανση ν παρατηρήσεων $t_1, t_2, \ldots, t_\nu$ μπορεί να πάρει την ισοδύναμη μορφή:

$$s^2 = \frac{1}{\nu}\left\{\sum_{i=1}^{\nu} t_i^2 - \frac{\left(\sum_{i=1}^{\nu} t_i\right)^2}{\nu}\right\}$$

που διευκολύνει τους υπολογισμούς όταν η μέση τιμή $\bar{x}$ δεν είναι ακέραιος αριθμός.

2. Όταν έχουμε κατανομή συχνοτήτων:

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

ή κατανομή σχετικών συχνοτήτων:

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

σε διακριτή μεταβλητή ή ομαδοποιημένα δεδομένα η διακύμανση δίνεται από τη σχέση:

$$s^2 = \frac{1}{\nu}\sum_{i=1}^{\kappa}(x_i - \bar{x})^2 \cdot \nu_i \qquad (1)$$

Στα ομαδοποιημένα δεδομένα οι τιμές x_i είναι τα κέντρα των κλάσεων.

Ο τύπος (1) μπορεί να μετασχηματιστεί στους τύπους:

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

όταν δίνονται οι σχετικές συχνότητες f_i.

ή

$$s^2 = \frac{1}{\nu}\left\{\sum_{i=1}^{\kappa} x_i^2 \nu_i - \frac{\left(\sum_{i=1}^{\kappa} x_i \nu_i\right)^2}{\nu}\right\} \qquad (3)$$

Ο τύπος (2) επίσης μπορεί να μετασχηματιστεί στον τύπο:

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

Πως ορίζεται η τυπική απόκλιση s ενός δείγματος;

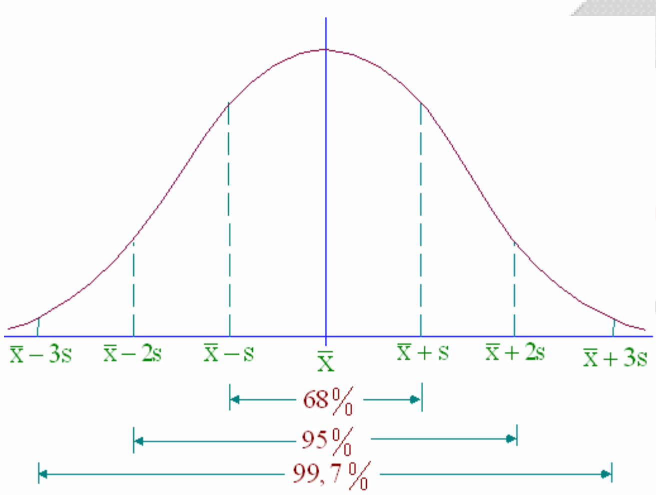
Απάντηση:

Η **τυπική απόκλιση** ορίζεται ως η θετική τετραγωνική ρίζα της διακύμανσης:

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

Παρατηρήσεις:

1. Αν μια κατανομή είναι κανονική ή περίπου κανονική η τυπική απόκλιση s έχει τις παρακάτω ιδιότητες:



- Το 68% των παρατηρήσεων βρίσκεται στο διάστημα: $\boxed{(\bar{x}-s, \bar{x}+s)}$
- Το 95% των παρατηρήσεων βρίσκεται στο διάστημα: $\boxed{(\bar{x}-2s, \bar{x}+2s)}$
- Το $99{,}7\%$ των παρατηρήσεων βρίσκεται στο διάστημα: $\boxed{(\bar{x}-3s, \bar{x}+3s)}$
- Το εύρος είναι περίπου ίσο με έξι τυπικές αποκλίσεις: $\boxed{R \simeq 6s}$

Με βάση τα παραπάνω ποσοστά, σε ασκήσεις που αναφέρονται σε κανονική κατανομή θα ζητηθεί να εκτιμήσουμε το ποσοστό των παρατηρήσεων που αναμένονται σε κάποιο διάστημα που θα είναι γραμμικός συνδυασμός των $\bar{x}$ και s.

Παράδειγμα:

Σε μια κανονική κατανομή (ή περίπου κανονική) το 50% των παρατηρήσεων είναι μικρότερες του $\bar{x}$.

Στο διάστημα $(\bar{x}, \bar{x}+s)$ θα βρίσκεται περίπου το 34% των παρατηρήσεων $\left(\frac{1}{2}\cdot 68\%\right)$

Στο διάστημα $(\bar{x}-2s, \bar{x}+s)$ αναμένεται να βρίσκεται το $81,5\%$ των παρατηρήσεων.

