Case Study: Statistical Data Analysis & Hypothesis Testing

From Raw Data to Business-Ready Insights

Tools Used: Python • Jupyter Notebook • Pandas • SciPy • Matplotlib

Project Overview

This project demonstrates a full end-to-end data analysis workflow focused on hypothesis testing and correlation analysis. The objective was to determine whether weekly self-study time has a statistically significant impact on academic performance.

This type of analysis is applicable to business analytics, training effectiveness studies, A/B testing, marketing performance analysis, and survey research.

Business Question

Is there a statistically significant relationship between weekly self-study hours and academic performance?

The question was answered through formal statistical testing to ensure conclusions are objective, defensible, and decision-ready.

Data Preparation & Quality Control

The dataset was cleaned and validated prior to analysis. This included verifying data types, checking for missing values, removing inconsistencies, and generating descriptive statistics to understand distributions and variability.

This step ensures that all statistical results are based on trustworthy and reliable data.

Exploratory Data Analysis (EDA)

Exploratory analysis was conducted to understand how academic scores and self-study hours were distributed. Key goals included identifying variability, detecting outliers, and visually assessing potential relationships.

Professional visualizations were created to make insights easy to communicate to non-technical stakeholders.

Statistical Assumption Testing

Before running hypothesis tests, a normality test was performed to determine whether parametric methods were appropriate. The results confirmed that a non-parametric approach should be used.

This step is critical for maintaining statistical validity.

Hypothesis Testing & Correlation Analysis

Spearman’s Rank Correlation Test was applied due to the non-normal data distribution.

Null Hypothesis (H■): No relationship exists between self-study time and academic performance.

Alternative Hypothesis (H■): A significant relationship exists between self-study time and academic performance.

The analysis produced a correlation coefficient, relationship direction, and statistical significance level.

Interpretation & Business Insights

Results were translated into clear business language to answer whether increased effort is associated with better performance, how strong the relationship is, and whether the effect can be trusted for decision-making.

Skills Demonstrated

- Data Cleaning & Validation
- Exploratory Data Analysis (EDA)
- Statistical Assumption Testing
- Hypothesis Testing
- Non-Parametric Correlation Analysis
- Data Visualization
- Business-Focused Insight Interpretation
- Reproducible Analytical Workflows

Client Value

- Statistically valid conclusions
- Transparent, auditable methodology
- Business-ready interpretation
- Executive-ready reporting
- Reusable analysis for future updates

Summary

This case study highlights the ability to transform raw data into trustworthy business insights through professional statistical analysis. The full workflow ensures accuracy, transparency, and decision relevance.