The Impact of Urbanicity and Class Size on School Performance

Ivan Karpov

March 17, 2025

Abstract

This report studies the effect of urbanicity and class size on math performance of 3rd grade students. The dataset used comes from Project STAR and Beyond, an extensive experimental study containing information on academic performance of elementary school children. A 2-Way ANOVA model is implemented with class type and urbanicity as factors and math scores as the response variable. The results show that these 2 variables do have a significant effect on 3rd grade math performance. The model follows all the necessary assumptions, as the Q-Q Plots and Residual plots display normal patterns. In conclusion, the study discovers that students in small classes score higher than other types, and that inner city students score lower than other areas.

1 Introduction

For children, performing well in school is very important for their development and future. However, not all students achieve the same level of success, and some students struggle. There are lots of factors that may affect how well a child is learning, and unfortunately some of them are not under their control. In particular, we want to look at 2 of those factors: urbanicity and class type and measure their effect on school performance.

Urbanicity is a categorical variable that describes how urban or rural the area in which the student goes to school is. It is divided into 4 categories: inner city, suburban, rural, and urban. Inner city schools were those located at the center of ciies and metropolitan areas. Suburban schools were located on the outskirts of those metropolitan areas. Schools in towns with population greater than 2500 were classified as urban, and schools in the countryside far away from cities were classified as rural.

Class type is a categorical variable that is split into 3 types: small, regular, and with aide.

The motivation to study urbanicity and class size is to see how much of a difference the school location and environment make on students' educational achievement. Many believe that small towns and rural areas are behind big cities when it comes to education, and that large classes also hinder performance so we want to investigate those claims. The results of this study will provide a picture of whether there are educational gaps between the different levels that need to be closed. This report will also provide a baseline of how well a certain student should be performing based on their class type and urbanicity.

2 Background

The dataset used for this project comes from a study called Project STAR and Beyond, where STAR stands for Student/Teacher Achievement Ratio. The study was conducted in Tennessee in the 1980s to see how class size affects academic achievement.

An experiment was conducted which involved collecting data from 11,601 students across 80 elementary schools in Tennessee. To avoid underrepresenting certain groups, they made sure to include children from both metropolitan and rural areas. To conduct the experiment, kindergarten students were assigned at random to 3 different class types: small (between 13-17 students), regular (between 22-25 students) and aide (between 22-25 students with a full-time RA). Regular classes were the control group, while small classes were the experimental group. Those students were monitored for 4 years and a large number of statistics were collected throughout and organized into the STAR dataset.

The dataset contains 379 columns with various kinds of information about each student. There are columns assigned to demographics such as race, gender, and urbanicity. Other variables include school

ID, class size, and teacher characteristics. To assess student achievement, test scores are reported for reading, math, science, reading, listening in each grade started from kindergarten to 4th grade. In addition, for some students data is available all the way through high school where GPA, SAT/ACT scores and graduation status are included.

The design of the experiment is very effective in the sense that a pretty large sample size is collected (N = 11601) and they conduct this test in the cities, suburbs, and rural areas which means they encompass children from all areas, and this should reduce the overall bias and make the results represent the total population.

However, there are also some potential shortcomings of the study. First, the study was conducted only in Tennessee, so it is hard to extrapolate it to the entire United States as the performance results might be different in other states. Also, there is the problem of too many missing values in the dataset. This incompleteness of data can cause bias as children that were left out of the study might perform differently on the same assessments.

In the initial analysis report, we studied the effect of class size and school ID on math scores. The problem was that the error terms were not normal for the 2-way ANOVA and school ID is not a statistic that can be well described to the public. So instead we will study the urban status, as this should allows us to obtain results that can be described to the policy makers to improve elementary school education.

3 Descriptive Analysis

The 3 variables that we will use in this report are class size, urbanicity, and math scores. Class type and urbanicity are both categorical variables with 3 categories for class type (1 = small, 2= regular, 3 = aide) and 4 categories for urban status(1 = inner city, 2 = suburban, 3 = rural, 4 = urban). The response variable is quantitative and is the computational math score on the SAT ranging from 451 to 739. We will take the math scores and urban status from 3rd grade and the class type as composite. The composite means that if the student has participated in a small class for at least 1 year by 3rd grade, the label will be small.

