Brandon Coates
Final Project
MA5701 R02

Student Performance Trends Versus School Funding and Typology Classification in Ohio

4/23/2021

Brandon Coates

1.0: Introduction

In the state of Ohio, school districts are funded using a combination of state funds, property taxes, income taxes, and federal funds. The funding relies on factors such as economic environment of the area that the school district covers, the number of students enrolled in the school district, and the property wealth of the district. Due to this funding method, not all schools are provided equal funding, which may impact various factors, but more importantly the academic achievements of the students in the lesser funded school districts. This statistical analysis will seek to examine the relationship between school district funding and the state of Ohio's standardized test scores of the school districts. By examining this relationship, inferences could be made about the funding method in place for Ohio school districts and if reform is needed in order to increase the academic success of students across the state of Ohio.

This observational study will randomly sample 140 school districts with 35 districts from each of the state designated typology categories (rural, small town, suburban, and urban). In order to observe the relation to the funding and student performance in the selected school districts, the total operational expenditure per pupil in each school district will be examined as well as the average Ohio State Test (OST) scores for the school district will be observed. The Ohio State Test is the standardized test that the state of Ohio uses to gauge the performance of each school district, it is broken up into four subjects which are English language arts, mathematics, science, and social studies. This study will examine the cumulate average test score in relation to the school district that is being examined. The study will also examine how the typology of a school district is related to the funding of the school as well as the student performance on the OST. The data for this analysis is provided by the state of Ohio, especially the Ohio Department of Education, who maintain and update data frames of the variables of interest.

Though this study will only look at two of the several socio-economic factors that may affect a student's academic performance, this study may expose how disproportionate school district funding or the typology of a school district can alter the performance of a student. The a priori hypothesis in this study is that low school district funding leads to lower average test scores of the students within the school district when compared to a better funded school. The performance of a student in high school has a major impact on the student's future, such aspects of their lives such as average income, college selection, and job aspects all rely on good performance in high school. With early schooling having such a large impact on a student's life, the funding of the school must be adequate to allow for proper preparation and education of the student. Therefore, this study will attempt to link these factors together in terms of school funding and average standardized test scores.

2.0: Data Methods

In order to conduct this observatory analysis on student performance and school funding in Ohio, data was collected from several databases maintained by the Ohio Department of Education (http://education.ohio.gov/). For the analysis, the sampling unit consists of a school district in Ohio which will be examined by the variables of total expenditure per pupil ($USD)[1], average Ohio State Test result, the average Ohio State Test result per subject on test (mathematics, science, social studies, and English language arts)[2], and the typology assigned to the school district by the State of Ohio (rural, small town, suburban, or urban)[3]. The data for all of the listed variables were found to be publicly available only via the Ohio Department of Education in the form of Microsoft Excel files, which allowed for the combination of the data set to be done within Microsoft Excel, which was then exported into a .csv file, which was then subsequently used in the R programming environment for statistical analysis. The specific websites where the data can be found can be found in the footnotes on this page.

The typology of the school district depended on several demographical and geographical variables that the State of Ohio considers before assigning its classification. The major variables that lead to the classification of the school district are the student poverty rate and the student population as well as the location of the school district. These variables are determined by economic and population statistics from the census are used. By combining these variables, the State of Ohio was able to classify each school district as rural, small town, suburban, or urban and allows them to examine similar districts within their studies.

The average total expenditure per pupil for each school district that is used within this study was determined by the State of Ohio. This expenditure includes the administration costs, building operation costs, instructional costs, pupil support costs, and staff support costs. This encompasses all of the expenses the school incurs for typical operation and is then ratioed to form a per pupil cost. This cost reflects the funding that the school is provided by the state and community taxes, so it provides a per student quantity which lends itself to the main concept of this study.

The data for every Ohio school district for the typology and total expenditure per pupil was straight forward as these variables were explicitly provided in the Excel files provided by the Ohio Department of Education, however, a method needed to be developed to assess the average Ohio State Test scores for each school district. The Ohio Department of Education only provides the percentage of student population of the school district that scored enough points on the test to be considered proficient, advanced, accelerated, basic, and limited. Given the data set available, it was decided that the percentage of students that scored enough points on the test to be considered proficient or above would be used to determine the academic success of the school district for each subject tested. So, in order to make an assessment on the test results, it was decided that each subject would be broken into the applicable tests and their percentage of

[1] http://education.ohio.gov/Topics/Finance-and-Funding/School-Payment-Reports/District-Profile-Reports/FY2018-District-Profile-Report
[2] http://education.ohio.gov/Topics/Testing/Testing-Results/Results-for-Ohios-State-Tests
[3] http://education.ohio.gov/Topics/Data/Report-Card-Resources/Report-Card-Data-Forms-and-Information/Typology-of-Ohio-School-Districts

students that scored to be proficient and above would be converted to a score out of 100 and this new calculated score would be the score used for this study. It should be noted as well that the tests for the subjects were not straight forward, the test results were given for each subject for several grades as well as concepts. It was decided that the high school level subjects would be considered, so for the mathematics scores Algebra 1 and Geometry were considered, for English language arts 8^{th} and 9^{th} grade scores were considered, science only consisted of high school physical science, and social studies scores consisted of both US Government and History. By considering all of the above results, composite scores for each subject were made for each Ohio school district and these scores were added together to obtain an average total Ohio State Test score for the respective school district.

Once these data points were collected, the school districts were sorted by their typology and then sorted alphabetically within the typological groups. A random number generator was then used to select 35 random Ohio school districts from each typology group, leading to a total sample size of 140 school districts. The decision to select schools depending on their typological category was rooted in the a priori belief that the funding of school districts in Ohio depended on how the school was classified in terms of their typology, so urban schools would be better funded than rural schools. The selected data points were then placed into a separate .csv file that was used in the statistical analysis.

3.0: Results

The project will now focus on analyzing the data collected so far and explore the relationships between the typology of a school district, the average total expenditure per pupil in a school district, and the average total Ohio State Test score of a pupil in the school district for the population of school districts located within the state of Ohio. There are three major aims of this project, the first aim is to use a one-way ANOVA model to determine if there is any difference in average total expenditure per pupil depending on the typology assigned to the school district. If it is found that there are differences in the average total expenditure per pupil among the typology groups, then a post hoc Tukey's HSD will be conducted to determine which two types of school typologies have different average expenditures per pupils. The second aim of this project will be to use a one-way ANOVA model to determine if there is any difference in average total OST test score depending on the typology assigned to the school district. If it is found that there are differences in the average total OST test scores among the typology groups, then a post hoc Tukey's HSD will be conducted to determine which two types of school typologies have different average total OST scores. The third aim will be to use a linear regression model to investigate the relationship between the average total expenditure per pupil and the average total OST test scores for school districts in Ohio. For all analyses, the significance level will be set to 0.05. The three major aims and their analyses are described in the following section (3.1-3.3).

