# The relationship between funding and test scores in the United States

By Debayan Sen

Email: dsen@falcon.bentley.edu

## Executive Summary:

Student success in the United States can be influenced by many factors, but one important factor is the amount of funding that schools receive. This report will focus on funding levels for primary and secondary schools in the United States. Do increased funding levels have a positive effect on student's test scores? Student success will be measured by standardized test scores, which research shows is a predictor of future student outcomes. This research is important because a good education can enable better opportunities for students once they go into the workforce, and better education creates a well-informed public that can bring about positive change.

## Part 1: Introduction and Literature Review

Education is what allows students to gain the skills they need to gain the skills necessary to eventually contribute to the economy and society. There are different measures that are used to measure student success ranging from standardized test scores to overall graduation rates and employment after high school. For my final project, I will investigate educational data and see what the effect of educational funding is on test scores in K-12 education systems. Test scores are an accurate measure of success for students, and there the literature supports more funding leads to better prospects for students.

More funding does lead to better schooling for students. Continued spending, especially in low-income areas has been documented to reduce education attainment gaps between low-income and poor families (Research Shows That When It Comes to Student Achievement, Money Matters 2018). Studies done to see the effect of funding have found increased test scores in both reading and math (Kreisman & Steinberg, 2019). Advancements in rural areas have been seen after a massive influx of funding (Burnette, 2019). Funding is what allows schools to have the resources to support their students. Especially during COVID-19, there is an increased need for technology such as laptops and good internet connection. A focus on funding is an important factor in determining student success.

Test scores are a good indicator of success for students. There is a difficulty in determining how to measure student outcome given there are many ways to measure success. Regression models do use academic test scores to measure results (Carhart 2016; Henry et al 2010). For standardized test scores, they can be a predictor of a student's first-year success in college (Camara & Croft, 2020). In schools, there are robust studies that indicate a clear link between higher test scores and later life prospects. These studies have found that for increased test scores there is also an increase in salary later in life (Camara & Croft, 2020). Test scores are also strongly associated with engaging in less risky behavior later in life such as smoking and participating in illegal activities. Thus, test scores are a useful measure in determining a student's success.

Funding is directly tied to improving test scores. Regression models that have looked at state and grade level data have found that there is a significant relationship between expenditure and student achievement (Carhart 2016; Jackson et al 2015). Furthermore, more expenditure can lead to better performance, specifically in disadvantaged districts (Henry et al 2010). There is past precedent for higher funding leading to higher student achievement. The data I am looking at will feature all 50 states, with data from each state. There exist interstate disparities between each state that influence educational inequality in the United States. Some states provide higher levels of funding than others, which can influence test scores and other outcomes. Looking into these interstate disparities using the data will provide insight into the differences that exist between states, instead of focusing on a singular state.

The data analysis I will conduct will add to this literature by looking at the relationship between funding and student test scores. Regression models will be built to determine how to best predict student scores based on funding. These regression models will be built primarily using the ordinary least squares method. Modeling student test scores based on funding is useful because it can be used to see the differences that exist between different funding allocations. Data analyses like this can be used to see where more funding is needed and the impact it will have on student performance. The literature often looks at legislation in tandem with regression data to see how laws can be changed to better influence student outcomes (Carhart, 2016; Chingos & Blagg, 2017). There is debate about whether funding leads to higher student outcomes. Sometimes funding can be misused, and not properly allocated to have any meaningful effect (Chingos & Blagg, 2017). The analysis can provide insight on this because it looks at multiple levels of funding, thus we can see which levels of funding (such as local vs state vs federal) are more important in determining student outcome.

Part 2: Data

The data is retrieved from Kaggle. There are three CSV files, two of which contain the finances of elementary and high schools from the United States Census Bureau. The third CSV file is a summary of data from the National Assessment of Educational Progress. The three CSV files are described below:

districts.csv A comma-separated spreadsheet containing revenues and expenditures for all U.S. school districts, 1992-2016.

states.csv A comma-separated spreadsheet containing state summaries of revenues and expenditures, organized by year.

naep.csv A comma-separated spreadsheet containing state performance on mathematics and reading tests, for 4th and 8th grade on a selection of years.