The distribution of math scores can be seen using the histogram below.

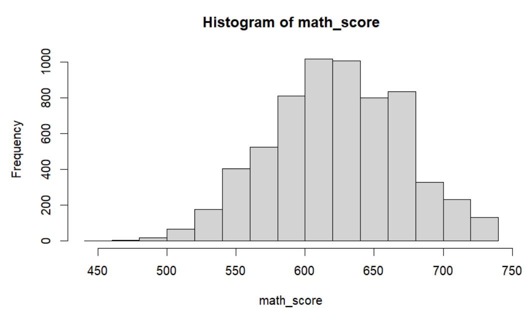

Figure 1:

As we can see, the distribution seems roughly symmetric with no significant skewness, so the mean is an appropriate summary statistic, with its value being 623.8. We can also see that the shape closely resembles a normal distribution.

Figure 2 shows the boxplot of math scores.

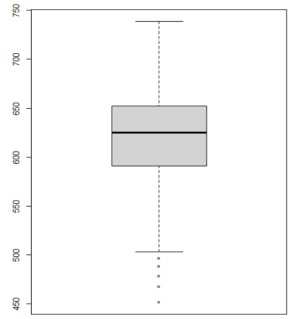

Figure 2:

As we can see we have some outliers on the left end, meaning there are a few students who achieved much lower scores.

Since class type and urbanicity are categorical variables, we will use a histogram to visualize their distributions, as seen in Figure 3.