3.1: Total Expenditure per Pupil and School District Typology

Before conducting an ANOVA test between the total expenditure per student and the typology of a school district, some assumptions of ANOVA testing must be checked to ensure that a one-way

ANOVA is fit to analyze the data set developed for this effort. The assumptions of having multiple samples drawn randomly from several distinct populations is valid and was addressed during the data collection step of this effort. The samples were also selected independently from one another. The two major assumptions that have not been assessed yet are the validity of the response variable having a normal distribution as well as the assumption of homoscedasticity which will be checked after an initial exploratory ANOVA analysis is performed, since the residuals from this analysis will be used in order to confirm or reject the assumptions. This initial ANOVA analysis was performed in R using the data collected for the typologies of the school districts sampled and the total expenditure per pupil data collected for these districts. The results from this analysis can be seen in **Table 1** and a box plot of this data can be seen in **Figure 1**.

Table 1: ANOVA Results for Comparison of Average Total Expenditure per Pupil and Typology

Source of Variability	*Df*	*SS*	*MS*	*F*	*p-value*
Typology	3	140,522,056	46,840,685	10.366	3.5e-6
Residuals	136	614,557,459	4,518,805		

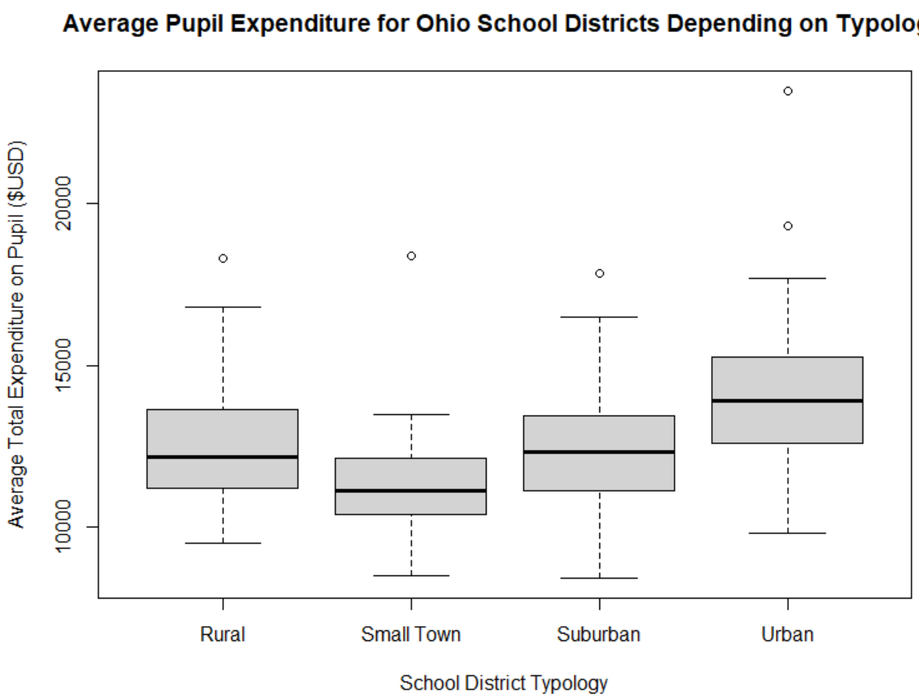

Figure 1: Average Total Expenditure per Pupil for Ohio School Districts Depending on Typology Group

From the results of the initial ANOVA model, the F value that was calculated was 10.366, which derives a p-value of 3.5e-6. This p-value is less than the significance level of 0.05, leading to the acceptance that there are differences between the total expenditure per pupil depending on the typology of the school district. The box plot that was created with the data set also confirms that there are visual differences that can be interpreted between the typological groups. This model also found that $\hat{\sigma}$ is $2,125.75 and the coefficient of determination was 0.223. The conclusion from this model, assuming the assumptions of normality and homoscedasticity are confirmed, shows that with 95% confidence, there is a difference in the total expenditure per pupil depending on the typology of the school district in the state of Ohio.

In order to check if the assumption of normality is valid for this analysis, a q-q plot of the residuals was constructed using R. Though it is typically assumed that a sufficiently large sample size is robust to this assumption, which this analysis would fit since the sample size is larger than 30 samples, it is still important to ensure the validity of this rule of thumb. The q-q plot of the residuals can be seen in **Figure 2**. A Shapiro-Wilk test was also conducted to examine the p-value, which is typically used to assess the normality of the residuals, if the p-value is low, it is usually assumed that the normality assumption should be rejected. The p-value was found to be 7.40e-6, which suggests that the assumption of normality should be rejected.

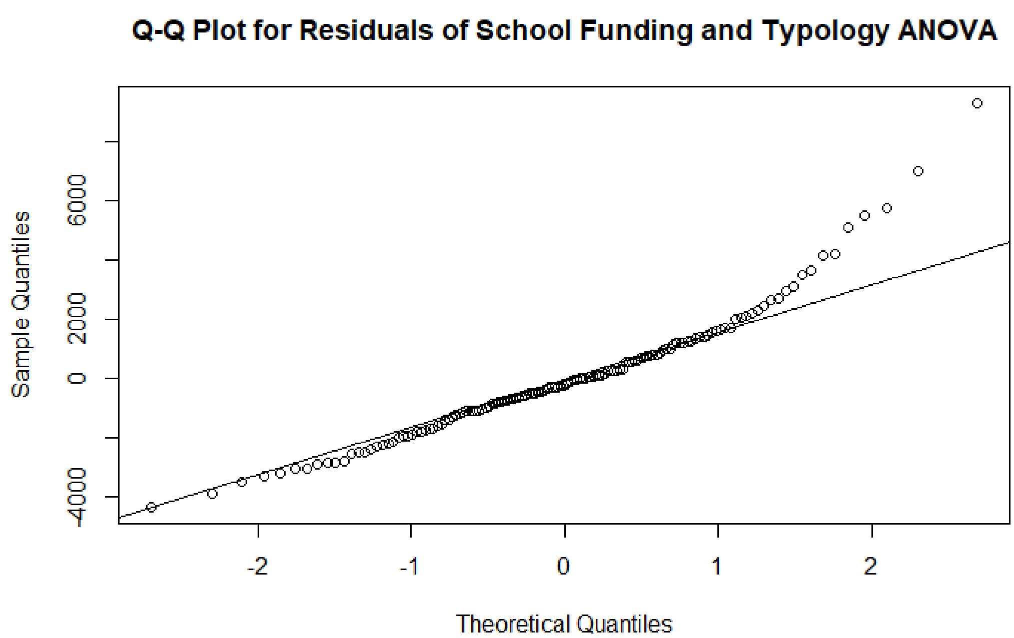

Figure 2: Q-Q Plot of Residuals for Cost vs Typology Relationship

As can be seen in the q-q plot, the residuals follow a fairly linear path, until the theoretical quantiles hit around 1, which is when the residuals start to increase in an exponential-type path. Due to the non-linear behavior of the higher theoretical quantiles, the assumption that the response variable has a normal distribution in each of the populations does not hold true for this analysis, going against the rule of thumb about how robust large samples are to normality.

