

STOR 664 Course Project

Yumo Bai Suhan Liu Xu Huang Beichen Wan

November 17, 2025

Project Question

Main Question:

- How do **study behaviors**—self-study hours, absence days, and extracurricular participation—relate to **academic performance** across seven subjects?

Specific Goals:

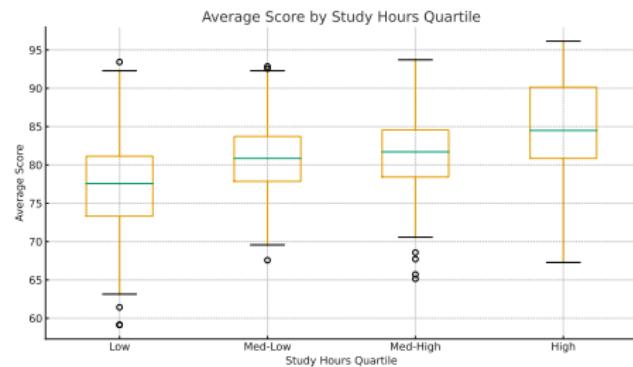
- Quantify the association between each behavioral factor and performance.
- Understand cross-subject correlations in exam scores.
- Identify which behaviors correlate most strongly with academic success.
- **Null hypotheses** $H_0^{(i)}$, $i = 1, 2, 3$: self-study hours, absence days, and extracurricular participation do not affect academic performance.

Data and Exploratory Analyses

Data Summary

- 2,000 students.
- Exam scores in 7 subjects: Math, History, Physics, Chemistry, Biology, English, Geography.
- Behavioral variables: weekly self-study hours, absence days, extracurricular participation.

EDA Highlight: Study Hours vs Average Score



Proposal for Analysis

Proposed Modeling Strategy

- Fit a multiple linear regression:

$$\text{AvgScore} = \beta_0 + \beta_1 \text{StudyHours} + \beta_2 \text{Absences} + \beta_3 \text{Extracurricular} + \varepsilon$$

- Add demographic controls (gender, age, aspiration) for robustness.
- Optionally estimate subject-specific regressions.

Rationale

- EDA suggests approximately linear relationships.
- Sample size ($n = 2000$) supports stable inference.

Extensions

- Interaction effects (e.g., $\text{StudyHours} \times \text{Absences}$)
- Hypothesis testing and multiple hypothesis testing
- Influence on each subjects