

The Evaluation of Education Equity in NYC

Ju-Eun Kim

April 19, 2021

* indicates: For more information, refer to Appendix.

Abstract

The SAT scores determine the education quality of a school. There has been lots of effort in the US for education equality. Using the SAT score of New York City, the results of the education policy for education equity will be examined. The data is collected from public schools in the 2014-15 school year with variables such as race, the average score of different categories of the exam, and region in NYC. The levelled education will have a low to no gap between the scores. A confidence interval of 95% will be used to examine the SAT result data. The result will determine if the policies were helpful for a better education environment. If the null hypothesis, stating there is no difference in SAT scores between regions, the NYC Department of Education has successfully made the education equity. They may continue with the existing policy. On the other hand, if significant differences were shown, new approaches will be required.

Introduction

The SAT is a standardized test used for university admission in the United States (Khandelwal, 2021). It stands for the Scholastic Aptitude Test, which measures essential skills required for academic success (Schalkwyk, 2011). High school students typically take the SAT for admissions to universities in the US. The score of students represents the education quality of each school. The difference in the average score of SAT between schools indirectly indicates that there exists a difference in education quality. The SAT scores gap occurred due to the racial discrepancies, poverty and the region they live in, such as Manhattan, Bronx, or Brooklyn. To reduce the education gap, the government has worked for a better environment for all students to learn. The students also participated in strikes and took action to make a difference in education (Conley, 2019). They argue that the students should not be limited to get education on race, sex, religion, or disability (Conley, 2019)

SAT scores can be used to examine how education equality has been accomplished while the education department in NYC worked for improvement. The research question is: Was the current effort effective enough to reduce the education quality? If it was more successful, the average score between different regions or races should have a lower score gap. The data that will be used is from Kaggle, **Average SAT Scores for NYC Public Schools**. It is from the 2014-15 school year. It includes various variables such as school name, city, phone number, student enrollment, percent White, Black, Hispanic and Asian, and average score (SAT Math, Reading, and Writing).

The education department but also the students tried for improvement for education equity. The hypothesis is that education equity is accomplished in NYC. The SAT score between regions will not be significant anymore. The SAT result data will be explored to test the hypothesis using the given variables. The result will be essential to examine if the government will need a new policy or continue with the current policy for a better learning environment.

Data

The data is from Kaggle, **Average SAT Scores for NYC Public Schools**. It has the variables of SAT scores of different categories, races, addresses, phone numbers and more. However, there are few crucial variables, and the irrelevant or unnecessary data will be cleaned out. The key variables that show a direct relationship to the score will be used for the analysis.

Key Variables

The key variables are introduced in the section. The **City** is an important variable used to compare the education equity between different regions in NYC. The four variables that give the racial demographic will be used how the education level differs based on the race. It will examine if certain races are getting a better education. If the result shows no difference in score based on race, it will prove that the education levelling has successfully done. Lastly, each category of scores' average scores is the key that will help determine if the education equity has been accomplished. The scores are the direct evidence that evaluates the education policies in NYC explicitly.

The schools that do not provide enough information will be removed from the analysis while they can cause bias on the result.

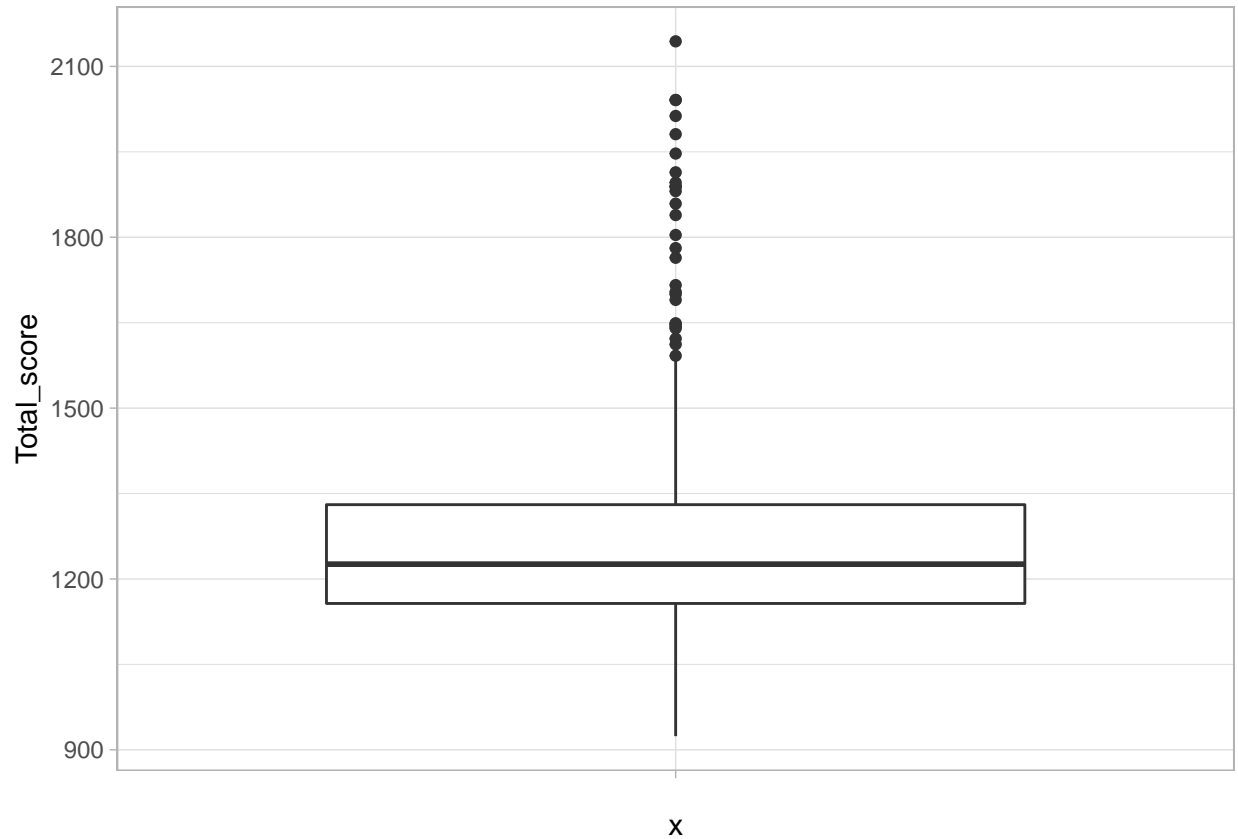
The new variables are created for more efficient analysis in this process. The new variable, the total score, will be added to the primary data. The total SAT score, the sum of math, writing, and reading, will be helpful for comparison. The total SAT score was out of 2400 in the school year 2014-15.

The overall education level of NYC is also essential to interpret before the education is divided into a **city**. The education department should aim to hit the average of the overall US means of SAT score.

The numerical summary is created using the total score of SAT. It seems there is a significant difference between the maximum and the minimum score. They both might be an outlier, but it is still critical to consider all the values.

```
## # A tibble: 1 x 5
##   min  mean median  max   sd
##   <dbl> <dbl> <dbl> <dbl> <dbl>
## 1   924 1276.  1226  2144  195.
```

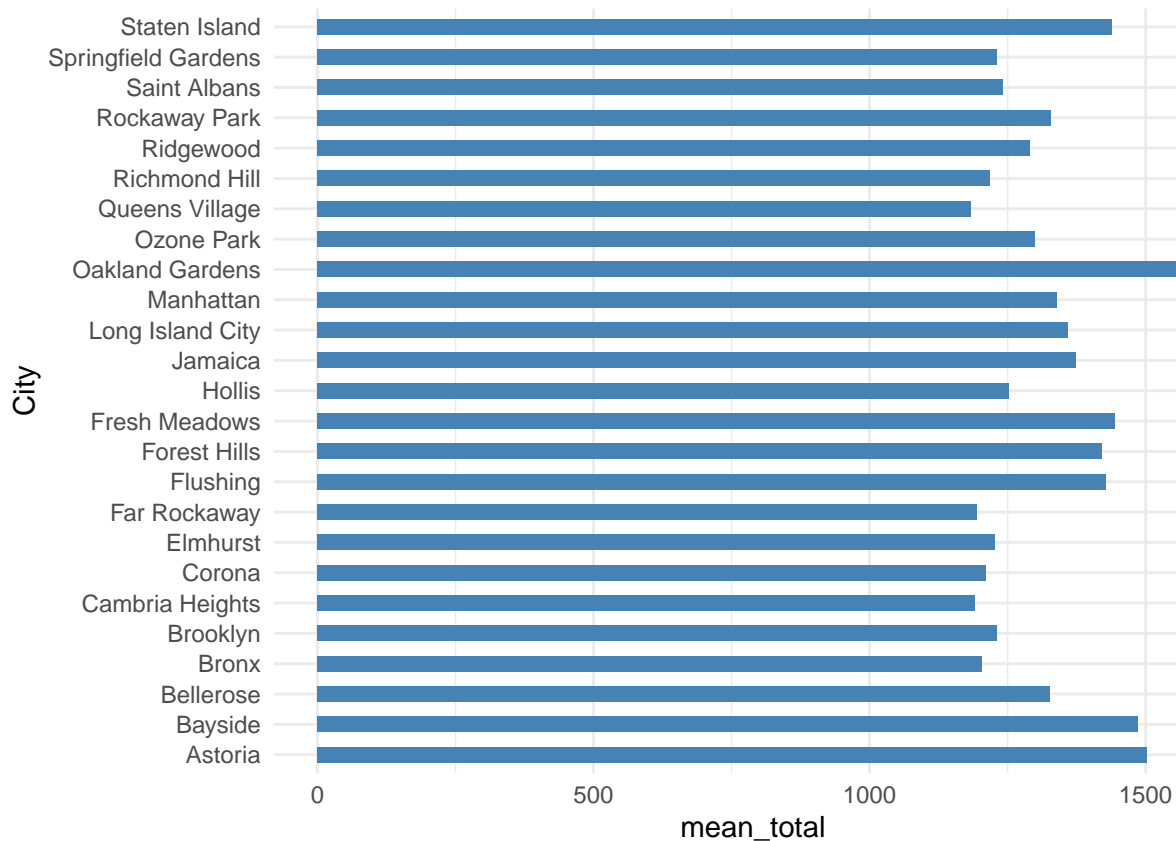
The distribution can be visualized using the box plot below. It summarizes the distribution of SAT total score quantitatively using five statistics. The middle line represents the median of the score. The first edge of the box and the last edge of the box represent 25% and 75% of each data value. The data defined in points are the outliers of the score. From this boxplot, the upper outliers were found more than there are outstanding schools that perform better than the median schools. Also, it shows that the mean of the score will be higher than the median. It indicates that more than half of the students perform worse than the mean of the total NYC SAT score. It indirectly states that there are apparent gaps between education accomplishments.



