

Στατιστική – Εισαγωγικές Έννοιες

➤ **Στατιστική:** η επιστήμη που παρέχει μεθόδους και εργαλεία για την οργάνωση, συστηματική περιγραφή και περιληπτική παρουσίαση δεδομένων, καθώς και για την ανάλυση της πληροφορίας που εμπεριέχεται στα δεδομένα, με τρόπο αντικειμενικό, ώστε να είναι δυνατή η εξαγωγή γενικότερων συμπερασμάτων από περιορισμένο αριθμό μετρήσεων ή παρατηρήσεων.

- Περιγραφική Στατιστική
- Στατιστική Συμπερασματολογία

Στατιστική – Εισαγωγικές Έννοιες

➤ **Πληθυσμός:** άτομα ή αντικείμενα συγκεκριμένου τύπου, για τα οποία θέλουμε να συνάγουμε συμπέρασμα σχετικά με κάποιο φαινόμενο που τα χαρακτηρίζει.

➤ **Δείγμα:** υποσύνολο του πληθυσμού που χρησιμοποιείται για συλλογή δεδομένων.

✂ Η αντικειμενικότητα της στατιστικής συμπερασματολογίας εξασφαλίζεται από τον κατάλληλο μηχανισμό επιλογής του δείγματος.

➤ **Πείραμα:** μηχανισμός επιλογής παρατηρήσεων που εισάγει τυχαιότητα στα δεδομένα και ταυτόχρονα επιτρέπει την εφαρμογή της θεωρίας των πιθανοτήτων τόσο για την επαγωγή γενικευμένων συμπερασμάτων όσο και για τη μέτρηση της αξιοπιστίας των συμπερασμάτων.

(Σχεδιασμός Πειραμάτων-ξεχωριστός κλάδος της στατιστικής)

Δειγματοληψία

➤ **Δειγματοληψία:** Η διαδικασία δημιουργίας δειγμάτων με τη βοήθεια πειράματος.

✿ «πιθανοκρατικός» μηχανισμός δημιουργίας του δείγματος, δηλ. κάθε στοιχείο του πληθυσμού θα περιέχεται στο δείγμα με γνωστή πιθανότητα.

✿ Ο πληθυσμός αφορά ένα σύνολο ατόμων ή αντικειμένων ή ακόμα και το σύνολο όλων των τιμών που μπορεί να πάρει ένα χαρακτηριστικό (π.χ. ο πληθυσμός όλων των πιθανών ετήσιων οικογενειακών εισοδημάτων στην Ελλάδα)

✿ Αναγκαιότητα δημιουργίας δειγμάτων:

- δεν είναι διαθέσιμος όλος ο πληθυσμός
- συχνά τα δείγματα καταστρέφονται μετά τη παρατήρηση
- δεν προϋπάρχουν απαραίτητα στον πληθυσμό όλες οι πιθανές τιμές

➤ **Δειγματοληψία με επανατοποθέτηση** - όλα τα στοιχεία έχουν την ίδια πιθανότητα να επιλεγούν και κάθε στοιχείο μπορεί να επιλεγεί περισσότερο από μία φορά.

➤ **Δειγματοληψία χωρίς επανατοποθέτηση** - τυχαία επιλογή μεταξύ των στοιχείων που δεν έχουν επιλεγεί άλλη φορά.

Η δειγματοληψία χωρίς επανατοποθέτηση δίνει περισσότερες πληροφορίες για τον πληθυσμό, όμως η δειγματοληψία με επανατοποθέτηση αναλύεται ευκολότερα.

Όταν το δείγμα είναι μικρό σε σχέση με τον πληθυσμό οι δύο τύποι δειγματοληψίας είναι προσεγγιστικά ισοδύναμοι.

Δειγματοληψία (Συνέχεια ...)

➤ Κατανομή Πληθυσμού:

X = η τιμή του χαρακτηριστικού που μελετάμε σε ένα συγκεκριμένο άτομο ή αντικείμενο του πληθυσμού (τυχαία μεταβλητή)

Κατανομή Πληθυσμού: η κατανομή της X που δίνεται από τη συχνότητα με την οποία η συγκεκριμένη τιμή συναντάται στον πληθυσμό.

➤ $X_1, X_2, \dots, X_n$: τιμές του ίδιου χαρακτηριστικού σε ένα δείγμα μεγέθους n (n τυχαίες μεταβλητές)

Δειγματοληψία με επανατοποθέτηση: $X_1, X_2, \dots, X_n$ ισόνομες και ανεξάρτητες τ.μ.

Δειγματοληψία χωρίς επανατοποθέτηση: $X_1, X_2, \dots, X_n$ ισόνομες αλλά εξαρτημένες τ.μ.

