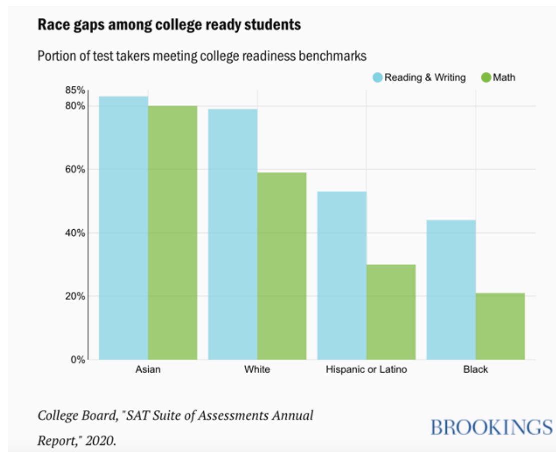
AVERAGE SAT SCORE ESTIMATION FOR SCHOOL BASED ON THE RACE OF THE STUDENT

Bahareh Zahirodini

Introduction

SAT was created in 1926 with the aim of providing equal opportunities for college admission and scholarships to talented students, regardless of their income. FrotLine(https://www.pbs.org/)

Many colleges use SAT scores for admissions and financial aid decisions. More selective institutions require high SAT scores for entry—and there are even bigger race gaps at the top of the score distribution. Bookings (https://Booking.edu)



The state of Massachusetts, like other states in the US, high schools must submit annual reports to their respective state education departments detailing their performance over the academic year. These reports typically include a range of metrics, including graduation rates, standardized test scores, reports on the schools' staff and teachers, and demographic information of students as proportion. The state of Massachusetts, like other states in the US, is committed to providing high-quality education to its students. In this report, We aimed to investigate whether the number of students of different races could be used to predict the average SAT score of a school in Massachusetts. Specifically, we aimed to determine if the race of students could explain a significant proportion of the variance in SAT scores.

Dataset

The Department of Elementary and Secondary Education (dese) is a government agency in Massachusetts responsible for overseeing the education of students from kindergarten through high school. They gather Information for schools and districts on the non-fiscal data and reports collected by the Department.

The dataset we used in this report was provided by the Department of Elementary and Secondary Education and published on their public online database in the Information Service section(https://profiles.doe.mass.edu/). We chose four reports to conduct the following analysis: the average high school SAT score from the SAT performance report, the Race ratio of schools from the Race and sex report, teachers' headcount from the teacher by grade, and the number of students from the enrollment by grade report for the educational year 2019-2020.to be noted, the data is on district level which might consist of one or group of high schools.

To ensure data cleanliness, we excluded schools with incomplete data or SAT scores outside the range of 400-1600.

Methodology

The question was, what proportion of SAT scores can be explained by the race of students? We took SAT score as the response variable.The total number of predictor variables was seven, the various races in dese reports, including Black, Asian, Hispanic, White, Native American, Native Hawaiian, and Multi-Race. For variable screening, WE adopted the brute force method to consider all combinations of variables for a model and find the best by higher adjusted R-squared qualification. Since WE applied multiple linear regression model, the general model would be:

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_7 + \varepsilon$$

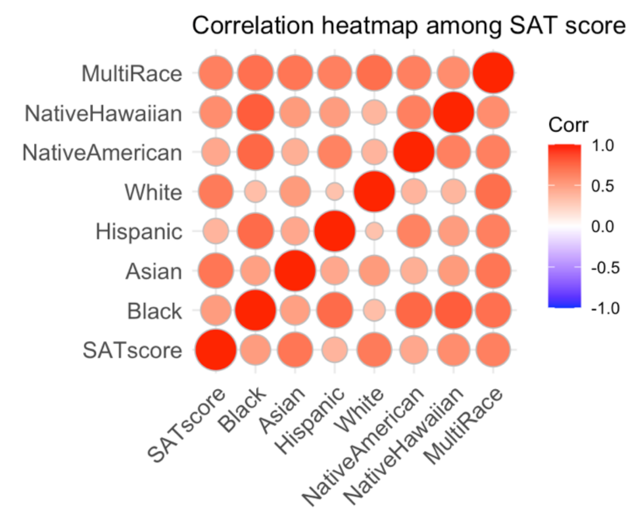
And the fitted model would be:

$$\hat{y} = \widehat{\beta_0} + \widehat{\beta_1} x_1 + \cdots + \hat{\beta}_k x_{k7}$$

Data Analysis

One limitation of our dataset was the race ratio of students reported on the school level and not specifically in twelfth grade. Therefore, we assumed the same ratio applied to twelfth-grade students.

Upon examining the data for potential multicollinearity, we observed that all variables had a VIF value of less than 10 and a conditional index of less than 30. Based on these findings, we determined no significant multicollinearity among the variables, and further investigation is unnecessary.



