Final Project Paper

Ekaterina Hofrenning, Lauren Lazaro, Arthur Starodynov, Yuki Low

## 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)

It is well-documented that students coming from families with higher socio-economic statuses see higher average test scores. But, there are other factors such as study time, sports participation, ethnic group, and number of siblings that may interact with other co-variates and impact test scores. We are interested in exploring the impact of various personal and socio-economic factors on test scores within a public school. The test scores in question are related to math, reading, and writing.

## Methods (data description and statistical methods)

The dataset is a sample of 948 individuals from a public school with math, reading, and writing test scores and the following variables of interest: gender, ethnic group, parental education, lunch type, test prep, parent’s marital status, sports participation, first child, number of siblings, means of transportation to school, number of weekly study hours. Due to significant missing data, we imputed the missing values with the average or most common value per variable.

First, we conducted a data description (Table 1), computing the averages and standard deviations for continuous variables and counts and percentages for categorical variables. Our sample is well-balanced across sex. The sample is composed of majority first born children, children on standard fee lunch, children with married parents, school bus riders, and children who study 5-10 hours. The average number of siblings is 2 and each test score (math, writing, reading) average ranges between 66-69 points.

| **Characteristic** | **N = 948**1 |
| --- | --- |
| Gender |  |
| female | 488 / 948 (51%) |
| male | 460 / 948 (49%) |
| EthnicGroup |  |
| group A | 80 / 948 (8.4%) |
| group B | 171 / 948 (18%) |
| group C | 336 / 948 (35%) |
| group D | 237 / 948 (25%) |
| group E | 124 / 948 (13%) |
| ParentEduc |  |
| some high school | 163 / 948 (17%) |
| high school | 176 / 948 (19%) |
| associate's degree | 198 / 948 (21%) |
| some college | 252 / 948 (27%) |
| bachelor's degree | 104 / 948 (11%) |
| master's degree | 55 / 948 (5.8%) |
| LunchType |  |
| free/reduced | 331 / 948 (35%) |
| standard | 617 / 948 (65%) |
| TestPrep |  |
| completed | 322 / 948 (34%) |
| none | 626 / 948 (66%) |
| ParentMaritalStatus |  |
| divorced | 146 / 948 (15%) |
| married | 565 / 948 (60%) |
| single | 213 / 948 (22%) |
| widowed | 24 / 948 (2.5%) |
| PracticeSport |  |
| never | 112 / 948 (12%) |
| sometimes | 493 / 948 (52%) |
| regularly | 343 / 948 (36%) |
| IsFirstChild | 634 / 948 (67%) |
| NrSiblings | 2 (1) |
| TransportMeans |  |
| private | 337 / 948 (36%) |
| school\_bus | 611 / 948 (64%) |
| WklyStudyHours |  |
| < 5 | 253 / 948 (27%) |
| 5-10 | 545 / 948 (57%) |
| > 10 | 150 / 948 (16%) |
| MathScore | 66 (16) |
| ReadingScore | 69 (15) |
| WritingScore | 68 (15) |
| 1n / N (%); Mean (SD) | |

After looking at the distribution of the continuous outcomes, we decided that the variables looked fairly normal so we did not use log transformations.

In order to predict test score from our variables of interest, we decided to use backwards stepwise regression for simplicity. For math score as the outcome, this gave the optimal predictors of Gender + EthnicGroup + ParentEduc + LunchType + TestPrep + ParentMaritalStatus + PracticeSport + IsFirstChild + WklyStudyHours.

Table 2. Math score backwards stepwise model

| term | estimate | p.value |
| --- | --- | --- |
| (Intercept) | 55.976 | <.001 |
| Gendermale | 4.981 | <.001 |
| EthnicGroup.L | 7.401 | <.001 |
| EthnicGroup.Q | 2.940 | 0.011 |
| EthnicGroup.C | 0.737 | 0.482 |
| EthnicGroup^4 | -0.985 | 0.251 |
| ParentEduc.L | 6.642 | <.001 |
| ParentEduc.Q | -0.045 | 0.972 |
| ParentEduc.C | -0.300 | 0.806 |
| ParentEduc^4 | 1.402 | 0.213 |
| ParentEduc^5 | -2.644 | 0.006 |
| LunchTypestandard | 11.155 | <.001 |
| TestPrepnone | -5.582 | <.001 |
| ParentMaritalStatusmarried | 3.876 | 0.002 |
| ParentMaritalStatussingle | 1.102 | 0.444 |
| ParentMaritalStatuswidowed | 5.066 | 0.087 |
| PracticeSport.L | 2.514 | 0.015 |
| PracticeSport.Q | -0.518 | 0.503 |
| IsFirstChildyes | 2.368 | 0.011 |
| WklyStudyHours.L | 2.621 | 0.008 |
| WklyStudyHours.Q | -1.032 | 0.158 |

Table 3. Reading score backwards stepwise model

| term | estimate | p.value |
| --- | --- | --- |
| (Intercept) | 68.495 | <.001 |
| Gendermale | -7.282 | <.001 |
| EthnicGroup.L | 4.149 | 0.001 |
| EthnicGroup.Q | 1.438 | 0.203 |
| EthnicGroup.C | -0.471 | 0.645 |
| EthnicGroup^4 | -0.942 | 0.261 |
| ParentEduc.L | 7.638 | <.001 |
| ParentEduc.Q | 1.535 | 0.22 |
| ParentEduc.C | 0.566 | 0.635 |
| ParentEduc^4 | 1.547 | 0.158 |
| ParentEduc^5 | -3.033 | 0.001 |
| LunchTypestandard | 7.494 | <.001 |
| TestPrepnone | -6.972 | <.001 |
| ParentMaritalStatusmarried | 4.113 | 0.001 |
| ParentMaritalStatussingle | 1.275 | 0.363 |
| ParentMaritalStatuswidowed | 4.645 | 0.106 |
| IsFirstChildyes | 2.446 | 0.007 |
| WklyStudyHours.L | 1.431 | 0.135 |
| WklyStudyHours.Q | -0.933 | 0.191 |

Table 4. Writing score backwards stepwise model

| term | estimate | p.value |
| --- | --- | --- |
| (Intercept) | 69.347 | <.001 |
| Gendermale | -9.209 | <.001 |
| EthnicGroup.L | 4.673 | <.001 |
| EthnicGroup.Q | 0.627 | 0.569 |
| EthnicGroup.C | -1.891 | 0.058 |
| EthnicGroup^4 | -1.649 | 0.043 |
| ParentEduc.L | 9.983 | <.001 |
| ParentEduc.Q | 1.365 | 0.265 |
| ParentEduc.C | 0.289 | 0.803 |
| ParentEduc^4 | 1.715 | 0.109 |
| ParentEduc^5 | -3.106 | 0.001 |
| LunchTypestandard | 8.388 | <.001 |
| TestPrepnone | -9.629 | <.001 |
| ParentMaritalStatusmarried | 4.135 | 0.001 |
| ParentMaritalStatussingle | 1.056 | 0.44 |
| ParentMaritalStatuswidowed | 3.950 | 0.16 |
| PracticeSport.L | 2.251 | 0.022 |
| PracticeSport.Q | -0.708 | 0.335 |
| IsFirstChildyes | 2.208 | 0.012 |
| WklyStudyHours.L | 1.338 | 0.152 |
| WklyStudyHours.Q | -0.960 | 0.167 |

## Results

## Conclusions/Discussion

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