There is a common identifier of "state" and "year" between the naep.csv and states.csv files. In my analysis, I merge these data files, and look at the following variables:

```
> colnames(df_merged)
 [1] "STATE"                        "YEAR"                         "ENROLL"
 [4] "TOTAL_REVENUE"                "FEDERAL_REVENUE"              "STATE_REVENUE"
 [7] "LOCAL_REVENUE"                "TOTAL_EXPENDITURE"            "INSTRUCTION_EXPENDITURE"
[10] "SUPPORT_SERVICES_EXPENDITURE" "OTHER_EXPENDITURE"            "CAPITAL_OUTLAY_EXPENDITURE"
[13] "AVG_SCORE"                    "TEST_SUBJECT"                 "TEST_YEAR"
```

The population of interest for this data set are students in fourth and eighth grade and we are looking at their average reading scores. The data is from 1992-2016 and has both reading and mathematics standardized scores. We will be able to use this data to see the relationship between funding and test scores in these grades and make conclusions about the relevance of funding.

# Part 3: Explanatory data analysis

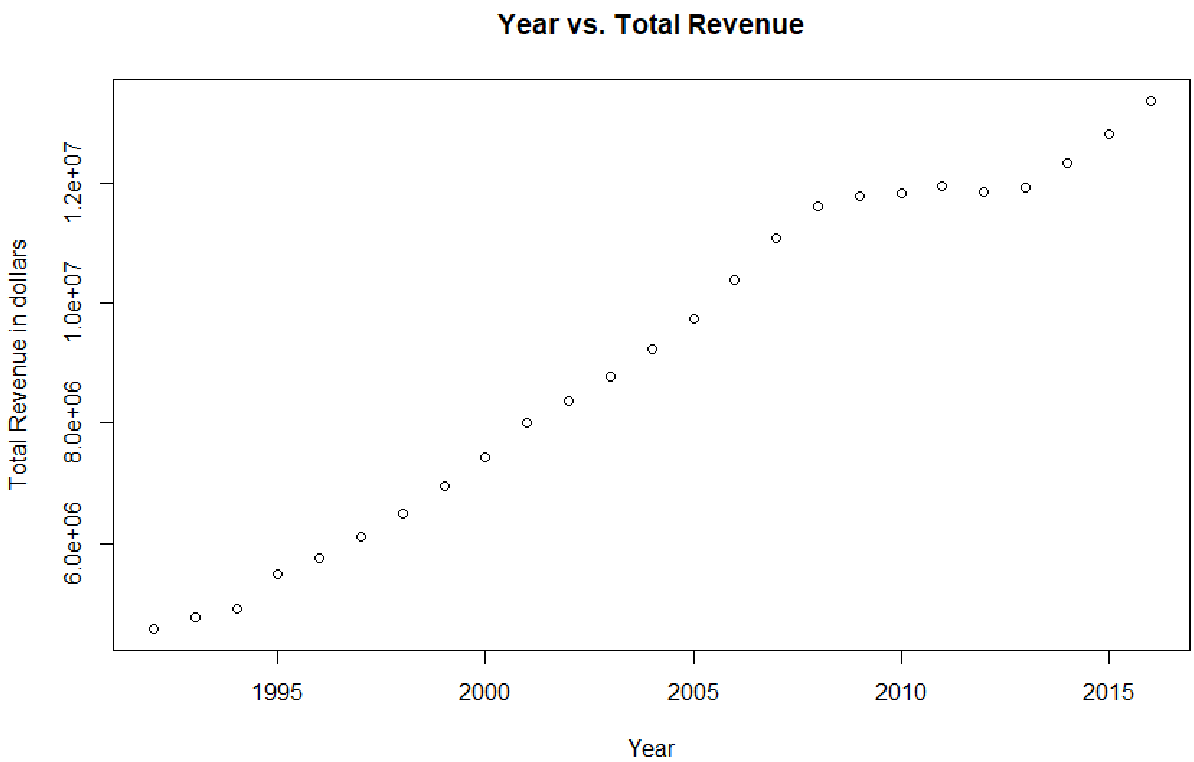
## Initial findings

First looking at the states.csv file, we are looking at the following variables:

```
                        name
                        YEAR
                      ENROLL
               TOTAL_REVENUE
             FEDERAL_REVENUE
               STATE_REVENUE
               LOCAL_REVENUE
           TOTAL_EXPENDITURE
     INSTRUCTION_EXPENDITURE
SUPPORT_SERVICES_EXPENDITURE
           OTHER_EXPENDITURE
  CAPITAL_OUTLAY_EXPENDITURE
```

A few questions to look at:

1. Since this data is from 1992-2016, how is the total amount of revenue changing per year?




We can see that the total revenue is increasing year. This means we need to be careful making over general claims for data that contains data over a long period of time as the impact of a certain amount now is not the same as it was 10 years ago.

2. Which states have the most/least revenue in 2016?

The states with the least amount of revenue in 2016 are the District of Columbia, South Dakota, North Dakota, Montana, Delaware, and Wyoming.

The states with the highest amount of revenue in 2016 are New Jersey, Pennsylvania, Illinois, Texas, New York, and California.

In our analysis, we can look at a specific state and see the impact of increased funding on test scores.

