Horace Liu and Joseph Rose

[redacted]

Professor Perez-Urdiales

ECO 321: Econometrics

30 October 2020

The Effect of Student Population Size and External Factors in Public New York City High Schools on the Schools' Average SAT Scores

1. Introduction

According to a study conducted during the late 80's known as Project STAR, students in smaller classes typically had a reading score roughly 8% higher than students in a medium-sized class, and a higher math score by 9% (Woods, 2015). Although this study was conducted on students in Tennessee who were in primary school, it has been able to serve as a benchmark for student-teacher ratios in all schools nationwide for the past thirty years (Woods, 2015). In addition, student-teacher ratios have been an important topic addressed in books such as Introduction to Econometrics by James Stock and Mark Watson, as well as Econometrics courses taught at Stony Brook University. When extrapolating this idea, it seems evident that in schools with a large population of students, the student-teacher ratios are significantly smaller than schools with a smaller population of students. As more and more students enroll in high school, schools typically seem hard-pressed to employ more teachers to keep the student-teacher ratio as balanced as schools with smaller populations. Therefore, it seems to be prudent that the population size of a high school is investigated against the students' average test scores, as well as any surrounding or external factors that may influence the average test scores of students in a school. Project STAR, as well as Econometrics courses that use student-teacher ratios as a tool also seem to largely ignore external influences, such as student ethnicities, number of students in an English Language Learners (ELL) program, as well as student backgrounds. The question therefore becomes: to what extent does the student population size and the surrounding external factors in a public New York City high school affect the schools' average SAT score?

By compiling the average SAT scores of all highschools in New York City from 2014-15, this data will help determine whether the population size of students in public New York City high schools actually affect the respective schools' average SAT scores or not, and if there are supporting factors to varying test scores such as student backgrounds or ethnicities. The data would aim to explore if these factors influence SAT math scores, reading scores, and writing scores respectively. When extrapolating and adding to the Project STAR assumption, the initial hypothesis would be that schools with a larger student population, schools with more students that come from lower socioeconomic backgrounds, or schools with more students who are in an ELL program would tend to have lower average SAT scores.

One of the external factors that can influence average SAT scores is the racial background of students. As 95% of colleges in the U.S. require an SAT score for admission purposes, studies have shown that average SAT scores differentiate among students by race. For example, according to a study conducted in North Carolina by the Center for Racial Equity in Education (CREED), out of around 61,000 students, Asian students had the highest average SAT scores during the 2016-2017 academic school year, followed by White students and other students of color, respectively (Ford, 2019). Furthermore, the results were similar when external factors such as gender, socioeconomic status, language status and special education status were controlled (Ford, 2019). Race therefore plays an important role when analyzing the average test scores of high schools. However, Pacific Islander students scored nearly the state average which was an exception when analyzing the significance of the differences between all student groups (Ford, 2019). In conclusion, Asian and White students scored above the state average and all other races except Pacific Islanders scored below average (Ford, 2019).

Another factor that can influence average SAT scores is the financial background of students. In general, it is believed that the more income the parent of a child has, the more opportunities the child will achieve in the future due to affordable higher education. This is emphasized in the book, "Race, Poverty and SAT Scores: Modeling the Influences of family Income on Black and White High School Students' SAT Performance," where authors Ezekiel J. Dixon-Roman, Howard T. Everson and John J. McArdle claim that the SAT does not measure academic success of students, but rather their wealth (Dixon-Roman et al., 2013). In addition, the relationship between family income and SAT performance is said to be concave and non-linear, showing an underestimation of the relationship between family income and academic success for students (Dixon-Roman et al., 2013). Financial aid is also another factor; the New Hope Program that provided low-income families financial support through the distribution of subsidized health insurance and subsidized childcare which improved the children's test scores compared to the controlled group (Dixon-Roman et al., 2013).

Moreover, students in ESL can also be at a disadvantage when taking the SAT. As the number of international and domestic students who acquire English as their second language has been increasing, a student in the process of learning the grammatical structures or the vocabulary of English can perform worse than a native English speaker due to the questions in the SAT being in English. A study by Joni M. Lakin, Daine Cardenas Elliott, and Ou Lydia Liu showed that the differences in correlations of SAT scores between ESL and non-ESL students were smaller in the math section of the SAT than the other parts: reading, writing and critical thinking (Lakin et al., 2015). This shows that English proficiency is important to achieving success on the SAT, as math requires less English than the other three sections of the test. In addition, the difference between the scores in the critical thinking skills was 3.6 scale points, showing that the credit hours improved more non-ESL students than the ESL students (Lakin et al., 2015). Math, however, had consistent growth between the two groups (Lakin et al., 2015).

Through these studies, it is therefore clear that external factors other than student-teacher ratios or student population in schools may interfere with test scores, consequently, these factors are evaluated alongside student population in schools in order to determine their correlation with SAT scores.

2. Data

Foremost, two datasets for public highschools in New York City were gathered. One dataset (scores.csv) contained all the data for each public highschool in New York City and the schools' corresponding SAT math, reading and verbal scores. Because this dataset was centered around the 2014-15 year, the SAT format used was out of 2400, with a maximum possible score of 800 in each category. SAT test scores were selected to measure test scores as this was a standardized test that most students in high schools are required to take if they want to consider post-secondary education, and so it was a standardized test that could be applied for every high school in New York City. In addition to containing SAT scores, this dataset also contained school ID, location, as well as percentages of students of different ethnicities, as well as total enrollment within the school. The other dataset (2013_-_2018_Demographic_Snapshot_School.csv) contained a list of every single public highschool in New York City as well, however, it contained school demographics for every single school year from 2013 to 2018 such as total enrollment per year, number of students in each grade (from Kindergarten to Grade 12) per year, gender population per year, students' ethnicities per year, the number of students with disabilities per year, the number of students in an ELL program per year, as well as the number of students in poverty per year for each respective school. By combining these two datasets together, it would be possible to correlate school demographics such as enrollment against SAT test scores, or the number of students in poverty against SAT test scores, or a variety of other factors such as students in an ELL program or students of different ethnicities against SAT test scores.

