Modeling SAT Scores Within Connecticut Districts

Nomi Vilvovsky

Abstract

In this paper we explore the different factors that may affect the percent of students who pass the SAT within districts in Connecticut. Using R, we create graphs and run tests to determine significance of factors, then, we build possible models that may explain the construct of the data, and finally pick between the models and interpret the results. We can determine which factors are important, and therefore which factors should be regarded when working on the education system and attempting to increase academic performance.

Introduction

The education system in the United States is considered a serious issue that faces both American work force and future generations. As the age of information grows and computers are considered a tool that will eliminate encyclopedic knowledge, the academic expectations for students relies more heavily on their ability to think critically, communicate well, and problem solve.[1] According to Collegeboard, the SATs are designed to represent a mark of success for students that graduate public high schools based off critical thinking, thus we will be exploring the SAT benchmark as an effect describing whether students possess the proper skills to excel in academia and careers in the Western world.

Collegeboard, the facilitator of the SATs, draws a specific line within SAT test scores of “success” or “failure”—this is the benchmark score, i.e., the score that determines which students thrive and require a challenge vs. which students require additional academic support. These scores are specifically 530 in Math and 480 in Reading and Writing at graduation for high school students. Collegeboard states that students that meet the SAT benchmark are 75% likely to earn at least a C or greater in college academic courses related to the test they’ve passed: quantitative classes grades are affected by Math SAT scores and writing/reading based classes are affected by Writing/Critical Reading SAT scores.[2] Taking into account Collegeboard’s statements as fact, SAT scores should be a direct factor for whether students succeed in college classes or in careers requiring them to work on a college-level of academic difficulty.

But what factors effect SAT scores in Connecticut schools? At first glance, income of counties is a huge factor when determining the success of students academically. However, lots of research has been done on this account, and in the end, wealth is determined a confounding variable which effects many other

---

[1] "Remaking the Academy in the Age of Information." DE ALVA, JORGE KLOR. *Issues in Science and Technology* 16, no. 2 (1999): 52-58. http://www.jstor.org/stable/43313988.
[2] "Benchmarks," SAT Suite of Assessments, March 07, 2018, , accessed April 28, 2018, https://collegereadiness.collegeboard.org/about/scores/benchmarks.

factors that therefore affect the academic performance of students. Looking at the numbers of median income within counties in Connecticut, we note that there is a significant difference in incomes between counties, from the richest Fairfield county ($82.3k) to the least wealthy Windham ($59.3k)[3]. However, instead of including wealth as a direct factor in our model, we'll be using county as a random effect and include variables that are heavily impacted by wealth (absentee percentage and participation in the SATs), in order to determine the actual factors that impact SAT performance. By using county as a random effect, we will control both for factors such as wealth which are different for each county, as well as control for geographic factors of districts. There are 8 counties that we will be using as a random categorical effect.

The percentage of chronic absentee students is determined by the percentage of students in the school district that miss over 15 days of school. We will be using 2012-2013 data within Connecticut school districts for our modeling. Absentee students often miss school due to lack of transportation or an outside job. When student misses 15 school days in one school year, they become seven times more likely to drop out of college. [4] We will model whether absentee students also effect whether a district will perform poorly on the SAT. Conversely, we will also examine whether a higher SAT participation rate also improves scores, or perhaps a higher participation rate actually decreases the percent meeting SAT benchmark, perhaps due to students being pushed to take the SAT without proper preparation.

We will also examine the percentage of change in participation from the previous year (current-2013, previous-2012), in order to test the effects of such change; perhaps districts where the number of participants is growing will do better on the SATs. We are essentially testing whether growth of academic interest is an indicator of academic performance. We will also test the affect of the number of total students at the school to see whether a larger school has an effect on SAT performance. Perhaps a larger school which allows for more diverse classes and different class levels produces higher academic achievement, or perhaps a smaller school where students receive more individualized attention will have the greater impact on academic performance.