Second, with the naep.csv file, the four columns are YEAR, STATE, AVG_SCORE, TEST_SUBJECT, and TEST_YEAR. We will merge the naep.csv file and the states.csv file on YEAR and STATE column. This results in a data frame with 2040 observations and 15 variables. We will use this merged dataframe to create our model.

Looking at the distribution of the variables:


Initially looking at the distribution of the numeric variables, it seems appropriate to use a log transformation for variables such as local revenue because the data looks right skewed.

## Model Creation

This section will look at models created from the data.

The first model looks at data only from the year 2015. This is because we know that the revenue is increasing every year, so it does not make sense to look at revenues from every year. Using a main effects model or a model with many interactions would be a poor choice because it would result in a model with a high variance inflation factor score, which indicates high collinearity among the variables.

```
Call:
lm(formula = AVG_SCORE ~ log(LOCAL_REVENUE) + log(CAPITAL_OUTLAY_EXPENDITURE) +
    (TEST_YEAR * TEST_SUBJECT) - TEST_SUBJECT, data = df_merged_2015)

Residuals:
     Min       1Q   Median       3Q      Max
-17.6957  -3.7778   0.1561   4.0617  13.9013

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                     236.3489     5.0343  46.947  < 2e-16 ***
log(LOCAL_REVENUE)                2.4369     0.5393   4.518 1.07e-05 ***
log(CAPITAL_OUTLAY_EXPENDITURE)  -2.4354     0.6231  -3.909 0.000127 ***
TEST_YEAR8                       41.5253     1.1239  36.949  < 2e-16 ***
TEST_YEAR4:TEST_SUBJECTReading  -18.1107     1.1239 -16.115  < 2e-16 ***
TEST_YEAR8:TEST_SUBJECTReading  -16.9054     1.1239 -15.042  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.675 on 198 degrees of freedom
Multiple R-squared:  0.9437,    Adjusted R-squared:  0.9422
F-statistic: 663.4 on 5 and 198 DF,  p-value: < 2.2e-16
```

This model was chosen with interaction in mind, as the effect of the test subject on average score depends on what grade a student is in. We see that all the variables are significant in explaining the response variable which is average scores. This model looks at local revenue, capital outlay expenditure, test year, and the interaction between test year and test subject. Capital outlay expenditure refers to the amount of money that is spent on fixed assets such as equipment or buildings. This model is important because the local government has a lot influence over schooling, thus it is meaningful to look at local revenue as an explanatory variable to see how that influences funding levels.