In order to combine these two datasets, the 2013-18 School Demographic dataset had to be first filtered to only include each school's demographics during the 2014-15 school year. This decreased the number of initial observations from 8972 to 1770. The two datasets were then combined using the school ID for each school as some school names are spelled differently between the two datasets. There were 435 initial observations in the "scores.csv" dataset, and after merging, 425 observations were found, leaving out 10 public New York City high schools that did not report their SAT scores. Finally, to reach the final dataset that was used, 6 repeated columns after the merging were removed such as school name, enrollment, and ethnicity percentages of students within the school as they were initially present in both datasets.

The final dataset that was used contained 425 observations (different public New York City highschools) with 54 variables. Some issues raised during this process was that all apostrophes, percents, and hashtags had to be manually removed from both datasets as there were errors that occurred during the comma separating process as a ".csv" file due to symbols that clashed with the formatting and processing. There were also 56 schools that were included within the observations that did not report their SAT scores.

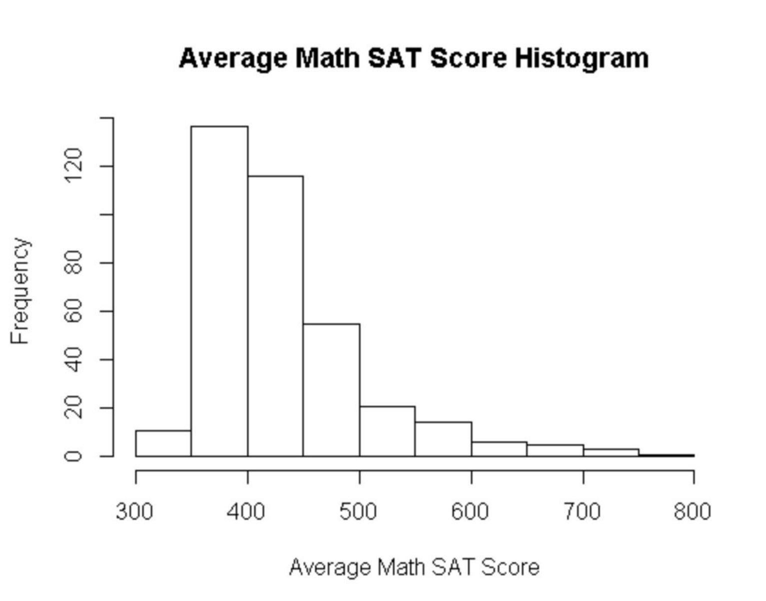
Dependant Variables	Descriptio n	Mean	Standard Deviation	Minimum	Maximum	N/A Value Count
SAT Math Score	Average SAT Math Score Per School (out of 800)	433.8	72.184	317.0	754.0	56
SAT Reading Score	Average SAT Reading Score Per School (out of 800)	425.2	62.075	302.0	697.0	56
SAT Writing Score	Average SAT Writing Score Per School (out of 800)	419.1	64.823	284.0	693.0	56

Independent Variables	Description	Mean	Standard Deviation	Minimum	Maximum
Total Enrollment	Dummy Variable of	0.188	0.391	0	1

(Dummy)	Total Enrollment where enrollment size > 700 = 1, otherwise, = 0				
English Language Learners	Number of students in an ELL program per school	75.86	129.633	0.0	1107.0
Students in Poverty	Number of students in poverty per school	512.4	507.208	37.0	3589.0
Asian	Number of asian students per school	119.8	322.269	0.0	3299.0
White	Number of white students per school	96.36	265.652	0.0	3190.0
Black	Number of black students per school	198.0	194.113	0.0	1351.0
Hispanic	Number of hispanic students per school	277.6	262.394	9.0	1684.0

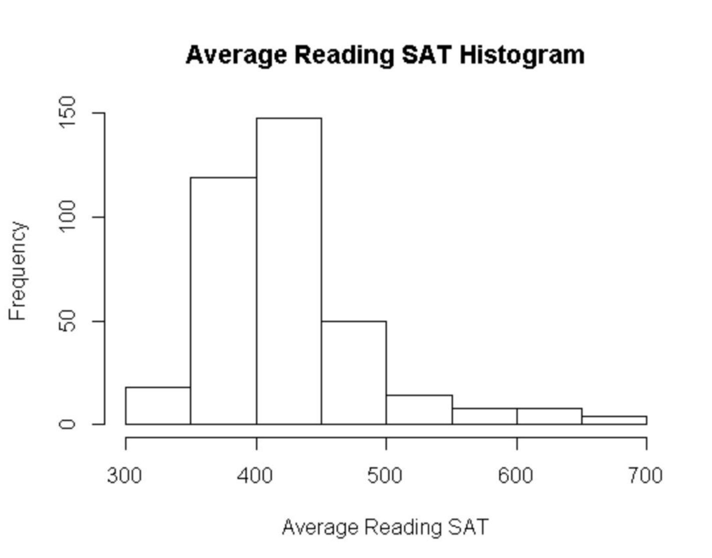
An exploratory data analysis of the average Math SAT score as the dependent variable showed that the mean SAT score across all public highschools in New York City was 433.8 out of a maximum possible score of 800, and that the minimum and maximum scores found were

317.0 and 754.0 respectively. By visualizing the data onto a histogram, it is able to be determined that the skewness is equal to 1.694283, which means that the data is skewed right, indicating that the mean of the data is larger than the median. Furthermore, it also indicates that the variable is not normally distributed as the skewness is not equal to 0.