The CAPT is a state-mandated standardized test in Connecticut public schools with 4 subject scores: Math, Science, Writing, and Reading. Students take the CAPT in March of their sophomore year, therefore in order to test its relationship with SAT benchmark scores, we will be looking at each district's percent of passing CAPT scores in 2012. It is noteworthy that the SAT does not include a science portion of the test, so the CAPT science subject may not be as significant as the other tests, and if it is, perhaps the test material itself is not the only factor in determining the district's SAT success. [5] It is important to note that the CAPT scores have an inherent bias, because students who do not meet the CAPT benchmark scores must retake the test until they do. Therefore, the percent of students who meet the CAPT benchmark score will include some form of resampling, which shouldn't impact our model, since this is the consistent measure that districts use.

---

[3] "Household Income in Connecticut (State)." *Household Income in Connecticut (State) - Statistical Atlas*, statisticalatlas.com/state/Connecticut/Household-Income#figure/county.

[4] "Chronic Absenteeism In The Nation's Schools". 2018. *Ed.Gov*. https://ed.gov/datastory/chronicabsenteeism.html.

[5] "CAPT Data Interaction (Public)". 2018. *Solutions1.Emetric.Net*. http://solutions1.emetric.net/captpublic/Index.aspx.

Research Question and Goal

What factors within a district in Connecticut can be used to determine the percent of students from said district meeting the SAT benchmark? Do certain factors require attention in order to increase the number of Connecticut students who do well on the SAT?

The research goal is to determine which factors are critical within a district in order to aid Connecticut schools in increasing academic achievement. Some factors cannot be controlled, such as school size or district, and others can be worked on, such as absentee percentage. CAPT scores can be evaluated to determine whether this is an accurate test to determine academic achievement, i.e., does it test students in the same way that the SAT does? It is important to remember that the SAT is not a determinant of the districts' academic program, however, according to Collegeboard, it is a determinant of first year college performance.

Variable Transformations

Before running correlation tests or modeling variables, we must first determine that the variables themselves are well adjusted and use similar scales, in order to make the model easy to interpret. A Shapiro test can be used to evaluate that *benchmarkpc* (percent of participating students in a district that pass the SAT benchmark scores) follows a normal distribution (p=.4475). We'll also create *benchmark* to use full percentage number rather than decimal. This transformation will keep the percentage scales equal. Next, we can create a new variable *ttchange2* (total SAT participant change from 2012 to 2013) that is a full number percentage rather than a decimal, keeping the same scale as other variables. We perform the same linear transformation to create a new variable *partic* which is *partrate* (participation rate in 2013) as a full percentage rather than a decimal. Although we can see through scatterplots that some of the variables are not normally distributed, normality of variables is not an assumption for linear modeling. In order to later interpret the results of the model, we refrain from transforming our variables unless we fail to meet a model assumption later on.

Testing Correlations

In order to build an effective model and narrow down which variables to use, we conduct a series of correlation tests. All relevant graphs and figures are available in the appendix.

Scatterplots and Pearson's correlation tests can be used to determine the correlation (R) of the variable and the effect. These metrics do not define the significance of the variable, but can be used to test if they should be included in the model.

A scatterplot (*Figure 1*) of *absent* appears to have a negative correlation with *benchmark*. We will test using this term in our full model. A scatterplot of *ttchange2* shows a weak correlation, and a Pearson's correlation test shows that p=.6095, therefore *ttchange2* does not appear to have a significant correlation. In the final model, then, it can only correlate to *benchmark* with a confounding variable, and we cannot use it to prove causation. A scatterplot (*Figure 2*) of *partic* along with a Pearson's correlation test reveal that *partic* is correlated with *benchmark*. The p-value is extremely small. We will use this term in our full model. CAPT scores all show a strong correlation on a scatterplot (*Figure 3*).