In this model, we see that the intercept is 236.3489. This intercept is not meaningful because it refers to when there is essentially zero local revenue or capital outlay expenditure, in which case a school most likely would not be able to operate. Looking at the categorical variables, we see that making the test year eight significantly increases the test score by about 41.52. Furthermore, we see that reading scores decrease for both $4^{th}$ and $8^{th}$ graders. This means that students in the $4^{th}$ and $8^{th}$ grade in 2015 are generally performing better in math than in reading.

Is this model overfitting the data? We can use an 80-20 split to evaluate this.

For our training data we get:

```
Call:
lm(formula = AVG_SCORE ~ log(LOCAL_REVENUE) + log(CAPITAL_OUTLAY_EXPENDITURE) +
    (TEST_YEAR * TEST_SUBJECT) - TEST_SUBJECT, data = train_2015)

Residuals:
     Min       1Q   Median       3Q      Max 
-18.0066  -3.8632   0.0701   4.3361  13.4542 

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)    
(Intercept)                     234.5390     6.0282  38.907  < 2e-16 ***
log(LOCAL_REVENUE)                2.5110     0.6462   3.886 0.000150 ***
log(CAPITAL_OUTLAY_EXPENDITURE)  -2.3878     0.7103  -3.362 0.000972 ***
TEST_YEAR8                       41.9955     1.3518  31.066  < 2e-16 ***
TEST_YEAR4:TEST_SUBJECTReading  -18.2759     1.2757 -14.326  < 2e-16 ***
TEST_YEAR8:TEST_SUBJECTReading  -17.6209     1.3533 -13.020  < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 5.908 on 157 degrees of freedom
Multiple R-squared:  0.9405,	Adjusted R-squared:  0.9386 
F-statistic: 496.2 on 5 and 157 DF,  p-value: < 2.2e-16
```

For the testing data, we get a root mean squared error (RSME) and R squared value of:

```
         RMSE        R2
[1,] 4.731301 0.9744249
```

Since our R squared value is high and RSME value is low, we are not at risk for over-fitting the data.

Now, with this data, can we see an increase in test scores after increasing funding? We want to see the effect of local funding on scores of students.

First, what are the average levels of funding and test scores in 2015? If we look at the average math score for eighth graders in 2015, this is 281.75. The average local revenue is 5837854. What if we increase this by a million dollars? If we increase the score by a million dollars, on average, the mean score for eighth grade mathematics scores will become 282.8273. While this point increase is not too large, we can see that there is a positive influence that local revenue has on test scores.

## Looking at a specific state

One state that performs poorly in both reading and math is New Mexico. They are in the top five lowest performing states for average scores in both reading and math in 2015. Thus, how would an increase in local funding affect test scores in New Mexico?

First, we create a model just for New Mexico:

```
Call:
lm(formula = AVG_SCORE ~ log(LOCAL_REVENUE) + log(CAPITAL_OUTLAY_EXPENDITURE) +
    (TEST_YEAR * TEST_SUBJECT) - TEST_SUBJECT, data = df_newMexico)

Residuals:
    Min      1Q  Median      3Q     Max
-8.7120 -1.9055 -0.0941  2.2654  8.2068

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                     111.7731    24.3947   4.582 7.09e-05 ***
log(LOCAL_REVENUE)                8.4970     4.0495   2.098   0.0441 *
log(CAPITAL_OUTLAY_EXPENDITURE)   0.2396     3.3481   0.072   0.9434
TEST_YEAR8                       41.6199     1.7743  23.458  < 2e-16 ***
TEST_YEAR4:TEST_SUBJECTReading  -18.0603     1.7336 -10.418 1.20e-11 ***
TEST_YEAR8:TEST_SUBJECTReading  -13.5116     1.7786  -7.597 1.45e-08 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 3.764 on 31 degrees of freedom
Multiple R-squared:  0.9798,    Adjusted R-squared:  0.9765
F-statistic: 300.4 on 5 and 31 DF,  p-value: < 2.2e-16
```

I chose the same variables as the previous model for consistency; it will be best to see how local revenue influences scores in New Mexico because we know from our previous model that it is significantly associated with average test scores. For this model, we see that log(local revenue), test year, and the interaction terms are all significantly explanatory of the response variable which is the average score. This model also has a high R squared value, however did we overfit the data? We can evaluate this using an 80-20 split.