95% confidence intervals were also found for the sample mean of Math SAT scores for public highschools in New York City where the variance is known, and the bounds were between 292.3349 and 575.2965. A test was also constructed for a null hypothesis that the true population mean is equal to 400 at the 5% level of significance. The alternative hypothesis was therefore that the true population mean is not equal to 400 at the 5% level of significance. By determining that the confidence interval was between 258.5192 and 541.4808, the null hypothesis was not rejected (at $\alpha = 5\%$) as the sample mean is 433.8, placing it within the confidence interval.

An exploratory data analysis was also conducted on the average SAT Reading score for public highschools in New York City. As another dependent variable, the mean was determined to be 425.2 out of a maximum possible score of 800, and that the minimum and maximum scores found were 302.0 and 697.0 respectively. Moreover, the skewness, after conducting data visualization as a histogram, was found to be 1.602759. This means that similar to the SAT Math score, the data is skewed right, the mean is larger than the median, and the variable is not normally distributed.



95% confidence intervals were also found for the sample mean of SAT Reading scores where the variance is known, and the bounds are between 303.5578 and 546.8921. By constructing a null hypothesis indicating that the true population mean for the SAT Reading scores is equal to 450 and an alternative hypothesis indicating that the true population mean for the SAT Reading scores is not equal to 450, the confidence interval for that test ended up having bounds of 328.3328 and 571.6672. This meant that the null hypothesis was not rejected as the sample mean of 425.2 could be found within those bounds.

Through basic data analysis of all the variables involved within the final regression analysis with multiple regressors, it helps to gain some basic understanding of how each variable stands on its own, and their roles as estimators within the OLS model.

3. Econometric Models

In order to carefully analyze the data, the use of several essential econometric models are necessary to verify the hypothesis. In general, econometric models use estimators to make statements, where estimators are functions of samples of data randomly chosen from a population. They use sample data to compute educated guesses of the value of a population parameter, such as the population mean. Furthermore, estimators in econometric models represent random data points whereas estimates are actual numerically computed values that can determine the sampling distribution of estimators that are tightly centered on the unknown variable. For this analysis, econometric models are required for three reasons: unbiasedness, consistency, and variance and efficiency. First, unbiasedness occurs when the mean of the sample distribution when evaluating an estimator multiple times on randomly drawn samples equals μ, or $E(\hat{\mu}) = \mu$. This eliminates under- or overestimation in standard errors. Second, it is said to be consistent when the probability approaches 1 as the sample size increases. The more consistent the data is, the less variance there is in the data as the number of samples increases, and the closer the estimator will be to the real population parameter. Moreover, not only will the data be consistent if there is less variance in it, but it will also be efficient as well. This would mean that the estimator is more efficient in using the information in the data.

One specific kind of an estimator is the Ordinary Least Squares, or OLS, estimator. It evaluates the minimization of the average of the differences between each scatter point on a graph displaying a sample data the actual values of the dependent variable Y_i and their corresponding predictions on the estimated line, $\overline{Y}_i = b_0 + b_1X_i$, which are called residuals, with respect to the intercept and the coefficient of the regressors. It essentially leads to determining the estimators of the regression intercept and slope(s). In addition, there are three conditions of OLS estimators, known as the Least Squares Assumptions: The mean of the conditional distribution of u given X is zero, or E(u|X = x) = 0, (X_i, Y_i), i = 1, …, n, are independent and identically distributed random variables (i.i.d), and large outliers in X and/or Y are rare because otherwise meaningless values of the estimators of a regressor will appear.

The OLS estimators in this research question are therefore prevalent due to the number of categories there are to analyze for the students: school size, financial and racial backgrounds, and ELL students. Since the dependent variable will rely on all of these factors, the multiple regression equation is $Y = \beta_0 + \beta_1X_1 + \beta_2X_2 + \beta_3X_3 + \beta_4X_4 + \beta_5X_5 + \beta_6X_6 + \beta_7X_7 + u_i$, with the estimators corresponding to each of their respective intercept and regressors: $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\beta}_2$, $\hat{\beta}_3$, $\hat{\beta}_4$, $\hat{\beta}_5$, $\hat{\beta}_6$, $\hat{\beta}_7$. In addition, the OLS estimator is $\min_{b0,\, b1,\, b2,\, b3,\, b4,\, b5,\, b6,\, b7} \sum[Y_i - (\beta_0 + \beta_1X_{1i} + \beta_2X_{2i} + \beta_3X_{3i} + \beta_4X_{4i} + \beta_5X_{5i} + \beta_6X_{6i} + \beta_7X_{7i})]^2$. The reason why OLS estimators are more fitting to the research question than other kinds of estimators is because the efficiency between the dependent and independent variables is key to analyzing the data, which has a lot of information stored given the amount of regressors. The more efficient the estimator is, the less variance there will be in the data. An idea extrapolating efficiency is the Gauss-Markov theorem, where the OLS estimator of β_1 is efficient out of all linear conditionally unbiased estimators, or the OLS estimator is the best linear unbiased estimator (BLUE).

OLS bias explains that the mean or the expectation of the OLS estimator is or is not equal to its corresponding regressor. If it is, where $E(\hat{\ }_0) = \beta_0$, $E(\hat{\ }_1) = \beta_1$,..., then it is said to be unbiased. If it is not, then it is said to be biased. In particular, the omitted variable bias measures the bias in the OLS estimator that occurs when a factor is omitted from the regression function. Two conditions exist for the omitted variable bias, where Z is a determinant of Y, or Z is part of u, and that the correlation between Z and X does not equal 0.