The comparison will be mainly made based on the city in NYC. This step creates a new data frame using the mean of each city (region). Twenty-five cities in NYC may all have different education levels.

```
## Rows: 25
## Columns: 9
## $ City      <chr> "Astoria", "Bayside", "Bellerose", "Bronx", "Brooklyn", ~
## $ mean_White <dbl> 23.580000, 18.100000, 8.400000, 3.121429, 7.177064, 1.35~
## $ mean_Black <dbl> 7.980000, 13.000000, 48.100000, 28.974490, 54.989908, 76~
## $ mean_Hispanic <dbl> 40.30000, 27.90000, 17.50000, 62.71122, 28.79083, 11.800~
## $ mean_Asian <dbl> 26.380000, 40.300000, 24.800000, 3.905102, 7.447706, 7.7~
## $ mean_math <dbl> 509.6000, 523.0000, 453.0000, 404.3571, 416.4037, 382.00~
## $ mean_reading <dbl> 494.8000, 479.0000, 434.0000, 402.5204, 410.7615, 413.50~
## $ mean_writing <dbl> 497.2000, 485.0000, 439.0000, 395.8469, 403.0917, 394.50~
## $ mean_total <dbl> 1501.600, 1487.000, 1326.000, 1202.724, 1230.257, 1190.0~
```

The mean of each city is represented in the bar graph below. The `mean_total` represents the mean total score of SAT in each city. The gap between the lowest and the highest score of each region has decreased than comparing the individual schools. It seems reasonable to test the education levelling was done successfully be-



tween NYC regions.

The distribution of race can be visualized. According to the previous research, “Race gaps in SAT scores highlight inequality and hinder upward mobility,” it states that the higher Asian student rates lead to the higher SAT average school score (Reeves and Halikias, 2017). In contrast, the Black and Latino tended to have a lower average SAT scores (Reeves and Halikias, 2017). The graph below is generated to show the racial demographic.



Mostly, in all the cities, black students were the majority, with some exceptions. However, the other race distributions were not uniform. Before the in-depth analysis, it is essential to recognize that the race demographic varies in all the cities.

Package Reference

The project was programmed using R version 1.2.5042.

Hadley Wickham (2007). Reshaping Data with the reshape Package. Journal of Statistical Software, 21(12), 1-20. URL <http://www.jstatsoft.org/v21/i12/>.

Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2021). dplyr: A Grammar of Data Manipulation. <https://dplyr.tidyverse.org>, <https://github.com/tidyverse/dplyr>.

Hadley Wickham (2021). tidyr: Tidy Messy Data. <https://tidyr.tidyverse.org>, <https://github.com/tidyverse/tidyr>.

Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, <https://doi.org/10.21105/joss.01686>

Methods

This section will introduce the statistical approach to the SAT data that cleaned from the previous section. Various methods will be used to examine if the current education policies are helpful for education equity for students in NYC.

Part 1. Bayesian Credible Interval

Suppose the education equity was accomplished. The score distribution should be almost symmetrical. Therefore, the probability of the schools' average being higher than the total mean of SAT score and the probability of the school's average being less than the mean SAT score should be the same. The simplification was made by dividing schools into only two groups, better or worse than the average. To examine this binomial problem, the beta a prior distribution will be used.

The beta distribution's domain is bounded between 0 and 1. It has two parameters, **a** and **b**. The parameter **a** is the ratio of successes, and **b** is the ratio of failures (Dekking, 2005).

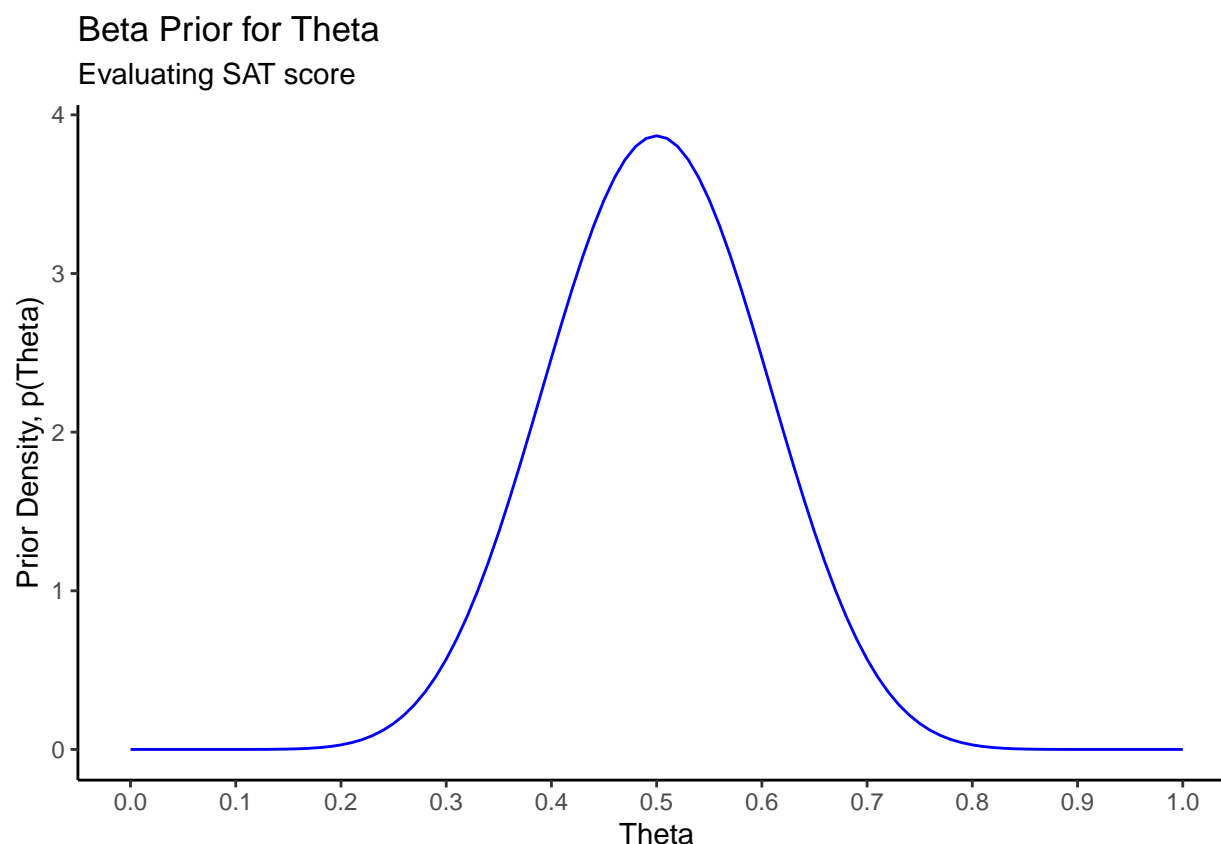
They can manipulate the shape of the beta distribution. The prior distribution, beta, measures how confident the parameters before seeing the SAT score results, that θ is near the centre of its possible values (Dekking, 2005).

The prior distribution is beta (12,12), where the parameters $a=12$ and $b=12$ are chosen. Therefore, while $a = b$, the probability of picking the school with the higher score than the average and the lower score than the average is the same.

*Refer to the Appendix for the details.

The shape below represents the prior distribution.

```
## Warning: 'data_frame()' was deprecated in tibble 1.1.0.  
## Please use 'tibble()' instead.
```



The prior probability is 0.957104.

The prior distribution is determined to be the beta distribution. The posterior distribution can be driven using the prior distribution.

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}$$

- Refer to the Appendix for the detailed derivation.

Using the posterior, the 95% credible interval will be calculated. It will help to assume how confident we are on the hypothesis that half of the school performs better on SAT than the average.

Part 2. Goodness of Fit Test

The goodness of fit test will be used to be more precise and advanced from the previous examination. In Bayesian credible interval, the assumption was simply made that the Education Equity was successful and the probability of having schools that obtained the higher average than the overall SAT score the US-wide.

However, in this case, the actual data will be used to examine if the data fit on the binomial distribution, with the probability of $\frac{1}{2}$. The mean of SAT US-wide will be used to see the overall NYC performance compared to the other regions in the US. Also, to know the education distribution within NYC, the mean of NYC's SAT score will be used as well. If the mean is close to the median, the schools are trying to improve the quality of education. The mean SAT score in NYC is 1275.907. Using the given data, the number of schools with a higher grade than the mean will be found. The mean of the SAT US-wide was 1497 in the 2014-15 school year (Anderson, 2014).

The Chi-squared test will be used. The Chi-square goodness of fit test determines if sample data fit the distribution (Dekking, 2005). It can be used for discrete distributions, such as binomial distribution (Dekking, 2005). The null and alternative hypotheses will be found. The null hypothesis, H_0 , claims that the distribution of SAT score fits the data, and the alternative hypothesis, H_A , claims that the distribution does not fit the data. Then the test statistics will be calculated for p-value. The p-values are the probability of observing that stat of something more extreme (Dekking, 2005). Using the p-value, the decision will be made either to reject or not reject the null hypothesis.

The Chi-square goodness of fit test has an assumption: The data needs to have independent observations (Gibbs and Stringer, 2021).

Part 3. Hypothesis Test of the Mean

In this section, the hypothesis test of the mean will be done. The mean of the NYC, Manhattan will be compared with the school in Bronx's mean score. They are the two regions with the most school that will have more students enrolled in the NYC education.

The assumption is that the data (the SAT scores) are independently and identically distributed (Gibbs and Stringer, 2021).

First, the hypothesis is stated. The null hypothesis is that the mean of Manhattan's SAT is the same as the random Bronx's mean of SAT.

$$H_0 : \mu_b = \mu_m$$

The alternative hypothesis is that they do not equal to each other.

$$H_0 : \mu_b \neq \mu_m$$

.

Then, the data of SAT score will be used to find a test statistic and p-value. For the test statistic, the distribution of the SAT score is assumed to be distributed normally. The test statistic can be used using the formula:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

The p-value is the probability of observing the test statistic (Dekking, 2005).

After calculating the p-value, the decision will be made. Suppose the p-value is small, usually less than 0.05, the evidence against H_0 . Then the null hypothesis would be rejected. If the p-value is significantly large, usually larger than 0.05, there is no evidence against H_0 . Then the null hypothesis would not be rejected.

