Factors Affecting Educational Success: A Deep Look into Gender, Ethnicity, and Socio-Economic Variables

# Introduction

Over the years, there has been a heavy focus by educational sectors on studies that moves the achievement of students. A study that should be looked at is how different regions, as well as economic circumstances impact the performance of students. The purpose of this study is to examine the relationship between gender groups, race/ethnicity, parent’s level of education,type of lunch taken by students, test prep and student’s scores in Math, Reading, as well as Writing.

The dataset under consideration for this research has 1000 observations with 8 variables. The “gender” variable is a variable with 2 levels: female and male. The “race.ethnicity” variable is also and it contains 5 levels: group A, group B, group C, group D, and group E .The “parental.level.of education” variable is a variable that represents the level of education that parents of the students have received. It has 6 levels which are some high school, some college, masters degree, bachelors degree, associates degree, and high school.

“Lunch” is also a variable that tells us the kind of lunch that a student takes in school. It has two levels: free/reduced and standard. The last variable, “test.preparation.course” is a variable that represents whether or not a student took a test preparation course. It has two levels: completed and none. The other three variables which are “math.score”, “reading.score,” and “writing.score” are continuous variables of the scores that the students got on these tests.

The purpose of this research project is to answer research questions. The first objective is to determine whether or not there is any difference among students based on gender in mathematics. Next, the paper will establish the relationship that exists between ethnicity groups and student performances .It also looks to determine the roles that education level of parents plays in influencing the learning outcomes of students. Finally, it will attempt to consider the impact that the type of lunch students take has on their academic performance and also find out if there is any difference in student performance when students complete a test preparation course or not.

For this analysis, we will use R. R as a statistical tool provides a series of data manipulation, data visualization , and modelling tools for statistical analyses. We will use R as a data mining tool to mine the dataset for relevant statistics and gain insightful and meaningful conclusions. The next part presents a review of some relevant literatures for each variable considered in the data set.

Research on gender has been a topic of interest in educational research over the years. A number of studies on gender gap in academic achievement propose different discoveries. While some studies show that males do better than women in mathematics and science subjects, others reveal that on the contrary females surpass males in reading and writing courses (Hyde,2014). Nevertheless, other studies have identified lack of gender inequality on educational performance (Halpern, 2012). If an institution wants to develop a useful strategy to improve the achievements of students, they need to understand the influence of gender on academic achievement.

Also, researchers have also thoroughly looked at another powerful variable – race/ethnicity. Research has made it known huge disparities exist within ethnic groups, in educational achievement (Fryer & Levitt, 2004). Diverse differences also exist among students which could be related with cultural and social economics, availability of resources and opportunities for education (Ogbu, 2003; Steele & Aronson, 1995). Investigating the relationship between race/ethnicity and the achievement gap promotes an understanding of how to we can close gaps in academic performance.

Regarding socio-economic variables, parent’s education level is one of the key variables that correlates well with the academic success of students. Previous studies have shown evidence in support of this claim. Moreover, other studies have indicated that students from homes where the parents have higher levels of education perform better in school than others (Duncan & Magnuson, 2012; Sirin, 2005). Understanding how parents’ level of education affects educational outcomes for all students could assist interventions that looks at improving these outcomes among the less privileged students.

Lunch status is another social economic aspect that shows up in relation to academic achievement. Studies indicate that students that are eligible for free or reduced lunch perform below expectation academically and are more likely to repeat a grade. In this sense, the kind of sandwich that a student eats reflects his or her socio-economic position and educational opportunities (Brooks-Gunn & Duncan, 1997).

Over the years, test preparation courses have become popular. They have been used to improve student performance in standardized tests. These courses intend to teach students how to perform well in their tests. It has been claimed that test preparation courses boost test performance in some studies (Hembree, 1988; Kuncel et al., 2001), but others say they have small, zero, or even negative impacts (Burgess & Greaves, 2013). Assessing how test preparation courses influence student’s performance can aid the educational process on test preparation itself.

Finally, this study is set out to examine the relationship between gender, race/ethnicity, parents’ level of education, lunch status, test preparation courses , and students’ performance in mathematics, reading, and writing. This study will add to the existing literature . It will do so by analysing a dataset of 1000. This study’s findings can guide the practices of educators and policies to improve education outcomes for all learners.

# Method

For this empirical study, R will be used to analyze the data set and find out relationsips between different variables. This will be based on descriptive statistics, data visualization, and inferential statistics methods.

Firstly , we’ll go through the data cleaning process by checking for structural errors, missing data, outliers and duplicates.

