

Introduction to Applied Statistics

Introduction to Applied Statistics

Key concepts, workflow, and practical tools for turning data into insight

1. What Is Applied Statistics?

Aspect	Description
Definition	The use of statistical methods to collect, analyze, interpret, and present data for real-world problems.
Goal	Turn raw observations into actionable information and decisions.
Typical Domains	Business analytics, public health, engineering, social sciences, sports, finance, environmental science, etc.
Contrast with Theoretical Statistics	Applied statistics focuses on implementation and interpretation; theoretical statistics develops the underlying probability models and proofs.

2. The Applied Statistics Workflow

1. Problem Definition

- Clarify the question: *What do we want to know or predict?*
- Identify success criteria (e.g., reduce defect rate by 10%).

2. Data Acquisition

- **Sources:** surveys, experiments, sensors, administrative records, web logs.
- **Design considerations:** random sampling, stratification, blocking, replication.

3. Data Cleaning & Preparation

- Handling missing values (deletion, imputation).
- Detecting and treating outliers.
- Variable transformation (log, sqrt, scaling).
- Creating derived variables (ratios, indicators).

Introduction to Applied Statistics

4. Exploratory Data Analysis (EDA)

- Summary statistics (mean, median, variance, quantiles).
- Visual tools: histograms, boxplots, scatterplots, heatmaps.
- Spotting patterns, trends, and anomalies.

5. Model Selection & Specification

- Choose a statistical model that matches the data structure and research question (e.g., linear regression, logistic regression, time-series, survival models).
- Check assumptions (linearity, independence, normality, homoscedasticity).

6. Estimation & Inference

- Estimate parameters (e.g., using least squares, maximum likelihood).
- Quantify uncertainty: standard errors, confidence intervals, p-values.

7. Model Diagnostics & Validation

- Residual analysis, influence measures, goodness-of-fit tests.
- Cross-validation, bootstrapping, out-of-sample prediction.

8. Interpretation & Communication

- Translate statistical results into plain-language conclusions.
- Use tables, graphs, and narrative that suit the audience.

9. Implementation & Monitoring

- Deploy models (e.g., scoring scripts, dashboards).
- Track performance over time and update as needed.

3. Core Statistical Concepts

Concept	Why It Matters	Typical Tools
Descriptive Statistics	Summarize central tendency & spread.	<code>mean()</code> , <code>median()</code> , <code>sd()</code> , <code>quantile()</code>

Introduction to Applied Statistics

Concept	Why It Matters	Typical Tools
Probability Distributions	Model randomness; foundation for inference.	Normal, Binomial, Poisson, Exponential, etc.
Sampling Theory	Guarantees that a sample represents a population.	Simple random sample, stratified, cluster sampling
Confidence Intervals	Provide a range of plausible values for parameters.	<code>confint()</code> in R, <code>statsmodels.stats</code> in Python
Hypothesis Testing	Formal decision rules for comparing groups or effects.	t-test, χ^2 test, ANOVA, non-parametric tests
Regression Analysis	Quantify relationships between variables.	Linear, multiple, logistic, Poisson regression
Analysis of Variance (ANOVA)		