Statistical Modeling & Hypothesis

Course: Data Analytics / Statistics —Project:Predictionofexam scores.

Prepared by: [Abhinav, Prashant, Srijan].

Objective: Describe model choices, document data exploration, and formulate precise hypotheses about exam performance.

Problem Statement

What is the problem?

Students show varying exam performance driven by study habits, preparation and background.

Main goal

Determine factor impacts on exam scores and use statistical models to analyse and forecast performance.

Dataset Overview

The dataset combines scholarly and demographic data and focuses on performance across three subjects.

- Gender
- Parental level of education
- Lunch type
- Test preparation course
- Math score, Reading score, Writing score

Variables

Dependent (Outcomes)

- Math_Score (numeric, continuous)
- Performance_Level (categorical: High vs Low — optional)

Predictors

- Reading_Score, Writing_Score
- Gender, Lunch_Type, Test_Preparation_Course, Parental_Education

Why Model Selection Matters

- **Different problems, different models**

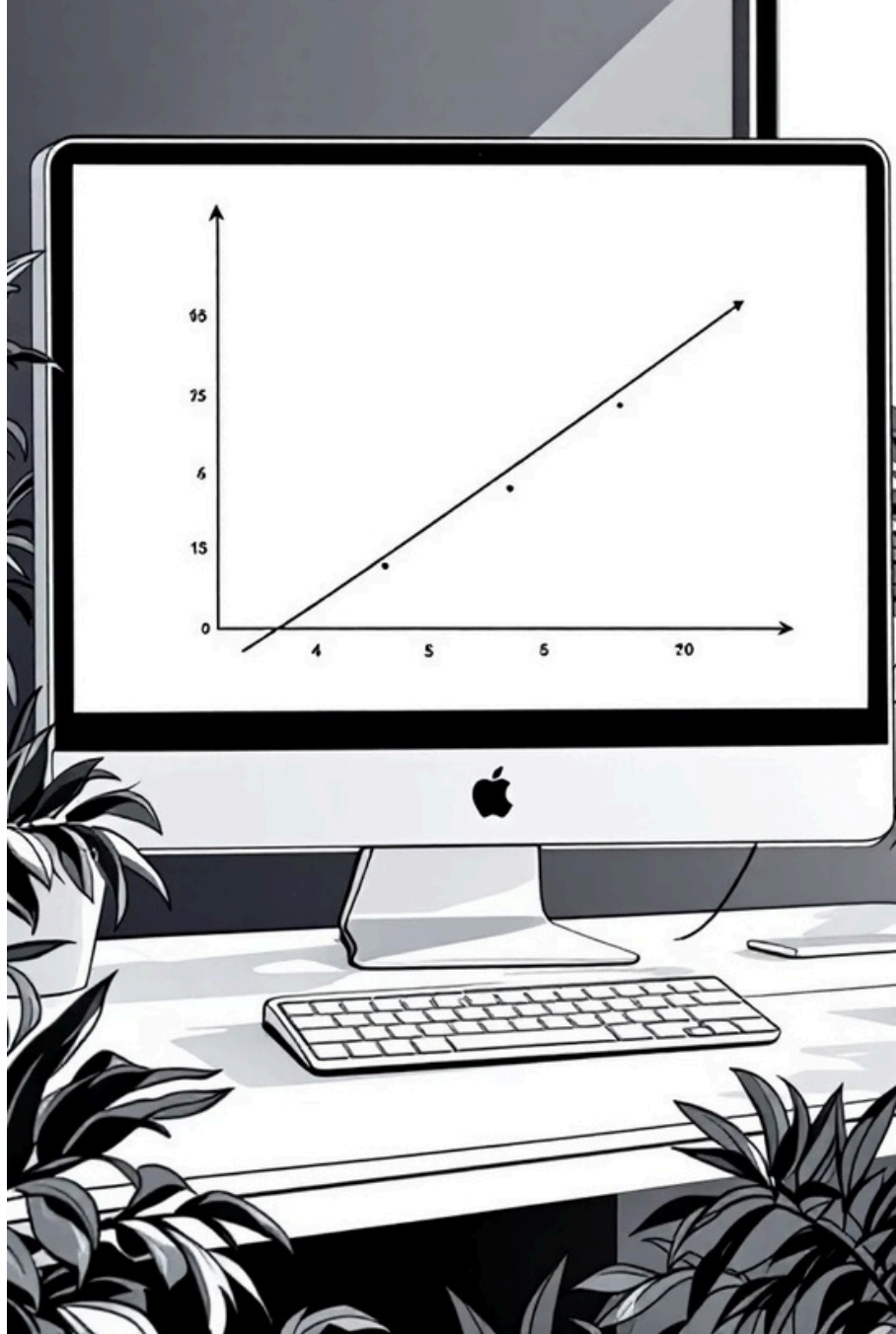
  Wrong model choice can lead to misleading conclusions.