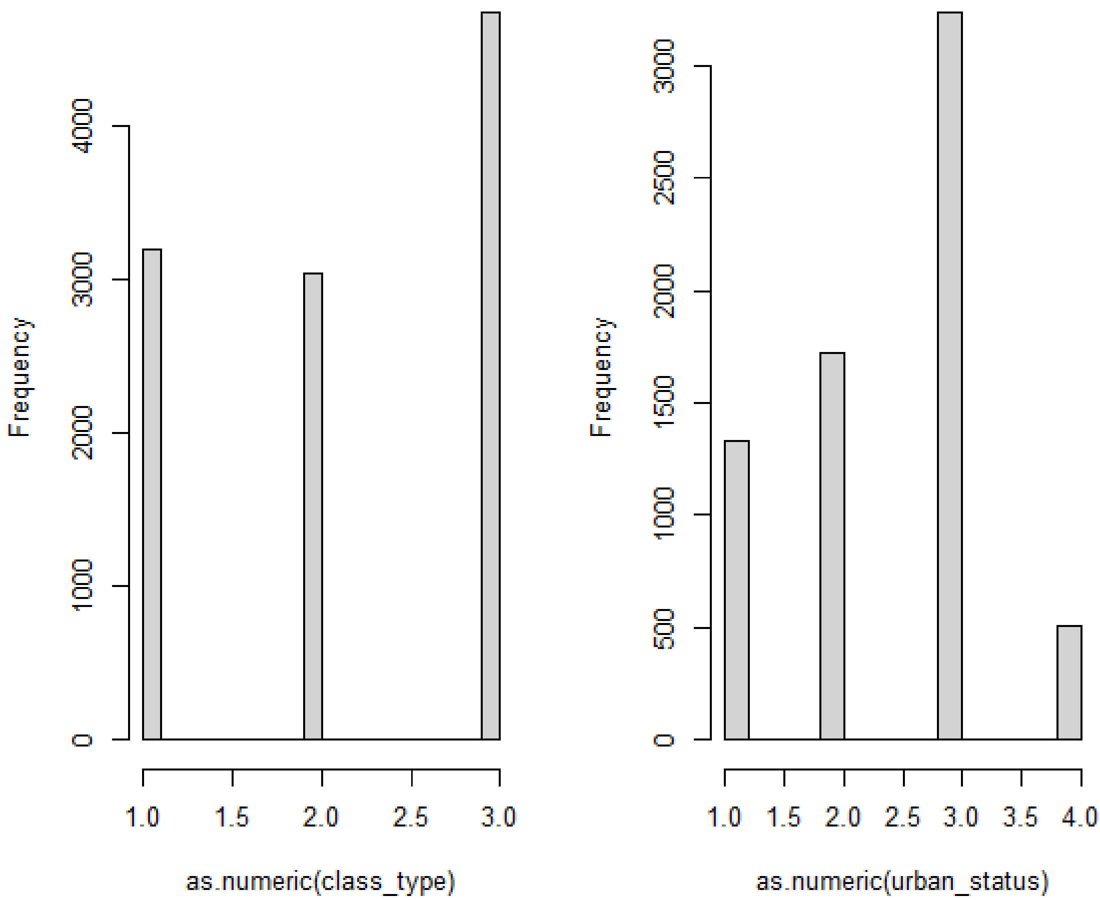

Figure 3:

As we can see that the biggest portion of study participants are from rural areas and in classes with aides.

It is worth mentioning that those 3 variables contain lots of missing values (5254 for math score, 4799 for urban status, and 613 for class type). These missing values may impact the results of the study, and can only be removed if they are missing completely at random. For this report, we will make the assumption that this is the case, so it should be noted that there is a higher chance of obtaining biased results because values are thrown out.

Lastly we want to look and see whether there exists an interaction between class type and urban status, and the relationship between the 2 variables is displayed in Figure 4.

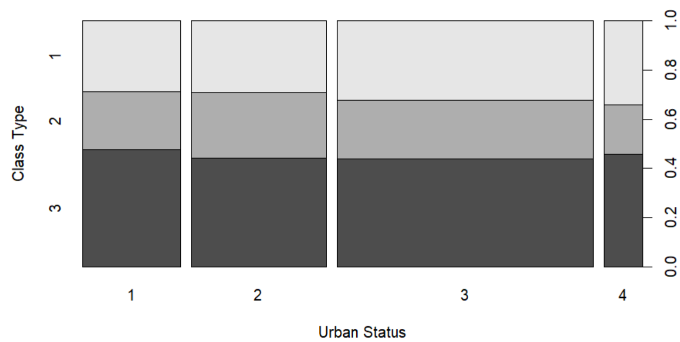

Figure 4:

The graph shows that these 2 variables have an effect on each other as schools in the inner city area have fewer small classes, but more classes with an aide, while the complete opposite is true for rural schools.

4 Inferential Analysis

We will now run a 2-way ANOVA model to study the effects of class type and urbanicity on math scores.

The 2-way ANOVA model has the form: $Y_{ijk} = \mu_{...} + \alpha_i + \beta_j + +(\alpha\beta)_{ij} + \epsilon_{ijk}$, where
Y_{ijk} denotes the math score of the kth student with the ith level of Urban Status and jth Class Type.
$\mu_{...}$ is the overall mean,
α_i is the main effect of factor A (Urban Status),
β_j is the main effect of factor B (Class Type),
$\alpha\beta_{ij}$ is the interaction between factors A and B,
ϵ_{ijk} is the error term.

The 2-way ANOVA model is subject to the following constraints:
1. $\sum \alpha_i = \sum \beta_j = 0$,
2. $\sum_{i=1}^{a} (\alpha\beta)_{ij} = 0, \forall j$,
3. $\sum_{j=1}^{b} (\alpha\beta)_{ij} = 0, \forall i$.
4. The error terms are independent, distributed $N(0, \sigma^2)$.