Tolerance and Variance Inflation Factor

	Variables	Tolerance	VIF
1	Black	0.1658401	6.029906
2	Asian	0.4811615	2.078304
3	White	0.3892948	2.568747
4	Hispanic	0.4139813	2.415568
5	NativeAmerican	0.3911131	2.556805
6	NativeHawaiian	0.3139306	3.185417
7	MultiRace	0.1755329	5.696937

Eigenvalue and Condition Index

	Eigenvalue	Condition Index	intercept
1	5.09015843	1.000000	0.004943956
2	1.22355852	2.039639	0.107575410
3	0.55679617	3.023553	0.076627487
4	0.44928146	3.365940	0.005725178
5	0.26162508	4.410887	0.151051172
6	0.21310159	4.887337	0.251056051
7	0.14256651	5.975261	0.268485974
8	0.06291224	8.994936	0.134534771

Continuing with the variable screening, the response variable is the average SAT score for a school, and the predictor variables are the number of students of each race in it. The brute force method revealed that the combination of three variables, namely Asian, White, and Native Hawaiian, had the lowest AIC value, while the combination of six variables had the highest adjusted R-squared value.

Best Subsets Regression

Model Index	Predictors
1	Asian
2	Asian White
3	Asian White NativeHawaiian
4	Asian White NativeHawaiian MultiRace
5	Black Asian White NativeHawaiian MultiRace
6	Black Asian White Hispanic NativeHawaiian MultiRace
7	Black Asian White Hispanic NativeAmerican NativeHawaiian MultiRace

Subsets Regression Summary

Model	R-Square	Adj. R-Square	Pred R-Square	C(p)	AIC	SBIC	SBC	MSEP	FPE	HSP	APC
1	0.4745	0.4727	0.399	133.9463	3596.8610	2772.4474	3607.8707	4078353.4053	14160.2734	49.0010	0.5328
2	0.5992	0.5964	0.5346	36.3115	3520.3195	2696.7494	3534.9990	3121615.8556	10875.4125	37.6366	0.4092
3	0.6415	0.6377	0.5377	4.5077	3489.9762	2667.0890	3508.3256	2801975.6818	9795.0136	33.9009	0.3686
4	0.6430	0.6380	0.5124	5.2732	3490.7240	2667.9048	3512.7433	2799726.7967	9820.3208	33.9925	0.3695
5	0.6456	0.6393	0.504	5.2397	3490.6494	2667.9503	3516.3386	2789592.1022	9817.8193	33.9887	0.3694
6	0.6471	0.6396	0.4748	6.0092	3491.3867	2668.7975	3520.7458	2787321.7375	9842.8473	34.0811	0.3704
7	0.6471	0.6384	0.4705	8.0000	3493.3773	2670.8452	3526.4062	2797149.9077	9910.6866	34.3225	0.3729

According to the brute force method result, examined the fitted model for combinations of three and six variables by least AIC and Highest adjusted R-squared qualifications.

Model 1: with six race

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)     1.092e+03  9.784e+00 111.563  < 2e-16 ***
Black           2.081e-02  1.177e-02   1.767  0.07824 .
Asian           1.195e-01  1.477e-02   8.090 1.77e-14 ***
White           5.163e-02  6.668e-03   7.743 1.74e-13 ***
Hispanic       -4.621e-03  4.158e-03  -1.111  0.26740
NativeHawaiian  4.196e+00  1.397e+00   3.004  0.00291 **
MultiRace      -1.183e-01  7.676e-02  -1.541  0.12447
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 98.03 on 283 degrees of freedom
Multiple R-squared:  0.6471,	Adjusted R-squared:  0.6396
F-statistic: 86.49 on 6 and 283 DF,  p-value: < 2.2e-16
```

Model 2: with 3 race

```
Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    1.093e+03  9.676e+00 112.955  < 2e-16 ***
Asian          1.077e-01  1.315e-02   8.194 8.53e-15 ***
White          4.415e-02  4.946e-03   8.926  < 2e-16 ***
NativeHawaiian 5.435e+00  9.356e-01   5.809 1.67e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 98.29 on 286 degrees of freedom
Multiple R-squared:  0.6415,	Adjusted R-squared:  0.6377
F-statistic: 170.6 on 3 and 286 DF,  p-value: < 2.2e-16
```

Comparing these two models, Model 1 provided a higher adjusted R-squared. However, it lost the significance of individual variables. While the reliability of the coefficient of variables Black, Hispanic, and Multi race are questionable, the negative coefficient for Hispanic ND Multi-race was something to notice. Model 2 satisfied the latter condition but had higher residual SE, which leads to a wider prediction interval. In addition, Model 1 was more comprehensive and kept more races inside.

Model 1:

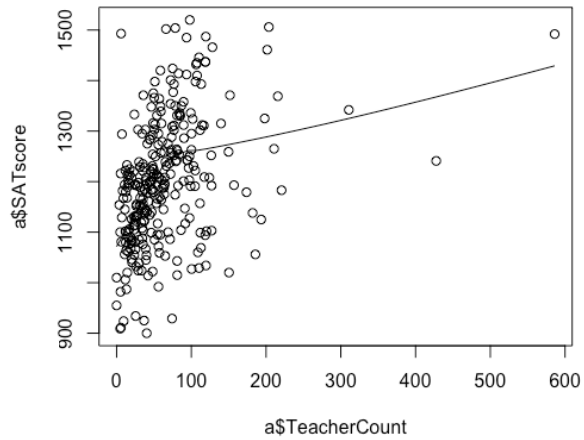
SAT score = 1092 + 0.208 (#Black) +.120 (#Asian) + 0.0052(#White) – 0.00046 (#Hispanic) + 4.196 (#Native Hawaiian) – 0.118 (#Multi-Race)

Model 2 :

SAT score = 1093 + 0.0108 (#Asian) + 0.0044 (#White) + 5.435 (#Native Hawaiian)

To search for improvement and examine how another variable that logically seemed as influential on the performance of students would improve the model and what the magnitude would be, if it is true.

We found the number of full-time equivalent teachers teaching in 9-12 grades that could explain a portion of variable response variation and created Model 3.