# Loading our data  
exams <- read.csv("~/Downloads/exams.csv")  
# Converting categorical variables to factor  
exams$gender <- as.factor(exams$gender)  
exams$race.ethnicity<-as.factor(exams$race.ethnicity)  
exams$parental.level.of.education<-as.factor(exams$parental.level.of.education)  
exams$lunch<-as.factor(exams$lunch)  
exams$test.preparation.course<-as.factor(exams$test.preparation.course)  
# Checking for missing values  
miss<-sum(is.na(exams))  
# Checking for duplicates  
dup<-sum(duplicated(exams))

After our data is clean and ready for analysis, we will use data visualization approaches like bar charts to show the average scores across each levels of the categorical variables.

# Plot1  
library(ggplot2)  
plot1<-ggplot(exams, aes(x = parental.level.of.education, y = math.score, fill = parental.level.of.education)) +  
 geom\_bar(stat = "summary", fun = "mean") +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1)) +  
 labs(title = "Average Math Scores by Parental Level of Education", x = "Parental Level of Education", y = "Average Math Score")  
  
  
# Plot 3  
plot3<-ggplot(exams, aes(x = gender, y = math.score, fill = gender)) +  
 geom\_bar(stat = "summary", fun = "mean") +  
 labs(title = "Average Math Scores by Gender", x = "Gender", y = "Average Math Score")  
  
# Plot 2  
plot2<-ggplot(exams, aes(x = race.ethnicity, y = math.score, fill = race.ethnicity)) +  
 geom\_bar(stat = "summary", fun = "mean") +  
 theme(axis.text.x = element\_text(angle = 45, hjust = 1)) +  
 labs(title = "Average Math Scores by ethnicity", x = "Ethnicity", y = "Average Math Score")  
  
  
# Plot 4  
plot4<-ggplot(exams, aes(x = lunch, y = math.score, fill = lunch)) +  
 geom\_bar(stat = "summary", fun = "mean") +  
 labs(title = "Average Math Scores by type of lunch taken", x = "Lunch", y = "Average Math Score")  
  
  
# Plot 5   
plot5<-ggplot(exams, aes(x = test.preparation.course, y = math.score, fill = test.preparation.course)) +  
 geom\_bar(stat = "summary", fun = "mean") +  
 labs(title = "Average Math Scores by preparation test completion status ", x = "Lunch", y = "Average Math Score")

“Plot 1” will show the distribution of math scores across all levels of the parental level of education.

“Plot 3” will show the distribution of math scores across both gender(male and female).

“Plot 2” will show the distribution of math scores across all levels of ethnicity groups.

“Plot 4” will show the distribution of math scores across all levels of the type of lunch taken by students.

“Plot 5” will show the distribution of math scores across all levels of test score completion status among students.

After making our plots, we will then proceed to using inferential statistics on our data, where we will use the one-way ANOVA test to determine whether or not there is a significant difference in mean math scores scores across all levels of parents education. Also, whether or not there is a significant difference in mean math scores scores across the ethnicity groups. We won’t just stop at the test of significance, we’ll also do a post-hoc test to see which of the groups are significantly different from each other

# ANOVA 1  
  
# Loading required libraries  
library(datarium)  
library(rstatix)  
# Selecting the data we need   
anovadata<-select(exams,parental.level.of.education,math.score)  
anovagrouping<-group\_by(anovadata,parental.level.of.education)  
# Getting descriptive statistics for our data  
stat1<-get\_summary\_stats(anovagrouping,type="mean\_sd")  
# Checking for outliers  
out1<-identify\_outliers(anovagrouping,math.score)  
# Checking for normality  
norm1<-shapiro\_test(anovagrouping,math.score)  
# Checking for homogeneity of variances  
vr1<-levene\_test(anovadata,math.score~parental.level.of.education)  
# ANOVA  
Ind.anova1<-aov(math.score~parental.level.of.education,anovadata)  
anova1<-Anova(Ind.anova1, type = "III")  
# Post-hoc test  
post1<-TukeyHSD(Ind.anova1)  
  
  
# ANOVA 2  
  
# Selecting the data we need  
racedata<-select(exams,race.ethnicity,math.score)  
anovagrouping2<-group\_by(racedata,race.ethnicity)  
# Getting descriptive statistics for our data  
stat2<-get\_summary\_stats(anovagrouping2,type="mean\_sd")  
# Checking for outliers  
out2<-identify\_outliers(anovagrouping2,math.score)  
# Checking for normality  
norm2<-shapiro\_test(anovagrouping2,math.score)  
# Checking for homogeneity of variances  
vr2<-levene\_test(racedata,math.score~race.ethnicity)  
# ANOVA  
Ind.anova2<-aov(math.score~race.ethnicity,racedata)  
anova2<-Anova(Ind.anova2, type = "III")  
# Post-hoc test  
post2<-TukeyHSD(Ind.anova2)

We’ll then use t-test to test the following:

1. Whether or not is a significant difference in average math scores scores between male and female gender.
2. Whether or not there is a significant difference in average math scores between students who receive “free/reduced” lunch and those who receive “standard” lunch.
3. Whether or not there is a significant difference in average math scores between students who have completed a test preparation course and those who have not.