Given that the normality assumption has been rejected, the assumption of homoscedasticity should also be confirmed or rejected. In order to test this assumption, a Levene's test was conducted, which performs a one-way ANOVA procedure on the absolute value of the residuals of this analysis. This test was conducted in R and the results can be found in **Table 2**.

Table 2: Results of Levene's Test for Cost vs Typology of School Districts

Response	*Df*	*Sum Sq*	*Mean Sq*	*F Value*	*p-value*
Typology	3	2.9e+14	9.8e+13	1.04	0.3772
Residuals	136	1.3e+16	9.4e+13		

From this test, it was found that the p-value of 0.3772 is greater than the significance level of 0.05, so we should retain the assumption homoscedasticity. As another quick check, a plot of residuals from the ANOVA model against fitted values was made as seen in **Figure 3**. It can be observed that the ranges for each associated x-value look roughly equal, which confirms the assumption of homoscedasticity.

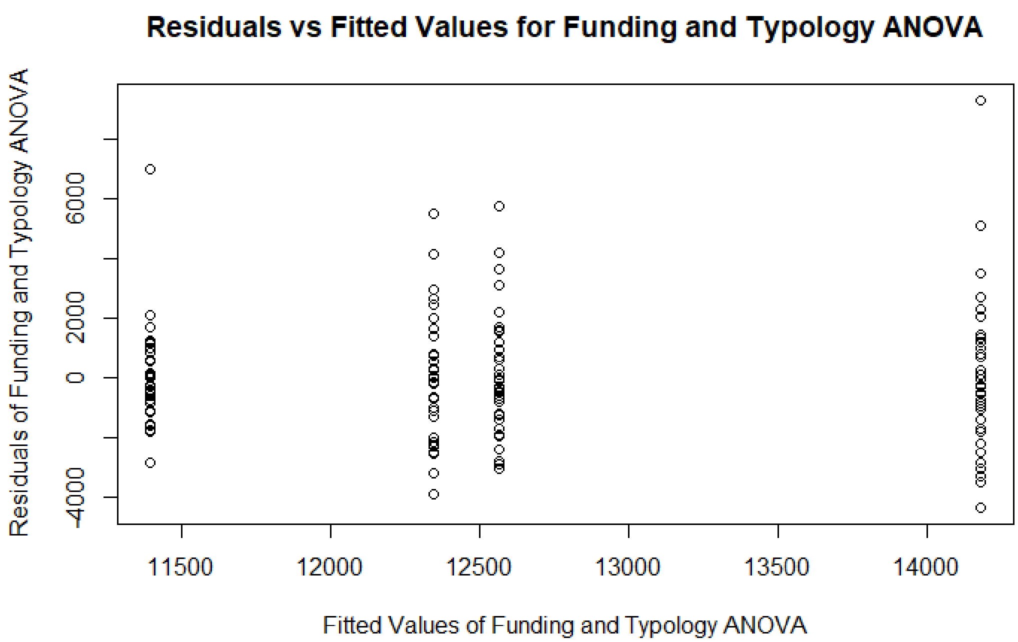

Figure 3: Residuals vs Fitted Values for Cost vs Typology Analysis.

Given that the assumption of normality was found to be violated, a typical ANOVA analysis of these variables is no longer warranted. This means that instead of conducting a one-way ANOVA, a Kruskal-Wallis test is the proper test to conduct. This test has the advantage of not assuming normality of the populations. This test was carried out in R and the results can be seen in **Table 3**. Once again, the test is being conducted to test the hypothesis that the populations have the same distribution.

Table 3: Kruskal-Wallis Test on the Relationship Between Total Expenditure per Student vs Typology of School District

Chi-squared	*Df*	*p-value*
28.822	3	2.4e-6

The Kruskal-Wallis test resulted in a p-value that is 2.4e-6, which is less than the significance level, which leads to the rejection of the null hypothesis, showing that there exist some differences between the total expenditure per pupil among the difference typologies of school districts. In order to determine which populations are shifted from each other, a post hoc pairwise Wilcox-Mann-Whitney test is conducted on each pair of the populations. This will allow for the testing of the null hypothesis from the last test but look at every pair to determine which relationship violates the hypothesis. The Bonferroni correction method to adjust the p-values was also utilized in this analysis in order to minimize the chance of committing a Type I experiment-wise error. The results from the Wilcox-Mann-Whitney tests with Bonferroni correction calculated from R can be seen in **Table 4**.

Table 4: P-values Derived from Post Hoc Wilcox-Mann-Whitney Test on all Populations Tested.

	Rural	*Small Town*	*Suburban*
Small Town	0.0512	-	-
Suburban	1.0000	0.1280	-
Urban	0.0183	2.9e-7	0.0070

From the results gathered, it was observed that when using a significance level of 0.05, the only two groups of populations that were found to show no differences between them were the rural and small town typologies and rural and suburban typologies, which means that the funding levels for these two groups of school districts in Ohio show no statistical difference in the amount of funding expended on a pupil, but every other relationship shows that there is enough statistical difference between them to reject the null hypothesis. This shows that typology is a decent identifier to show how school funding may affect test scores when using a categorical approach which will be examined next.

3.2: Average Total Ohio State Test Score and School District Typology

Once again, an initial one-way ANOVA model was created in R in order to examine the relationship between the typology assigned to a school district and the average total Ohio State Test score for the school district. This model will determine if there are differences in the average total OST test scores and the typologies assigned to school districts in Ohio. The result from this initial analysis can be seen in **Table 5** and a boxplot of the data analyzed can be seen in **Figure 5**.

