

ΦΥΛΛΑΔΙΟ 1**ΘΕΩΡΙΑ ΣΤΑΤΙΣΤΙΚΗΣ**

- 1 *Τι ονομάζεται πληθυσμός, τι άτομα του πληθυσμού και τι μέγεθος του πληθυσμού;*

Απάντηση:

Στην στατιστική, για την ανάπτυξη πινάκων, συλλέγονται στοιχεία από ένα σύνολο αντικειμένων (εμφύχων ή αφύχων).

Το σύνολο αυτών των αντικειμένων ονομάζεται πληθυσμός. Κάθε στοιχείο του πληθυσμού ονομάζεται άτομο.

Το πλήθος των ατόμων ενός πληθυσμού ονομάζεται μέγεθος του πληθυσμού και συμβολίζεται με V .

- 2 *Τι ονομάζεται μεταβλητή; Ποια τα είδη των μεταβλητών;*

Απάντηση:

Οι μετρήσεις που παίρνουμε από τα άτομα ενός πληθυσμού, αφορούν κάποιο χαρακτηριστικό του πληθυσμού. Αυτό το χαρακτηριστικό ονομάζεται **ΜΕΤΑΒΛΗΤΗ**.

Ποιοτική μεταβλητή είναι αυτή που δεν επιδέχεται μέτρηση (π.χ. χρώμα ματιών, φύλο, θρήσκευμα κλπ).

Ποσοτική μεταβλητή είναι αυτή που επιδέχεται μέτρηση (ύψος, βάρος, μισθός κλπ).

Οι ποσοτικές μεταβλητές 2 ειδών είναι:

- i. Διακριτές μεταβλητές στις οποίες κάθε άτομο μπορεί να πάρει μόνο διακεκριμένες τιμές (πχ αριθμός παιδιών οικογένειας)
- ii. Συνεχείς μεταβλητές στις οποίες κάθε άτομο μπορεί να δώσει μέτρηση οποιοδήποτε πραγματικό αριθμό από ένα διάστημα ή μια ένωση διαστημάτων.

- 3 *Τι ονομάζεται δείγμα;*

Απάντηση:

Ένα μέρος του πληθυσμού λέγεται δείγμα.

- 4 *Τι ονομάζεται συχνότητα της τιμής X_i μιας μεταβλητής X ;*

Απάντηση:

Συχνότητα της τιμής X_i μιας μεταβλητής X ονομάζεται το πλήθος των ατόμων του πληθυσμού (ή του δείγματος) για τα οποία η μεταβλητή παίρνει τιμή X_i και συμβολίζεται με V_i .

Παρατήρηση: Ισχύει $V_1 + V_2 + V_3 + \dots + V_k = V$

- 5 Τι ονομάζεται σχετική συχνότητα της τιμής X_i μιας μεταβλητής X ;

Απάντηση:

Σχετική συχνότητα της τιμής X_i μιας μεταβλητής X ονομάζεται ο λόγος της συχνότητας προς το μέγεθος του δείγματος και συμβολίζεται με f_i .

Δηλαδή $f_i = V_i / V$

Παρατήρηση: Ισχύει $f_1 + f_2 + f_3 + \dots + f_k = 1$

- 6 Τι ονομάζεται αθροιστική συχνότητα της τιμής X_i μιας μεταβλητής X ;

Απάντηση:

Σε ΠΟΣΟΤΙΚΗ (μόνο) μεταβλητή, αθροιστική συχνότητα της τιμής X_i λέγεται το άθροισμα των συχνοτήτων V_i των τιμών που είναι μικρότερες ή ίσες με την τιμή αυτή.

- 7 Τι ονομάζεται σχετική αθροιστική συχνότητα της τιμής X ;

Απάντηση:

Σε ΠΟΣΟΤΙΚΗ (μόνο) μεταβλητή, σχετική αθροιστική συχνότητα της τιμής X_i λέγεται το άθροισμα των σχετικών συχνοτήτων f_i των τιμών που είναι μικρότερες ή ίσες με την τιμή αυτή.

- 8 Τι ονομάζεται επικρατούσα τιμή μιας μεταβλητής;

Απάντηση:

Επικρατούσα τιμή μιας μεταβλητής ονομάζεται η τιμή με τη μεγαλύτερη συχνότητα. Αν δύο ή περισσότερες τιμές έχουν την μέγιστη συχνότητα τότε υπάρχουν περισσότερες από μια επικρατούσες τιμές.

