Revenue and Education: Does More Money Lead to Greater Student Success?

A Reflection on the Data Analytics Process

Cristian Mojica

December 6, 2020

BUSI 12

**Project Summary**

**Domain Frame**

I am a mathematics tutor at Modesto Junior College in Modesto, California. The tutoring program there is a known contributor to its students’ success, but there are other economic and financial factors outside of our control that also determine adult student success. The same is to be expected for children in K-12 public school districts. In the case of educating children, which is the focus of this paper, the academic and dietary needs of students are recognized, met, and coordinated by teachers, parents, and administration. In a sense, student success is dependent on this coordination. Part of this coordination is the funding that a school receives to help meet students’ academic and dietary needs. For student success to rise, there must be an efficient allocation of the financial wherewithal that obtains the resources that students will need to perform well in school.

**Problem Statement/Hypothesis**

This problem of financial resource allocation needs to be solved if a district’s administration hopes to efficiently allocate its financial resources to its schools and thereby set them up for student success. It seems to me that with more funding, a school district will have more financial wherewithal to provide resources to teachers to set up their classes for success. Therefore, I expect to find that with more revenue, academic performance will go up. I also expect to find a greater effect of revenue due to the Child Nutrition Act if nutrition plays a role in student success. (I will study these relationships at the state level, the average of district performances across a given state.) These insights can be used by college admissions boards to determine where to accept students from (if approaching admissions from an equitable perspective), and it can also be used to make a case at the district, state, and federal government levels to provide more funding to schools if funding has significant bearing on academic performance.

**Data**

I found my data (from 2017) on Kaggle, and it comes from The Nation's Report Card and the U.S. Census Bureau. There is a sheet with 14,306 school district records. Columns include state\_code, idcensus, school\_district, nces\_id, yr\_data, t\_fed\_rev, c14, and c25. As most of these columns serve identification purposes, and as I am interested in two types of funding (and not all total revenue available to a district), I will only be looking into state\_code, school\_district, c14, and c25. The state\_code is an identification number for each state (including Washington D.C.), c14 represents Title 1 (No Child Left Behind Act) revenue through the state, and c25 represents Child Nutrition Act federal revenue through the state. A separate sheet has a list explaining the state codes as well as a column for the number of records (i.e., number of districts) per state. Another sheet gives the average state scores on the National Assessment of Educational Progress (NAEP) math exam given to eighth graders in 2017. The NAEP math exam is structured such that the scores range from 0 to 500. For this study, the independent variable of interest is revenue stream, and there are two levels: No Child Left Behind Act (Title 1) funding and Child Nutrition Act funding. The dependent variable of interest is the state average NAEP math score.

**Key Stakeholders**

Stakeholders of low power and low interest include students, less-involved parents, secretaries, and clerks. Students may not even be aware that their test scores have been analyzed, nor parents realize that their children’s academic performance is being evaluated in a standardized fashion. Stakeholders of low power and high interest are members of parent-teacher associations (PTAs), taxpayers, and teachers; parents and teachers need to work together in order to keep certain school operations running, and teachers educate students while subject to their principal and superintendent. Other people in this category may include tutors, the financial department of a school district, and dieticians (since the Child Nutrition Act deals with providing access to healthy meals). Stakeholders of high power and low interest include policymakers and legislators, college admissions boards, politicians, and state and federal departments of education. Stakeholders of high power and high interest include superintendents, principals and vice principals, and other district officials, as they make important decisions for the district and are involved in daily management, operations, training, and education.

**Findings**

I found that with an increase in either type of funding, scores also increased, although with weak correlation. Statistical testing for linearity between funding type and average scores, however, confirms that there is a linear relationship (i.e., positive correlation) between both kinds of funding and scores, suggesting that as funding goes up, so do average scores. There was a stronger relationship between Title 1 finding and scores, so stakeholders will be informed that there needs to be a higher priority on the efficient allocation of Title 1 funds for each district.

**Reflection: The Process**

**Select a domain area of research**

I knew that I would want to research education. I initially considered studying the effects of tutoring on the success of community college students in mathematics courses, but realized that data for such a project would be highly specific and most likely unavailable unless I created my own survey, which would not allow time for proper data collection and analysis. I thought then to look at a K-12 district’s financial wherewithal that can be allocated to schools for the sake of student success. From there, I had to operationalize financial wherewithal and student success.

**Formulate a question or hypothesis**

From having observed the positive effects of present, able, and willing instructors on their students, and from hearing of the consequences of lack of resources on students, it made sense to ask “Does more money lead to greater student success?” It also made sense to hypothesize that if a school has more money, then this money can be used to obtain resources in such a way that students find themselves in a better place academically. The only major update since the previous stage was when I decided to study math performance for a specific grade level.

**Frame the question or hypothesis according to your domain**

I tied the funding types and math performance together in a hypothesis regarding students of a specific grade level in the United States of America. In keeping with the operational definitions of the variables, my hypothesis reads that if revenue increases, then math scores will increase.