Table 5: ANOVA Results for Comparison of Average Total OST Test Score and Typology

Source of Variability	*Df*	*SS*	*MS*	*F*	*p-value*
Typology	3	510,874	170,291	25.38	4.2e-13
Residuals	136	912,361	6,709		

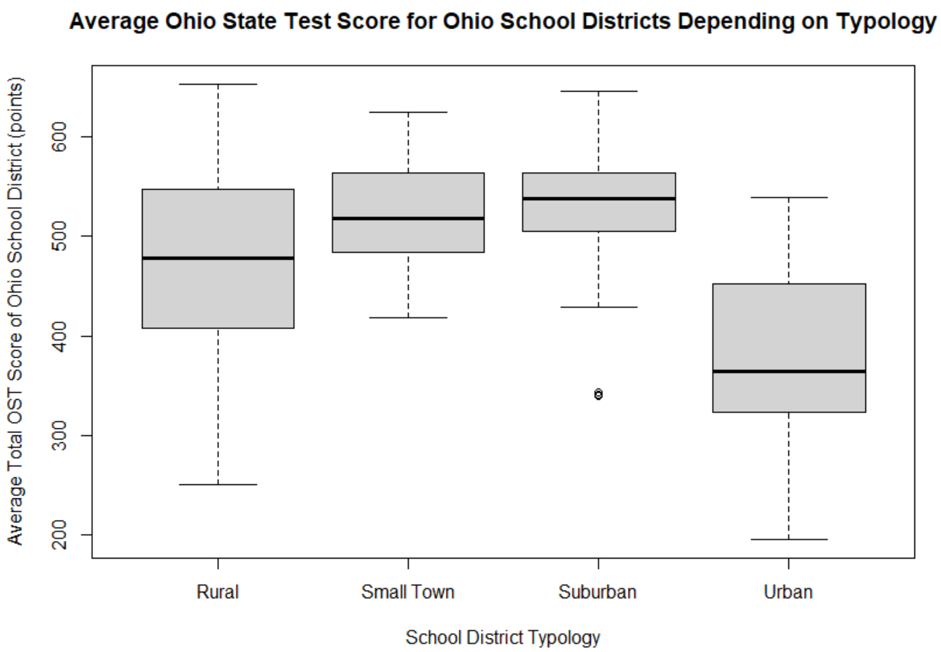

Figure 4: Average Total OST Score for Ohio School Districts Depending on Typology Group

From the results of the initial ANOVA model, the F value that was calculated was 25.38, which derives a p-value of 4.2e-13. This p-value is less than the significance level, leading to the acceptance that there are differences between the average total OST test scores depending on the typology of the school district. The box plot that was created with the data set also confirms that there are visual differences that can be interpreted between the typological groups. This model also found that $\hat{\sigma}$ is 81.91 points and the coefficient of determination was 0.359.

Once again, the assumptions of normality and homoscedasticity must be checked for this ANOVA analysis. For these checks, the same procedure as was done in the first analysis was followed. The q-q plot of the residuals for this analysis can be seen in **Figure 5**, once again this plot shows that there is some variation from the linear fit line, but this time the variation is at both the high and low values, so this plot suggests that the assumption should be rejected. A Shapiro Wilk test was also conducted on the residuals, which yielded a p-value of 0.177, which does not show much confidence in making an assumption of normality for this analysis.

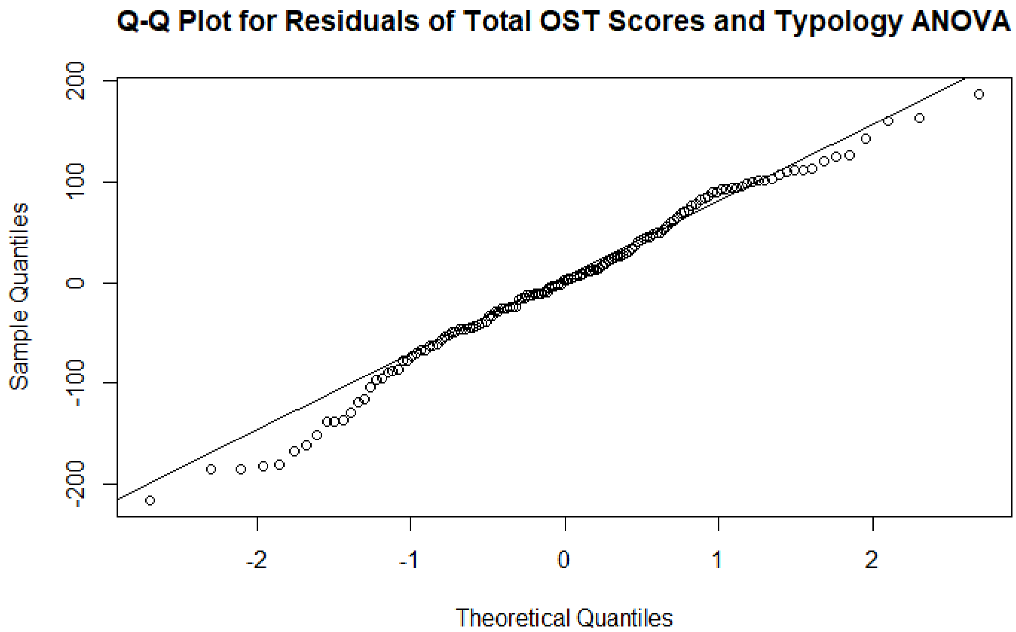

Figure 5: Q-Q Plot of Residuals for Typology and OST Test Score Results Analysis

Next to test the homoscedasticity assumption, a Levene's test was conducted and a plot of the residuals against fitted values was made. The results of the Levene's test can be seen in **Table 6** and the plot can be seen in **Figure 6**. Once again, the Levene's test resulted in a p-value of 0.027, which is smaller than the significance level of 0.05, leading to the rejection of the assumption of homoscedasticity. However, the plot of the residuals versus the fitted values suggests that this assumption is not violated since the spread between all of the x values are similar to each other. So for this model, it can be assumed that the sample is large enough to be robust to this assumption, allowing for the ANOVA analysis to be used.