We’ll also find the effect sizes of these differences

# T-test 1  
  
# Loading required libraries  
library(datarium)  
library(rstatix)  
# Selecting the data we need  
tdata<-select(exams,gender,math.score)  
# Getting descriptive statistics for our data  
anovagrouping3<-group\_by(tdata,gender)  
stat3<-get\_summary\_stats(anovagrouping3,type="mean\_sd")  
# Checking for outliers  
out3<-identify\_outliers(anovagrouping3,math.score)  
# Checking for normality  
norm3<-shapiro\_test(anovagrouping3,math.score)  
# Checking for equality of variances  
vr3<-levene\_test(tdata,math.score~gender)  
# T-test   
test3<-t\_test(exams, math.score ~ gender, var.equal=TRUE)  
# Effect size calculation  
effsize3<-cohens\_d(math.score ~ gender,data=exams,paired = FALSE)  
  
# T-test 2  
  
# Selecting the data we need  
lunchdata<-select(exams,lunch,math.score)  
# Getting descriptive statistics for our data  
anovagrouping4<-group\_by(lunchdata,lunch)  
stat4<-get\_summary\_stats(anovagrouping4,type="mean\_sd")  
# Checking for outliers  
out4<-identify\_outliers(anovagrouping4,math.score)  
# Checking for normality  
norm4<-shapiro\_test(anovagrouping4,math.score)  
# Checking for equality of variances  
vr4<-levene\_test(lunchdata,math.score~lunch)  
# T-test  
test4<-t\_test(exams, math.score ~ lunch, var.equal=TRUE)  
# Effect size calculation  
effsize4<-cohens\_d(math.score ~ lunch,data=exams,paired = FALSE)  
  
# T-test 3  
  
# Selecting the data we need  
coursedata<-select(exams,test.preparation.course,math.score)  
# Getting descriptive statistics for our data  
anovagrouping5<-group\_by(coursedata,test.preparation.course)  
stat5<-get\_summary\_stats(anovagrouping5,type="mean\_sd")  
# Checking for outliers  
out5<-identify\_outliers(anovagrouping5,math.score)  
# Checking for normality  
norm5<-shapiro\_test(anovagrouping5,math.score)  
# Checking for homogeneity of variances  
vr5<-levene\_test(coursedata,math.score~test.preparation.course)  
# T-test  
test5<-t\_test(exams, math.score ~ test.preparation.course, var.equal=TRUE)  
# Effect size calculation  
effsize5<-cohens\_d(math.score ~ test.preparation.course,data=exams,paired = FALSE)

After the t-test, we’ll use regression to determine which of “gender”, “race.ethnicity”,“parental.level.of.education”,“lunch”,“test.preparation.course” is a significant predictor of math scores.

model1<-lm(math.score~gender+race.ethnicity+parental.level.of.education+lunch+test.preparation.course,data=exams)  
summary<-summary(model1)

# Results

Our data contained no missing values,outliers and duplicates, so we didn’t have to do much in cleaning our data. We’ll now take a look at the most important results from the analyses we have done.

### Table 1.

Descriptive statistics for ANOVA 1

# A tibble: 6 × 5  
 parental.level.of.education variable n mean sd  
 <fct> <fct> <dbl> <dbl> <dbl>  
1 associate's degree math.score 204 70.3 14.8  
2 bachelor's degree math.score 105 69.9 14.3  
3 high school math.score 215 65.4 16.0  
4 master's degree math.score 75 71.0 14.2  
5 some college math.score 224 68.6 14.6  
6 some high school math.score 177 64.2 15.7

### Table 2.

Test result for ANOVA 1

Anova Table (Type III tests)  
  
Response: math.score  
 Sum Sq Df F value Pr(>F)   
(Intercept) 1009565 1 4438.9504 < 2.2e-16 \*\*\*  
parental.level.of.education 6267 5 5.5113 5.162e-05 \*\*\*  
Residuals 226069 994   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

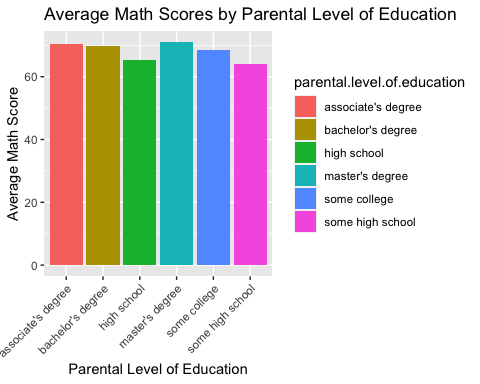
### Table 3.