The errors could be made depending on the result. When the null hypothesis is rejected when H_0 is true, it is a type 1 error (Dekking, 2005). In contrast, when the null hypothesis is not rejected when H_A is true, it is called the type 2 error (Dekking, 2005).

Part 4. Confidence Interval

In this section, a confidence interval will be used to examine the research question once again. The cleaned data does not have all the schools in New York City. For example, some schools did not provide some sections of SAT scores. It might be misleading only to consider a small number of samples. Therefore, the bootstrap method will be used. It draws 1000 bootstrap samples of size n with replacement from the original sample. For each bootstrap sample, the statistic will be calculated.

The SAT score data collected from observations are assumed to be from an independent and identically distributed population.

The empirical bootstrapping method is chosen because the dataset is given randomly in this case. The parametric bootstrapping method deals with the known distribution, which is not appropriate for this analysis.

The first step is to generate the dataset from the original data set. Then, the studentized mean (centred sample mean) for bootstrap will be calculated many times (Dekking, 2005). The simulation will be done using `r`.

The population mean of SAT score of 2014-15 school year was 1497 nationwide in the US (Anderson, 2014). If the education equality was successful, all the mean of NYC regions should be similar to the nationwide average score.

Part 5. Maximum Likelihood Estimator Calculation

Maximum likelihood is choosing the parameter in such a way that the data are most likely. In this case, the mean will be a parameter of interest (Dekking, 2005).

The goal is to find the maximum likelihood estimate that maximizes the likelihood function. The maximum likelihood estimator will be noted as θ (Dekking, 2005).

If the education equity has been successfully accomplished, there will be a normal distribution of the SAT score. The students who can score the average on the test will be the most. Only a few students can achieve the extremely high or low score given that the same education is done to all students in NYC.

The score dataset will be the continuous distribution, following the normal distribution. The assumption is that the data are independent and identically distributed random variables (Gibbs and Stringer, 2021).

The parameter of the normal distribution is

$$x \sim N(1275.907, 194.9063^2)$$

Mean and the variance was obtained in the previous (`data`) section.

- MLE formula derivation is included in the Appendix.

The maximum likelihood estimate can be found after finding the maximum likelihood function. The score that is most likely to be obtained by the students in NYC can be calculated.

Part 6. Linear Regression Method

The linear regression model will be used to identify any direct relationship in the data. The simple linear regression model will be used. The dataset is bivariate with independent and dependent variables (Dekking, 2005).

The percentage of the race, which is the independent variable, is non-random and the SAT scores are the realization of random variables. The equation is $Y_i = \alpha + \beta x_i + U_i$ for $i = 1, 2, \dots, n$. The error term, U_i are independent random variable. The expectation of the error term is 0 and the variance will be σ^2 . α represents the intercept and β represents the slope.

Linear regression will help provide if there is any relationship between the SAT score and the race. Will including more of the particular race lead to higher or to lower ACT average?

This is helpful because if the race demographics do not matter, no matter what percentage of racial is the majority of the school, the linear regression slope will be flat, rather than positive or negative slope. If there's a strong correlation (both positive and negative), it indicates that race is one of the factors that determine the education level. In this case, rather than dividing into the different regions, the entire NYC will be observed with total SAT scores. The data provides four different races, and the linear regression will be done four times.

Results

In this section, the results of the statistical analyses will be delivered. The results of the six different methodologies are included in the report. Also, the interpretation of the result will be discussed.

Part 1. Bayesian Credible Interval

The posterior distribution is driven in the `method` section and will be used. There are approximately 400 schools in NYC. Therefore, $n = 400$ will be used for different cases of the score distribution.

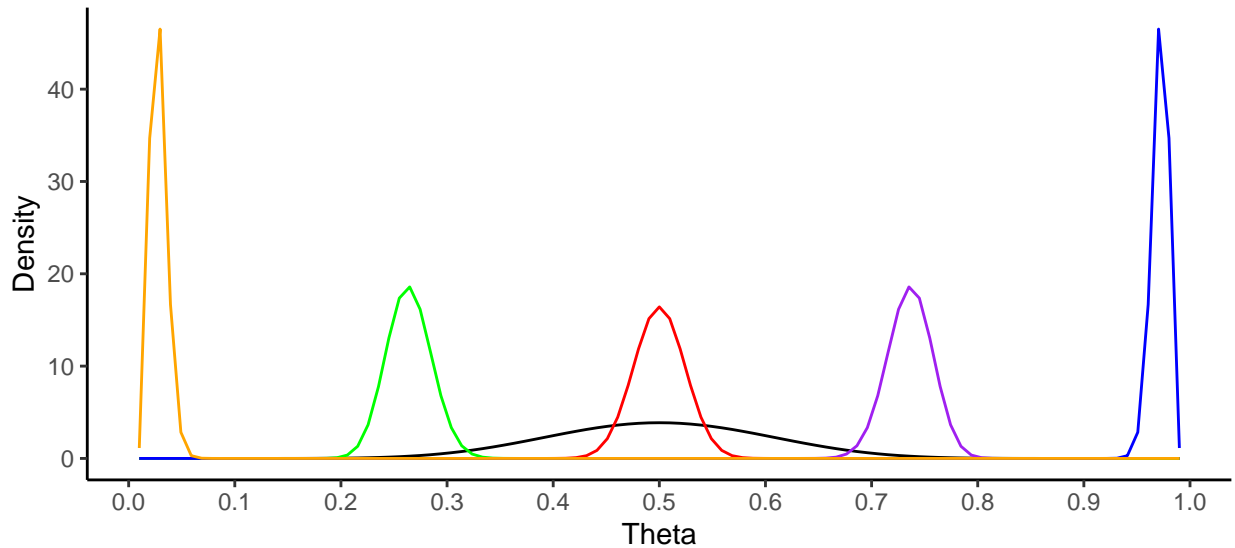
Prior vs Posterior for Theta in Beta Distribution,

Choosing SAT score randomly among 400 NYC School

black: Prior Purple: 300 over mean score

Red: 200 over mean score Orange: 0 over mean score

Green: 100 over mean score blue: 400 over mean score



The black line, which is the prior distribution graph and the red graph share a similar part. The place where the prior has the peak, and the most probable, the posterior becomes more peaked around that value, which is 0.5. As the weights get more extreme, the frequentist inference becomes extreme as well.

The 95% interval will be calculated for the Bayesian Credible Interval. The confidence level most commonly used is 95%. The range calculated explains uncertainty by giving a range of values on the posterior probability distribution. The range is called the 95% Bayesian credible interval (Dekking, 2005).

The Bayesian Credible Interval of 95% can be calculated as:

- When $n = 400$ and the number of School that has higher SAT score than the average is 0: (0.0147, 0.0461)
- When $n = 400$ and the number of School that has higher SAT score than the average is 100: (0.223, 0.307)
- When $n = 400$ and the number of School that has higher SAT score than the average is 200: (0.452, 0.548)
- When $n = 400$ and the number of School that has higher SAT score than the average is 300: (0.693, 0.705)
- When $n = 400$ and the number of School that has higher SAT score than the average is 400: (0.954, 0.985)

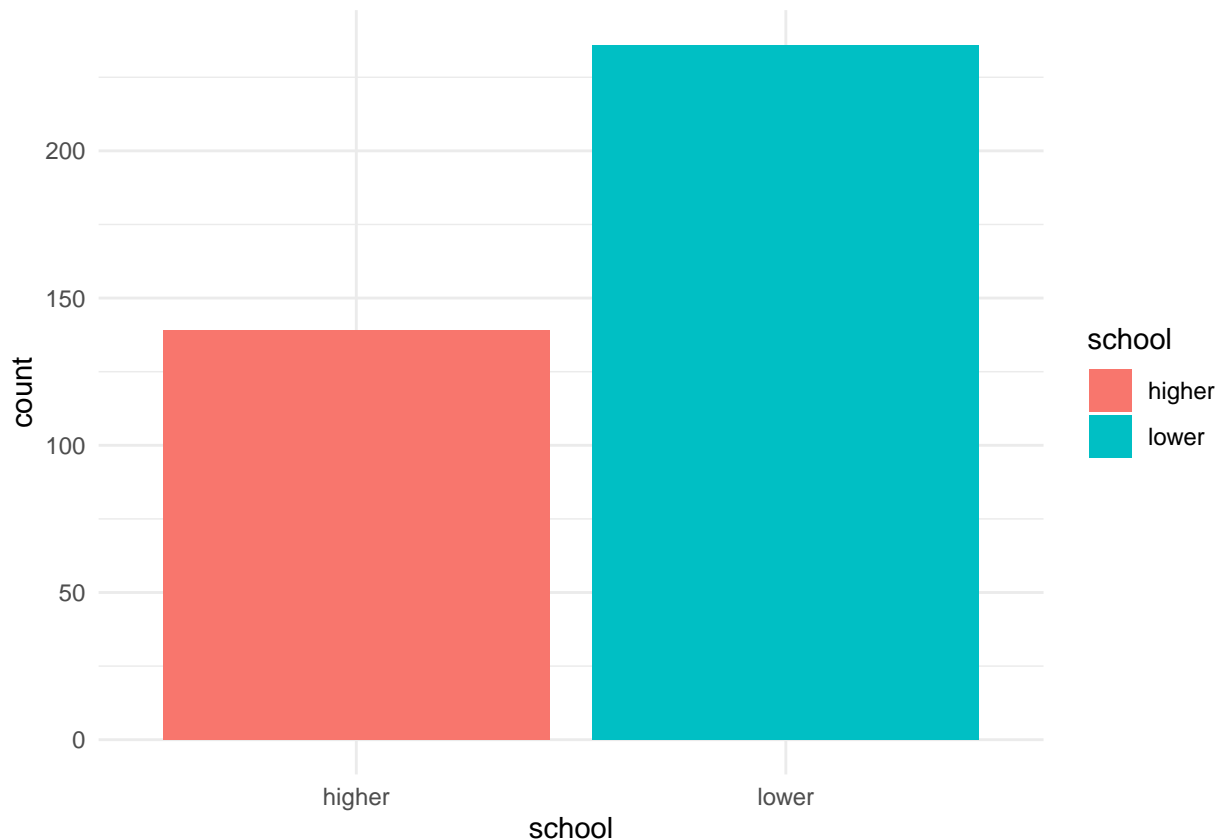
The 95% credible interval was all calculated with various cases. Our goal was to find the interval that has more than 200 schools that are over the mean. Therefore, the interval for 95% is (0.452, 0.548). It is near 0.5, which is 1/2 probability. Therefore, it indicates that the probability of having half of the schools with better scores than the mean is between 0.452 and 0.548 with the confidence of 95%.