- **Selection depends on:**
  - Typeof dependentvariable
  - Nature of predictors
  - Research objective

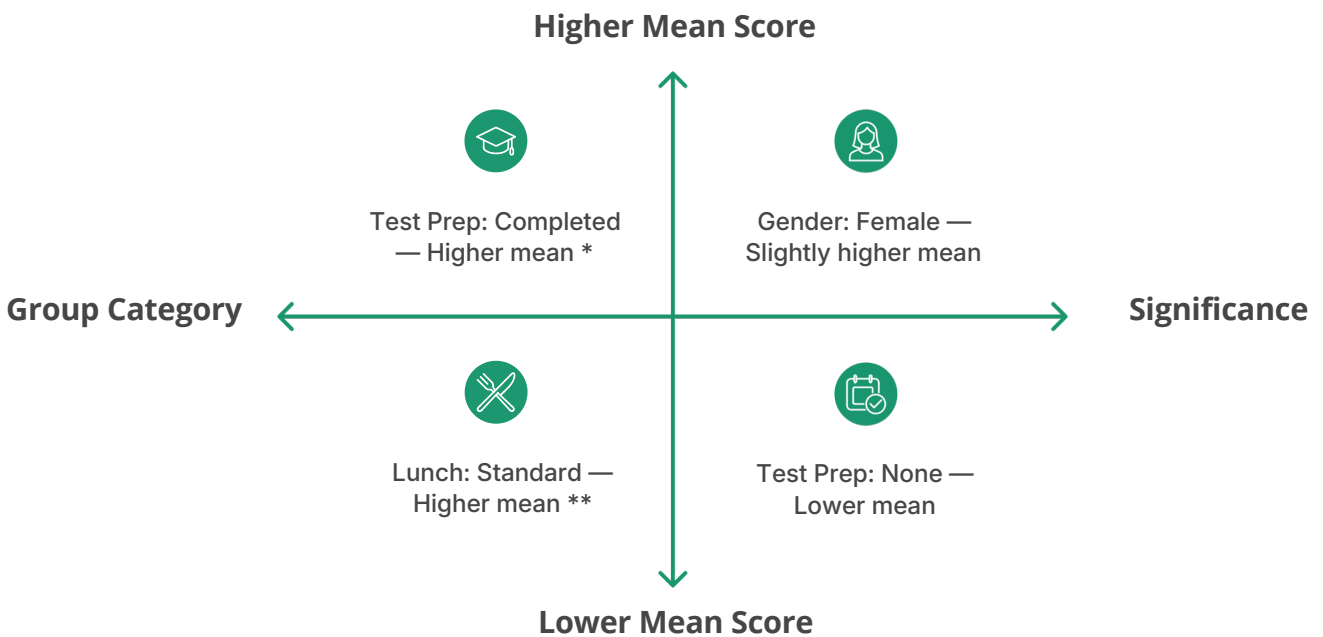
Multiple Linear Regression

Use line a regression to predict continuous Math_Scorefrom academic and demographic variables and identifythe mostinfluential factors.

Rationale: Math score is continuous; regression quantifies factor effects.



ANOVA for Group Comparisons



ANOVA compares mean exam scores across categorical groups (test preparation, gender, lunch type) to detect statistically significant differences.

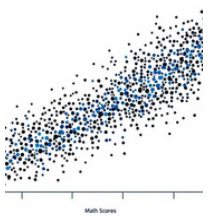
Purpose: Distinguish whether group means differ meaningfully.

Logistic Regression for Categories

When performance is classified (High vs Low), logistic regression estimates probability of high performance using academic and background variables.

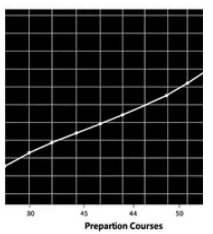
Purpose: Rank students by predicted probability to support decision-making.

Feature Selection



ReadingScore

Strong relationship with math score observed in scatter plots.



WritingScore

Box plots reveal differences across preparation groups.

TestPreparation, LunchType, Gender

Chosenfor statistical relevance and educational meaning.

Hypotheses & Summary

01

Regression Hypotheses

H0:Reading score has no effect on math score. H1: Higher reading score increases math score.

02

Regression Hypothesis 2

H0:Writings core has no effect on math score. H1: Higher writing score increases math score.

03

ANOVA Hypothesis

H0:Mean math scores equal across test preparation groups. H1: Means differ between groups.

04

Logistic Hypothesis

H0:Academic and demographic variables do not predict performance category. H1: They are significant predictors (key: Reading, Writing, TestPreparation).

Conclusion: Reading, writing and preparation strongly influence performance. Appropriate model selection (regression, ANOVA, logistic) and careful feature choice from EDA ensure reliable predictions of academic performance.