

ΣΤΑΤΙΣΤΙΚΗ

Πληθυσμός : είναι το σύνολο του οποίου θέλουμε να εξετάσουμε τα στοιχεία του ως προς ένα ή περισσότερα χαρακτηριστικά του

Μεταβλητή : είναι το χαρακτηριστικό ως προς το οποίο εξετάζουμε έναν πληθυσμό και τη συμβολίζουμε με τα γράμματα X, Ψ, Z, B,...

Ποιοτικές ή κατηγορικές: λέγονται οι μεταβλητές των οποίων οι τιμές δεν είναι αριθμοί

Ποσοτικές : λέγονται οι μεταβλητές των οποίων οι τιμές είναι αριθμοί και διακρίνονται σε:

διακριτές που παίρνουν μόνο μεμονωμένες τιμές και σε **συνεχείς** που μπορούν να πάρουν οποιαδήποτε τιμή σε ένα διάστημα (α, β)

Απόλυτη συχνότητα v_i : μιας τιμής x_i της εξεταζόμενης μεταβλητής X, λέγεται ο φυσικός αριθμός που δείχνει πόσες φορές εμφανίζεται η x_i στο δείγμα.

Σχετική συχνότητα f_i μιας τιμής x_i της εξεταζόμενης μεταβλητής X, λέγεται το πηλίκο της απόλυτης συχνότητας της x_i προς το μέγεθος του δείγματος δηλ. $f_i = v_i/n$

Αν x_1, x_2, \dots, x_k οι τιμές μιας μεταβλητής X, ν.δ.ο. α) $0 \leq f_i \leq 1$ και β) $f_1 + f_2 + \dots + f_k = 1$

Απόδειξη: α) έχω: $f_i = v_i/n$ και επειδή $0 \leq v_i \leq n$ η ανίσωση ισχύει

$$\beta) \quad \text{έχω: } f_1 + f_2 + \dots + f_k = \frac{v_1}{n} + \frac{v_2}{n} + \dots + \frac{v_k}{n} = \frac{v_1 + v_2 + \dots + v_k}{n} = \frac{v}{n} = 1$$

Αθροιστική συχνότητα ή αθροιστική σχετική συχνότητα μιας τιμής x_i της εξεταζόμενης μεταβλητής X είναι το πλήθος και το ποσοστό αντίστοιχα των παρατηρήσεων που είναι μικρότερες ή ίσες από την x_i

Δηλ. $N_i = v_1 + v_2 + \dots + v_i$ και $F_i = f_1 + f_2 + \dots + f_i$

Το ραβδόγραμμα χρησιμοποιείται για τη γραφική παράσταση των τιμών μιας ποιοτικής μεταβλητής.

Το κυκλικό διάγραμμα χρησιμοποιείται για τη γραφική παράσταση των ποιοτικών και των ποσοτικών δεδομένων όταν οι διαφορετικές τιμές της μεταβλητής είναι σχετικά λίγες

και αν α_i η γωνία που αντιστοιχεί στην τιμή x_i , ισχύει: $\alpha_i = v_i \frac{360^\circ}{n} = f_i \cdot 360^\circ$

Το σημειόγραμμα χρησιμοποιείται όταν οι παρατηρήσεις είναι λίγες και οι τιμές της μεταβλητής παριστάνονται με σημεία πάνω από ένα οριζόντιο άξονα.

Το χρονόγραμμα χρησιμοποιείται για τη γραφική απεικόνιση της διαχρονικής εξέλιξης ενός οικονομικού, δημογραφικού ή άλλου μεγέθους.

Πολύγωνο συχνοτήτων ή σχετικών συχνοτήτων είναι η ένωση το σημείων (x_i, v_i) ή (x_i, f_i) αντίστοιχα .

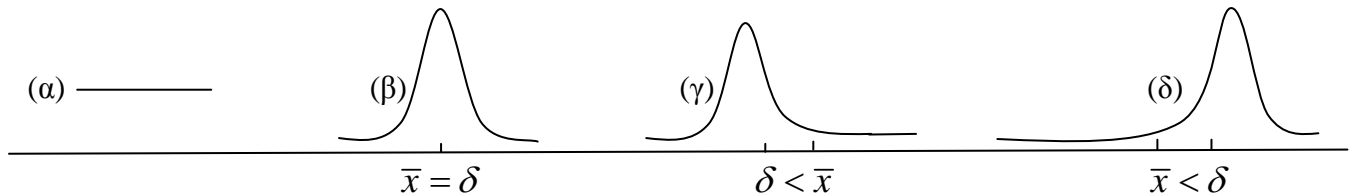
Ιστόγραμμα συχνοτήτων είναι γραφική παράσταση ενός πίνακα συχνοτήτων με ομαδοποιημένα δεδομένα όπου τα ορθογώνια που κατασκευάζουμε έχουν βάση όσο το πλάτος της κλάσης και ύψος τέτοιο ώστε το εμβαδόν του ορθογωνίου να είναι ίσο με τη συχνότητα της κλάσης αυτής.