On top of the three least squares assumptions for the OLS estimator, there is one other assumption, which is that there is no perfect multicollinearity, or one of the regressors is not an exact linear function of the other regressors. Within the analysis, however, there is a possibility of imperfect multicollinearity existing in the regression model because two or more regressors can be correlated with each other. For example, an ELL student and a student in poverty can possibly be correlated. Furthermore, under the four total least squares assumptions, assumption three, or the fact that large outliers in X and/or Y are rare, can possibly be violated in this research because of the removal of the outliers. Therefore, no observation can be made on how many outliers are spotted in the scatter plot.

Furthermore, in terms of standard errors, the standard errors of OLS estimators behave differently under homoskedastic-only standard errors and heteroskedasticity-robust standard errors; homoskedastic-only standard errors are standard errors for the OLS estimator when the variances are the same for every data point whereas heteroskedasticity-robust standard errors are standard errors where the variances can be (nut not necessarily) different for every data point, which can make it homoskedastic as well. Observing the heteroskedasticity-robust standard errors is more appropriate for analyzing the average SAT scores because there are different variances for every single data point along the linear regression.

Another element in the regression analysis is R-Squared (R^2), which is used to represent the portion of the variance for the dependent variables that are explained by the independent variables. Its value ranges between 0 and 1, where the closer it is to 1, the higher the percentage of the variance of Y explained by the regressors, and the closer it is to 0, the less explanation there is. Additionally, the adjusted R^2 corrects R^2 with the degrees of freedom for estimation ambiguity, and it is used to avoid the increase in R^2 due to less degrees of freedom when adding more regressors. Lastly, the F-statistic is used to test a joint hypothesis with respect to the regression. Since the data uses multiple regressors, using joint hypothesis and setting restrictions on regression coefficients would help in determining if those extra coefficients would have an impact on the dependent variable. Heteroskedasticity is also important in the F-statistics because the data uses heteroskedasticity-robust standard errors.

4. OLS Regression Data Analysis

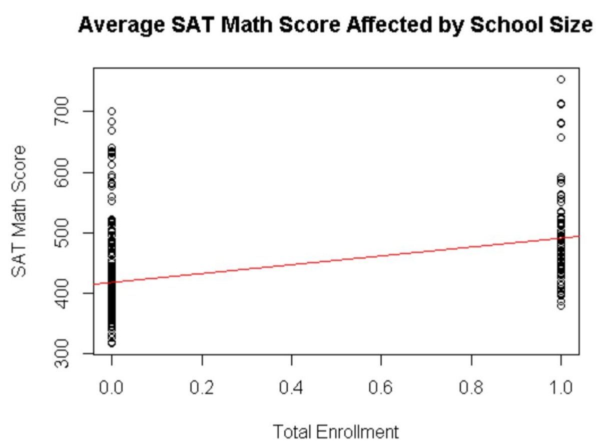
To start conducting the regression analysis investigating the extent to which the student population size and the surrounding external factors in a public New York City high school affect the schools' average SAT score, an OLS model first needs to be constructed. To begin, initial regression equations were constructed that correlated only the size of the school population with SAT scores, and assuming that everything else, all other factors, were contained in the error term.

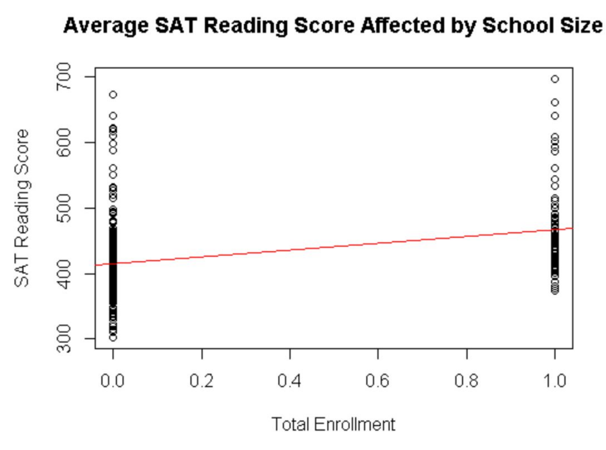
$$\text{SAT Math} = \beta_0 + \beta_1(\text{Total Enrollment}) + u_i$$

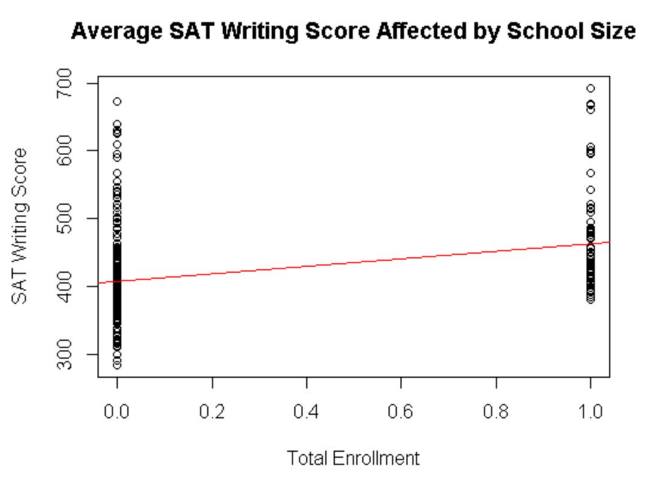
$$\text{SAT Reading} = \beta_0 + \beta_1(\text{Total Enrollment}) + u_i$$

$$\text{SAT Writing} = \beta_0 + \beta_1(\text{Total Enrollment}) + u_i$$

The variable of total enrollment, which represented the size of the student population at each school, was reduced to a dummy variable, where it was considered as a large school and the variable was set to 1 if the school had a population larger than 700 students, and the variable was set to 0 if it was considered as a small school and had a population smaller than 700 students. 700 was chosen as an arbitrary number approximate to the mean of the total enrollment variable as it would therefore give an even balance between the number of large schools and the number of small schools. All outliers were assumed to be removed. By determining the value of the regressors in the regression model, scatter plots were able to be generated as a form of data visualization as to the extent to which enrollment can affect average SAT scores of each school.