Post-hoc result for ANOVA 1

Tukey multiple comparisons of means  
 95% family-wise confidence level  
  
Fit: aov(formula = math.score ~ parental.level.of.education, data = anovadata)  
  
$parental.level.of.education  
 diff lwr upr  
bachelor's degree-associate's degree -0.4813725 -5.6532187 4.6904736  
high school-associate's degree -4.9666439 -9.1753518 -0.7579359  
master's degree-associate's degree 0.6786275 -5.1361417 6.4933966  
some college-associate's degree -1.7051821 -5.8725216 2.4621575  
some high school-associate's degree -6.1502991 -10.5735020 -1.7270962  
high school-bachelor's degree -4.4852713 -9.6119631 0.6414204  
master's degree-bachelor's degree 1.1600000 -5.3500911 7.6700911  
some college-bachelor's degree -1.2238095 -6.3165950 3.8689760  
some high school-bachelor's degree -5.6689266 -10.9731201 -0.3647330  
master's degree-high school 5.6452713 -0.1293729 11.4199155  
some college-high school 3.2614618 -0.8497052 7.3726288  
some high school-high school -1.1836552 -5.5539755 3.1866650  
some college-master's degree -2.3838095 -8.1283732 3.3607542  
some high school-master's degree -6.8289266 -12.7617182 -0.8961349  
some high school-some college -4.4451170 -8.7756130 -0.1146211  
 p adj  
bachelor's degree-associate's degree 0.9998209  
high school-associate's degree 0.0101350  
master's degree-associate's degree 0.9994571  
some college-associate's degree 0.8518793  
some high school-associate's degree 0.0010828  
high school-bachelor's degree 0.1255741  
master's degree-bachelor's degree 0.9958481  
some college-bachelor's degree 0.9834947  
some high school-bachelor's degree 0.0281921  
master's degree-high school 0.0596590  
some college-high school 0.2094640  
some high school-high school 0.9719649  
some college-master's degree 0.8441831  
some high school-master's degree 0.0133826  
some high school-some college 0.0403236

### Plot 1.

Distribution of average math scores across parental levels of education



### Table 4

Descriptive statistics for ANOVA 2

# A tibble: 5 × 5  
 race.ethnicity variable n mean sd  
 <fct> <fct> <dbl> <dbl> <dbl>  
1 group A math.score 79 65.7 12.5  
2 group B math.score 198 64.1 14.6  
3 group C math.score 323 65.5 14.6  
4 group D math.score 257 68.9 15.8  
5 group E math.score 143 77.4 13.9

### Table 5.

Test result for ANOVA 2

Anova Table (Type III tests)  
  
Response: math.score  
 Sum Sq Df F value Pr(>F)   
(Intercept) 340963 1 1585.404 < 2.2e-16 \*\*\*  
race.ethnicity 18347 4 21.328 < 2.2e-16 \*\*\*  
Residuals 213989 995   
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

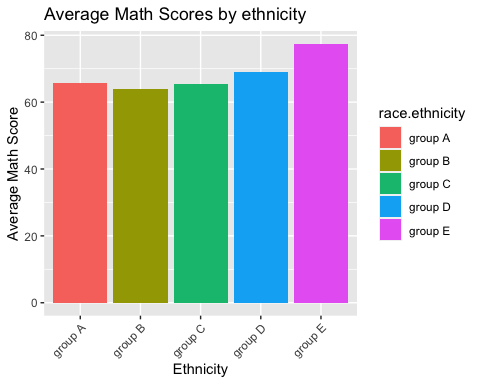
### Table 6.

Post-hoc test result for ANOVA 2

Tukey multiple comparisons of means  
 95% family-wise confidence level  
  
Fit: aov(formula = math.score ~ race.ethnicity, data = racedata)  
  
$race.ethnicity  
 diff lwr upr p adj  
group B-group A -1.6254955 -6.95863750 3.707647 0.9203892  
group C-group A -0.1853666 -5.21559411 4.844861 0.9999768  
group D-group A 3.1831749 -1.97242181 8.338772 0.4422778  
group E-group A 11.7303709 6.11233733 17.348404 0.0000002  
group C-group B 1.4401288 -2.17708690 5.057345 0.8127882  
group D-group B 4.8086704 1.01904788 8.598293 0.0049592  
group E-group B 13.3558664 8.95775742 17.753975 0.0000000  
group D-group C 3.3685415 0.01861269 6.718470 0.0479636  
group E-group C 11.9157375 7.89030004 15.941175 0.0000000  
group E-group D 8.5471960 4.36615119 12.728241 0.0000003