Table 6: Results of Levene's Test for Average Total OST Test Score vs Typology of School Districts

	Df	*Sum Sq*	*Mean Sq*	*F Value*	*p-value*
Typology	3	7.22e8	2.41e8	3.15	0.0270
Residuals	136	1.04e10	7.64e7		

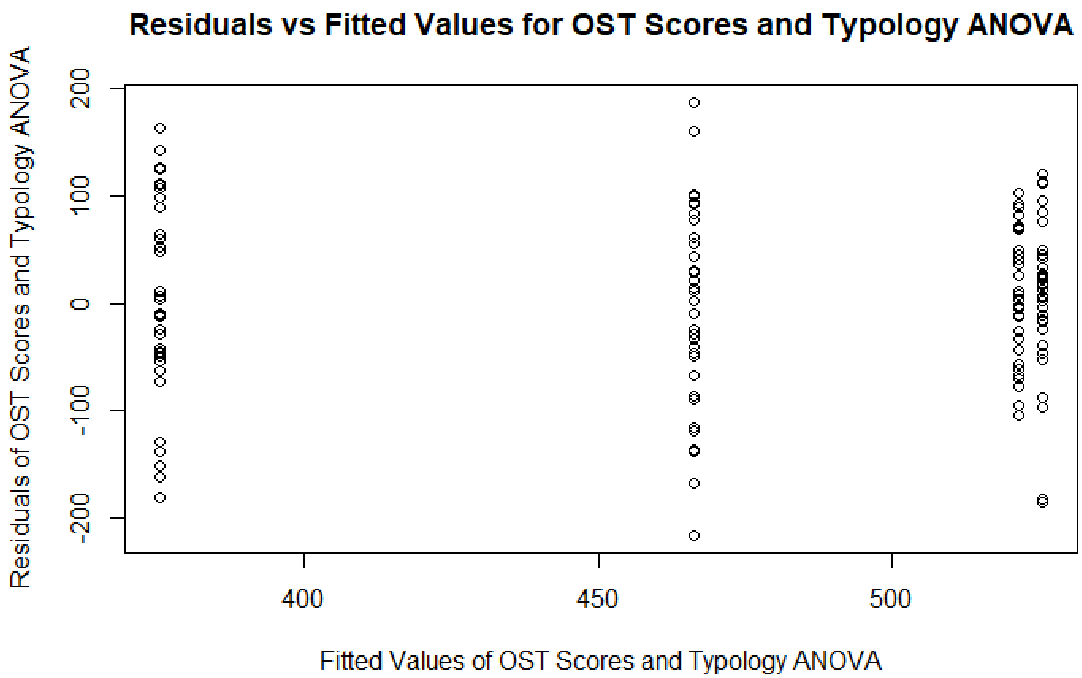

Figure 6: Residuals vs Fitted Values for Average OST Total Test Scores vs Typology Analysis.

In order to refine the understanding of the relationships that were examined in the ANOVA analysis, post hoc Fisher's Least Significant Difference (LSD) tests are conducted on all of the possible pairs of the typologies. The Bonferroni Correction method is also implemented to adjust the p-values calculated to avoid the chance to make an experiment-wise Type I error. These tests were conducted once again in R and the results can be seen in **Table 7**.

Table 7: P-values Derived from Post Hoc Fisher's Least Significant Difference Test on all Populations Tested for OST Test Scores

	Rural	*Small Town*	*Suburban*
Small Town	0.0330	-	-
Suburban	0.0180	1.0000	-
Urban	5.2e-5	6.1e-11	2.0e-11

From the results of the Fisher's Least Significant Difference tests, it was found that the only two typologies that did no showcase a statistically relevant difference in the mean average total OST test score was small town school districts and suburban school districts. The rest of the pairs examined in this study resulted in rejection of the null hypothesis since the p-values were all lower than the significance level of 0.05. So, from this analysis it can be concluded that with 95% confidence that the only two school district typologies that have similar average total OST test scores in Ohio are small town and suburban school districts.

A series of post hoc Tukey's Honestly Significant Difference (Tukey's HSD) tests were also carried out to gather more information through the means of 95% confidence intervals of the possible groupings of typologies as well take a less conservative approach than what was found in the Fisher's LSD tests. This analysis was performed in R software and the results of these tests can be found in **Table 8**.

Table 8: Post Hoc Tukey's Honestly Significant Difference Tests on all Populations Tested for OST Test Scores

	Diff	*Lower Confidence Interval Bound*	*Upper Confidence Interval Bound*	*Adjusted p-value*
Small Town-Rural	55.20	4.27	106.12	0.0280
Suburban-Rural	59.18	8.25	110.10	0.0157
Urban-Rural	-90.50	-141.43	-39.57	5.1e-5
Suburban-Small Town	3.98	-46.95	54.91	0.9970
Urban-Small Town	-145.70	-196.62	-94.77	<1.0e-6
Urban-Suburban	-149.68	-200.60	-98.75	<1.0e-6

From the results of these tests, the results of the Fisher LSD test are confirmed. It was once again found that with 95% confidence, the only group of school districts that did not have different mean total average OST scores was the suburban and small town group. More specifically, the results from the Tukey's HSD tests also show the 95% confidence intervals for the differences in the group means, which can be useful information in future analyses. For example, it was found that the mean average total OST test scores of school districts classified as small town is anywhere from 4.27 to 106.12 points greater than that of rural school districts in Ohio. It should also be noted that the means of the two populations that are being tested have a statistically significant difference, which can be observed in every test except for the suburban-small town test, which is reflected in the p-value. The confidence intervals for the tests in which the urban typology group is considered population A (first population labeled in **Table 8**) have both negative values in the 95% confidence intervals, meaning that the mean scores reported for these school districts are lower than any of the typological groups it is compared with (population B). This observation was also seen in the boxplot of this relation; however, this test allows for the calculation of exact numbers in which with 95% confidence, the scores in the urban areas can be less than the other population of the test. For example, it was found that, with 95% confidence, when testing the urban and suburban typological groupings, the mean average total OST score for the urban school districts can be anywhere from 98.75 to 200.60 points lower than the mean average OST scores in the suburban school districts. From the results, it was observed that the largest possible gap between mean average total OST scores between two typologies of school districts were found between the urban and suburban typologies with the urban and small town test very close, about a 4 point difference in confidence bounds.

3.3: Linear Regression Model of Total Expenditure per Pupil vs Average Total OST Score

Since the above tests do not indicate that using typology to examine the relationship between school district funding and OST test scores leads to an obvious conclusion, a linear regression model between the two was developed. This model was developed to determine if there is a linear correlation between these two variables that could be used to retain or reject the a priori hypothesis of this study. The initial plot of the two variables with the means shown can be seen in **Figure 7**. There seems to be an initial negative correlation between these variables when considering the average total OST score as the response variable and the total expenditure per pupil as the independent variable.