First, we create the trained model:

```
Call:
lm(formula = AVG_SCORE ~ log(LOCAL_REVENUE) + log(CAPITAL_OUTLAY_EXPENDITURE) +
    (TEST_YEAR * TEST_SUBJECT) - TEST_SUBJECT, data = train_mexico)

Residuals:
    Min      1Q  Median      3Q     Max
-4.7698 -1.5469  0.3316  2.1000  3.8784

Coefficients:
                                Estimate Std. Error t value Pr(>|t|)
(Intercept)                      132.835     21.109   6.293 2.47e-06 ***
log(LOCAL_REVENUE)                 9.534      3.443   2.769   0.0112 *
log(CAPITAL_OUTLAY_EXPENDITURE)   -2.225      2.757  -0.807   0.4283
TEST_YEAR8                        38.499      1.512  25.455  < 2e-16 ***
TEST_YEAR4:TEST_SUBJECTReading   -20.602      1.484 -13.879 2.32e-12 ***
TEST_YEAR8:TEST_SUBJECTReading   -13.271      1.454  -9.125 6.22e-09 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 2.657 on 22 degrees of freedom
  (53 observations deleted due to missingness)
Multiple R-squared:  0.9901,    Adjusted R-squared:  0.9879
F-statistic: 441.9 on 5 and 22 DF,  p-value: < 2.2e-16
```

Then, using this model on the new data that the model has not seen yet we get:

```
          RMSE        R2
[1,] 22.76848 0.9397764
```

We see that this has a relatively high R squared value, meaning we did not overfit the data.

Now, how much do local funding increases in New Mexico increase test scores? The score for 8th graders taking a mathematics exam with the average amount of local revenue and capital outlay expenditure, on average, scores 267.44. If funding were increased by a million dollars, on average, the score for 8th graders taking a mathematics exam would increase to 277.1661. This is a significant bump in score than before, although it would require lots of funding to achieve.

# Part 4: Inference – Model Assumptions

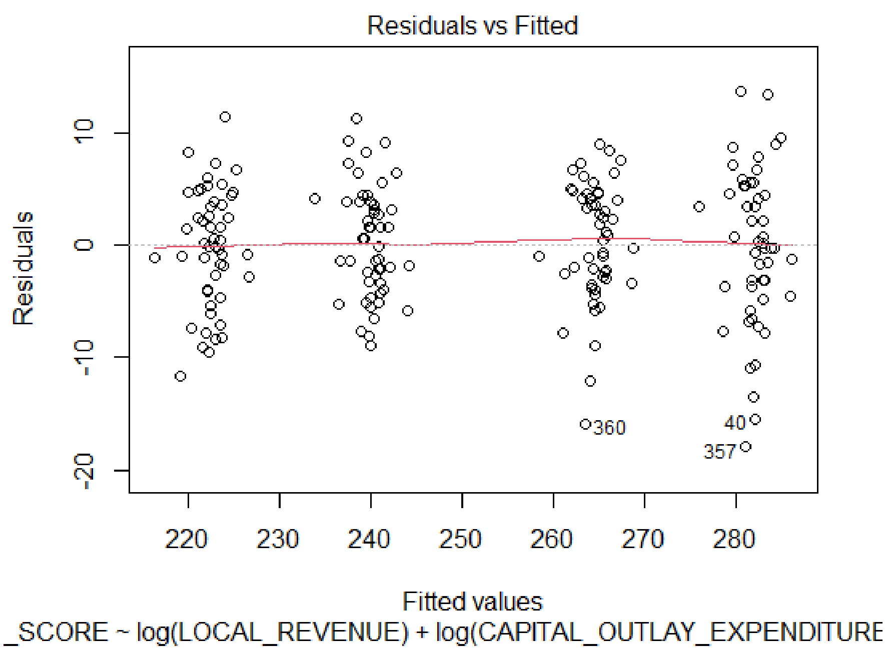
## Conditions for the first model:

1) Independence

The collection of the data and funding levels are independent.