The OLS estimator was conducted under heteroskedastic robust standard error as error variances were most likely unequal for the data. Under heteroskedasticity, the Gauss-Markov theorem would not apply, and the OLS estimator would not be the best linear unbiased estimator.

In terms of total enrollment possibly affecting the average SAT Math score, $\hat{\beta}_1$ was found to be 72.2950, while $\hat{\beta}_0$ was found to be at 418.3379. This means that the average SAT Math score per school would increase by 1 as the size of the school increased by approximately 72 students. Furthermore, under the null hypothesis that β_1 is equal to 0, the test statistic found was 7.6539, and when α is set to 0.01, the null hypothesis is rejected as the absolute value of the test statistic was greater than the quantile function of the normal distribution of α at 0.01. The P-value is also very small at $1.74 * 10^{-13}$, which means that at most significance levels, the null hypothesis would be rejected as the P-value is smaller which means that the confidence interval would not contain 0. That would provide statistical evidence that β_1 is not equal to 0, therefore indicating a correlation between school size and SAT Math scores. The confidence interval at 95% for $\hat{\beta}_1$ was also found, and the bounds were between 53.78209 and 90.80787. Furthermore, the R^2 was found to be 0.1692, which means that the fit is not very good, suggesting the probability that there are more factors in play other than just school size. The adjusted R^2 was found at 0.167, which will be contrasted against when external factors are involved. Overall, these results mean that in general, there is evidence that the size of the student population at a school is correlated with the schools average SAT Math score, and as the student population increases by approximately 70 students, the average SAT Math score of that school would be higher by 1.

As for total enrollment possibly affecting the average SAT Reading score, $\hat{\beta}_1$ was found to be 52.1566, while $\hat{\beta}_0$ was found to be at 414.0586. This means that on average, the SAT Reading score of each school would increase by one as the school population increased by 52 people. Furthermore, in order to find if β_1 is statistically correlated with β_0, a null hypothesis was set where β_1 was equal to 0, and α was set to 0.01. The test statistic when β_1 is equal to 0 was found to be 6.248434, and when compared against the quantile function of the normal distribution at $\alpha = 0.01$, it was larger. This means that β_1 is not equal to 0, therefore indicating a correlation between school size and SAT Reading scores. However, the R^2 was found to be at 0.1191, which indicates that although there is a correlation, it is not a very strong fit or correlation. The adjusted R^2 was also found, and at 0.1167, it will be used when comparing with multiple regressors. Lastly, the 95% confidence interval for $\hat{\beta}_1$ was found between 35.79647 and 68.51667. Overall, the results meant that the overall enrollment at a school is statistically significant enough to correlate with the average SAT Reading scores at each school, and that as the student population increases by approximately 50 students, the average SAT Reading score would increase by 1.

One more category within each school is the average SAT Writing score. $\hat{\beta}_1$ was found to be equal to 55.6596 and $\hat{\beta}_0$ was found to be equal to 407.2138. As with previous results, this means that as the student population increases by 55 people, the average SAT Writing score at that school would increase by 1. Under the null hypothesis that β_1 is equal to 0 and α is set to 0.01, the test statistic was found to be equal to 6.368. As it was larger than the quantile function of the normal distribution when the significance level is equal to 1%, the null hypothesis was rejected. Similar to the SAT Reading and Math scores, it demonstrates that school size does statistically correlate with SAT Writing scores at each school. The R^2 value was found to be at 0.1244, and the adjusted R^2 was 0.122, meaning that there is not a very strong fit between the independent and dependent variable, suggesting that there might be more factors at play. Moreover, the 95% confidence interval was found for $\hat{\beta}_1$, and they were between 38.52853 and 72.79072. Therefore, there seems to be statistical evidence that there is a correlation between school size and average SAT Writing scores for each school, and that as the size increases by approximately 56 students, the average SAT Writing score would increase by 1.

In brief, it seems that school size (student population size) correlates with all three sections of SAT scores. Approximately 20 more students on average were needed to improve each school's average SAT Math score by 1 as compared to improving the SAT Reading or Writing score, where around 70 students were needed to increase the SAT Math score by 1 whereas it was approximately 50 students for both the SAT Reading and Writing scores. So far, the positive correlation between school size and all three SAT scores indicates that large schools tend to have higher SAT scores than smaller schools. However, there are still external factors that were observed and have yet to be included in the regression equation. This was shown as the R^2 values for all three SAT sections were fairly low, indicating that there may be external factors that influence SAT scores more than just school size, and that will be able to be determined through the differences between the adjusted R^2.