After running the ANOVA model in R we get the following output:

```
                    Df   Sum Sq Mean Sq F value   Pr(>F)    
g3surban             1   266590  266590  117.29  < 2e-16 ***
cmpstype             1    58815   58815   25.88 3.75e-07 ***
g3surban:cmpstype    1    29542   29542   13.00 0.000314 ***
Residuals         5871 13343803    2273                     
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
5726 observations deleted due to missingness
```

Figure 5: 2 Way ANOVA

As we can see the p-values for all 3 coefficients are very low, indicating that we should keep all 3 terms. Just in case, we want to test whether to keep the interaction term. To do this, we will use the F-test. The hypotheses are:
$H_0 : (\alpha\beta)_{ij} = 0$, $H_A : (\alpha\beta)_{ij} \neq 0$

After running the test in R, we get the p-value = 0.0003144, so at any significance level the null hypothesis is rejected. Thus, we conclude that the interaction between urbanicity and class size is important in determining math scores.

Since we are using the ANOVA model, there are certain assumptions that need to be made. The error terms have to be independent, identically distributed, have to follow a normal distribution, and have equal variance. In the next section if this report, we will analyze how well these assumptions hold for our model.

We now want to see which differences in means for urbanicity and class size are statistically significant. For that, we will run the TukeyHSD(Honest Significant Difference) Test for comparison between pairwise means. For this test, there are 3 key assumptions: the observations are independent within and among groups, the groups for each mean are distributed normally, and there is equal variance between the groups.

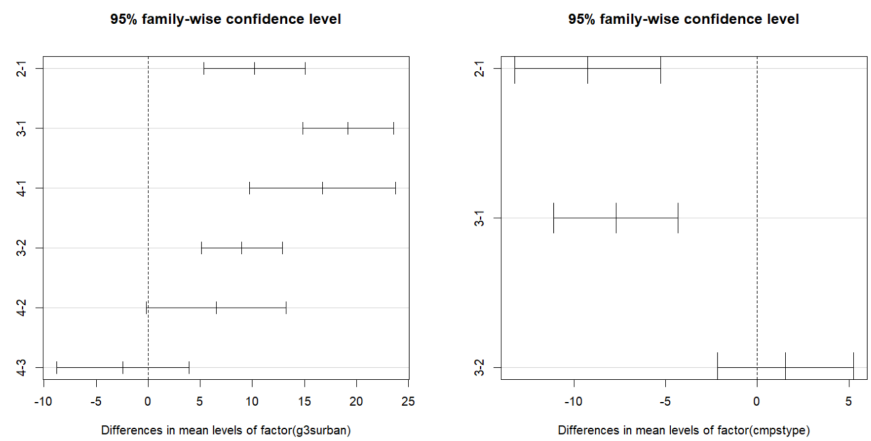

Figure 6: Tukey Confidence Intervals

The results of the Tukey's HSD demonstrate that there is a significant difference in math scores between these categories: inner city and suburban, inner city and rural, inner city and urban, suburban and rural. There are also notable differences between small classes and classes that are either regular or aided.

We also were able to get fitted results for the purpose of trying to predict a student's math scores given their class type and urbanicity, as seen in Figure 7.

```
Coefficients:
                                   Estimate Std. Error t value Pr(>|t|)    
(Intercept)                        622.3943     2.6691 233.182  < 2e-16 ***
factor(g3surban)2                    7.5239     3.5010   2.149  0.03167 *  
factor(g3surban)3                    9.1617     3.0841   2.971  0.00298 ** 
factor(g3surban)4                   13.3557     4.7311   2.823  0.00477 ** 
factor(cmpstype)2                  -12.8924     3.9976  -3.225  0.00127 ** 
factor(cmpstype)3                  -18.3218     3.3989  -5.391 7.30e-08 ***
factor(g3surban)2:factor(cmpstype)2  -0.2974    5.1813  -0.057  0.95423    
factor(g3surban)3:factor(cmpstype)2   6.9838    4.6788   1.493  0.13559    
factor(g3surban)4:factor(cmpstype)2   1.5601    7.7348   0.202  0.84016    
factor(g3surban)2:factor(cmpstype)3   5.8954    4.4767   1.317  0.18792    
factor(g3surban)3:factor(cmpstype)3  17.6085    3.9642   4.442 9.08e-06 ***
factor(g3surban)4:factor(cmpstype)3   5.1368    6.1728   0.832  0.40535    
```

Figure 7: Coefficients

As we can see, the average math score is 622.3943 for someone living in the inner city studying in a small class. To determine the predicted math score for all other categories, we take the intercept and add the coefficient for the urbanicity and class type effects, and then add the interaction term. For example, to calculate the math score for someone in a regular class living in a rural area, it would be 622.3943 + 9.1617 - 12.8924 + 6.9838 = 625.6474. It is worth noting that urbanicity and class type significantly affect the predicted outcome, in many cases changing the math score by more than 10 points.