### Plot 2

Distribution of average math scores across ethnicity groups



### Table 7

Descriptive statistics for t-test 1

# A tibble: 2 × 5  
 gender variable n mean sd  
 <fct> <fct> <dbl> <dbl> <dbl>  
1 female math.score 492 64.8 15.1  
2 male math.score 508 70.8 14.8

### Table 8

Test result for t-test 1

# A tibble: 1 × 8  
 .y. group1 group2 n1 n2 statistic df p  
\* <chr> <chr> <chr> <int> <int> <dbl> <dbl> <dbl>  
1 math.score female male 492 508 -6.31 998 4.08e-10

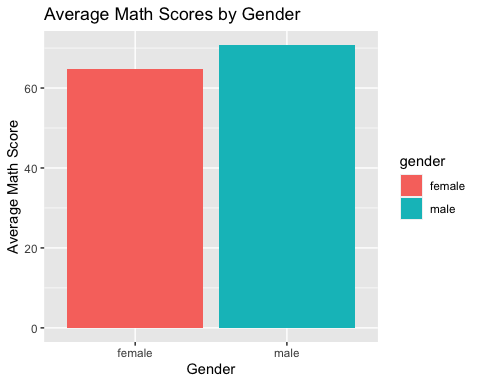
### Table 9

Effect size result for t-test 1

# A tibble: 1 × 7  
 .y. group1 group2 effsize n1 n2 magnitude  
\* <chr> <chr> <chr> <dbl> <int> <int> <ord>   
1 math.score female male -0.399 492 508 small

### Plot 3

Distribution of average math scores across the male and female gender



### Table 10

Descriptive statistics for t-test 2

# A tibble: 2 × 5  
 lunch variable n mean sd  
 <fct> <fct> <dbl> <dbl> <dbl>  
1 free/reduced math.score 340 59.9 14.0  
2 standard math.score 660 71.9 14.3

### Table 11

Test result for t-test 2

# A tibble: 1 × 8  
 .y. group1 group2 n1 n2 statistic df p  
\* <chr> <chr> <chr> <int> <int> <dbl> <dbl> <dbl>  
1 math.score free/reduced standard 340 660 -12.7 998 2.90e-34

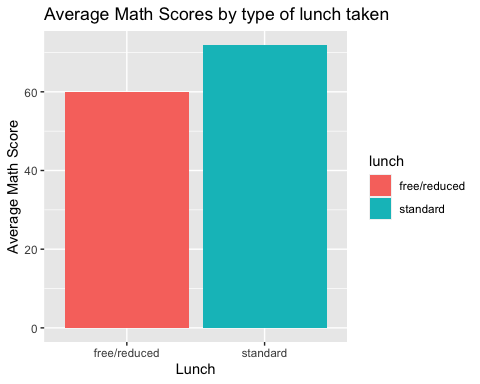
### Table 12

Effect size results for t-test 2

# A tibble: 1 × 7  
 .y. group1 group2 effsize n1 n2 magnitude  
\* <chr> <chr> <chr> <dbl> <int> <int> <ord>   
1 math.score free/reduced standard -0.849 340 660 large

### Plot 4

Distribution of average math scoes acros the types of lunch taken by students



### Table 13

Descriptive statistics for t-test 3

# A tibble: 2 × 5  
 test.preparation.course variable n mean sd  
 <fct> <fct> <dbl> <dbl> <dbl>  
1 completed math.score 344 70.3 14.7  
2 none math.score 656 66.5 15.4

### Table 14

Test results for t-test 3

# A tibble: 1 × 8  
 .y. group1 group2 n1 n2 statistic df p  
\* <chr> <chr> <chr> <int> <int> <dbl> <dbl> <dbl>  
1 math.score completed none 344 656 3.82 998 0.000144

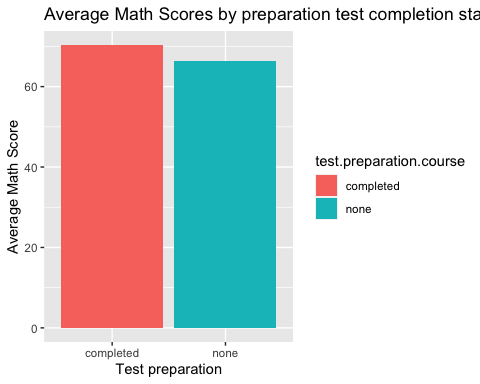
### Table 15

Effect size result for t-test 3

# A tibble: 1 × 7  
 .y. group1 group2 effsize n1 n2 magnitude  
\* <chr> <chr> <chr> <dbl> <int> <int> <ord>   
1 math.score completed none 0.256 344 656 small

### Plot 5

Distribution of average math scores across the levels of test preparation



### Table 16

Regression result

summary(model1)