```
Residuals:
    Min      1Q  Median      3Q     Max 
-299.34  -52.44   -4.14   60.41  399.92 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept)   1086.24888    9.94457 109.230  < 2e-16 ***
TeacherCount     0.64553    0.15285   4.223 3.24e-05 ***
Asian            1.24487    0.18085   6.883 3.74e-11 ***
Hispanic        -0.24201    0.06459  -3.747 0.000217 ***
White            0.47113    0.06903   6.825 5.31e-11 ***
NativeHawaiian  38.09384   14.00666   2.720 0.006937 ** 
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 96.49 on 284 degrees of freedom
Multiple R-squared:  0.6569,	Adjusted R-squared:  0.6509 
F-statistic: 108.8 on 5 and 284 DF,  p-value: < 2.2e-16
```

Stepwise Selection Summary

Step	Variable	Added/ Removed	R-Square	Adj. R-Square	C(p)	AIC	RMSE
1	Asian	addition	0.475	0.473	148.9350	3596.8095	118.5782
2	White	addition	0.599	0.596	47.7830	3520.3021	103.7469
3	NativeHawaiian	addition	0.634	0.631	20.5870	3495.5896	99.2501
4	TeacherCount	addition	0.640	0.635	18.0390	3493.1980	98.6741
5	Hispanic	addition	0.657	0.651	6.0000	3481.2051	96.4915

In presence of teachers' count, Hispanic race came into the model.

Model 3:

SAT score = 1092 + 0.646 (Teacher Count) +1.255 (#Asian) + 0.471 (#White) – 0.242 (#Hispanic)+38.1 (#Native Hawaiian)

External model validation

To ensure the external validity of the Models, we performed PRESS statistics and Jackknife analysis. The PRESS statistics helped us assess the predictive performance of the model on new data by omitting one observation at a time, and the Jackknife analysis helped us determine the stability and reliability of our model.

Model 1: PRESS = 4109464, Jackknife R-squared = 0.467
Model 2: PRESS = 3718130, Jackknife R-squared = 0.518
Model 3: PRESS = 3668772, Jackknife R-squared = 0.524

Model 2 would perform better on prediction for other samples or the future, however, compared to the adjusted R-squared of the model.

Summary of improvements in the models:

	R_a^2	Residual SE (√MSE)	Model Sig.	Variables Sig.	R_{jack}^2	PRESS
Model 1	.6396	98.03	✓	-	0.467	4109464
Model 2	.6377	98.29	✓	✓	0.518	3718130
Model 3	.6509	96.49	✓	✓	0.524	3668772

Conclusion and limitation:

In conclusion, it should be noted that this analysis may have limitations in fully exploring the subject. However, one result suggests that the enrollment of students of certain races may have a marginally positive effect on the average SAT score for a school, while the enrollment of students of other races may have a negative effect.

It is also crucial to gather more data on the academic performance of these students prior to twelfth grade and identify potential barriers that hindered them from achieving higher scores. Such barriers could include economic, social, and cultural factors that disproportionately affect these groups. By identifying and addressing these barriers, educators and policymakers can develop more effective strategies to improve the academic outcomes of low performed students and promote a more equitable education system as according to Ezekiel J. Dixon-Román in his 2013 study on the influences of family income on Black and White high school students' SAT performance, there has been a removal of race-conscious policies such as affirmative action over the past three decades.

Moreover, in order to gain a more comprehensive understanding of the relationship between race and SAT scores, we included the proportion of twelfth-grade students of each race who took the SAT in the model. It should be noted that personal preferences may influence whether a student takes the exam, which may affect the results.

Finally, it is essential to acknowledge that other critical aspects must be explored to obtain more meaningful insights. Gathering and analyzing data on factors such as the number of students per teacher, access to educational facilities, and equal opportunities for students of different races can provide a more comprehensive understanding of the factors influencing SAT scores.