When considering the extent to which student enrollment and school size actually affect SAT scores, it seemed that the model that was previously arrived at was almost a restricted regression equation, where it omitted external factors that were able to be observed and had data collected from. Taking into consideration multiple external factors such as ethnicity, number of students in poverty, and number of students in an ELL program, the OLS model was generated.

$$\text{SAT Math} = \beta_0 + \beta_1(\text{Total Enrollment}) + \beta_2(\text{English Language Learners}) + \beta_3(\text{Students in Poverty}) + \beta_4(\text{Asian}) + \beta_5(\text{White}) + \beta_6(\text{Black}) + \beta_7(\text{Hispanic}) + u_i$$

$$\text{SAT Reading} = \beta_0 + \beta_1(\text{Total Enrollment}) + \beta_2(\text{English Language Learners}) + \beta_3(\text{Students in Poverty}) + \beta_4(\text{Asian}) + \beta_5(\text{White}) + \beta_6(\text{Black}) + \beta_7(\text{Hispanic}) + u_i$$

$$\text{SAT Writing} = \beta_0 + \beta_1(\text{Total Enrollment}) + \beta_2(\text{English Language Learners}) + \beta_3(\text{Students in Poverty}) + \beta_4(\text{Asian}) + \beta_5(\text{White}) + \beta_6(\text{Black}) + \beta_7(\text{Hispanic}) + u_i$$

This took into consideration that it was not only the size of the student population within each school that determines each schools' average SAT score, but also additional factors that due to studies found, were possibly influencing the average SAT score within each school. These factors were observed, had data collected from, and were included in the regression. All other control variables were assumed to be included in the error term, u_i.

As already determined, there is a correlation between school size and average SAT Math scores. Therefore, it is more important not to redetermine the 95% confidence interval of a single regressor such as $\hat{\beta}_1$, but to determine joint hypothesis tests and joint confidence intervals for multiple coefficients. Foremost, after finding the OLS model under heteroskedastic robust standard error, the following was found.

```
                            Estimate  Std. Error t value  Pr(>|t|)    
(Intercept)              445.8580870   6.1447042 72.5597 < 2.2e-16 ***
TotalEnroll               64.0787267   9.9040661  6.4699 3.195e-10 ***
NumEnglishLanguageLearners  -0.0971713   0.0300875 -3.2296 0.0013532 ** 
NumPoverty                -0.1029622   0.0293842 -3.5040 0.0005159 ***
NumAsian                   0.1787421   0.0234651  7.6174 2.296e-13 ***
NumWhite                   0.0590146   0.0223865  2.6362 0.0087468 ** 
NumBlack                  -0.0034482   0.0258179 -0.1336 0.8938251    
NumHispanic                0.0330137   0.0288640  1.1438 0.2534772    
---
```

As compared to average SAT Math scores with only total enrollment as the regressor, $\hat{\beta}_1$ is determined to be lower with multiple regressors as opposed to when it was the only regressor. $\hat{\beta}_1$ is found here to be 64.08, which is a significant decrease from 72.295. This would mean that the school size was initially overestimated, and that as other factors were introduced, the significance of school size went down. Furthermore, by using the F-test (under heteroskedastic robust standard errors) on regressors $\hat{\beta}_2$ to $\hat{\beta}_7$, under the null hypothesis that they are all equal to 0, the significance level came out to be $2.2 * 10^{-16}$. This shows that under a 1% significance level, the null hypothesis would be rejected, meaning that there is statistical evidence that there are external factors that are correlated with the average SAT Math score. Moreover, $\hat{\beta}_2$ and $\hat{\beta}_3$ show that the number of ELL students and the number of students in poverty at a school decreased as the SAT score increased, albeit not by much. Despite that, β_2 and β_3 are still shown to have statistical correlations with the average SAT Math score, as the absolute value of the test statistic found for both (3.2296 and 3.5040) are both greater than the quantile function of the normal distribution at $\alpha = 0.01$ when presented with the null hypothesis that β_2 and β_3 are equal to 0. There also seems to be statistical evidence that there is a correlation between the number of Asian and White students with the average SAT scores of each school, but there is no statistical evidence showing a correlation between the number of Black or Hispanic students with the average SAT scores of each school as the absolute value of the test statistic of β_6 and β_7 when the null hypothesis is that they are equal to 0 is 0.1336 and 1.1438 respectively. As they are smaller than the quantile functions of the normal distribution at $\alpha = 0.01$, it would imply that the confidence interval may encompass when β_6 and β_7 are equal to 0, meaning that they would not affect the average Math SAT score. A single restriction test was also conducted, where the null hypothesis was $\beta_2 = \beta_3$ at the 1% significance level. As the F-statistic turned out to be fairly small (0.0131) and the P-value turned out to be fairly large (0.9088), it is fair to say that at the 1% significance level, the null hypothesis would not be rejected. This would mean that there is a very likely chance that β_2 and β_3 are very similar to each other. When examining this fact, it can be concluded that typically, students in ELL programs are also students who are in poverty, and vice versa.

```
Residual standard error: 51 on 361 degrees of freedom
  (56 observations deleted due to missingness)
Multiple R-squared:  0.5103,    Adjusted R-squared:  0.5008
F-statistic: 53.74 on 7 and 361 DF,  p-value: < 2.2e-16
```

Finally, as shown, the adjusted R^2 was found to be at 0.5008. When this is compared to the adjusted R^2 when total enrollment was the only regressor (0.167), it is evident that the independent variables fit the dependent variables much better, and this is consistent with earlier results noting that school size was overestimated as there were other factors at play. Overall, it seems that there are more factors than just school size that can vary the average SAT Math score at a school.

