

Introduction to Statistical Analysis

Introduction to Statistical Analysis

Key concepts, steps, and tools for turning data into insight

1. What Is Statistical Analysis?

- **Definition:** The systematic collection, exploration, and interpretation of data to discover patterns, test hypotheses, and support decision-making.
 - **Purpose:**
 - Summarize large data sets in a digestible form.
 - Identify relationships among variables.
 - Quantify uncertainty and assess the reliability of conclusions.
 - **Fields of Application:** Business, medicine, engineering, social sciences, public policy, sports, etc.
-

2. Types of Statistics

Category	Description	Typical Techniques
Descriptive	Summarizes data without inferring beyond the sample.	Measures of central tendency (mean, median, mode), dispersion (range, variance, standard deviation), frequency tables, histograms, boxplots.
Inferential	Draws conclusions about a population from a sample.	Confidence intervals, hypothesis testing, regression, ANOVA, chi-square tests, Bayesian inference.

3. The Statistical Analysis Workflow

1. Define the Problem

- Clarify the research question or business objective.
- Identify the population of interest and the variables to measure.

Introduction to Statistical Analysis

2. Collect Data

- Choose a sampling design (simple random, stratified, cluster, systematic).
- Ensure data quality (validity, reliability, completeness).

3. Explore & Clean Data

- Detect missing values, outliers, and inconsistencies.
- Perform data transformation (scaling, encoding categorical variables).
- Visual exploration: scatter plots, histograms, heatmaps.

4. Describe the Data

- Compute summary statistics.
- Produce visual summaries (bar charts, box plots, density plots).

5. Model & Infer

- Select appropriate statistical models (e.g., t-test, linear regression, logistic regression).
- Check model assumptions (normality, independence, homoscedasticity).
- Estimate parameters and calculate confidence intervals.

6. Validate & Interpret

- Use diagnostics (residual plots, goodness-of-fit tests).
- Perform cross-validation or hold-out testing for predictive models.
- Translate statistical findings into actionable insights.

7. Communicate Results

- Prepare clear tables, graphs, and narrative explanations.
- Emphasize practical significance, not just statistical significance.
- Document methodology and limitations.

4. Core Concepts & Formulas

Introduction to Statistical Analysis

4.1 Measures of Central Tendency

- **Mean (? or $\bar{x}$)**: ($\displaystyle \bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$)
- **Median**: Middle value after sorting; robust to outliers.
- **Mode**: Most frequently occurring value.

4.2 Measures of Dispersion

- **Variance**: ($\displaystyle s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$)
- **Standard Deviation (SD)**: ($s = \sqrt{s^2}$)
- **Interquartile Range (IQR)**: ($Q_3 - Q_1$)

4.3 Probability Foundations

- **Probability of event A**: ($P(A) = \frac{\text{Number of favorable outcomes}}{\text{Total outcomes}}$) (for equally likely outcomes).