- 9 Τι ονομάζεται μέση τιμή μιας μεταβλητής; Πως υπολογίζεται η μέση τιμή μιας διακριτής μεταβλητής;

Απάντηση:

Η μέση τιμή διαφόρων τιμών είναι γνωστή ως το ηλικό του αθροίσματος των τιμών δια το πλήθος τους. Συμβολίζεται \bar{x}

Αν μια διακριτή μεταβλητή X έχει τιμές $X_1, X_2, X_3, \dots, X_k$

με αντίστοιχες συχνότητες $V_1, V_2, V_3, \dots, V_k$

τότε η μέση τιμή δίνεται από τον τύπο:

$$\bar{x} = \frac{V_1X_1 + V_2X_2 + \dots + V_kX_k}{V_1 + V_2 + V_3 + \dots + V_k}$$

10 Τι ονομάζεται διάμεσος δ ενός δείγματος n παρατηρήσεων;

Απάντηση:

(Θέμα 2010)

Διάμεσος δ ενός δείγματος n παρατηρήσεων που έχουν διαταχθεί σε αύξουσα σειρά ονομάζεται:

- Η μεσαία παρατήρηση αν το πλήθος των παρατηρήσεων είναι περιττό
- Το ημίαθροισμα των μεσαίων παρατηρήσεων αν το πλήθος των παρατηρήσεων είναι άρτιο

11 Τι ονομάζεται εύρος των τιμών μιας μεταβλητής;

Απάντηση:

Εύρος είναι η διαφορά της μικρότερης τιμής από την μεγαλύτερη.

12 Τι ονομάζεται διακύμανση των τιμών $t_1, t_2, t_3, \dots, t_n$ μιας μεταβλητής;

Απάντηση:

Αν μια μεταβλητή παίρνει τις n τιμές $t_1, t_2, t_3, \dots, t_n$ που έχουν μέση τιμή \bar{x} , τότε διακύμανση της μεταβλητής ονομάζεται το πηλίκο

$$S^2 = \frac{(\bar{x} - t_1)^2 + (\bar{x} - t_2)^2 + \dots + (\bar{x} - t_n)^2}{n} = \frac{\sum_{i=1}^n (\bar{x} - t_i)^2}{n}$$

13 Τι ονομάζεται διακύμανση μιας μεταβλητής με τιμές X_1, X_2, \dots, X_k και αντίστοιχες συχνότητες V_1, V_2, \dots, V_k ;

Απάντηση:

Αν μια μεταβλητή παίρνει τιμές X_1, X_2, \dots, X_k με αντίστοιχες συχνότητες V_1, V_2, \dots, V_k που έχουν μέση τιμή \bar{x} , τότε διακύμανση της μεταβλητής ονομάζεται το πηλίκο

$$S^2 = \frac{v_1 \cdot (\bar{x} - x_1)^2 + v_2 \cdot (\bar{x} - x_2)^2 + \dots + v_k \cdot (\bar{x} - x_k)^2}{n}$$

!! Αν μας ρωτήσουν απλά τι ονομάζεται διακύμανση, τότε γράφουμε ή την απάντηση 12 ή την απάντηση 13 διότι σύμφωνα με το βιβλίο οι δύο ορισμοί είναι ισοδύναμοι.

- 14 Τι ονομάζεται τυπική απόκλιση των τιμών $t_1, t_2, t_3, \dots, t_n$ μιας μεταβλητής;

Απάντηση:

Αν μία μεταβλητή παίρνει τις n τιμές $t_1, t_2, t_3, \dots, t_n$ που έχουν μέση τιμή \bar{x} , τότε τυπική απόκλιση της μεταβλητής ονομάζεται το

$$S = \sqrt{\frac{(\bar{x} - t_1)^2 + (\bar{x} - t_2)^2 + \dots + (\bar{x} - t_n)^2}{n}}$$

- 15 Τι ονομάζεται τυπική απόκλιση των τιμών X_1, X_2, \dots, X_k μιας μεταβλητής με αντίστοιχες συχνότητες V_1, V_2, \dots, V_k ;

Απάντηση:

Αν μία μεταβλητή παίρνει τιμές X_1, X_2, \dots, X_k με αντίστοιχες συχνότητες V_1, V_2, \dots, V_k που έχουν μέση τιμή \bar{x} τότε τυπική απόκλιση της μεταβλητής ονομάζεται το

$$S = \sqrt{\frac{(\bar{x} - t_1)^2 + (\bar{x} - t_2)^2 + \dots + (\bar{x} - t_n)^2}{n}}$$

!! Αν μας ρωτήσουν απλά τι είναι τυπική απόκλιση τότε γράφουμε ή την απάντηση 14 ή την απάντηση 15 διότι οι δύο ορισμοί είναι ισοδύναμοι.

- 16 Τι ονομάζουμε συντελεστή μεταβλητότητας CV;

Απάντηση:

Αν ένα δείγμα εξεταζόμενο ως προς μία ποσοτική μεταβλητή του, παρουσιάζει μέση τιμή \bar{x} και τυπική απόκλιση S , τότε συντελεστής μεταβολής ή συντελεστής μεταβλητότητας ονομάζεται το ηλίκο

$$CV = \frac{\text{τυπική απόκλιση}}{\text{μέση τιμή}} = \frac{S}{\bar{x}} \cdot 100\%$$

- 17 Πότε ένα δείγμα θεωρείται ομοιογενές;

Απάντηση:

Εάν $CV < 10\%$ ο πληθυσμός του δείγματος θεωρείται ομοιογενής.