The probability of having half of the schools with better scores than the mean is between 0.452 and 0.548, with the confidence of 95% given that the prior distribution was correct. However, more examination is required to see if the prior assumption was valid. In the next section, using the goodness of fit test, it will be examined.

Part 2. Goodness of Fit Test

Using the NYC Mean

From the cleaned data, 139 schools have higher or equal to the NYC SAT mean score. 236 schools have lower SAT scores than the NYC SAT mean score. From this, it looks like the mean is greater than the median and will be right-skewed.



The Chi-squared test will be used. Chi-square goodness of fit test determines if sample data fit the distribution. It can be used for discrete distributions, such as binomial distribution.

The conducted null hypothesis is that the SAT score data comes from a binomial distribution. The alternative hypothesis is that the data does not come from a binomial distribution.

The goal is to see if half of the score is higher than the mean. Therefore, the binomial distribution is used with $p = 0.5$.

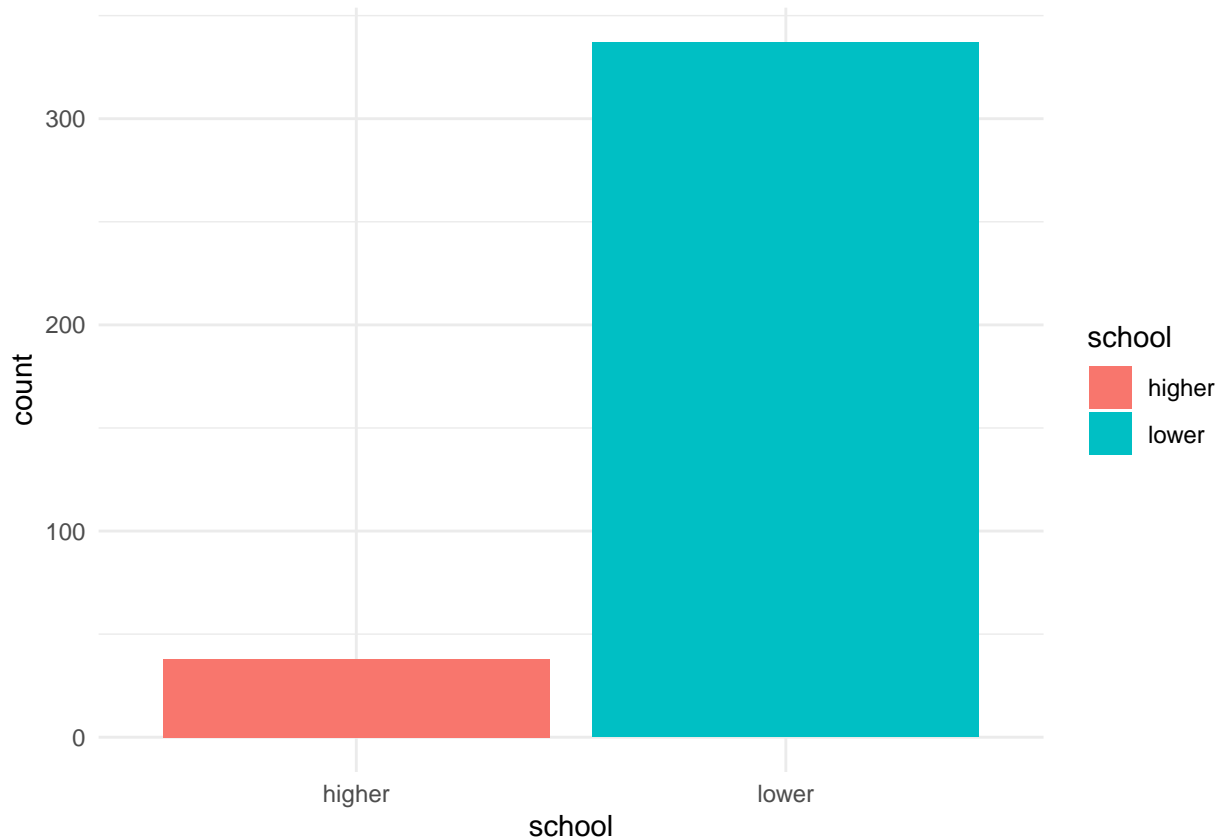
```
##  
## Chi-squared test for given probabilities  
##  
## data: c(236, 139)  
## X-squared = 25.091, df = 1, p-value = 5.47e-07
```

The p-value is too small, which is less than the significance level (0.05), the null hypothesis is rejected. The data does not come from the binomial distribution with $p = 0.5$. Therefore, there is sufficient evidence to claim that half of the SAT scores are not higher than the mean score of NYC.

Using the US-wide Mean

The US-wide SAT mean score was 1497 in 2014.

There are 38 schools that have higher or equal to the US-wide SAT mean score. There are 337 schools that have lower SAT scores than the US-wide SAT mean score.



The Chi-squared test was once again used for this part to see if the data fits the binomial distribution with $p = 0.5$.

```
##
## Chi-squared test for given probabilities
##
## data: c(337, 38)
## X-squared = 238.4, df = 1, p-value < 2.2e-16
```

The p-value is extremely small, which is far less than the significance level (0.05), the null hypothesis is rejected. The data does not come from the binomial distribution with $p = 0.5$. Therefore, there is sufficient evidence to claim that half of the SAT scores are not higher than the mean score US-wide.

The education equity was not sufficient, while more than half of the students could not reach the mean SAT score US-wide in NYC. The schools in NYC had lower scores overall compared to the other regions in the US. The NYC education department will have to continue to work on the balance of score US-wide. Also,

more than half of the students could not reach the mean SAT score of NYC. It indicates that the scores are right-skewed. The mean score is higher than the median. The more students are obtaining, the lower scores than the higher scores. Therefore, education needs to be improved for the overall NYC region.

Part 3. Hypothesis Test of the Mean

Two regions, Bronx and Manhattan, that have more students enrolled in the NYC education compared to the other areas will be tested.

The null hypothesis is that the mean of Manhattan's SAT is the same as the random Bronx's mean of SAT.

$$H_0 : \mu_b = \mu_m$$

The alternative hypothesis is that they do not equal to each other.

$$H_0 : \mu_b \neq \mu_m$$

.

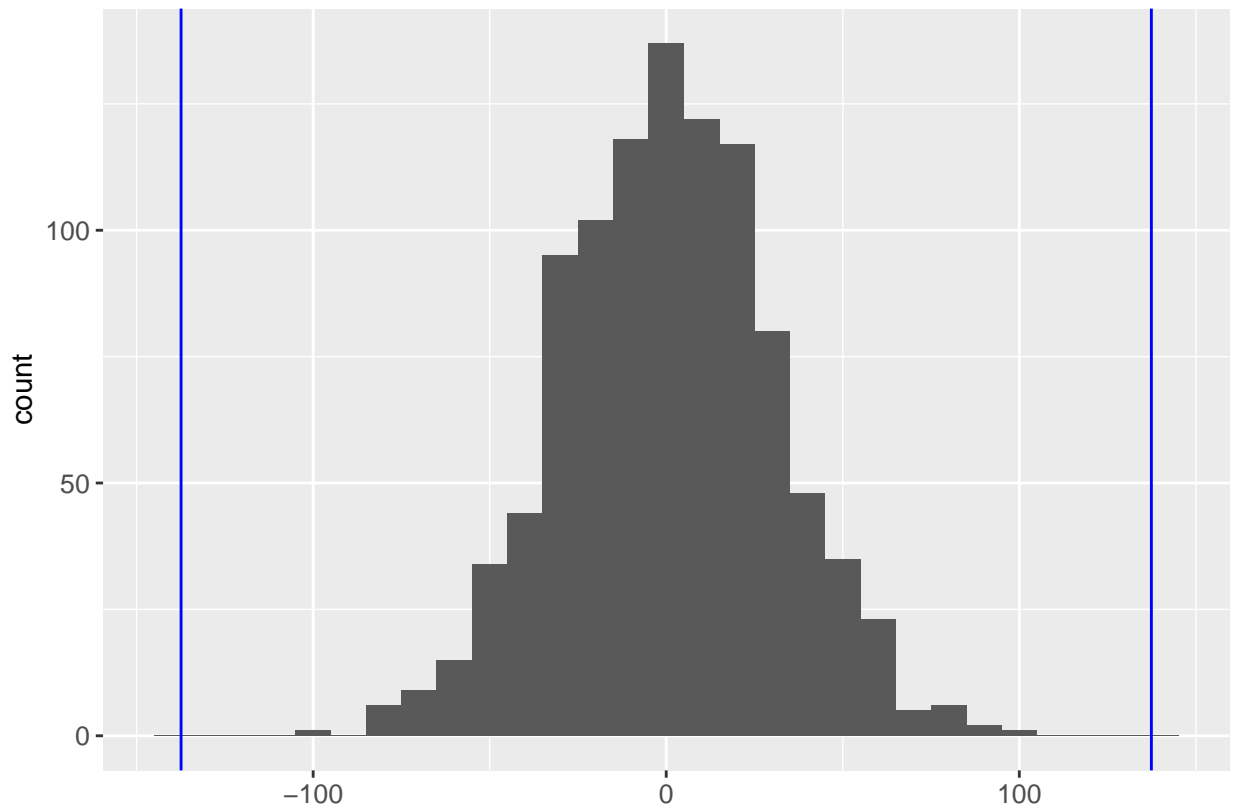
The mean SAT score from each region is compared. The mean and sd of SAT scores of Bronx are 1202.724 and 150.393901 each. The mean and sd of SAT scores of Manhattan are 1340.135 and 230.294140 each.

```
## # A tibble: 2 x 4
##   City      mean    sd     n
##   <chr>    <dbl> <dbl> <int>
## 1 Bronx    1203.   150.    98
## 2 Manhattan 1340.   230.    89
```

Then, the test statistics are calculated. The test statistic is the difference between the sample mean of Bronx's SAT score and the sample mean of Manhattan's SAT score. While $H_0 : \mu_b = \mu_m$, it can be also written as $\mu_b - \mu_m = 0$. The test statistics is $\bar{x}_b - \bar{x}_m$. Using the R program, the test statistics are calculated, and the result is 137.4103.

The simulation will be done for test statistics under H_0 . The simulation will be done using 1000 repetitions to examine how the test statistic might have looked if the null hypothesis was true. It will estimate the distribution of its possible values.