$\frac{1}{2}\cdot 95\%$ στο $(\bar{x}-2s, \bar{x})$ και 34% στο $(\bar{x}, \bar{x}+s)$

2. Σύγκριση μέτρων διασποράς

Εύρος

Πλεονεκτήματα

- Είναι πολύ απλό στον υπολογισμό.
- Χρησιμοποιείται στον έλεγχο ποιότητας.
- Μπορεί να χρησιμοποιηθεί για την εκτίμηση της τυπικής απόκλισης.

Μειονεκτήματα

- Δεν θεωρείται αξιόπιστο αφού υπολογίζεται μόνο από τις δυο ακραίες τιμές.
- Δε χρησιμοποιείται για περαιτέρω στατιστική ανάλυση.

Διασπορά και τυπική απόκλιση

Πλεονεκτήματα

- Λαμβάνονται υπ' όψιν για τον υπολογισμό τους όλες οι παρατηρήσεις.
- Έχουν μεγάλη εφαρμογή στη στατιστική συμπερασματολογία
- Σε πληθυσμούς που ακολουθούν την κανονική κατανομή μπορούμε να εκτιμήσουμε τα ποσοστά των ατόμων του πληθυσμού που βρίσκονται σε διαστήματα της μορφής:

$$(\bar{x}-\kappa\cdot s, \bar{x}+\lambda\cdot s)$$

$$\kappa = 0,1,2,3 \qquad \lambda = 0,1,2,3$$

Μειονεκτήματα

- Η διασπορά δεν εκφράζεται με τις ίδιες μονάδες που εκφράζονται οι μεταβλητές. Το πρόβλημα λύνεται με τη χρήση της τυπικής απόκλισης.
- Απαιτούνται αρκετές πράξεις για τον υπολογισμό τους.

◈ **Πως ορίζεται ο συντελεστής μεταβολής** (CV) **ενός συνόλου παρατηρήσεων.**

Απάντηση:

Ο συντελεστής μεταβολής ενός συνόλου παρατηρήσεων ορίζεται ως το πηλίκο της τυπικής απόκλισης s προς την μέση τιμή $\bar{x}$ των παρατηρήσεων.

Αν είναι $\bar{x} < 0$, τότε χρησιμοποιούμε την $|\bar{x}|$. Δηλαδή έχουμε:

$$\boxed{CV = \frac{s}{|\bar{x}|}, \quad \bar{x} \neq 0}$$

Παρατηρήσεις:

1. Λέγεται και **συντελεστής μεταβλητότητας.**

2. Δεχόμαστε ότι ένα δείγμα τιμών μιας μεταβλητής είναι **ομοιογενές**, αν:

$$\boxed{CV \leq 0{,}1 \quad \text{ή} \quad CV \leq 10\%}$$

3. Ο συντελεστής μεταβολής CV είναι ένα μέτρο σχετικής διασποράς. Είναι **καθαρός αριθμός**. Χρησιμοποιείται ως **μέτρο σύγκρισης** της **μεταβλητότητας** διαφορετικών δειγμάτων ομοειδών μεταβλητών με ίδιες ή διαφορετικές μονάδες μέτρησης.

   Αν CV_A και CV_B είναι οι συντελεστές μεταβολής δυο δειγμάτων A και B και ισχύει:

$$CV_A < CV_B$$

   τότε λέμε ότι το δείγμα A είναι περισσότερο ομοιογενές από το δείγμα B.

4. Ο συντελεστής μεταβολής **δεν ορίζεται** όταν η μέση τιμή $\bar{x}$ του δείγματος είναι κοντά στο μηδέν.

ΜΕΤΡΑ ΑΣΥΜΜΕΤΡΙΑΣ

◈ **Τι ονομάζονται μέτρα ασυμμετρίας;**

Απάντηση:

Όταν ένας πληθυσμός είναι κατά προσέγγιση **συμμετρικός** οι τιμές της μέσης τιμής, της διαμέσου και της επικρατούσας τιμής (αν έχει μια κορυφή) δεν διαφέρουν πολύ μεταξύ τους.

Ο πληθυσμός που δεν είναι συμμετρικός λέγεται **ασύμμετρος.**

Τα μέτρα που μετρούν την ασυμμετρία μιας κατανομής λέγονται **μέτρα ασυμμετρίας.**

Παρατηρήσεις:

1. Ένα μέτρο ασυμμετρίας είναι ο συντελεστής ασυμμετρίας του Peason:

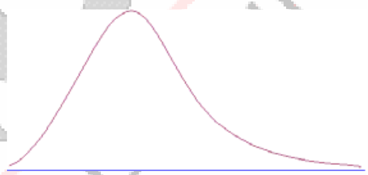
$$\boxed{\frac{(\text{μέση τιμή})-(\text{επικρατούσα τιμή})}{\text{τυπική απόκλιση}}}$$