**Obtain data for your project**

I looked for data on various sites, although Kaggle caught my attention the most. However, it did not have much specialized data except for eighth grade NAEP scores. I began experiencing problems with this data set when I saw that the only math exam scores were state averages, and I tried to find district-level average scores to have more low-level data to analyze. I could not find scores for all districts in all states, and I was in a panic for some time about how to proceed. I was also worried when I found that if I wanted to demonstrate the ability to scrub data, I would have to delete district-level data, but that would not work because district-level average scores would not be removed with them since they were unavailable to begin with. When I accepted the fact that I would have to devise a method to adjust state average scores, I kept this data and proceeded to scrub it.

**Scrub the data**

I had no district-level data apart from financial figures, so I had to figure out a way to calculate adjusted state average scores. I moved my data to Google Sheets and created a pivot table where the rows were the states (and Washington D.C.) and the columns were the average No Child Left Behind funding figures and the average Child Nutrition funding figures. I filtered the data so that only the schools without combined Title 1 and Child Nutrition funding being $0 was considered, so the pivot table did not consider 1,095 districts out of all 14,306. This effectively hides just under 8% of the data. The considered data will still have districts that lack none or exactly one of these sources of funding, but not both. Although I did get the average funding amounts I needed from this table, I was still figuring out a way to come up with a way to adjust average test scores since I lack that data at the district level. After experimenting with sample mean formulas and quartile calculations, I devised a method to adjust average scores. To calculate the adjusted state average scores after deleting districts, apply this formula for each state as necessary:

(282 was the national median NAEP math score for eighth grade in 2017.) I decided to remove the state of Vermont from the dataset because the new adjustment took out nearly 80% of the school’s districts and made the average test score negative, which is not appropriate for interpretation. This was the only state where the supermajority (beyond 67%) of school districts did not have either source of funding considered in this study.

**Explore the data**

I studied the distribution of the test scores and saw that they were left skewed (see Figure 1). This was not surprising considering how the NAEP scores are assigned. I plotted Title 1 funding against scores, and Child Nutrition Act funding against scores. I realized that I needed to remove states that were outliers. Going back to five-number summaries of both funding types, I created lower and upper fences to determine which states were outliers. The five states that were outliers in both types of funding were Washington D.C., Florida, Hawaii, Maryland, and Nevada. Upon removing those states by updating the pivot table with a filter that did not consider those states, I visualized the data again (see Figures 2 and 3) and checked for relationships between funding and state average test scores. I studied 45 states. I tested for linearity between Title 1 funding and average scores at the level of significance α = 0.05. The hypothesis test reads: H0: ρ = 0, H1: ρ ≠ 0. I then tested for linearity between Child Nutrition Act funding and state average test scores at the level of significance α = 0.05. The hypothesis test reads: H0: ρ = 0, H1: ρ ≠ 0. The results inform the findings to be presented.

**Present your findings**

This stage was rather underwhelming considering that most of my time and energy went into scrubbing the data and gathering descriptive statistics on it. It will only be challenging to determine how to present the findings to stakeholders in an accessible yet succinct manner. I found that despite a weak correlation (r=0.33), there is a linear relationship between Title 1 funding and state average test scores (t(43) = 2.29, p = 0.027). Despite another weak correlation (r=0.31), there is also a linear relationship between Child Nutrition Act funding and state average test scores (t(43) = 2.14, p = 0.038). There is more evidence in favor of the strength of the linear relationship between Title 1 funding and test scores than of the strength of the linear relationship between Child Nutrition Act funding and test scores, which goes against my initial expectations for the effect of funding type on test scores. Regardless, as funding of either type goes up, scores are expected to increase as well. As there are linear relationships present (albeit weak), I am glad to have provided conscionable correlation figures that do not account for outliers.

**Reflection: Key Learnings**

First, data does not always come neatly packaged or exactly match one’s desired format. In that case, it may help to re-operationalize the variables to be studied, or to change the guiding question slightly. Second, it helps to have clear guiding questions. The analytical process is already messy, especially with the cyclical nature of exploratory data analysis, but it can be made more manageable with direction. Third, statistical procedures can be improperly applied and abused. This occurred to me as I came up with a way to adjust state average test scores. I know how to manipulate the formulas for average and mathematical expectation, but I wonder if I could have found a better way to adjust average scores. It is my hope that I did not abuse the average or its properties. Fourth, it helps to study mathematical techniques and some data cleaning techniques. If data analytics is primarily scrubbing and preparation for analysis, then it may be worthwhile to find the best ways to do that depending on the nature of the data. Finally, in accepting the fact that I was denied granularity in this project, I realized that a data analytics project is a balancing act, or a storybook being written continuously, where its ever-present morals are to beware of distorted motives in analysis and to not forget the individuality of the experimental units considered in analysis. It is my hope to bring these points into future work as a guide and for future colleagues to be cognizant of these points in their own work.