MODEL

In order to determine which model best describes the effects of the variables on the benchmark SAT percentage, we build multiple models and test them against each other to find the most significant model. We test interactions and well as the significance of the random effect of county. The full linear model with each variable but with no interaction is Model 5: *benchmark= ttchange2 + partic + absent + readdpi + mathdpi + writedpi + sciencedpi + (1|county).* This model includes no interactions and meets the assumption of normality of residuals (p= .6026). Model 6 is Model 5 with the random effect of county removed.

Next, we attempt to build multiple models with different interactions stemming from our investigation of correlation as well as intuitive knowledge of factors that interact. We include all variables which showed a correlation in our Pearson and scatterplot tests as was done in Model 5/6. We assume county as a random effect for all the models, because the districts are nested within the different counties and it also insures we control for geographic interaction between districts, and we also build models that do not include this random effect variable to test the significance of county.

We build Model 1 including the most intuitive interaction of the rate of participation in SATs and the rate of absentees (lack of participation in schooling). Districts with a high absentee rate should theoretically have lower participation rates, and thus these two factors interact. We build Model 1: *benchmark = partic + absent + readdpi + mathdpi + writedpi + sciencedpi + partic*absent + (1|county).* When testing the normality of residuals assumption, the Shapiro-Wilk test shows normality, with p=.9625. We also create a sibling Model 2 which is simply Model with the random effect of county removed, to test the significance and necessity of this variable.

Next, we construct Model 3 similarly to Model 2, but with an additional interaction variable between change in participation and current participation, since those two factors are inherently related (a large increase in participation would be correlated with a large current participation rate, i.e., a district with a small participation rate id not likely to have experienced a large increase in participation). Since we add this interaction term, we therefore add *ttchange2* (the change in participation) as an additive factor. The Normality assumption of this model is met (p=.8909). We also create a sibling Model 4 which is Model 3 but simply with the random effect factor removed.

Using an ANOVA comparison of the models, we pick Model 1 with p=.00000124, verifying this is the best option wince Model 1 has both the lowest AIC and BIC. Model 1 is the strongest model from Models 1-6. We now create a new model, Model 7, where we remove the “insignificant” variables from the output of Model 1 (testing at 95%: t=1.984 at df=100). When we run another ANOVA to test Model 1 versus Model 7, Model 1 remains the best model.

Now we use an interaction plot to check whether there is true interaction between Absentee and Participation rate. This figure can be seen in the appendix (fig. 5). Since the lines do not cross, there is not true interaction in the model. This means that absentee, participation rate, or both, is significant, which caused Model 1 to be significant against other models. We build Model 9, 10, and 11, testing

Model 1 without the interaction term but with absentee, participation, and both. Using an ANOVA, we compare these models and distinguish Model 11 as the best, meaning absentee is significant.

We now confirm the assumption that the mean of the residuals is very close to 0 for Model 11.

INTERPRETATION

Our model has been built and tested against other models. We can confidently determine that the best model to predict SAT scores of a district is the following.

Model: *percent of students who meet SAT benchmark in district i in county j =-35.2928 + .3742(Absentee percentage in district i) + 0.8959(percent students who pass Reading CAPT in i) + 0.588(percent students who pass Math CAPT in i) - 0.6896(percent students who pass Writing CAPT in i) + .3423(percent students who pass Science CAPT in i) + (1|county j)*

The intercept value, -35.2928, does not reveal much about the dependent variable, as it is unlikely for all variables to be 0, however we can see that other factors will add percentages, as the SAT benchmark cannot be negative.

Contrary to common sense, the absentee percentage is actually a positive effect toward the SAT benchmark score. As illustrated by Fig. 1, this does not mean that the absentee percentage has a positive correlation with the SAT benchmark score. The interpretation here is that the majority of the effect of absentee is actually taken up by other variables in the model, most likely the random effect of county, illustrated in Fig. 6.