ενώ στην περίπτωση που υπάρχουν δυο ή περισσότερες κορυφές (επικρατούσες τιμές) ο συντελεστής δίνεται από τον τύπο:

$$\boxed{\frac{(\text{μέση τιμή})-(\text{διάμεσος})}{\text{τυπική απόκλιση}}}$$

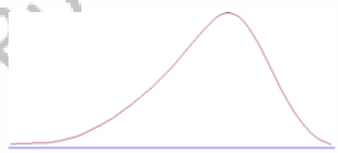
2. Αν η καμπύλη συχνοτήτων παρουσιάζει μακριά ουρά προς το μέρος που οι τιμές της μεταβλητής μεγαλώνουν, η κατανομή ονομάζεται **θετικά ασύμμετρη**, δηλαδή όταν ισχύει:

$$\bar{x} > \delta \quad \Leftrightarrow \quad \bar{x} - \delta > 0$$



Αν η καμπύλη συχνοτήτων παρουσιάζει μακριά ουρά προς το μέρος που οι τιμές της μεταβλητής μικραίνουν η κατανομή ονομάζεται **αρνητικά ασύμμετρη**, δηλαδή όταν:

$$\bar{x} < \delta \quad \Leftrightarrow \quad \bar{x} - \delta < 0$$



ΓΡΑΜΜΙΚΟΙ ΜΕΤΑΣΧΗΜΑΤΙΣΜΟΙ

◈ **Δίνονται ν παρατηρήσεις $x_1, x_2, \ldots, x_\nu$ με μέση τιμή $\bar{x}$ και τυπική απόκλιση s_x. Πως μεταβάλλεται η μέση τιμή, η τυπική απόκλιση και ο συντελεστής μεταβολής;**

α. **Αν σε όλες τις παρατηρήσεις προσθέσουμε μια σταθερά c.**

β. **Αν πολλαπλασιάσουμε όλες τις παρατηρήσεις με τον ίδιο αριθμό α.**

Απάντηση:

α. Έστω:

$$y_1 = x_1 + c,\ y_2 = x_2 + c, \ldots, y_\nu = x_\nu + c$$

οι νέες τιμές των παρατηρήσεων. Τότε η νέα μέση τιμή θα είναι:

$$\boxed{\bar{y} = \bar{x} + c}$$

Η νέα τυπική απόκλιση θα είναι:

$$\boxed{s_y = s_x}$$

παραμένει δηλαδή αμετάβλητη.

Ο νέος συντελεστής μεταβολής θα είναι:

$$\boxed{CV_y = \frac{s_y}{|\overline{y}|} = \frac{s_x}{|\overline{x}+c|}, \qquad \overline{x}+c \neq 0}$$

β. Έστω:

$$\omega_1 = \alpha \cdot x_1,\ \omega_2 = \alpha \cdot x_2, ..., \omega_\nu = \alpha \cdot x_\nu$$

οι νέες τιμές των παρατηρήσεων. Τότε η νέα μέση τιμή θα είναι:

$$\boxed{\overline{\omega} = \alpha \cdot \overline{x}}$$

Η νέα τυπική απόκλιση θα είναι:

$$\boxed{s_\omega = |\alpha| \cdot s_x}$$

ενώ ο νέος συντελεστής μεταβολής θα είναι:

$$\boxed{CV_\omega = \frac{s_\omega}{|\overline{\omega}|} = \frac{|\alpha| \cdot s_x}{|\alpha \cdot \overline{x}|} = \frac{|\alpha| \cdot s_x}{|\alpha| \cdot |\overline{x}|} = \frac{s_x}{|\overline{x}|} = CV_x, \qquad \alpha \neq 0, \overline{x} \neq 0}$$

Ο συντελεστής μεταβολής παραμένει αμετάβλητος.

Παρατήρηση:

Αν οι τιμές $x_1, x_2, ..., x_\nu$ μιας μεταβλητής που έχουν μέση τιμή $\overline{x}$, διάμεσο δ_x τυπική απόκλιση s_x και συντελεστή μεταβολής CV_x αντικατασταθούν από:

$$y_1 = \alpha \cdot x_1 + c,\ y_2 = \alpha \cdot x_2 + c, ..., y_\nu = \alpha \cdot x_\nu + c$$

τότε για τις παρατηρήσεις y_i, $i = 1, 2, ..., \nu$ έχουμε:

- Μέση τιμή:

$$\boxed{\overline{y} = \alpha \cdot \overline{x} + c}$$

- Διάμεσος:

$$\boxed{\delta_y = \alpha \cdot \delta_x + c}$$

- Τυπική απόκλιση:

$$\boxed{s_y = |\alpha| \cdot s_x}$$

- Συντελεστή μεταβολής:

$$\boxed{CV_y = \frac{|\alpha| \cdot s_x}{\alpha \cdot \overline{x} + c}, \qquad \alpha \cdot \overline{x} + c \neq 0}$$