2) Linearity

We can use a residuals vs fitted plot to evaluate this assumption.



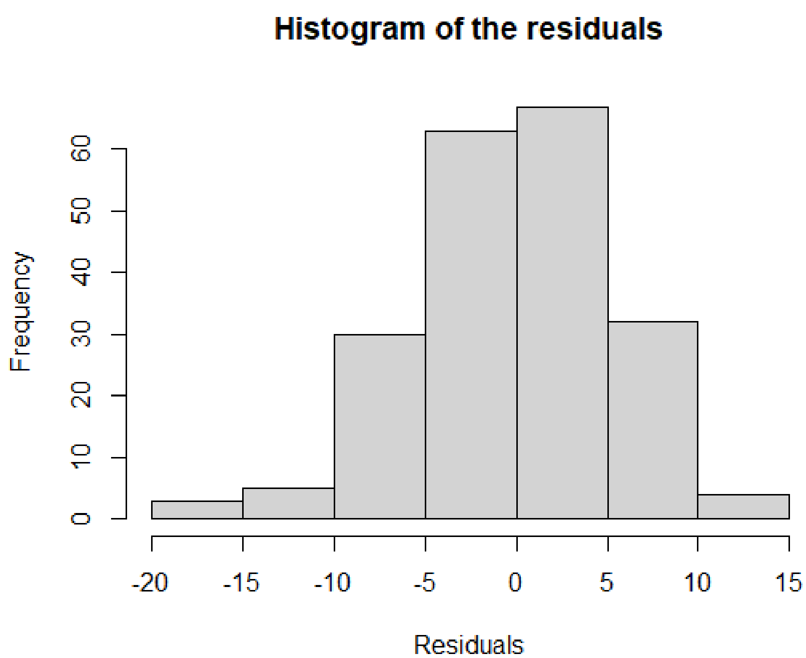

We do not see any overwhelming pattern in this plot.

3) Constant Variance

We can also look at the residuals vs. fitted plot to evaluate this assumption. Since there is no clear fanning out, this assumption is satisfied

4) Normality of residuals

We can look at a histogram of the residuals to evaluate this assumption:




There is no overwhelming skew, therefore this assumption is satisfied.

5) Collinearity

We can look at the variance inflation factor to determine if there is a multi-collinearity issue in our model:

```
log(LOCAL_REVENUE) log(CAPITAL_OUTLAY_EXPENDITURE)                       TEST_YEAR
          2.964392                       2.964392                         1.000000
      TEST_SUBJECT
          1.000000
```

Since these VIF values are less than 5, there is no collinearity issue in our model.

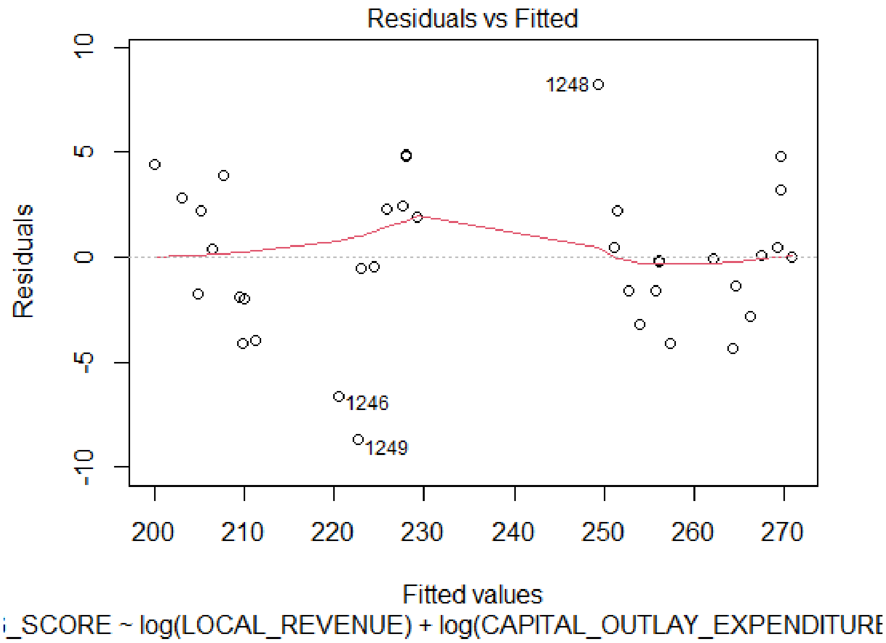
## Conditions for the New Mexico model:

1) Independence

The collection of the data and funding levels are independent.

2) Linearity

Using a residuals vs fitted plot to evaluate this assumption:



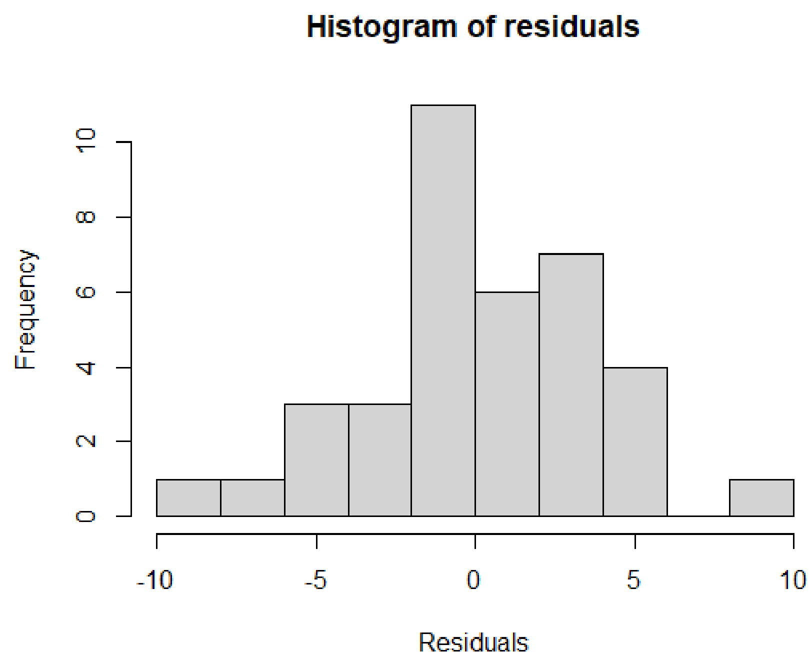

There is no overwhelming pattern, meaning that this assumption is satisfied.

3) Constant Variance

Although there is slight fanning out, it is not overwhelming in nature, thus we can treat this assumption as true.

4) Normality of Residuals

We can look at a histogram of the residuals to evaluate this assumption:




The data does not seem to be very skewed in either direction, meaning that this assumption is satisfied.

5) Collinearity

We can look at the variance inflation factor to determine if there is a multi-collinearity issue in our model:

```
> vif(model_mexico)
                                   GVIF Df GVIF^(1/(2*Df))
log(LOCAL_REVENUE)             4.815331  1        2.194386
log(CAPITAL_OUTLAY_EXPENDITURE) 4.831195  1        2.197998
TEST_YEAR                      2.054054  1        1.433197
TEST_YEAR:TEST_SUBJECT         2.074397  2        1.200115
```

Since these VIF values are less than 5, we can say that there is no multi-collinearity issue in our model.