The Reading, Math, and Science CAPT scores each increase the model by .8959, .588, and .3423 respectively. This intuitively makes sense, especially since Reading and Math match the subjects of the SATs and thus they increase the SAT benchmark score the most.

Interestingly, for each additional percent of the student body that passes the Writing portion of the CAPT, there is actually a decrease of .6896 percent of students in that district that pass the SAT benchmark. This would be an interesting topic to investigate. Perhaps the writing CAPT tests different writing skills than the SAT, and therefore trains students poorly for the SAT. Or, perhaps students gifted in writing overthink or run out of time on the SAT. This causation itself is unclear.

The additional random effect in this model, the given county where the districts lies, controls for outside factors that may occur from geography. Different counties may encounter different weather, different accessibility to on-site learning (museums, parks, etc), or simply a unifying factor of the county. We can also now factor in the differences of variable effects within the counties (i.e., perhaps a high absentee percentage would be more significant in Hartford than in Litchfield).

NOTES

- The document containing information about SAT scores was divided by district and then by school. Since the other data was only determined by district, collapsing the SAT data with a weighted average of the scores (weighted by population) helped achieve a smooth data table.

- Our random effect *county* has unequal variances within the groups (Fig. 4). This makes our model lose power, however, although the variances are different, the means are fairly similar and there are seldom outliers.