The simulation is done. Then, the result can be visually interpreted using the histogram data. Now, it is possible to assess evidence of the null hypothesis.



ferences in mean SAT scores between Bronx and Manhattan, assuming no difference in scores between the two regions.

The p-value* is 1. It means that there is strong evidence that agrees with the hypothesis of no difference in mean SAT scores depending on the region.

The result suggests that there is almost no difference between SAT scores of the two regions. There was no evidence found to reject the null hypothesis, where it stated the average SAT scores in the Bronx and Manhattan are the same. From this simulation, at least two cities in NYC are obtaining a similar level of education. The education department has put effort into reducing the gap between the regions in NYC.

If this is not true, the type 2 error could have occurred by not rejecting the null hypothesis when the null hypothesis is false. The null hypothesis may not be true.

Part 4. Confidence Interval

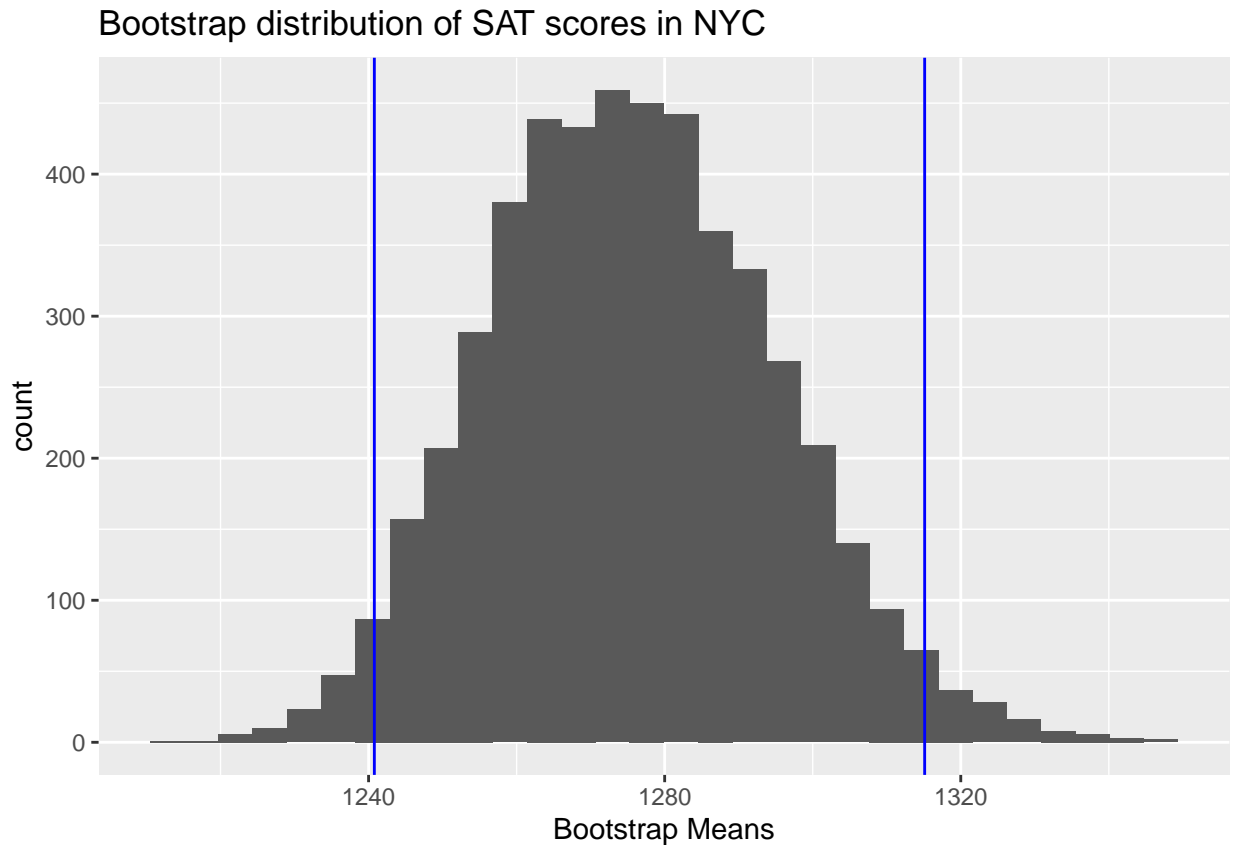
If the education equality was successful, all the mean of NYC regions should be similar to the nationwide average score.

For the first step, the bootstrap data will be generated.

A confidence level of 95% was used for this section. It means that 95% of the intervals should contain the population mean (Dekking, 2005). A 95% confidence level for a population parameter was calculated from the previous sample data. It gives a range of plausible values for the true parameter based on the limited information provided by the sample from the population.

```
##      2.5%      97.5%
## 1240.779 1315.121
```

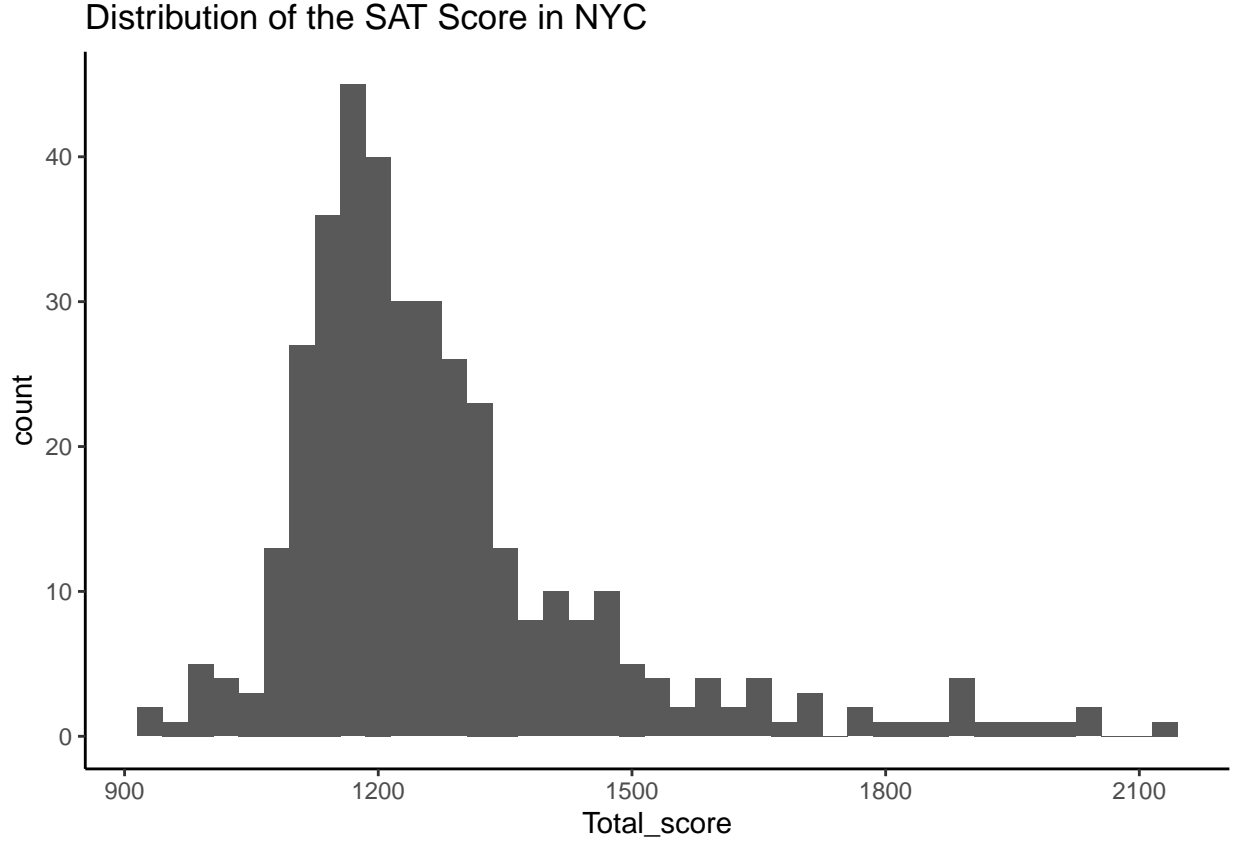
```
## 'stat_bin()' using 'bins = 30'. Pick better value with 'binwidth'.
```



The results can be interpreted as that we are 95% confident that the mean of the SAT score is between 1240.779 and 1315.121. However, the US's population average in total was 1497 in the same school year (Anderson, 2014). Overall, NYC is behind the entire education level in the US. They should work on education equality more for the students in all NYC regions to obtain a better education.

Part 5. Maximum Likelihood Estimator Calculation

First, the given data will be visualized in the histogram. All the scores will be graphed.



The graph looks like a normal distribution. It is right-skewed, but it is unimodal that most of the school's SAT score is distributed between 1000 and 1500. Therefore, the assumption will be made that the score of SAT data follows the normal distribution.

The parameter of the normal distribution is

$$x \sim N(1275.907, 194.9063^2)$$

The mean and the variance were obtained in the previous (**data**) section.

- MLE formula derivation is included in the Appendix.

The normal distribution has the maximum likelihood's sd when:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$$

The normal distribution has the maximum likelihood's mean when:

$$\hat{\mu}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} (X_{ij})$$

Using the parameters given as $x \sim N(1275.907, 194.9063^2)$.

Therefore, the MLE (maximum likelihood estimator)'s mean is 1275.907, and the standard deviation is 194.9063^2 for the normal distribution.

Therefore, the SAT score is most likely to occur at the mean of 1275.907 with a standard deviation of 194.9036. The mean of NYC is slightly lower than the worldwide SAT average score, which was 1497. The standard deviation was relatively high as well, meaning there are gaps between schools. Therefore, the department of education should aim to reduce the gap and push the overall NYC public school quality for overall better results.