## Part 5: Conclusion

From our analysis, we can see that increases in local funding do result in increased test scores for eighth and fourth grade standardized scores. The data we looked at comes from 2015 for our first model where we looked at test scores in general, and the second model created looks at a New Mexico, a state that is struggling in terms of their standardized test scores. Although there is an increase in test scores, it is not too significant for the first model, which is an increase of about a point. This suggests that while local revenue does increase scores, it is just one of many factors that plays a factor in a student's performance. In the second model, we see that an increase of a million dollars would increase an eight grader's math score by about 10 points. It is important to take into consideration that the value of money is not the same throughout the states and a million dollars might be a significant amount of money in New Mexico for educational funding. There should be careful consideration in using these models to determine funding levels because they do not account for other factors that a state's funding scheme relies on. Models should take into consideration the impact of multiple levels of funding (federal, state, and local) to see where energy should be focused to increase student scores.

An important variable of interest for future consideration that can be looked at to measure student performance are high school graduation rates. Looking at high school graduation rates across the country would give insight into the potential interstate disparities that exist in education and could be an important factor when considering how funding determines a student's success. Using graduation rates might be better than using test scores to measure student success given that scores from the fourth and eighth grade are not the only determination of how a student performs after they finish education.

## Part 6: References

Works Cited


Burnette, D. (2019, June 4). Student Outcomes: Does More Money Really Matter? EducationWeek. Retrieved November 22, 2021, from https://www.edweek.org/policy-politics/student-outcomes-does-more-money-really-matter/2019/06.

Carhart, A. E. (2016). School Finance Decisions And Academic Performance: An Analysis Of The Impacts Of School Expenditures On Student Performance (thesis).

Camara, W., & Croft, M. (2020, April 9). *Research tells us standardized admissions tests benefit under-represented students*. EdSource. Retrieved November 22, 2021, from https://edsource.org/2020/research-tells-us-standardized-admissions-tests-benefit-under-represented-students/628611.

Chingos, M. M., & Blagg, K. (2017, May). Do Poor Kids Get Their Fair Share of School Funding? Urban Institute. Retrieved December 16, 2021, from https://www.urban.org/sites/default/files/publication/90586/school_funding_brief.pdf

Henry, G. T., Fortner, C. K., & Thompson, C. L. (2010). <strong>Targeted Funding for Educationally Disadvantaged Students: A Regression Discontinuity Estimate of the Impact on High School Student Achievement</strong>. Educational Evaluation and Policy Analysis, 32(2), 183–204. http://www.jstor.org/stable/40732417

Jackson, C. K., Johnson, R. C., & Persico, C. (2015). The Effects of School Spending on Educational and Economic Outcomes: Evidence from School Finance Reforms. doi:10.3386/w20847

Kreisman, D., & Steinberg, M. P. (2019, February). *The Effect of Increased Funding on Student Achievement: Evidence From Texas's Small District Adjustment*. Retrieved November 22, 2021, from http://www.dkreisman.com/research/Kreisman_Texas.pdf.

*Research Shows That When it Comes to Student Achievement, Money Matters*. Learning Policy Institute. (2018, July 17). Retrieved November 22, 2021, from https://learningpolicyinstitute.org/press-release/research-shows-student-achievement-money-matters.

# Appendix

## Appendix 1 – Highest and lowest performing states for average score in 2015

In R, these are the highest and lowest performing states for average scores in 2015:

Highest in math:

| STATE <chr> |
|---|
| Massachusetts |
| New Hampshire |
| Minnesota |
| New Jersey |
| Vermont |
| Wisconsin |

Highest in reading:

| STATE <chr> |
|---|
| New Hampshire |
| Massachusetts |
| Vermont |
| Connecticut |
| New Jersey |
| Minnesota |

Lowest in math:

| STATE <chr> |
|---|
| Mississippi |
| Nevada |
| California |
| District of Columbia |
| New Mexico |
| Alabama |

Lowest in reading:

| STATE <chr> |
|---|
| Nevada |
| Mississippi |
| Alaska |
| California |
| District of Columbia |
| New Mexico |

## Appendix 2 – The districts CSV File

The districts.csv file is not used due to time limitations and purpose of this report, here are some of the variables found in that datasheet:

```
STATE,ENROLL,NAME,YRDATA,TOTALREV,TFEDREV,TSTREV,TLOCREV,TOTALEXP,TCURINST,TCURSSVC,TCURONON,TCAPOUT
```