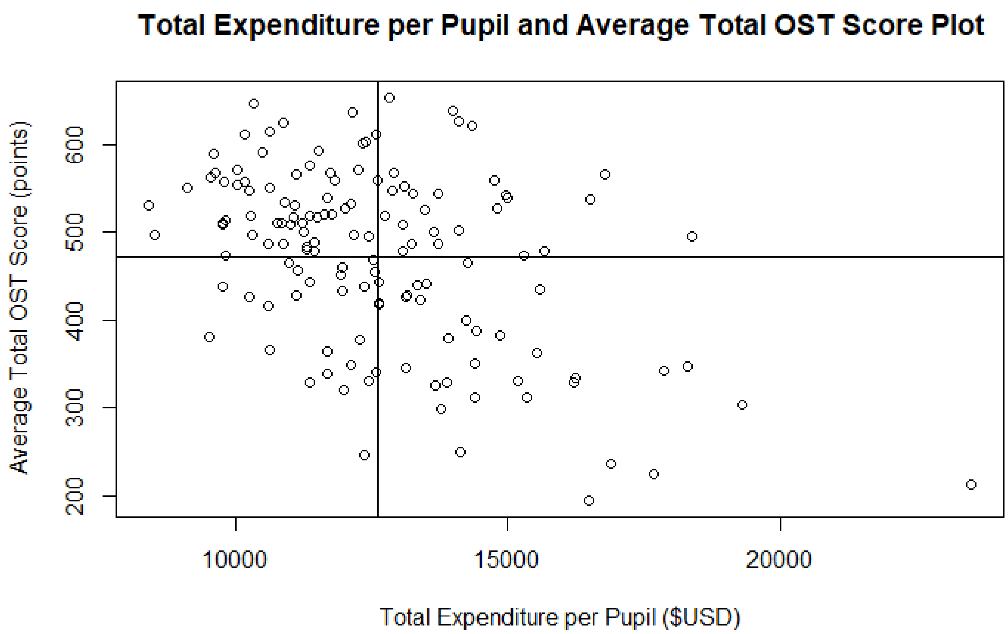

Figure 7: Total Expenditure per Pupil vs Average Total OST Score

By using R to determine the coefficients for the linear regression model of the two variables, it was found that the intercept was 729.14 points, and the slope is -0.0204 (points/$USD). This estimate for the slope suggests that for every dollar increase in expenditure per pupil a school district implements, the average total OST score drops by 0.0204 points. So, it would take roughly a $50 increase in expenditure per pupil to drop the average total OST score by a single point. Though it should be considered that there are several other factors that influence the test results of a school district, so this relationship should be used cautiously. However, before confirming these results, the residuals should be examined to assess if the assumptions of linear regression should be accepted. A q-q plot of the residuals can be seen in **Figure 8**, which shows a fairly linear relationship between the residuals, with slight variations in the upper and lower theoretical quantiles. A plot of the residuals and the fitted values can also be seen in **Figure 9**, which seems to confirm the assumption of homoscedasticity with no extreme outliers present and most equal variances. A histogram of the residuals is also found in **Figure 10**, which suggests that there is a slight skew to the left, but not a large enough one to reject the normality assumption.

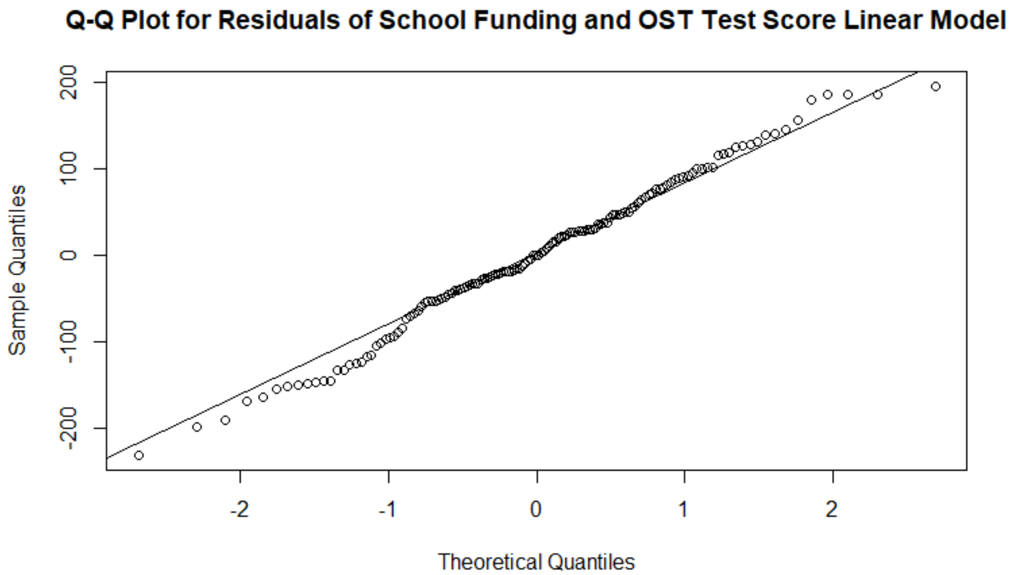

Figure 8: Q-Q Plot of Residuals for Linear Regression Model of Expenditure vs OST Scores