**Conclusion**

The OSEMN framework was adhered to in this study. I obtained state data for eighth grade students and their performance on a standardized math exam, as well as district data on Title 1 funding and Child Nutrition Act funding. I scrubbed the data primarily to get rid of districts with lacking or outlying levels of revenue. I explored the data and found linear relationships between funding and test scores. I found that the scores as plotted against funding fit to a linear model and predictions can be made reasonably using that model; this is the only part of the framework not pursued in depth. I interpreted the findings of my study and found that as funding increases, so will test scores.

This was not a perfect project. There is low generalizability of the results since I cannot apply them across all grades of a K-12 district. Also, academic success is more than just a standardized math test score, so my operationalization of success was rather limited. From the standpoint data analysis, however, this was a *tour de force* in introductory data analysis, and I find myself to be a better analyst because of it.

**Appendix**

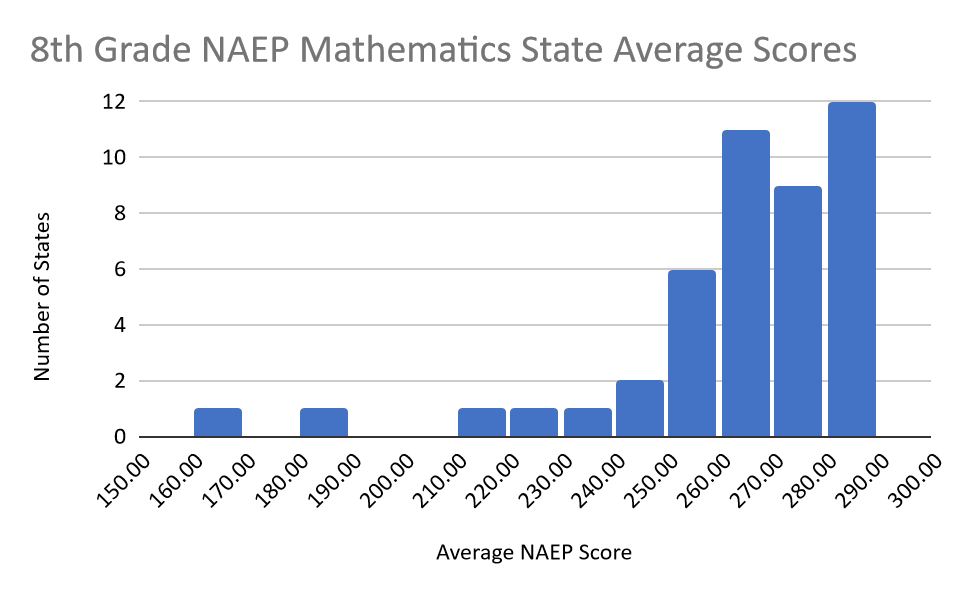


Figure 1: 8th Grade NAEP Mathematics State Average Scores

Figure 2: Title 1 Funding and Average State Scores

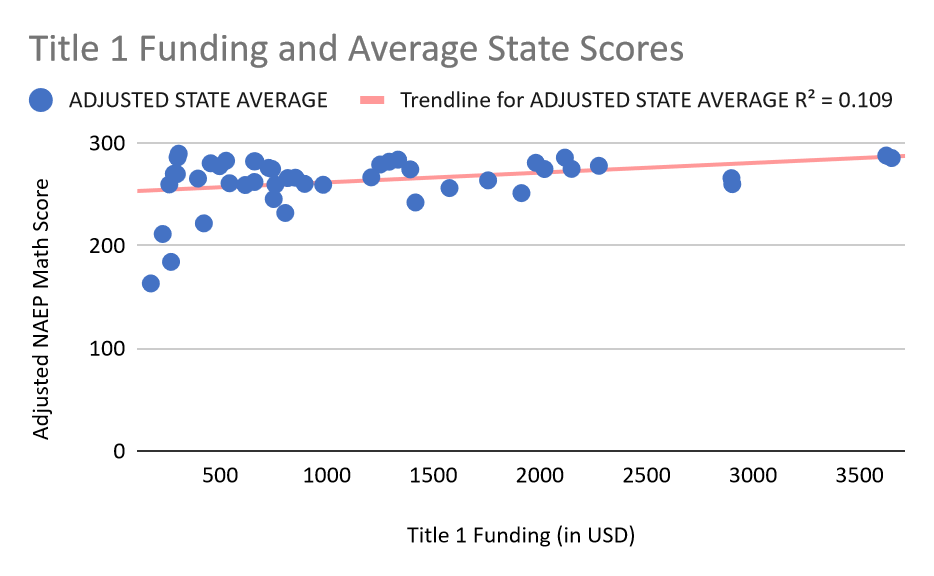


Figure 3: Child Nutrition Act Funding and Average State Scores