Δειγματοληψία με επανάληψη: $X_1, X_2, \dots, X_n$ ισόνομες και ανεξάρτητες τ.μ, όπου X_i το αποτέλεσμα της i επανάληψης.

➤ Τυχαίο δείγμα μεγέθους n : όταν υπάρχουν οι τ.μ. $X_1, X_2, \dots, X_n$ που προκύπτουν από το δείγμα είναι ανεξάρτητες και ισόνομες.

➤ Απλό τυχαίο δείγμα μεγέθους n : όταν υπάρχουν οι τ.μ. $X_1, X_2, \dots, X_n$ που προκύπτουν από το δείγμα είναι απλά ισόνομες.

✿ Για τη δημιουργία τυχαίων ή απλών τυχαίων δειγμάτων χρησιμοποιούμε πίνακες τυχαίων αριθμών ή γεννήτριες τυχαίων αριθμών.

Τρόποι Περιγραφής Δεδομένων

➤ Διαγράμματα / Γραφικές Παραστάσεις

Τα δεδομένα ομαδοποιούνται (εφόσον χρειάζεται) σε μη επικαλυπτόμενες και ισομήκεις ομάδες (κλάσεις). Ο αριθμός των κλάσεων *Κ* ορίζεται από τον εμπειρικό κανόνα $2^{k-1} < n < 2^k$, όπου *n* το σύνολο των δεδομένων.

Πίνακας Συχνοτήτων: Για κάθε κλάση βρίσκουμε τον αριθμό των δεδομένων που περιέχονται σ' αυτήν (συχνότητα κλάσης) ή τον αριθμό των δεδομένων που περιέχονται σ' αυτήν και τις προηγούμενες (αθροιστική συχνότητα κλάσης).

Πίνακας Σχετικών Συχνοτήτων: Για κάθε κλάση βρίσκουμε τη σχετική συχνότητα της, δηλαδή τη συχνότητα προς τον συνολικό αριθμό των δεδομένων ή την αθροιστική σχετική συχνότητα δηλαδή την αθροιστική συχνότητα κλάσης προς τον συνολικό αριθμό των δεδομένων.

Ιστόγραμμα: Για κάθε κλάση δημιουργούμε ορθογώνιο με βάση όλο το μήκος της και ύψος τη συχνότητα της (ιστόγραμμα συχνοτήτων) ή την σχετική συχνότητα της (ιστόγραμμα σχετικών συχνοτήτων). Δίνει την εικόνα της εμπειρικής κατανομής των δεδομένων και βοηθάει στην εύρεση της κατανομής του πληθυσμού.

Όταν τα δεδομένα είναι ποιοτικά τότε οι κλάσεις αντιπροσωπεύουν τις τιμές του χαρακτηριστικού.

Τρόποι Περιγραφής Δεδομένων

➤ Αριθμητικά μεγέθη (Μέτρα Θέσεως)

Δειγματική μέση τιμή: $\bar{X} = \sum_{i=1}^n X_i / n$ (αριθμητικός μέσος)

Σταθμισμένη μέση τιμή: $\bar{X} = \sum_{i=1}^n w_i X_i / \sum_{i=1}^n w_i$, όπου w_i τα βάρη

Δειγματική διάμεσος: Διατεταγμένο δείγμα- $X_{(1)}, X_{(2)}, \dots, X_{(n)}$

$$m = \begin{cases} X_{(\frac{n+1}{2})} & n \text{ περιττός} \\ \frac{1}{2} \left[X_{(\frac{n}{2})} + X_{(\frac{n}{2}+1)} \right] & n \text{ άρτιος} \end{cases}$$

Δειγματική κορυφή: η τιμή που εμφανίζεται περισσότερες φορές στα δεδομένα.

Δειγματική μέση έκταση: $[X_{(1)} + X_{(n)}] / 2$

Τρόποι Περιγραφής Δεδομένων

➤ Αριθμητικά μεγέθη (Μέτρα μεταβλητότητας)

Δειγματική έκταση: $r = X_{(n)} - X_{(1)}$

Δειγματική διασπορά:

$$\frac{\sum_{i=1}^n |X_i - \bar{X}|}{n}$$

Δειγματική μέση απόκλιση:

Δειγματική διάμεσος απόκλιση:

$$\frac{\sum_{i=1}^n |X_i - m|}{n}$$

$$s^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n-1} = \frac{\sum_{i=1}^n X_i^2 - \frac{\left(\sum_{i=1}^n X_i\right)^2}{n}}{n-1}$$

Δειγματικά ποσοστιαία σημεία.

Συντελεστής διακύμανσης.

Δειγματικός συντελεστής λοξότητας.

Δειγματικός συντελεστής κύρτωσης.