Call:  
lm(formula = math.score ~ gender + race.ethnicity + parental.level.of.education +   
 lunch + test.preparation.course, data = exams)  
  
Residuals:  
 Min 1Q Median 3Q Max   
-42.485 -8.618 0.446 9.600 34.498   
  
Coefficients:  
 Estimate Std. Error t value  
(Intercept) 60.02668 1.83477 32.716  
gendermale 5.73387 0.81536 7.032  
race.ethnicitygroup B 0.08959 1.71562 0.052  
race.ethnicitygroup C 0.79180 1.61825 0.489  
race.ethnicitygroup D 3.55649 1.65452 2.150  
race.ethnicitygroup E 12.47640 1.80350 6.918  
parental.level.of.educationbachelor's degree -0.68397 1.54502 -0.443  
parental.level.of.educationhigh school -4.60603 1.25646 -3.666  
parental.level.of.educationmaster's degree 0.01050 1.73550 0.006  
parental.level.of.educationsome college -2.45955 1.24447 -1.976  
parental.level.of.educationsome high school -7.26432 1.31870 -5.509  
lunchstandard 12.32913 0.86257 14.294  
test.preparation.coursenone -5.09191 0.86046 -5.918  
 Pr(>|t|)   
(Intercept) < 2e-16 \*\*\*  
gendermale 3.79e-12 \*\*\*  
race.ethnicitygroup B 0.95837   
race.ethnicitygroup C 0.62474   
race.ethnicitygroup D 0.03183 \*   
race.ethnicitygroup E 8.23e-12 \*\*\*  
parental.level.of.educationbachelor's degree 0.65808   
parental.level.of.educationhigh school 0.00026 \*\*\*  
parental.level.of.educationmaster's degree 0.99518   
parental.level.of.educationsome college 0.04839 \*   
parental.level.of.educationsome high school 4.61e-08 \*\*\*  
lunchstandard < 2e-16 \*\*\*  
test.preparation.coursenone 4.50e-09 \*\*\*  
---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
  
Residual standard error: 12.82 on 987 degrees of freedom  
Multiple R-squared: 0.3016, Adjusted R-squared: 0.2931   
F-statistic: 35.52 on 12 and 987 DF, p-value: < 2.2e-16

# Discussions

We’ll start by discussing the results of our ANOVA tests. We ran the first ANOVA test to determine to determine whether or not there is a significant difference in mean math scores scores across all levels of parental level of education.

From the 1000 students under consideration, 204 students have parents with associate’s degree,105 students have parents with bachelor’s degree, 215 students have parents with some high school certificate, 75 students have parents with masters degree, 224 students have parents with some college degree and 177 students have parents with some high school certificate. The data contained no extreme outliers. Shapiro test demonstrated the data to be normal by group. Levene-test concluded that the the variances are homogeneous. As seen in Table 1.The average math scores for students whose parents have an associate’s degree is 70.348(SD= 14.822). The average math scores for students whose parents have an bachelor’s degree is 69.867(SD= 14.262).The average math scores for students whose parents have an high school certificate is 65.381(SD= 15.971). The average math scores for students whose parents have a masters degree is 71.027(SD= 14.190). The average math scores for students whose parents have some college degree is 68.643(SD= 14.553). The average math scores for students whose parents have some high school certificate is 64.198(SD= 15.740).From this we can conclude that students with parents having a “master’s degree” tend to have the highest average math score. As seen in table 2, “parental.level.of.education” has a very small p-value of 0.00005162. It is safe to reject the null hypothesis and conclude that there’s a significant difference in average math scores scores across all levels of parental education. Even though we know that there’s a significant difference across al levels of parental education, we didn’t just stop there. We ran a post-hoc test to determine which levels have a significant difference. Table 3 shows us exactly that.

We can see that there’s a significant difference among the “high school” and “associate’s degree” levels, since they have a p-value of 0.01. There’s a significant difference among the “some high school” and “associate’s degree” levels, since they have a p-value of 0.001. There’s a significant difference among the “some high school” and “bachelor’s degree” levels, since they have a p-value of 0.028. There’s also a significant difference among the “some high school” and “masters degree” levels, since they have a p-value of 0.013,and lastly there’s a significant difference among the “some high school” and the “some college” levels, since they have a p-value of 0.04. We can also take a look at plot 1 for the distribution of math scores across the parental levels of education. It is also clear from plot 1 that students whose parents have a masters degree tend to have more math scores. This supports the claim from (Duncan & Magnuson, 2012; Sirin, 2005) that students from homes where the parents have high levels of education perform better academically compared to others.

We’ll now move to the second ANOVA test which we ran to determine whether or not there is a significant difference in mean math scores scores across the ethnicity groups.