Part 6. Linear Regression Method

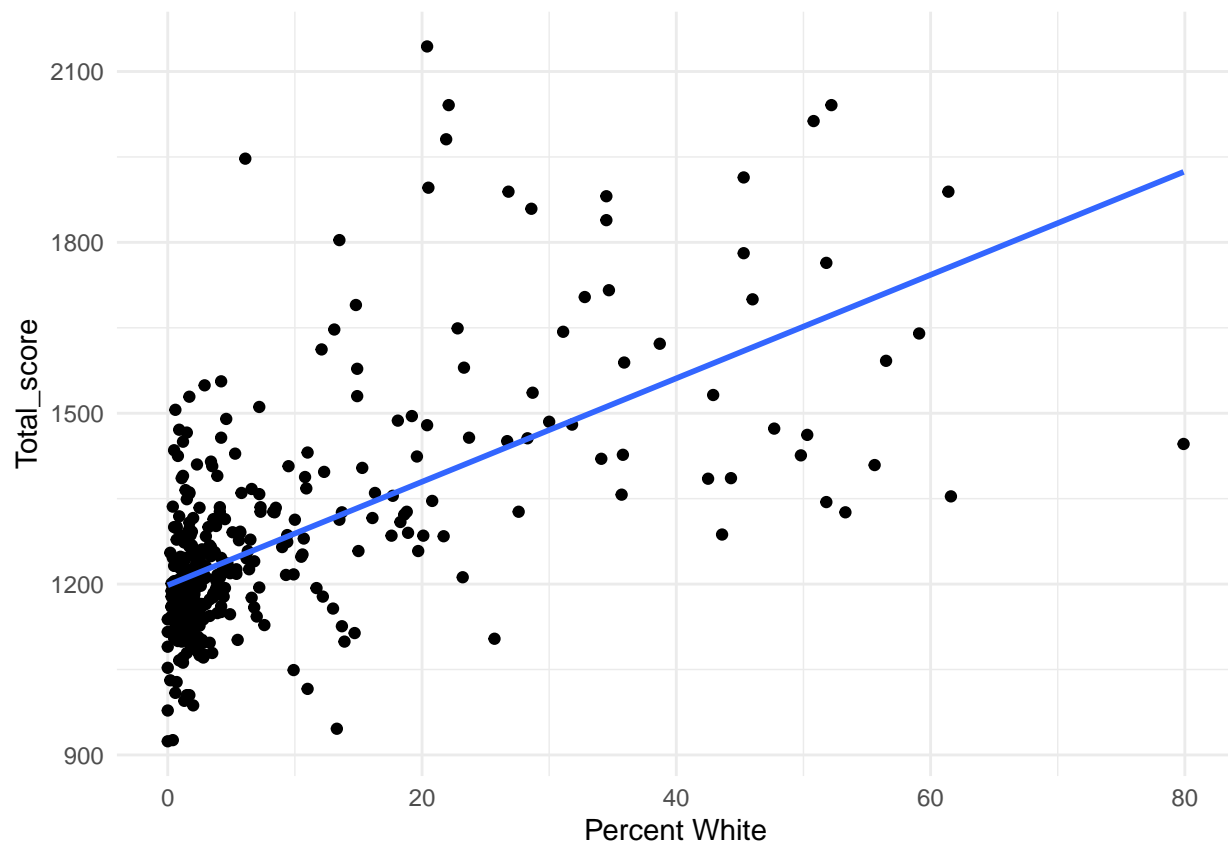
White Students

The chart indicates the summary of linear regression and introduces the expected intercepts *.

```
##               Estimate Std. Error   t value    Pr(>|t|)
## (Intercept)    1197.956384  9.3594926 127.99373 2.751099e-310
## 'Percent White'  9.083138  0.5891351  15.41775 7.669199e-42
```

It can now be represented in the graph visually with the linear regression line. It has a positive relation, where the slope is positive. It indicates that the more White students in the school, the higher the SAT average is.

```
## 'geom_smooth()' using formula 'y ~ x'
```

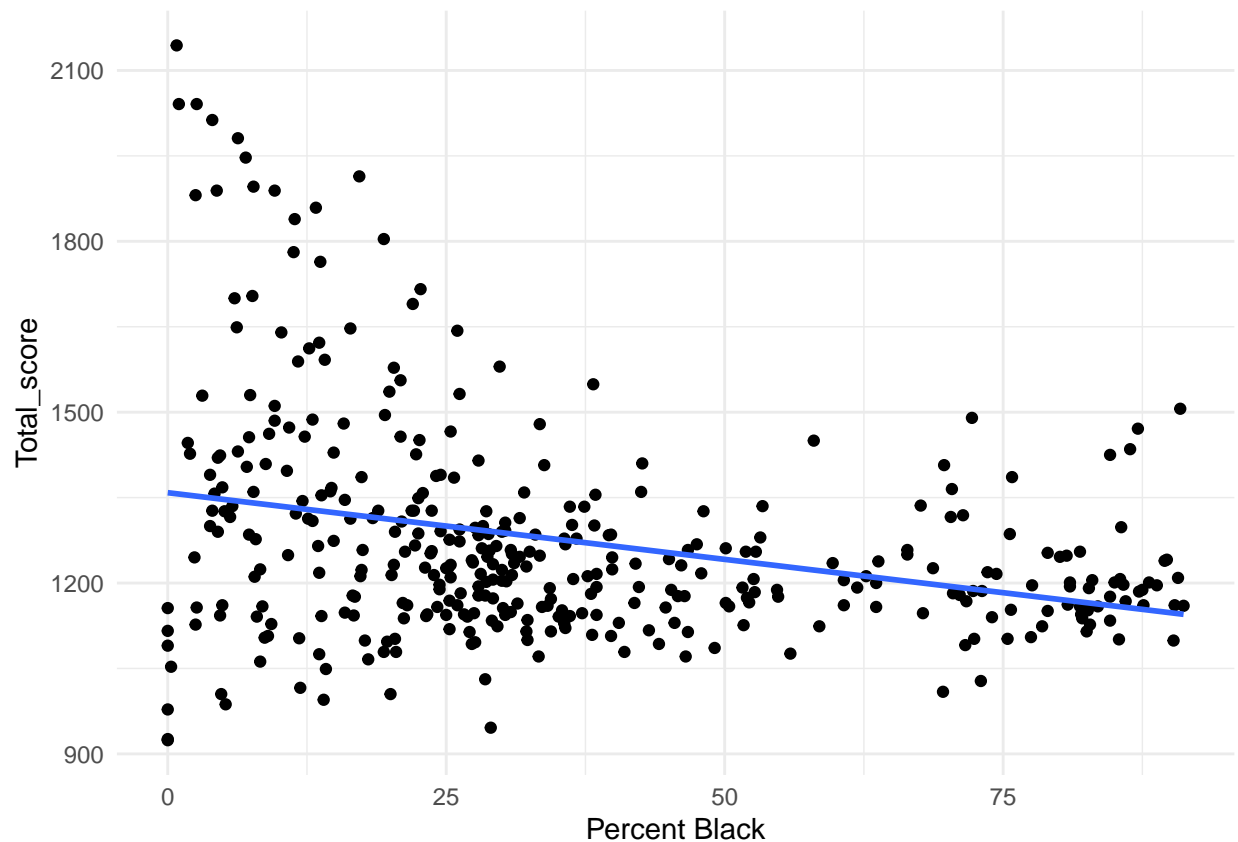


With similar steps, the test for the other races will be done. ### Black Students

```
##               Estimate Std. Error   t value    Pr(>|t|)
## (Intercept)    1358.590155  16.468856  82.494505 1.266184e-241
## 'Percent Black' -2.341088   0.378898 -6.178676 1.693413e-09
```

The slope is slightly negative. Therefore, the more Black students there are, the average SAT score tended to be lower.

```
## 'geom_smooth()' using formula 'y ~ x'
```

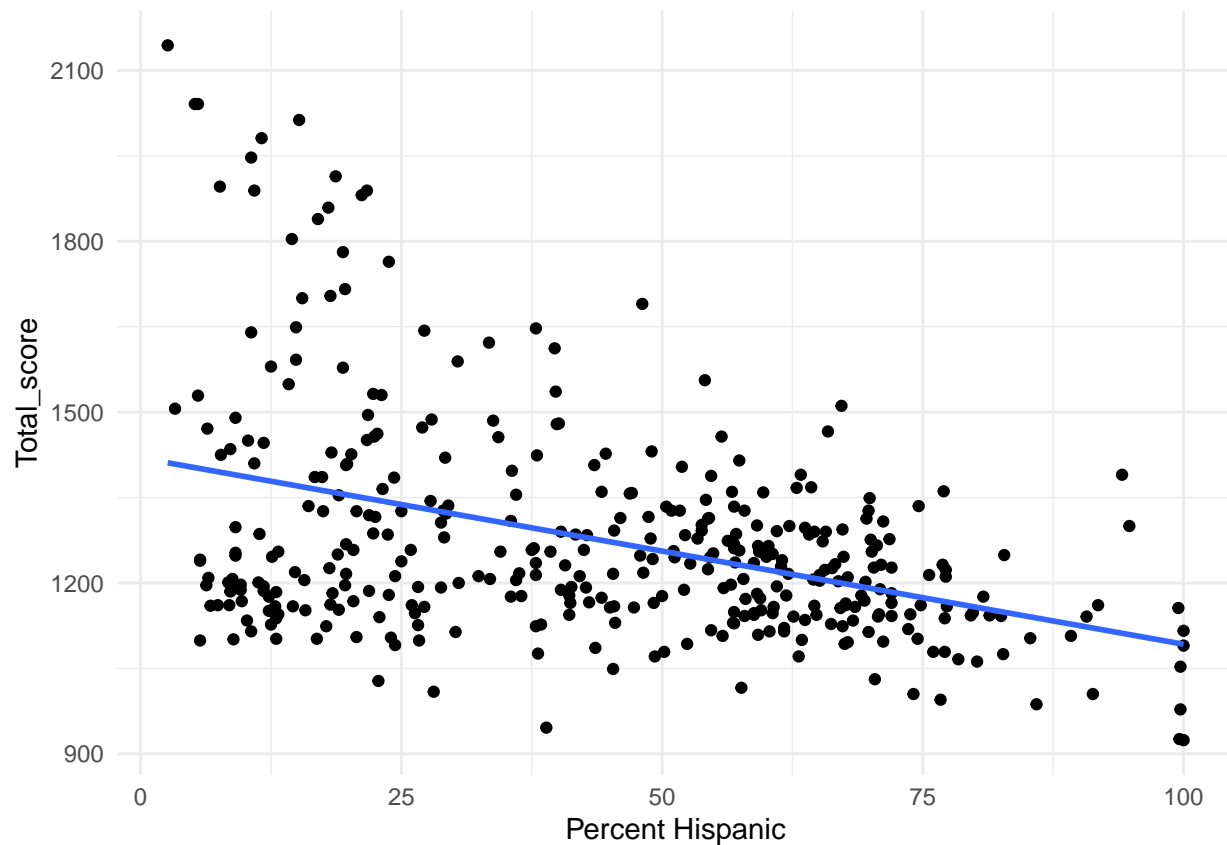


Hispanic Students

```
##               Estimate Std. Error   t value    Pr(>|t|)
## (Intercept)    1419.516494  18.895300  75.125373 2.519655e-227
## 'Percent Hispanic' -3.271096   0.376064 -8.698242 1.091451e-16
```