As for the initial correlation between the school size and average SAT Reading scores, it was determined that there was a correlation, but as the R^2 was low, it was evident that there were possibly multiple regressors at play. Foremost, after finding the OLS model under heteroskedastic robust standard error, the following was found.

```
                            Estimate Std. Error t value  Pr(>|t|)    
(Intercept)               434.582136   6.144704 70.7247 < 2.2e-16 ***
TotalEnroll                45.642104   9.904066  4.6084 5.636e-06 ***
NumEnglishLanguageLearners  -0.185730   0.030087 -6.1730 1.803e-09 ***
NumPoverty                 -0.094646   0.029384 -3.2210  0.001394 ** 
NumAsian                    0.143490   0.023465  6.1150 2.509e-09 ***
NumWhite                    0.062558   0.022387  2.7944  0.005477 ** 
NumBlack                    0.031042   0.025818  1.2023  0.230019    
NumHispanic                 0.053590   0.028864  1.8567  0.064175 .  
```

When $\hat{\beta}_1$ is compared to average SAT Reading scores with only total enrollment as the regressor, it is much higher than $\hat{\beta}_1$ with multiple coefficients introduced. With multiple regressors, $\hat{\beta}_1$ is found to be 45.642, whereas when $\hat{\beta}_1$ was the only regressor, it was 52.1566. This meant that the school size coefficient was initially overestimated, and as other factors were introduced, the significance and importance of school size in determining SAT Reading scores decreased. Furthermore, by using the F-test (under heteroskedastic robust standard errors) on the other regressors under the null hypothesis that they are all equal to 0 at $\alpha = 0.01$, the F-statistic turned out to be large at 22.063, and the significance level was very small at $2.2 * 10^{-16}$, meaning that the null hypothesis would be rejected and that there is statistical evidence that these coefficients do have a correlation with the average SAT Reading score. Moreover, further investigation showed that as the number of ELL students and the number of students in poverty at a school decreased, the average SAT Reading score at each school increased, but not by much despite having statistical evidence that there is a correlation due to the test statistic being larger than the quantile function of the normal distribution at $\alpha = 0.01$ with the null hypothesis for each individual coefficient being that they are equal to 0. There is also evidence that the number of Asian students play more of a role in increasing the average SAT Reading score than students of other ethnicities, due to the test statistic shown when the null hypothesis involves that ethnicity coefficient being equal to 0 at the 1% level of significance. A single restriction test was also conducted, where the null hypothesis was $\beta_2 = \beta_3$ at the 1% significance level. The F-statistic was found to be 3.5121, while the P-value was 0.06173, meaning that at the 1% significance level, the null hypothesis would not be rejected. This determines that there is a likely chance, similar to the results reached for average SAT Math scores, that typically, students in ELL programs are also students who are in poverty, and vice versa.

```
Residual standard error: 43.94 on 361 degrees of freedom
  (56 observations deleted due to missingness)
Multiple R-squared:  0.5085,    Adjusted R-squared:  0.499
F-statistic: 53.36 on 7 and 361 DF,  p-value: < 2.2e-16
```

The adjusted R^2 value for the average SAT Reading score with multiple regressors was found to be 0.499, which makes the fit relatively stronger as opposed to when school size was the only regressor (0.1167). This is consistent with results noting that school size was overestimated, and that there are more factors involved in determining average SAT Reading scores at schools than just school size.

In terms of the initial correlation between the school size and average SAT Writing scores, similar to the average SAT Math and Reading scores, school size was not a very good fit for the SAT Writing scores, with a very low R^2. It meant that there were possibly external factors that needed to be taken into account such as the number of ELL students and the number of students in poverty at each school, as well as the ethnicities of the students at each school. To investigate this, the OLS model was found under heteroskedastic robust standard error.

```
                             Estimate Std. Error t value  Pr(>|t|)
(Intercept)                430.870354   5.773569 74.6281 < 2.2e-16 ***
TotalEnroll                 51.795528   9.728948  5.3239 1.789e-07 ***
NumEnglishLanguageLearners  -0.185420   0.030748 -6.0304 4.048e-09 ***
NumPoverty                  -0.095823   0.030333 -3.1590  0.001717 **
NumAsian                     0.147794   0.023061  6.4089 4.581e-10 ***
NumWhite                     0.064418   0.024940  2.5830  0.010189 *
NumBlack                     0.018831   0.026023  0.7236  0.469758
NumHispanic                  0.049493   0.030475  1.6241  0.105234
```

When $\hat{\beta}_1$ was compared in this model to the one where $\hat{\beta}_1$ was the only regressor, the $\hat{\beta}_1$ found with multiple other regressors was found to be lower. With an estimate of 51.7955, this was a decrease from 55.6596, meaning that school size was overestimated in the linear equation with only school size as the estimator. Moreover, the F-test was conducted under heteroskedastic robust standard errors on regressors $\hat{\beta}_2$ to $\hat{\beta}_7$, under the null hypothesis that they are all equal to 0. The F-statistic turned out to be large at 20.811, while the P-value turned out to be very small at $2.2 * 10^{-16}$, showing that under a 1% significance level, the null hypothesis would be rejected. This would mean that there is statistical evidence indicating external factors that correlate with the average SAT Writing score. The results found for the other regressors were also similar to the results found for the average SAT Reading score, where $\hat{\beta}_2$ and $\hat{\beta}_3$ show that the number of ELL students and the number of students in poverty at a school decreased as the SAT Writing score increased, albeit not by much. Because $\hat{\beta}_2$ and $\hat{\beta}_3$ were found to be very similar, a single restriction test was conducted where the null hypothesis was $\beta_2 = \beta_3$ at the 1% significance level. The F-statistic was found to be 3.0973, while the P-value was 0.07927, which meant that at the 1% significance level, the null hypothesis would not be rejected. This would mean that β_2 is, at the 1% significance level, not very different from β_3, meaning that students in ELL programs could be students in poverty, and vice versa.

```
Residual standard error: 45.67 on 361 degrees of freedom
  (56 observations deleted due to missingness)
Multiple R-squared:  0.513,     Adjusted R-squared:  0.5036
F-statistic: 54.33 on 7 and 361 DF,  p-value: < 2.2e-16
```

The F-statistic was found on 7 restrictions (7 regressors) and it was very large at 54.33, indicating that it is fair to assume the regressors do have a correlation with the dependent variable. The adjusted R^2 was also found, and at 0.5036, it is significantly higher than the adjusted R^2 when school size was the only regressor (0.122), meaning that there is a better fit when there are more regressors. This is consistent with results noting that more regressors explain the average SAT Writing score better, and that there are factors other than solely school size which correlate with the average SAT Writing score at each school.