From the 1000 students under consideration, 79 students are from the “Group A” ethnicity group,198 students are from the “Group B” ethnicity group, 323 students students are from the “Group C” ethnicity group, 257 students are from the “Group D” ethnicity group, while 143 students are from the “Group E” ethnicity group. The data contained no extreme outliers. Shapiro test demonstrated the data to be normal by group. Levene-test concluded that the the variances are homogeneous. As seen in Table 4, the average math scores for students in the “Group A” ethnicity group is 65.696(SD= 12.48).The average math scores for students in the “Group B” ethnicity group is 64.071(SD= 14.603).The average math scores for students in the “Group C” ethnicity group is 65.511(SD= 14.585). The average math scores for students in the “Group D” ethnicity group is 68.879(SD= 15.793). The average math scores for students in the “Group E” ethnicity group is 77.427(SD= 14.912). From this we can conclude that students from the “Group E” ethnicity group tend to have the highest average math score.

We can see from table 5 that “race.ethnicity” has a very small p-value of 2.2e-16. It is safe to reject the null hypothesis and conclude that there’s a significant difference in average math scores scores across ethnicity groups. Even though we know that there’s a significant difference across the ethnicity groups, we didn’t just stop there. We ran a post-hoc test to determine which groups have a significant difference. Table 6 shows us exactly that.It ts clear that there’s a significant difference among the “group E” and “group A” ethnicity groups, since they have a p-value of 0.0000002. There’s a significant difference among the “group D” and “group B” ethnicity groups, since they have a p-value of 0.0049592. There’s a significant difference among the “group E” and “group B” ethnicity groups, since they have a p-value of 0. There’s a significant difference among the “group D” and “group C” ethnicity groups, since they have a p-value of 0.0479636.There’s a significant difference among the “group E” and “group C” ethnicity groups, since they have a p-value of 0, and lastly there’s a significant difference among the “group E” and “group C” ethnicity groups, since they have a p-value of 0.0000003.This supports the notion of (Fryer & Levitt, 2004) that huge disparities in educational attainment exist within various racial and ethnic groups. We can also take a look at plot 2 for the distribution of math scores across the ethnicity groups. From plot 2, it is also clear that students from the “Group E” ethnicity group tend to have the highest average math score.

Now, we’ll move to the t-tests. We performed the first t-test to determine whether or not there is a significant difference in mean math scores scores between male and female gender.

From the 1000 students under consideration, 492 students are of the female gender while 508 students are of the male gender.Our data contained no extreme outliers. A Shapiro-Wilks test demonstrated normality by group, Levene’s test demonstrated homogeneity of variance. The average math score of the female students is 64.774 (SD = 15.079) while the average math score of male students is 70.750 (SD = 14.847). This means that the male students have more average math score than the female students.We can see that we got a very small p-value of 4.08e-10. It is safe to reject the null hypothesis and conclude that there’s a significant difference in average math scores scores between the male and female gender. We didn’t just stop at the test for the significant difference, we have also the effect size. It is shown on table 9.We have an effect size of . According to Cohen’s (1988) conventions, this is a small effect. Plot 3 shows the distribution of math scores across both gender groups.We can see that male students have more average math score than the female students. This supports the claim from ( Hyde, 2014) that males on the whole do better than women in mathematics and science subjects.

For the second t-test,we determined whether or not there is a significant difference in average math scores between students who receive “free/reduced” lunch and those who receive “standard” lunch.

From the 1000 students under consideration, 340 students take free/reduced lunch while 660 students take standard lunch .Our data contained no extreme outliers. A Shapiro-Wilks test demonstrated normality by group, Levene’s test demonstrated homogeneity of variance. The average math score of students that the free/reduced lunch is 59.9 (SD = 13.965), while the average math score of sudents that take standard lunch is 71.885 (SD = 14.259). This means that the students that take standard lunch tend to have more average math score than students that take free/reduced lunch. We can see that we have a very small p-value of 2.9e-34. It is safe to reject the null hypothesis and conclude that there’s a significant difference in average math scores scores between students take take free/reduced lunch and students that take standard lunch. We didn’t just stop at the test for the significant difference, we have also the effect size. It is shown on table 12 that we have an effect size of . According to Cohen’s (1988) conventions, this is a large effect. Plot 4 shows the distribution of math scores across both types of lunch taken by the students. Obviously, students who take standard lunch have more average math score than students who take free/reduced lunch. We can relate this to the claim from (Brooks-Gunn & Duncan, 1997) that the kind of sandwich that a student eats reflects his or her socio-economic position and educational opportunities. Also, in my opinion, taking a standard lunch will obviously get you a better grade than taking a reduced lunch.