The slope is steep negative. Therefore, the more Hispanic students there are, the average SAT score tended to be lower. The score decreases at a faster rate than Black students.

```
## 'geom_smooth()' using formula 'y ~ x'
```

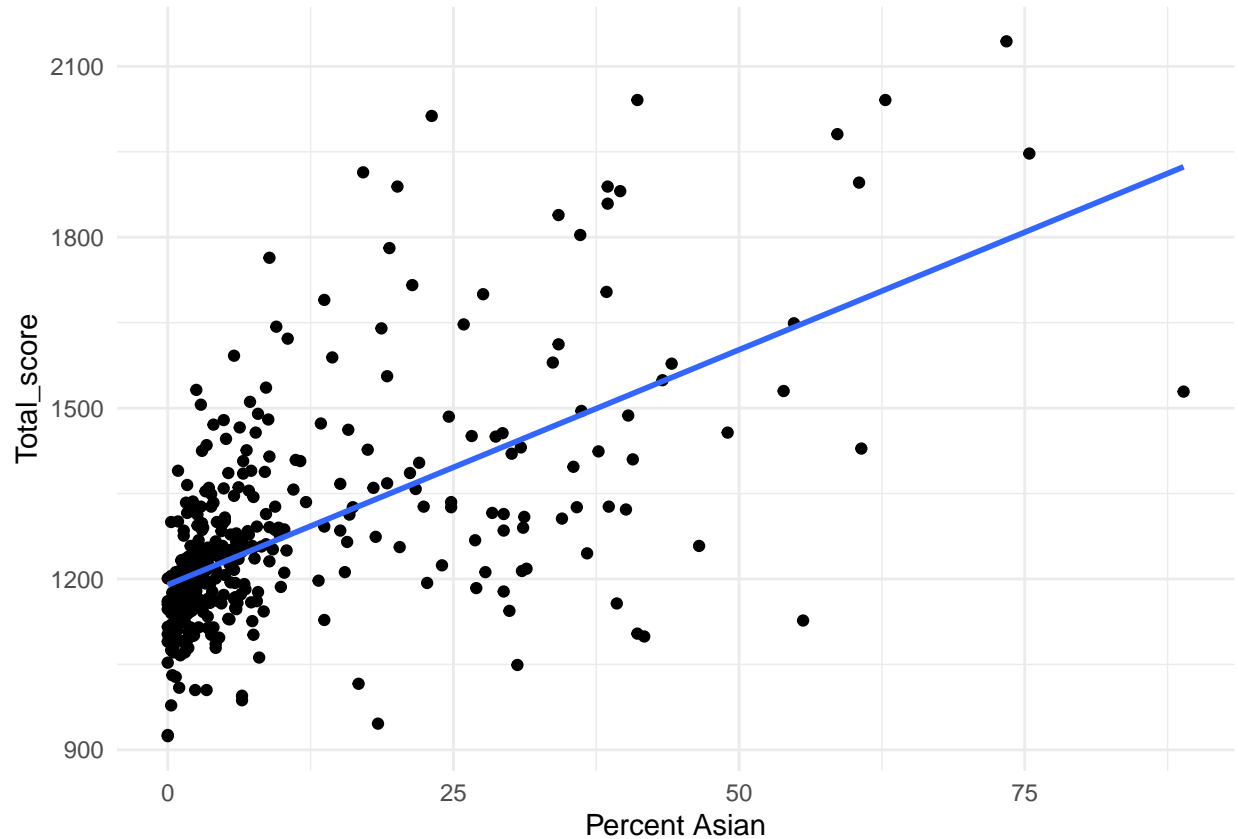


Asian Students

```
##               Estimate Std. Error  t value    Pr(>|t|)
## (Intercept)   1189.624274   9.8722017 120.50243 9.533448e-301
## 'Percent Asian' 8.256372   0.5553405  14.86722 1.368911e-39
```

The slope is strongly negative. Therefore, the more Asian students there are, the average SAT score tended to be higher.

```
## 'geom_smooth()' using formula 'y ~ x'
```



From the result of linear regression, the education levelling has not done a great job of reducing the education gap between the races. The slopes were either positive or negative, that the SAT average score distribution was highly dependent on the racial demographics. From this section, the conclusion is that the improvement in NYC education is still required.

Conclusion

The project's goal was to examine if NYC had improved on education equity for all the students who attend the public school in NYC. The evaluation was done with the average scores of SAT from public school. Some schools did not provide enough information for the entire SAT average score. Therefore, they were not considered in the data. The null hypothesis that education's equity is accomplished, various methods were used to test the hypothesis.

Results and Weakness

The first method was the Bayesian Credible interval with a confidence of 95%. The goal was to find the interval that has more than 200 schools that are over the mean. Therefore, the interval for 95% is (0.452, 0.548). The probability of having half of the schools with higher scores than the mean is between 0.452 and 0.548. However, the method's weakness is that the assumption was made that the prior distribution was beta distribution without the exact prove. This result may not be exact, while the strong assumption for prior distribution was made. According to this method's result, NYC seems to accomplished education equity in 2014-15.

The second method was the goodness of fit test. The actual data was used to examine if the data fit the binomial distribution, with the probability of $\frac{1}{2}$. The mean of SAT US-wide was used to see the overall NYC

performance compared to the other regions outside NYC. The mean of NYC's SAT score was used to know the education distribution within NYC. The Chi-squared test was being used in the section. The conducted null hypothesis is that the SAT score data comes from a binomial distribution. The alternative hypothesis is that the data does not come from a binomial distribution. Therefore, NYC did not perform enough to satisfy the overall education equity.

The third method was the hypothesis test of the mean. The mean of Manhattan was compared with the school in Bronx's mean score. The null hypothesis was that the mean of Manhattan's SAT is the same as the random Bronx's mean of SAT $H_0 : \mu_b = \mu_m$. The alternative hypothesis was that they do not equal to each other. $H_0 : \mu_b \neq \mu_m$. There was no evidence found to reject the null hypothesis because the p-value was significantly large. It indicates that the average SAT scores in the Bronx and Manhattan are the same (similar). From this simulation, it can be found that at least two cities in NYC are obtaining a similar level of education. The weakness is that the results of a hypothesis test are based on probabilities. For example, the null hypothesis was not rejected in this section. However, if the simulation could be biased and not representative of the population, then simulated data may not represent the population, which may cause type 2 error.

The fourth method was the confidence interval. The empirical bootstrap method was used. It drew 1000 bootstrap samples of size n with replacement from the original SAT score sample. The confidence interval of 95% was that the mean of the SAT score is between 1240.779 and 1315.121. However, the population average of the US in total was 1497 in the same school year (Anderson, 2014). The whole confidence interval is below the US total mean. Therefore, NYC overall has a lower score of SAT. The weakness is that the test does not produce a numeric measure of the degree of significance and indicates whether P is more or less than 0.05. Also, bootstrapping depends on the representative sample. If the sample schools had a biased score, likely, the result is also biased.

The fifth method was maximum likelihood estimator calculation. The assumption was that if equity has been successfully accomplished, there will be a normal distribution of the SAT score. The likelihood function and the maximum likelihood estimator were found using the normal distribution. The SAT score is most likely to occur at the mean of 1275.907. The mean of NYC is slightly lower than the worldwide SAT average score, which was 1497. Thus, the department of education in NYC should increase the quality of education of the overall NYC public school for overall better results. The weakness was that the distribution was estimated as the normal distribution, which may not be correct. If the distribution set was wrong, the entire calculation might be wrong. Therefore, choosing the density function carefully would be the most crucial step for an accurate result.

The last method was the linear regression method. It was used to examine the race demographics. If the race's score is affected, it would have a positive or negative slope rather than a flat slope. As a result, the slopes of scores in all the races were either positive or negative, that the SAT average score distribution was highly dependent on the racial demographics. Therefore, the equity of education in the race was not accomplished yet in NYC. The weakness of this method is assuming the linearity between the dependent variable and the independent variables. They may have a non-linear relationship. Therefore, it might sometimes lead to making the incorrect or irrelevant relation.

Next Steps

NYC Department of education should adjust the policy first while the current results did not satisfy the education equity. In the future, after adjusting the education policy, the data from the same schools can be collected again. It would be best to have a term of 2 years so that the new education can settle down and be effective in students' performance. Then, using the same six methods, the tests can be conducted again. The analysis can be done in precisely the same way how it was done this project. If it indicates that NYC has overall the better mean and no gaps between the regions, it can be considered that the new policy was more helpful than the current policy. Then, it should keep going with the new guidelines. However, if the results worsen, NYC should decide to return to the current policy or make improvements once again.

Discussion

Overall, NYC did not accomplish the education equity enough. Therefore, the NYC education department will need a better policy for a higher quality of education. NYC should aim to reduce the difference of score due to the racial disparity. The racial demographic should not affect the quality of the education that the students can get. Also, the overall NYC education quality should be improved. It would be best to have a similar mean with the overall US-wide SAT score. The gap between the NYC mean and the US-wide mean should be reduced. Lastly, the gap within the NYC region should be reduced. Even though the two largest cities, Manhattan and Bronx, did not have differences, the improvements should continue to be made so that all the cities do not have differences in education quality. While having the education is all students' right, it would be best to provide a high quality of education.

Bibliography

1. Anderson, A (October 7, 2014). *SAT scores for Class of 2014 show no improvement from previous marks*. The Washington Post. https://www.washingtonpost.com/local/education/sat-scores-for-class-of-2014-show-no-improvement-from-previous-marks/2014/10/06/80beb554-4d5b-11e4-aa5e-7153e466a02d_story.html
2. Conely, J. (November 25, 2019). *NYC Students Strike to Demand Racial Equity in Nation's Largest—and Most Segregated—School District*. Common Dreams. <https://www.commondreams.org/news/2019/11/25/nyc-students-strike-demand-racial-equity-nations-largest-and-most-segregated-school>
3. Dekking, F. M., et al. (2005) *A Modern Introduction to Probability and Statistics: Understanding why and how*. Springer Science & Business Media.
4. Gibbs, A and Stringer, A (January 20, 2021). *Probability, Statistics, and Data Analysis*. Github. <https://awstringer1.github.io/sta238-book/index.html>
5. Khandelwal, A. (March 15, 2021). *Planning to study in the US? The SATs have changed and this is what it means for you*. The Economic Times. <https://economictimes.indiatimes.com/nri/study/planning-to-study-in-the-us-the-sats-have-changed-and-this-is-what-it-means-for-you/articleshow/81323605.cms>
6. NYC Open Data. (2017). *Average SAT Scores for NYC Public Schools*. Kaggle. <https://www.kaggle.com/nycopendata/high-schools>
7. Reeves, R and Halikias, D (February 1, 2017). *Race gaps in SAT scores highlight inequality and hinder upward mobility*. Brookings. <https://www.brookings.edu/research/race-gaps-in-sat-scores-highlight-inequality-and-hinder-upward-mobility/>
8. Schalkwyk G.J. (2011) *Scholastic Aptitude Test*. In: Kreutzer J.S., DeLuca J., Caplan B. (eds) *Encyclopedia of Clinical Neuropsychology*. Springer, New York, NY. https://doi.org/10.1007/978-0-387-79948-3_1487

Appendix

Calculating a,b for beta distribution

Before looking at the actual data, we assumed that at least half of the students could get a more outstanding grade than the mean. Also, the other half of the students can get a lower grade than the mean SAT score.