5. Conclusion

When adding in external factors such as student ethnicities, number of students in an ELL program, and the number of students in poverty, there is shown to be a much better fit and correlation between the regressors and the SAT scores, with the adjusted R^2 value typically increasing by 0.4 or more. The "unrestricted" regression equation allowed for more exploration into various observed factors that correlated with SAT scores, and it resulted in a more accurate derivation of factors that affect SAT scores at each school.

Ultimately, the initial hypothesis was incorrect, and it seemed that with a positive correlation, larger schools would have higher average SAT scores, suggesting that in order to increase the average SAT score at each school, the student population size should increase. However, the extent to which school size actually affected SAT scores was low, and that more emphasis was placed on external factors such as student ethnicity, the number of students in poverty, and the number of students in an ELL program. In addition to that, as derived from the results of the OLS model, there is also a much larger and more significant positive correlation between average SAT test scores and the number of Asian students (as compared to other ethnicities), as well as a negative correlation between average SAT test scores and the number of students in poverty and the number of students in an English Language Learner program at each school. However, additional biases could be included for students who are of mixed ethnicities at each school, as well as various other control factors such as parental influence which could not be measured. More investigation would be needed into other unmeasurable control factors, however, a new policy that could be derived from the results of this study could be for New York City to reduce the number of schools, and to focus on schools with larger student populations with more introduced scholarships for impoverished students.

Works Cited

Dixon-Roman, E. J., Everson, H., & Mcardle, J. J. (2013, May). (PDF) Race, Poverty and SAT Scores: Modeling the Influences of Family Income on Black and White High School Students' SAT Performance. Retrieved October 29, 2020, from https://www.researchgate.net/publication/280232788_Race_Poverty_and_SAT_Scores_Modeling_the_Influences_of_Family_Income_on_Black_and_White_High_School_Students'_SAT_Performance

Ford, J. (2019, December 19). E(race)ing Inequities: The influence of race on SAT scores. Retrieved October 29, 2020, from https://www.ednc.org/eraceing-inequities-the-influence-of-race-on-sat-scores/

Lakin, J. M., Elliott, D. C., & Liu, O. L. (2015, September 27). Investigating ESL Students' Performance on Outcomes Assessments in Higher Education. Retrieved October 29, 2020, from https://uncw.edu/assessment/documents/lakinetal2012.pdf

Woods, D. (2015, September 23). The Class Size Debate: What the Evidence Means for Education Policy (Darian Woods). Retrieved October 29, 2020, from https://gspp.berkeley.edu/research/featured/the-class-size-debate-what-the-evidence-means-for-education-policy

Bibliography

Dixon-Roman, E. J., Everson, H., & Mcardle, J. J. (2013, May). (PDF) Race, Poverty and SAT Scores: Modeling the Influences of Family Income on Black and White High School Students' SAT Performance. Retrieved October 29, 2020, from https://www.researchgate.net/publication/280232788_Race_Poverty_and_SAT_Scores_Modeling_the_Influences_of_Family_Income_on_Black_and_White_High_School_Students'_SAT_Performance

Ford, J. (2019, December 19). E(race)ing Inequities: The influence of race on SAT scores. Retrieved October 29, 2020, from https://www.ednc.org/eraceing-inequities-the-influence-of-race-on-sat-scores/

Hu, W. (2012, January 25). City Students at Small Public High Schools Are More Likely to Graduate, Study Says. Retrieved October 29, 2020, from https://www.nytimes.com/2012/01/26/education/new-york-city-students-at-small-public-high-schools-are-more-likely-to-graduate-study-finds.html

InsideSchools. (n.d.). Large and Small High Schools. Retrieved October 29, 2020, from https://www.nyxt.nyc/insideschools/large-and-small-high-schools/

Lakin, J. M., Elliott, D. C., & Liu, O. L. (2015, September 27). Investigating ESL Students' Performance on Outcomes Assessments in Higher Education. Retrieved October 29, 2020, from https://uncw.edu/assessment/documents/lakinetal2012.pdf

Page, M., & Jackson, E. (n.d.). Small Class Sizes Yield Higher Test Scores Among Young Children. Retrieved October 29, 2020, from https://poverty.ucdavis.edu/policy-brief/small-class-sizes-yield-higher-test-scores-among-young-children

Schwartz, A., Stiefel, L., & Wiswall, M. (2016, March 26). Are all schools created equal? Learning environments in small and large public high schools in New York City. Retrieved October 29, 2020, from https://www.sciencedirect.com/science/article/abs/pii/S0272775716301467

Woods, D. (2015, September 23). The Class Size Debate: What the Evidence Means for Education Policy (Darian Woods). Retrieved October 29, 2020, from https://gspp.berkeley.edu/research/featured/the-class-size-debate-what-the-evidence-means-for-education-policy

Datasets Used

https://www.kaggle.com/nycopendata/high-schools/notebooks
https://data.cityofnewyork.us/Education/2013-2018-Demographic-Snapshot-School/s52a-8aq6