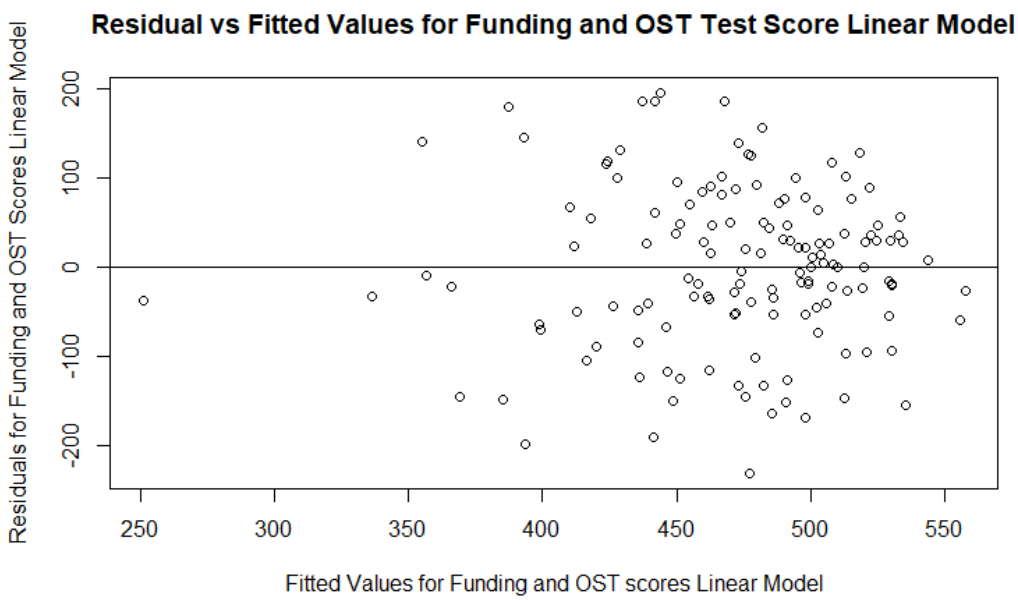

Figure 9: Plot of Residuals against Fitted Values for Linear Regression Model

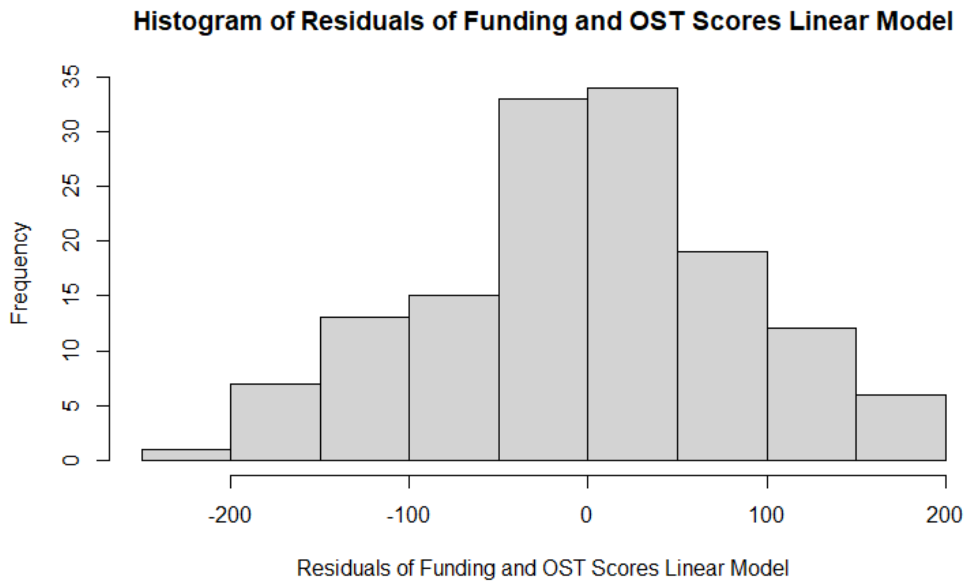

Figure 10: Histogram of Residuals for Linear Regression Model

It was also checked to see if transformations could be used to ensure no violations have been committed of the assumptions needed for the linear regression model. The first one that was checked was the log transformation. The q-q plot of the residuals transformed using a log function can be seen in **Figure 11**, which shows that there is not much change thus this transformation will not be used. The second transformation that was checked was the square root function, which the q-q plot of the residuals post-transformation can be seen in **Figure 12**. Once again it was found that no transformations seem needed for this model.

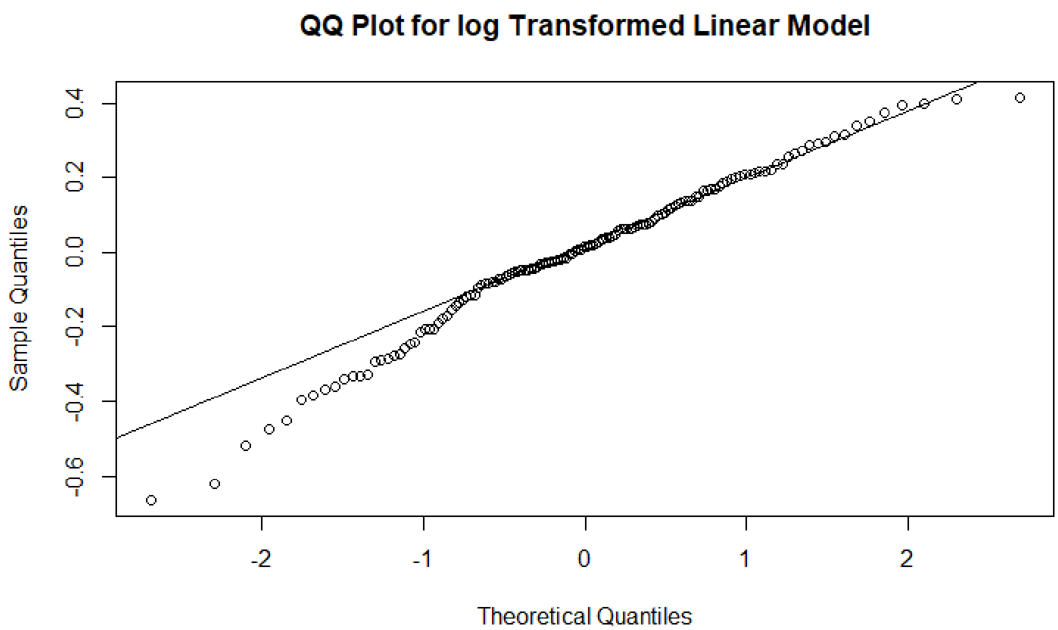

Figure 11: Q-Q Plot of Residuals of Linear Regression Model after Log Transformation

