

Introduction to Statistics

Course Description: This course teaches you statistical thinking concepts that are essential for learning from data and communicating insights. By the end of the course, you will be able to perform exploratory data analysis, understand key principles of sampling, and select appropriate tests of significance for multiple contexts. You will gain the foundational skills that prepare you to pursue more advanced topics in statistical thinking and machine learning.

Topics include Descriptive Statistics, Sampling and Randomized Controlled Experiments, Probability, Sampling Distributions and the Central Limit Theorem, Regression, Common Tests of Significance, Resampling, Multiple Comparisons.

Duration: 3 Days

What you will learn from this course

1. Introduction and Descriptive Statistics for Exploring Data

This module provides an overview of the course and a review of the main tools used in descriptive statistics to visualize information.

2. Producing Data and Sampling

In this module, you will look at the main concepts for sampling and designing experiments. You will learn about curious pitfalls and how to evaluate the effectiveness of such experiments.

3. Probability

In this module, you will learn about the definition of probability and the essential rules of probability that you will need for solving both simple and complex challenges. You will also learn about examples of how simple rules of probability are used to create solutions for real-life complex situations.

4. Normal Approximation and Binomial Distribution

This module covers the empirical rule and normal approximation for data, a technique that is used in many statistical procedures. You will also learn about the binomial distribution and the basics of random variables.

5. Sampling Distributions and the Central Limit Theorem

In this module, you will learn about the Law of Large Numbers and the Central Limit Theorem. You will also learn how to differentiate between the different types of histograms present in statistical analysis.

6. Regression

This module covers regression, arguably the most important statistical technique based on its versatility to solve different types of statistical problems. You will learn about inference, regression, and how to do regression diagnostics.

7. Confidence Intervals

In this module, you will learn how to construct and interpret confidence intervals in standard situations.

8. Tests of Significance

In this module, you will look at the logic behind testing and learn how to perform the appropriate statistical tests for different samples and situations. You will also learn about common misunderstandings and pitfalls in testing.

9. Resampling

This module focuses on the two main methods used in computer-intensive statistical inference: The Monte Carlo method, and the Bootstrap method. You will learn about the theoretic principles behind these methods and how they are applied in different contexts, such as regression and constructing confidence intervals.

10. Analysis of Categorical Data

This module focuses on the three important statistical analysis for categorical data: Chi-Square Goodness of Fit test, Chi-Square test of Homogeneity, and Chi-Square test of Independence.

11. One-Way Analysis of Variance (ANOVA)

This module covers the basics of ANOVA and how F-tests work on one-way ANOVA examples.

12. Multiple Comparisons

In this module, you will learn about very important issues that have surfaced in the era of big data: data snooping and the multiple testing fallacy. You will also explore the reasons behind challenges in data reproducibility and applicability, and how to prevent such issues in your own work.