

Final Report

Huanyu Chen, Yifei Liu, Xiaoting Tang, Longyu Zhang

2023-12-16

not exceed 5 double-spaced pages using 11 or 12-point font EXCLUDING figures and tables, references, appendix, that can be placed at the end of the main text.

Abstract

(condenses a brief introduction, brief description of methods, and main results into a one-paragraph summary)

Introduction

(brief context and background of the problem)

- use predictors to predict Math, Reading, and Writing scores
- determine which factors affect each score

Methods

(data description and statistical methods)

- data description: This dataset includes three test scores of students at a public school and a variety of personal and socioeconomic factors
- changed the categorical data from words to numbers
- for NAs, take the mean of the column, and replace the NAs with mean value.
- *a descriptive table with summary statistics for all variables created, but has a rounding problem*
- distribution of the response variable coded in histogram and boxplot
- in diagnostics, all linear
- no transformation needed for all three
- no outliers showed in plot, but for influential points
- multicollinearity checked. no correlated variables, VIF all small
- two models created for each response. one backward (better compared to forward), one lasso

Results

- final model for each test score
- analyze between models

Conclusions/Discussion

A brief summary on each group member's contribution (method, data analysis, writing, etc).

Appendix

Analytical questions (you may only answer some of them but properly addressing more will improve the rate of your report): 1. Using variables 1-11 as the covariates to predict Math, Reading and Writing scores. 2. Which factors (features) affect each test score significantly? Are there interacting effects? 3. Are the optimal prediction models similar or different across the three test scores? Is it possible to leverage one score as the auxiliary information to learn the model for another score (still its model against variables 1-11) better?

describe your final model and interpret its parameters. examine the marginal distributions and pairwise relationships between variables (e.g., nonlinearities) explore several candidate models, and explain why you selected your model. be clear about your motivation for carrying out certain analyses. Your report should include a table summarizing parameter estimates associated with your final fitted model, characterizing predictor variables.

- Data exploration: descriptive statistics and visualization.
- In your regression model, be watchful for variables that are highly correlated and be selective in the variables you will include in your analysis. selective interactions between variables.

Grading: Method and interpretation: Understand the background and task of the project; Preprocess of the data; Proper choice and decision on models' assumption, specification, diagnostic, selection, comparison, and validation; Strategies and reasons of your designs and decisions should be included. Correct and informative interpretation of the results.

Writing and result display: Well-organized report with professional writing and structuring; Proper introduction of the method with helpful and informative explanations; Display the data and analysis results with tables and plots. Be concise and avoid displaying redundant or useless results; Do not exceed the page limit.