Therefore, it is equally likely to pick a school that is a school with a “good quality” of education or a school with an “unsatisfying quality” of education. However, it cannot be precisely 0.5 of probability. The

assumption is that the parameter θ is between 0.3 and 0.7. While the mean is 0.5, the expectation and the variance can be calculated.

The expectation is 0.5, and the variance is 0.01, where the standard deviation is 0.1.

The formula for finding the expectation in beta distribution is:

$$E(\theta) = \frac{a}{a+b}$$

While the expectation is calculated, the equation can be written as:

$$E(\theta) = \frac{a}{a+b} = 0.5$$

(Gibbs and Stringer, 2021). The formula for finding the variance in beta distribution is:

$$Var(\theta) = \frac{ab}{(a+b)^2(a+b+1)}$$

While the expectation is calculated, the equation can be written as:

$$Var(\theta) = \frac{ab}{(a+b)^2(a+b+1)} = 0.1^2$$

(Gibbs and Stringer, 2021).

To find a and b, the 2 equations will be used. The variance equation is divided into two:

$$\frac{a}{a+b} \frac{b}{(a+b)(a+b+1)} = 0.1^2$$

While

$$\frac{a}{a+b} = 0.5$$

, it can be substituted.

$$0.5 \frac{b}{(a+b)(a+b+1)} = 0.1^2$$

$$\frac{b}{(a+b)(a+b+1)} = 0.02$$

From the expectation formula,

$$\frac{a}{a+b} = 0.5$$

$$a = 0.5(a+b)$$

$$a = 0.5a + 0.5b$$

$$0.5a = 0.5b$$

$$a = b$$

Lastly, substituting $a = b$, the equation can be solved as:

$$\frac{a}{(a+a)(a+a+1)} = 0.02$$

$$\frac{a}{(2a)(2a+1)} = 0.02$$

$$\frac{a}{4a^2 + 2a} = 0.02$$

$$a = 0.08a^2 + 0.04a$$

$$0.08a^2 = 0.96a$$

$$a = 12, b = 12$$

Therefore, the parameter of the beta equation is (12,12)

Posterior Distribution from Beta Prior Distribution

The prior beta distribution has the density of:

$$p(\theta) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{(a-1)} (1-\theta)^{(b-1)}$$

(Gibbs and Stringer, 2021) The likelihood function is:

$$p(X|\theta) = \theta^{\sum_{i=1}^n x_i} (1-\theta)^{(n-\sum_{i=1}^n x_i)}$$

Using Bayes' Rule, the posterior distribution of beta prior distribution can be found using this equation:

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{P(X)}$$

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{\int_0^1 P(X|\theta)P(\theta)d\theta}$$

From the above equation, $P(X) = \int_0^1 P(X|\theta)P(\theta)d\theta$ Therefore,

$$P(X) = \int_0^1 \theta^{\sum_{i=1}^n x_i} (1-\theta)^{(n-\sum_{i=1}^n x_i)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{(a-1)} (1-\theta)^{(b-1)}$$

$$p(X) = \int_0^1 \theta^{\sum_{i=1}^n x_i + a - 1} (1-\theta)^{(n-\sum_{i=1}^n x_i + b - 1)} d\theta \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$

$$p(X) = \frac{\Gamma(\sum_{i=1}^n x_i + a) \Gamma(n - \sum_{i=1}^n x_i + b)}{\Gamma(n + a + b)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}$$

Putting them altogether to find the posterior using Bayes' Rule,

$$P(\theta|X) = \frac{P(X|\theta)P(\theta)}{\int_0^1 P(X|\theta)P(\theta)d\theta}$$

$$P(\theta|X) = \frac{\theta^{\sum_{i=1}^n x_i} (1-\theta)^{(n-\sum_{i=1}^n x_i)} \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{(a-1)} (1-\theta)^{(b-1)}}{\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \frac{\Gamma(\sum_{i=1}^n x_i + a) \Gamma(n - \sum_{i=1}^n x_i + b)}{\Gamma(n + a + b)}}$$

After cleaning the equation, the equation is:

$$P(\theta|X) = \frac{\Gamma(n + a + b)}{\Gamma(\sum_{i=1}^n x_i + a) \Gamma(n - \sum_{i=1}^n x_i + b)} \theta^{\sum_{i=1}^n x_i + a - 1} (1-\theta)^{(n-\sum_{i=1}^n x_i + b - 1)}$$

It seems complicated. However, looking at the structure of the equation, it looks similar to the beta distribution again. The parameter, $a = a + \sum_{i=1}^n x_i$ and $b = b + n - \sum_{i=1}^n x_i$ Therefore, the posterior distribution is: $Beta(a + \sum_{i=1}^n x_i, b + n - \sum_{i=1}^n x_i)$

MLE Derivation for Normal Distribution

The probability density function of normal distribution is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2} \left(\frac{x-\mu}{\sigma}\right)^2\right)$$

The likelihood function of normal distribution is identified:

$$L(\mu, \sigma^2) = f(x_1)f(x_2)...f(x_n)$$

$$L(\mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x_1 - \mu}{\sigma}\right)^2\right) \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x_2 - \mu}{\sigma}\right)^2\right) \dots \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x_n - \mu}{\sigma}\right)^2\right)$$

$$L(\mu, \sigma^2) = \frac{1}{\sigma^n \sqrt{2^2 \pi^2}} \exp\left(-\frac{1}{2\sigma^2} \left(\sum_{i=1}^n (X_i - \mu)^2\right)\right)$$

Then, the loglikelihood is found for a more straightforward derivative calculation.

$$\ln(f_{\mu, \sigma^2}) = \ln\left(\frac{1}{\sigma^n \sqrt{2^2 \pi^2}} \exp\left(-\frac{1}{2\sigma^2} \left(\sum_{i=1}^n (X_i - \mu)^2\right)\right)\right)$$

$$l(\mu, \sigma^2) = -\ln(\sigma^n \sqrt{2^2 \pi^2}) + \ln(\exp\left(-\frac{1}{2\sigma^2} \left(\sum_{i=1}^n (X_i - \mu)^2\right)\right))$$

$$l(\mu, \sigma^2) = -\ln(\sigma^n 2^{\frac{n}{2}} \pi^{\frac{n}{2}}) + \left(-\frac{1}{2\sigma^2} \left(\sum_{i=1}^n (X_i - \mu)^2\right)\right)$$

$$l(\mu, \sigma^2) = \frac{-n}{2} \ln \sigma^2 - \frac{n}{2} \ln(2\pi) - \left(\frac{1}{2\sigma^2} \left(\sum_{i=1}^n (X_i - \mu)^2\right)\right)$$

The derivative was found to find the maximum.

$$\frac{\partial l}{\partial \sigma^2} = \frac{-n}{\sigma^2} + \frac{1}{\sigma^4} \left(\sum_{i=1}^n (X_i - \mu)^2\right)$$

The maximum may exist when the result of derivative equals zero.

$$\frac{\partial l}{\partial \sigma^2} = 0$$

$$\frac{-n}{\sigma^2} + \frac{1}{\sigma^4} \left(\sum_{i=1}^n (X_i - \mu)^2\right) = 0$$

$$\frac{-n}{(\sigma^2)^2} (\sigma^2 - \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2) = 0$$

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2$$

The maximum likelihood for the normal distribution is equal to the variance.

The second derivative test was used to prove if it is the maximum value because it can also be minimum when the derivative equals zero.

$$\frac{\partial^2 l}{\partial (\sigma^2)^2} = \frac{n}{(\sigma^2)^2} - \frac{2}{(\sigma^2)^3} \left(\sum_{i=1}^n (X_i - \mu)^2\right)$$

$$\frac{\partial^2 l}{\partial (\sigma^2)^2} = \frac{n}{(\sigma^2)^2} \left(1 - \frac{2}{\sigma^2} \frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2\right)$$

Since $\frac{1}{n} \sum_{i=1}^n (X_i - \mu)^2 = \hat{\sigma}^2$:

$$\frac{\partial^2 l}{\partial (\sigma^2)^2} = \frac{n}{(\sigma^2)^2} \left(1 - \frac{2}{\sigma^2} \hat{\sigma}^2\right)$$

$$\frac{\partial^2 l}{\partial(\sigma^2)^2} = \frac{n}{(\sigma^2)^2}(1-2)$$

$$\frac{\partial^2 l}{\partial(\sigma^2)^2} = \frac{n}{(\sigma^2)^2}(-1)$$

While $n > 0$ and $(\sigma^2)^2 > 0$:

$$\frac{\partial^2 l}{\partial(\sigma^2)^2} < 0$$

The second derivative was negative at the maximum likelihood. Therefore, the graph of the likelihood was concave down at the point, which makes the maximum likelihood.

Linear Regression

Linear regression assumes the best linear line along the numerous data points. It is assumed that the independent variables are nonrandom and dependent variable values are realization of random variables. U_i is the independent random variables with $E[U_i] = 0$ and $Var(U_i) = \sigma^2$ (Dekking, 2005) In equation, it can be written as:

$$Y_i = \alpha + \beta x_i + U_i$$

α is a intercept parameter. β is a slope parameter

Regression Coefficients: Estimate is the value of $\hat{\beta}$. The Std. Error represent the estimated standard deviation of $\hat{\beta}$. The t value is the number of $\hat{\beta}$ is away from 0. $Pr(>|t|)$ is the p-value for a hypothesis test of H_0 vs. H_A .

P-values

The p-value is the probability of obtaining results at least as extreme as the observed results of a statistical hypothesis test, assuming that the null hypothesis is correct. A smaller p-value means that there is stronger evidence in favor of the alternative hypothesis (Dekking, 2005).