Appendix

Fig. 1

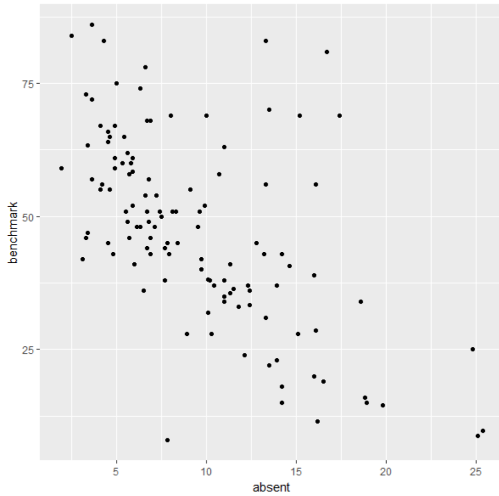


Fig. 2

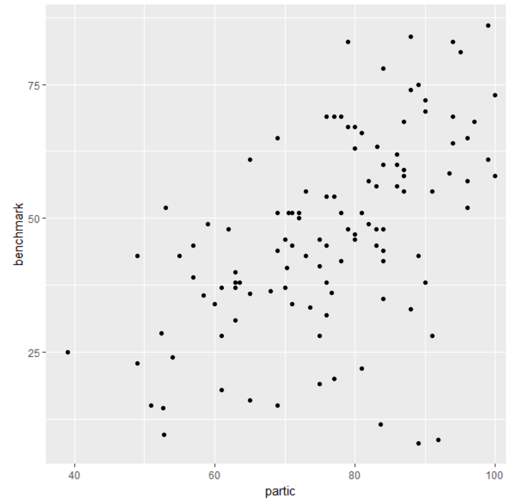


Figure 3

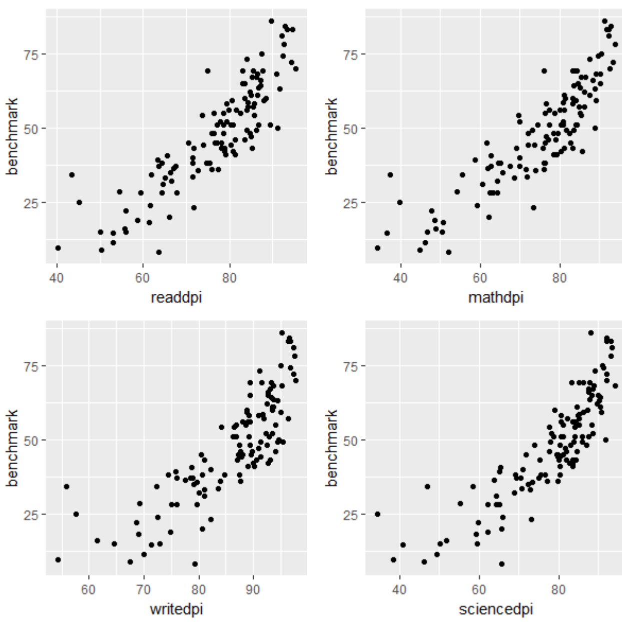
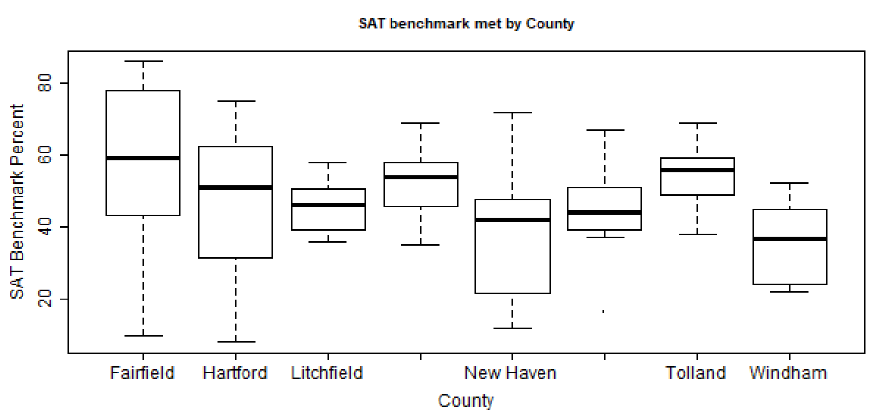


Fig. 4



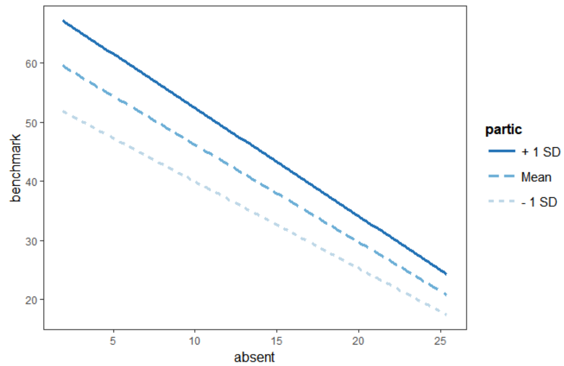
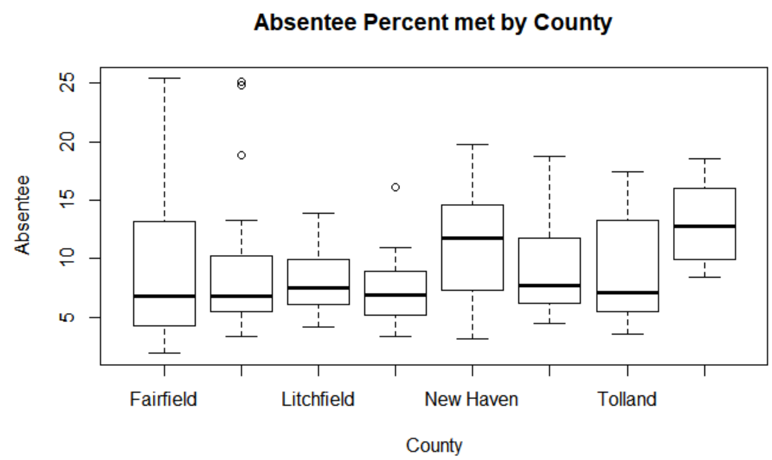

benchmark
absent
partic
+ 1 SD
Mean
- 1 SD

Fig. 5

Fig. 6


Absentee Percent met by County
Absentee
County
Fairfield
Litchfield
New Haven
Tolland