Το εμβαδόν του χωρίου που ορίζεται από το πολύγωνο συχνοτήτων και από τον οριζόντιο άξονα είναι ίσο με το μέγεθος του δείγματος n.

Το εμβαδόν του χωρίου που ορίζεται από το πολύγωνο σχετικών συγνοτήτων και από τον οριζόντιο άξονα είναι ίσο με 1.

Καμπύλη συγνοτήτων είναι η πολυγωνική γραμμή η οποία τείνει να πάρει τη μορφή ομαλής καμπύλης, εάν υποθέσουμε ότι ο αριθμός των κλάσεων είναι αρκετά μεγάλος (τείνει στο άπειρο) και το πλάτος των κλάσεων είναι αρκετά μικρό (τείνει στο 0).

Τα κυριότερα είδη κατανομής είναι: (α) η ομοιόμορφη, (β) η κανονική, (γ) η ασύμμετρη με θετική ασυμμετρία και (δ) η ασύμμετρη με αρνητική ασυμμετρία



Μέτρα θέσης είναι τα αριθμητικά μεγέθη που μας δίνουν τη θέση του κέντρου των παρατηρήσεων στον οριζόντιο άξονα και τα οποία είναι: **η μέση τιμή, η διάμεσος και η επικρατούσα τιμή**

Μέση τιμή n παρατηρήσεων t_i είναι ο μέσος όρος τους δηλ.
$$\bar{x} = \frac{1}{v} \sum_{i=1}^v t_i = \frac{1}{v} \sum_{i=1}^k v_i x_i = \sum_{i=1}^k f_i x_i$$

όπου x_i οι διαφορετικές τιμές των t_i

Σταθμικός μέσος είναι ο μέσος όρος των n παρατηρήσεων όταν δίνεται διαφορετική βαρύτητα w_i στις τιμές

τους. δηλ.
$$\bar{x} = \frac{x_1 w_1 + x_2 w_2 + \dots + x_v w_v}{w_1 + w_2 + \dots + w_v}$$

Διάμεσος (δ) ενός δείγματος n παρατηρήσεων που έχουν διαταχθεί κατά αύξουσα σειρά, είναι η μεσαία παρατήρηση αν n περιττός και το ημιάθροισμα των δύο μεσαίων αν n άρτιος

Μέτρα διασποράς είναι τα αριθμητικά μεγέθη που μας δίνουν τη διασπορά των παρατηρήσεων δηλ. πόσο αυτές εκτείνονται γύρω από το κέντρο τους και οι οποίες είναι: **το εύρος, η διακύμανση και η τυπική απόκλιση**

Εύρος (R) ενός συνόλου παρατηρήσεων είναι η διαφορά της ελάχιστης από τη μεγαλύτερη παρατήρηση.

Διακύμανση ή διασπορά (s^2) n παρατηρήσεων με τιμές t_i είναι ο μέσος όρος των τετραγώνων των αποκλίσεών τους από τη μέση τους τιμή δηλ.

$$s^2 = \frac{1}{v} \sum_{i=1}^v (t_i - \bar{x})^2 = \frac{1}{v} \sum_{i=1}^k (x_i - \bar{x})^2 = \sum_{i=1}^k f_i (x_i - \bar{x})^2 \quad \text{όπου } x_i \text{ οι διαφορετικές των } t_i$$

Τυπική απόκλιση (s) n παρατηρήσεων είναι η τετραγωνική ρίζα της διασποράς τους δηλ.
$$s = \sqrt{s^2}$$

Συντελεστής μεταβολής ή μεταβλητότητας (CV)

είναι το πηλίκο:
$$s/\bar{x}, \text{ αν } \bar{x} > 0 \quad \text{και} \quad s/|\bar{x}|, \text{ αν } \bar{x} < 0$$
 και παριστάνει ένα μέτρο σχετικής διασποράς των τιμών και όχι την απόλυτη διασπορά. Εκφράζει δηλ. την μεταβλητότητα των δεδομένων απαλλαγμένη από τη μέση τιμή, είναι ανεξάρτητος από τις μονάδες και εκφράζεται σε επί τοις εκατό.

Ομοιογενές χαρακτηρίζεται ένα δείγμα για το οποίο: $CV \leq 10\%$