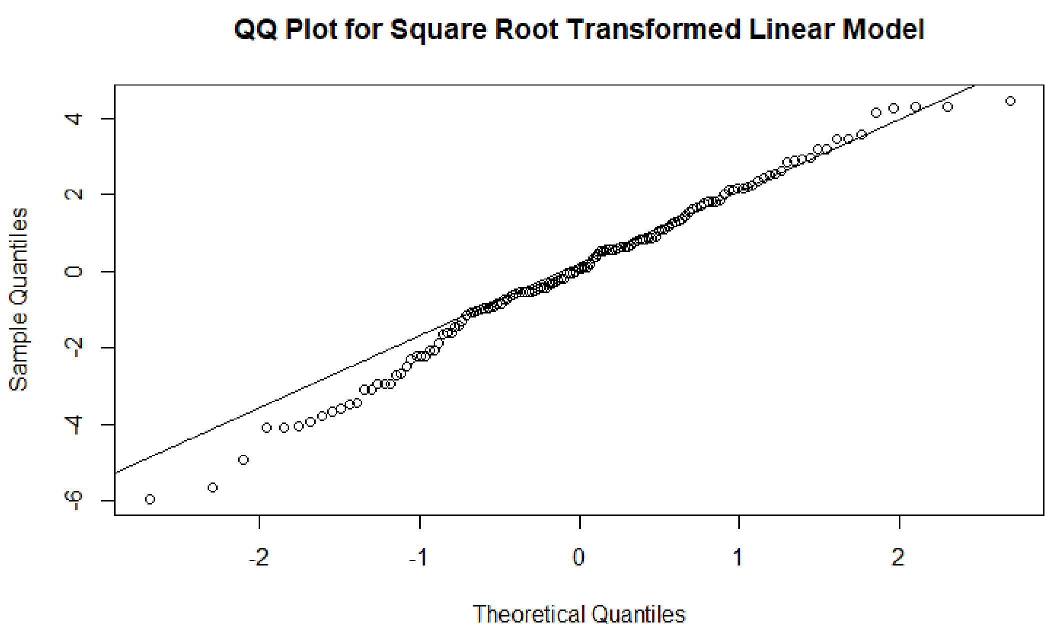

Figure 12: Q-Q Plot of Linear Regression Model after Square Root Transformation

Since It was found that the transformations did not affect the data in a significant amount, the linear regression model will be based on the original data with no transformations. The summary of the untransformed linear regression model can be seen in **Table 9**. It was found that the p-values for the intercept and the slope suggests that there is a strong linear relationship between the total expenditure per pupil and the average Ohio State Test score. More specifically, it was found that there is a negative correlation between these two variables, suggesting that the higher the expenditure per pupil, the lower the average total OST test score became as was discussed earlier.

Table 9: Linear Regression Model Coefficients for Relationship between Total Expenditure per Pupil and Average Total OST Test Score for Ohio School Districts

	Estimate	*Std. Error*	*T-value*	*p-value*
(Intercept)	729.14	41.89	17.406	<2e-16
Cost	-0.0204	0.0033	-6.236	5.12e-9

By using the results from the test, a plot of the linear regression model along with the confidence interval in blue and prediction interval in red can be seen in **Figure 13**. The 95% confidence limits for this model were also calculated and can be seen in **Table 10**. The value of $\hat{\sigma}$ was found to be 89.7 on 138 degrees of freedom and the coefficient of determination was 0.2199.

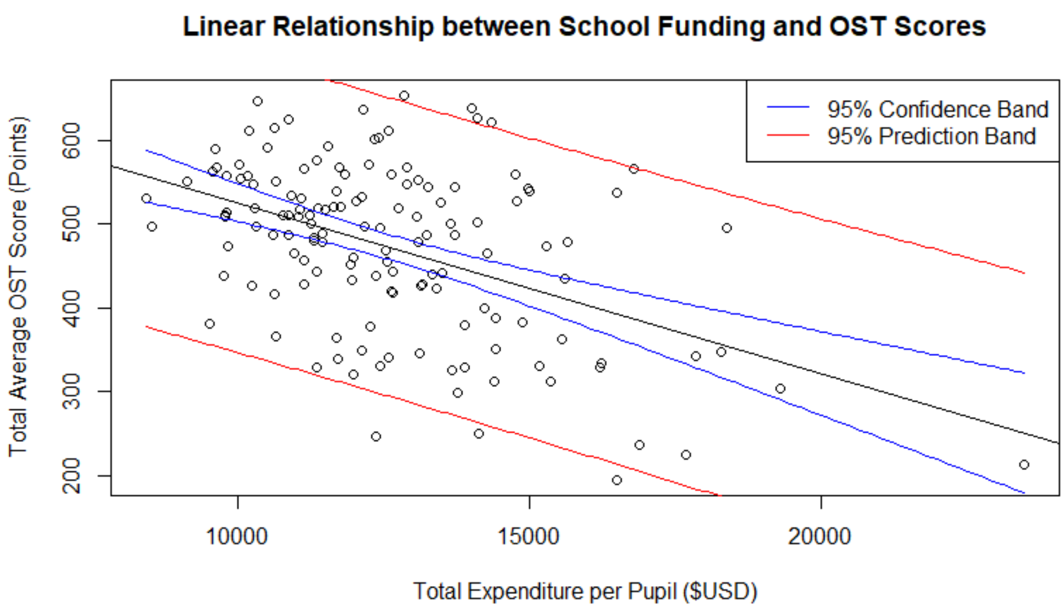

Figure 13: Linear Regression Model for Relationship between Total Expenditure per Pupil and Average Total OST Score

Table 10: 95% Confidence Limits of Linear Regression Model of Total Expenditure per Pupil and Total Average OST Test Score

	2.5% Confidence Limit	*97.5% Confidence Limit*
Intercept	646.315	811.971
Cost	-0.0268	-0.0139

4.0: Conclusion

From the performed statistical analyses, it was found that the a priori hypothesis should be rejected. Through the development of a linear regression model, it was found that the sampled school districts had a negative correlation between the total expenditure per pupil and the average total Ohio State Test score reported for said school districts in the state of Ohio. More specifically the linear regression model found that with a 95% confidence, an increase in expenditure per pupil will lead to a decrease in the school district's average total OST score. The linear model found the slope to be -0.0204 points/$USD, which provides a direct relation between funding and student test scores which can predict how much the test score will be lowered when funding is increased by a certain amount. This statement seems counter-intuitive, since it is typically thought that underfunded school districts would perform worse due to the lack of resources available to prepare the students for these standardized tests. However, it is advised to take this result with a grain of salt due to the fact it only examines the relationship between the school district's funding and test results in Ohio. There are several other socio-economic issues in play that may be the cause behind this relationship. It was also found that there is a relationship between the typology assigned to a school district and the total expenditure per pupil of the district. It was found that with 95% confidence, suburban and rural schools do not exhibit statistically significant differences in the average reported expenditure per pupil from the sample examined in this study. The rest of the typology pairings exhibited statistically significant differences in the reported expenditures per pupil with small town schools exhibiting the lowest funding levels and urban schools exhibiting the highest funding levels on average.

The analyses that were conducted also examined how the school district typology affected the performance of a student on these standardized tests. From the analyses conducted, it was found that with 95% confidence, the only typological pairing of school districts that did not exhibit a statistically significant difference in mean average total OST score was small town and suburban school districts within the State of Ohio. It was found that districts classified as urban exhibited the lowest average total OST scores, while small town and suburban school districts exhibited the highest OST scores. The relation in typology appears to play a slight role in the performance of a student since even though suburban school districts exhibit both the highest test scores and expend the most money on their pupils, there is no other apparent correlation between student performance and funding with consideration of the school district typology. To better understand how the various factors affect the performance of a student a much larger and robust statistical analysis would be recommended.