15.0.2 Module 15 Roadmap

Looking Ahead

In this module, you'll apply your understanding of statistics and hypothesis testing to analyze a series of datasets from the automotive industry. Your analysis will include visualizations, statistical tests, and interpretation of the results. All of your statistical analysis and visualizations will be written in the R programming language.

Throughout the module, you'll extract, transform, and load (ETL) data; visualize the data; and analyze the data using R. Additionally, you'll learn a variety of statistical tests, their real-world application in data science, and their implementation in R. The goal is for you to apply these statistical concepts beyond this module, to any dataset, using any programming language—including Python.

What You Will Learn

By the end of this module, you will be able to:

- Load, clean up, and reshape datasets using tidyverse in R.
- Visualize datasets with basic plots such as line, bar, and scatter plots using ggplot2.
- Generate and interpret more complex plots such as boxplots and heatmaps using ggplot2.
- Plot and identify distribution characteristics of a given dataset.
- Formulate null and alternative hypothesis tests for a given data problem.

- Implement and evaluate simple linear regression and multiple linear regression models for a given dataset.
- Implement and evaluate the one-sample t-Tests, two-sample t-Tests, and analysis of variance (ANOVA) models for a given dataset.
- Implement and evaluate a chi-squared test for a given dataset.
- Identify key characteristics of A/B and A/A testing.
- Determine the most appropriate statistical test for a given hypothesis and dataset.

Planning Your Schedule

Here's a quick look at the lessons and assignments you'll cover in this module. You can use the time estimates to help pace your learning and plan your schedule.

- Introduction to Module 15 (15 minutes)
- Getting Started with R (30 minutes)
- Programming and ETL in R (2 hours)
- Visualize Your Data Using ggplot2 (2 hours)
- Introduction to Statistical Tests (1 hour)
- Introduction to Hypothesis Testing (30 minutes)
- Perform an Analysis of Means in R (1 hour)
- Correlation and Regression in R (1 hour)
- Characterize Categorical Data (30 minutes)
- Getting Real with A/B Testing (15 minutes)
- Choose the Right Test for Your Data (15 minutes)
- Application (5 hours)