

DATA SCIENCE

INTERMEDIATE

5 DAYS

Statistics for Data Science and Business Analysis

Build a strong foundation in statistics for data science by mastering data types, distributions, and visualization techniques. Learn how to apply hypothesis testing, regression analysis, and statistical measures using Python and R to support accurate, data driven insights.

Why Business Statistics?

Data-centric decision-making

Equipping your employees with statistical knowledge will eventually reduce reliance on guess-based, intuitive decision-making. It will enable them to uncover hidden patterns in raw business data, resulting in evidence-based business strategies.

Operational Efficiency

Trained employees can identify and address discrepancies in your business data through efficient and prompt analysis. Over time, reduced errors lead to optimal resource allocations and operations, leading to faster project completion and significant cost savings.

Market analysis proficiency

Employees with statistically sound backgrounds can significantly enhance competitor and market analysis by working with complex consumer data. Insights from this analysis improve product development and marketing strategies, giving you a competitive edge even in a crowded market.

Thorough reporting

Statistical knowledge also enhances visualization and reporting capabilities. Employees with this knowledge can furnish accurate and insightful reports, providing a granular view of your performance.

What you'll be able to do

- ✓ Understand the fundamentals of statistics and different types of data
- ✓ Master plotting various data types and visualizing data effectively
- ✓ Calculate central tendency, asymmetry, variability, correlation, and covariance measures
- ✓ Distinguish between different types of distributions and estimate confidence intervals
- ✓ Perform hypothesis testing and interpret statistical significance for business decisions
- ✓ Grasp the mechanics of regression analysis
- ✓ Use and interpret dummy variables in regression models
- ✓ Apply statistical concepts in data science with practical applications using Python and R

Who this is for

- Data and business analysts
- Sales and marketing teams

Curriculum

01 Introduction

- A brief overview of the course

02 Sample or Population Data

- Understanding the difference

03 The Fundamentals of Descriptive Statistics

- Types of data and levels of measurement
- Categorical variables and corresponding visualization techniques
- Numerical variables and frequency distribution tables
- Histogram charts
- Cross tables
- Scatter plots

04 Measures of Central Tendency, Asymmetry, and Variability

- Fundamentals of central tendency measures - mean, median, and mode
- Measuring skewness
- Measuring data spread through standard deviation and variance
- Analyzing relationships between variables through covariance and correlation

05 Distributions

- Fundamentals of inferential statistics
- What is a distribution
- Normal and Standard Normal Distribution
- The Central Limit Theorem
- Standard Error

06 Estimators and Estimates

- Basics of estimators and estimates
- Confidence intervals and their calculation with known population variance
- Student's T Distribution
- Using Student's T-score to calculate confidence intervals with unknown population variance
- The Margin of Error and its significance

07 Confidence Intervals: Advanced Topics

- Confidence intervals for two means with dependent samples
- Confidence intervals for two means with independent samples

08 Hypothesis Testing: Introduction

- Null and alternative hypotheses
- Rejection region and significance level
- Type 1 Error vs Type II Error

09 Hypothesis Testing: Let's Start Testing!

- P-value
- Testing for the mean with known and unknown population variance
- Testing for the mean with dependent and independent samples

10 The Fundamentals of Regression Analysis

- Fundamentals of regression analysis
- Relationship between correlation and causation
- The Linear Regression Model (LRM)
- Correlation vs Regression
- Geometrical representation of LRM

11 Subtleties of Regression Analysis

- R-squared and its utility
- The Ordinary Least Squares (OLS) method
- Regression tables
- Adjusted R-squared
- F-statistic and its utility

12 Assumptions for Linear Regression Analysis

- Assumptions of the Ordinary Least Squares method
- Linearity
- No endogeneity
- Normality and homoscedasticity
- No autocorrelation
- No multicollinearity

13 Dealing with Categorical Data

- Dummy variables

14 Bonus Lecture

- What's next