A third t-test was performed to determine whether or not there is a significant difference in average math scores between students who have completed a test preparation course and those who have not.From the 1000 students under consideration, 344 students completed a test preparation course while 656 students did not take a test preparation course .Our data contained no extreme outliers. A Shapiro-Wilks test demonstrated normality by group, Levene’s test demonstrated homogeneity of variance. The average math score of students that completed a test preparation course is 70.334 (SD = 14.694), while the average math score of sudents that who did not take a test preparation course is 66.486 (SD = 15.380). This means that the students that completed a test preparation course tend to get more scores in math that students who did not take a test preparation course.We have an effect size of . According to Cohen’s (1988) conventions, this is a small effect. Plot 5 shows the distribution of math scores between students that completed a test preparation course and those that did not take any. We can see that students who have completed a test preparation course tend to have more scores in math than students that did not take a test preparation course.

Lastly, we performed a regression test to determine which of”gender”,“race.ethnicity”, “parental.level.of.education”,“lunch”,“test.preparation.course” is a significant predictor of math scores.

From the results given by Table 16, we can make the following conclusions:

1. Male students are expected to get an increase of 5.73 in math scores in comparison with female students.
2. Students in ethnicity group B are expected to have a math score that is, on average, 0.09 units higher than students in ethnicity group A.
3. Students in ethnicity group C are expected to have a math score that is, on average, 0.79 units higher than students in ethnicity group A.
4. Students in ethnicity group D are expected to have a math score that is, on average, 3.56 units higher than students in ethnicity group A.
5. Students in ethnicity group E are expected to have a math score that is, on average, 12.48 units higher than students in ethnicity group A.
6. Students whose parents have a bachelors degree are expected to have a math score that is, on average, 0.68 units lower than students whose parents have an associate’s degree.
7. Students whose parents have a high school certificate are expected to have a math score that is, on average, 4.6 units lower than students whose parents have an associate’s degree.
8. Students whose parents have a masters degree are expected to have a math score that is, on average, 0.01 units higher than students whose parents have an associate’s degree.
9. Students whose parents have some college degree are expected to have a math score that is, on average, 2.46 units lower than students whose parents have an associate’s degree.
10. Students whose parents have some college degree are expected to have a math score that is, on average, 2.46 units lower than students whose parents have an associate’s degree.
11. Students whose parents have some high school certificate are expected to have a math score that is, on average, 7.26 units lower than students whose parents have an associate’s degree. 12.Students who take standard lunch are expected to have a math score that is, on average, 12.33 units higher than students who take free/reduced lunch. 13.Students who do not take a test preparation course are expected to have a math score that is, on average, 5.0 units lower than students who complete a test preparation course.

The multiple R-squared value of 0.3016 indicates that the model explains about 30.16% of the variance in math scores. It explains a moderate proportion of the variability in math scores, but there are other factors of variation that is not explained by the variables we have considered. The p-value of 2.2e-16 suggests that the model is statistically significant. From the model, it is safe to conclude that gender, race/ethnicity, parental level of education, lunch type, and test preparation course are significant predictors of math scores.

In conclusion, the findings of this research relates back to the literatures we have reviewed earlier on. The literature review gave us some understanding of how the different factors we have considered influence the academic performance of students, math scores specifically. From the impact of parental education levels on academic performance to the gender factor, race/ethnicity, test preparation tests, standard of lunch taken by students.

# References

1. Burgess, S., & Greaves, E. (2013). Test Scores, Subjective Assessment, and Stereotyping of Ethnic Minorities. *Journal of Labor Economics*, 31, 535 - 576.
2. Brooks-Gunn, J., & Duncan, G. J. (Eds.). (1997). *Consequences of growing up poor*. Russell Sage Foundation.
3. Duncan, G. J., Magnuson, K., Kalil, A., & Ziol-Guest, K. (2012). The Importance of Early Childhood Poverty. *Social Indicators Research*, 108(1), 87–98.
4. Fryer, Roland, G. Jr., and Steven D. Levitt. 2010. “An Empirical Analysis of the Gender Gap in Mathematics.*” American Economic Journal: Applied Economics*, 112(2), 225-253.
5. Halpern, D. F. (2012). *Sex differences in cognitive abilities* (4th ed.). Psychology Press.
6. Hyde, J. S. (2014). Gender similarities and differences. *Annual Review of Psychology*, 65, 381-405.
7. Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Annual Review of Psychology*, 39(1), 81-109.
8. Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2001). A comprehensive meta-analysis of the predictive validity of the graduate record examinations: implications for graduate student selection and performance. *Psychological bulletin*, 127(1), 162–181.
9. Ogbu, J. U. (2003). *Black American students in an affluent suburb: A study of academic performance and the culture of poverty*. Routledge.
10. Steele, C. M., & Aronson, J. (1995). Stereotype threat and the intellectual test performance of African Americans. *Journal of Personality and Social Psychology*, 69(5), 797-811.
11. Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research*, 75(3), 417-453.