Εκτός από τους προηγούμενους ορισμούς, τους οποίους πρέπει να ξέρουμε όσο το δυνατόν πιο καλά χρήσιμες είναι και οι εξής γνώσεις για τις ερωτήσεις αντιστοιχίσης – συμπλήρωσης και σωστού – λάθους.

A. Οι αθροιστικές συχνότητες – σχετικές συχνότητες – σχετικές συχνότητες % χρησιμοποιούνται μόνο στις ΠΟΣΟΤΙΚΕΣ μεταβλητές.

B. Η γωνία ω_i που αντιστοιχεί στην τιμή x_i σε κυκλικό διάγραμμα είναι ίση με $\omega_i = \frac{v_i}{v} \cdot 360^\circ$ ή αλλιώς $\omega_i = f_i \cdot 360^\circ$

Γ.	<u>Παράμετροι θέσης</u>	<u>Παράμετροι διασποράς</u>
	Μέση τιμή \bar{x}	Εύρος
	Επικρατούσα τιμή	Διακύμανση s^2
	Διάμεσος δ	Τυπική απόκλιση s

Μόνο η επικρατούσα τιμή χρησιμοποιείται ΚΑΙ για ποιοτικές μεταβλητές.

Όλες οι άλλες είναι αποκλειστικά για ποσοτικές μεταβλητές.

Δ. Σύγκριση παραμέτρων θέσης

- Η μέση τιμή επηρεάζεται από τις ακραίες τιμές και εξαρτάται από όλες τις τιμές της μεταβλητής.
- Η επικρατούσα τιμή εξαρτάται μόνο από τη μεγαλύτερη τιμή.
- Η διάμεσος δεν επηρεάζεται από τις ακραίες τιμές και εξαρτάται από όλες τις τιμές της μεταβλητής.

Ε. Το εύρος χρησιμοποιεί μόνο τις ακραίες τιμές της μεταβλητής και καθόλου τις υπόλοιπες. (μειονέκτημα)

ΣΤ. Η διακύμανση s^2 χρησιμοποιεί όλες τις τιμές της μεταβλητής αλλά παρουσιάζει το εξής πρόβλημα: Οι μονάδες της διακύμανσης είναι τα τετράγωνα των μονάδων της αντίστοιχης μεταβλητής.

Ζ. Η τυπική απόκλιση s χρησιμοποιεί όλες τις τιμές της μεταβλητής και εκφράζεται στις ίδιες μονάδες με τη μεταβλητή.

Η. Ο συντελεστής μεταβολής χρησιμεύει για να συγκρίνουμε τη μεταβλητότητα δύο ή περισσότερων πληθυσμών ως προς κάποιο (ποσοτικό) χαρακτηριστικό ακόμα και αν είναι μετρημένο σε διαφορετικές μονάδες ή διαφορετικές κλίμακες.

- 18 Πότε χρησιμοποιούμε το σταθμικό μέσο, αντί του αριθμητικού μέσου και ποιος είναι ο τύπος του;

Απάντηση:

Το σταθμικό μέσο ή σταθμισμένο αριθμητικό μέσο τον χρησιμοποιούμε στις περιπτώσεις που δίνεται διαφορετική βαρύτητα στις τιμές X_1, X_2, \dots, X_n ενός συνόλου δεδομένων. Αν οι συντελεστές στάθμισης (βαρύτητας) είναι w_1, w_2, \dots, w_n τότε

$$\bar{x} = \frac{x_1 w_1 + x_2 w_2 + \dots + x_n w_n}{w_1 + w_2 + w_3 + \dots + w_n} = \frac{\sum_{i=1}^n x_i w_i}{\sum_{i=1}^n w_i}$$

- 19 Ποιες είναι οι ιδιότητες της τυπικής απόκλισης s αν η καμπύλη συχνοτήτων για το χαρακτηριστικό που εξετάζουμε είναι κανονική ή περίπου κανονική;

Απάντηση:

- i. Το 68% περίπου των παρατηρήσεων είναι στο διάστημα $(\bar{x} - s, \bar{x} + s)$
- ii. Το 95% περίπου των παρατηρήσεων είναι στο διάστημα $(\bar{x} - 2s, \bar{x} + 2s)$
- iii. Το 99,7% περίπου των παρατηρήσεων είναι στο διάστημα $(\bar{x} - 3s, \bar{x} + 3s)$
- iv. Το εύρος ισούται περίπου με έξι τυπικές αποκλίσεις δηλ. $R \approx 6s$