5 Sensitivity Analysis

We will now plot the diagnostics to test whether the assumptions hold for our 2-Way ANOVA Model, as seen in Figure 8.

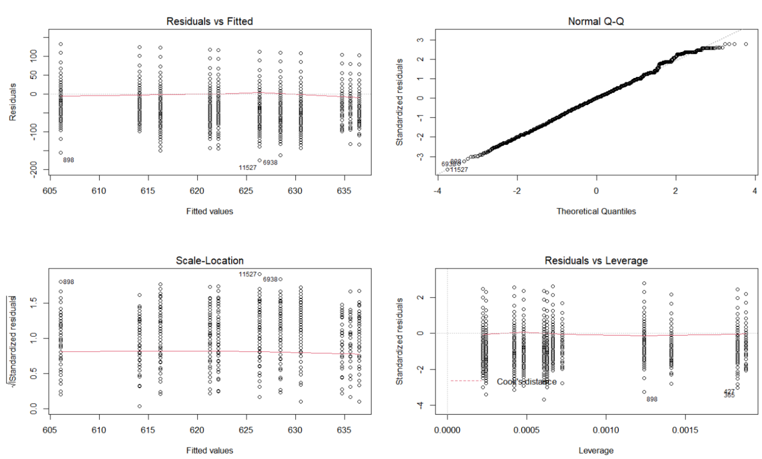

Figure 8: Diagnostic Plots

The first plot is the residuals vs. fitted plot and it is used to test whether there is equal variance. Since the line is straight and the range of the residuals remains roughly the same regardless of the fitted values, we can conclude that there is equal variance of the error terms. The values follow the line on the Normal Q-Q plot, indicating that the errors are normally distributed. There also does not seem to be too many outliers, and none of the outliers are that extreme. So overall, the model assumptions are met very well.

6 Discussion

In this project, a 2-Way ANOVA model was conducted to see how math scores in 3rd grade are affected by how urbanized their location is (city vs. suburb vs. urban vs. rural area) and by class type (small vs. regular vs. aide). As a result of this study, we have determined that these 2 variables do have a significant effect on a student's math score. As we can see from the fitted results and Tukey's HSD, small class size has a positive effect on math scores. Also, we could see that students in the inner city perform significantly worse than the other regions, especially compared to the urban (small town) and rural areas.

Based on these results, policymakers should consider making the class size smaller, since there is an obvious positive effect on the math computation score. They should also consider closing the educational gaps between inner cities and less densely populated areas, as the students living in cities are behind everyone else.

A good direction for further research would be to study whether the results would be the same for other subjects, because it is possible that class type and urbanicity might not have the same effect on reading scores, for example.

One of the main caveats of this report is that there are a lot of missing values, so the data used does not contain the information about every student that went through this study. Also, it should be noted that this study specifically focuses on 3rd grade math scores, so you cannot take the results of this study and extrapolate them to overall academic performance prior to testing the results for a different subject and grade level.

7 Acknowledgement

I have collaborated with students Erin Moffat and Samantha DeMars to discuss and brainstorm ideas for the project.

8 References

Bevans, R. (2023, June 22). ANOVA in R: A complete step-by-step guide with examples. Scribbr. https://www.scribbr.com/statistics/anova-in-r/

Tukey test / tukey procedure / honest significant difference. Statistics How To. (2024, August 21). https://www.statisticshowto.com/probability-and-statistics/statistics-definitions/post-hoc/tukey-test-honest-significant-difference/