Analyzing the Effect of Class Type on First-Grade Students' Math Scores

Liela Meng, 917843295, Team #2

1/27/2021

Abstract

In this report, data from the project STAR was analyzed to investigate the relationship between class type and first-grade students' math scores. After initial descriptive data analysis, a two-way ANOVA model was conducted to answer the question of interest. Furthermore, causal assumptions were discussed. The results show a significant effect of class type on students' math scores but fail to draw a causal conclusion.

Introduction

The Tennessee Student/Teacher Achievement Ratio study was a randomized, longitudinal class-size study, which aims to compare student's academic achievements based on the class type they attended in. There were three class types: small (13-17 students per teacher), regular (22 to 25 students per teacher), and regular with an aide (22 to 25 students with a full-time teacher's aide).

Class size reduction (CSR) has been advocated for years, and the small class has been promoted to improve students' enthusiasm and test scores immediately. (https://en.wikipedia.org/wiki/Class-size_reduction)

This analysis tends to investigate the relationship between first-grade students' total STA math score and class types. In other words, we would like to know whether there are any differences in the math scores across three class types for 1-st grade students.

Thus, this analysis investigates whether a small class advantage exists for first-grade students' math performance. If the result corresponds to known facts, this analysis can serve as a support for CSR.

Background

The data file "STAR_students.tab" is downloaded from the Harvard database website [**?**]. This random experiment was conducted with 11,601 students from kindergarten to grade 3. Within each school, all students and teachers were randomly assigned to a class. Each teacher is involved in one and only one class, and the class type remains unchangeable.

The database records students' demographic information (e.g., gender, race, etc.), their yearly academic achievements in various areas (e.g., SAT score, reading score, engagement assessment scores, high school graduation, etc.), and education information (e.g., school location, teacher degree, teacher race, etc.).

Descriptive analysis

Only first-grade student participated in the experiment were included in this analysis. Selected variables are teacher id, school id, student's math score, and class type.

Three hundred thirty-nine teachers from 76 schools participated in the stage of the experiment we are interested in (grade 1). Each school has 3 to 12 teachers and the mean number around 5. Basically, three class types have similar number of classes: $36.6\%(n = 124)$ small classes, $33.9\%(n = 115)$ regular classes, and $29.5\%(n = 100)$ regular+aide classes. Notice, only 72 schools have class type regular+aid.

The total number of students one teacher has, regardless of class type, has a mean $20(sd = 4)$. Particularly, mean±sd for number of students in small class is 15 ± 1.7, regular class 22 ± 2.3, regular with aid class 22 ± 2.1 [see table 1].

Table1: Summary regardless of class type	data (N = 339)
median math score per teacher	
minimum	465.00
median (IQR)	532.00 (512.00, 549.00)
mean (sd)	530.64 ± 26.74
maximum	619.50
number of students per teacher	
minimum	11
median (IQR)	20 (16.00, 23.00)
mean (sd)	19.46 ± 3.92
maximum	29
class type	
SMALL CLASS	124/339 (36.6%)
REGULAR CLASS	115/339 (33.9%)
REGULAR + AIDE CLASS	100/339 (29.5%)

Since the median is more robust than the mean when the total number of students in a class is relatively small, I use the median math score among students as the summary measure for each teacher/class.

From now on, I will refer to this median math score per teacher as "math scores" for simplicity. Class or teacher are reviewed as study units and I will use these two interchangeably. The math scores vary between teachers, mean±sd=530.6 ± 26.7 with a minimum 465 and a maximum 619.5.

Both mean and median value of math for the small class is higher than regular and regular+aide types. Such a result is consistent with our question of whether students in small class types perform better regarding math scores than other class types. Table 2 shows mean±sd for small, regular, regular with aid class types are 538 ± 28, 526 ± 25, 527 ± 25 correspondingly. From the box plot (see figure1), we can draw a similar conclusion. One thing to notice is that the small class type's dispersion is larger than the other two.

Table2: Summary by class type	SMALL CLASS (N = 124)	REGULAR CLASS (N = 115)	REGULAR + AIDE CLASS (N = 100)
median math score per teacher			

Table2: Summary by class type	SMALL CLASS (N = 124)	REGULAR CLASS (N = 115)	REGULAR + AIDE CLASS (N = 100)
minimum	465.00	468.00	470.00
median (IQR)	538.00 (518.00, 557.00)	523.00 (508.50, 540.00)	527.50 (512.00, 545.00)
mean (sd)	538.10 ± 28.10	525.56 ± 25.19	527.25 ± 24.88
maximum	619.50	588.00	572.00
number of students per teacher			
minimum	11	15	18
median (IQR)	15.00 (14.00, 16.00)	22 (20.00, 23.00)	22.00 (21.00, 23.00)
mean (sd)	15.06 ± 1.69	21.80 ± 2.29	22.24 ± 2.07
maximum	20	28	29
class type			
SMALL CLASS	124/124 (100.0%)	0/115 (0.0%)	0/100 (0.0%)
REGULAR CLASS	0/124 (0.0%)	115/115 (100.0%)	0/100 (0.0%)
REGULAR + AIDE CLASS	0/124 (0.0%)	0/115 (0.0%)	100/100 (100.0%)

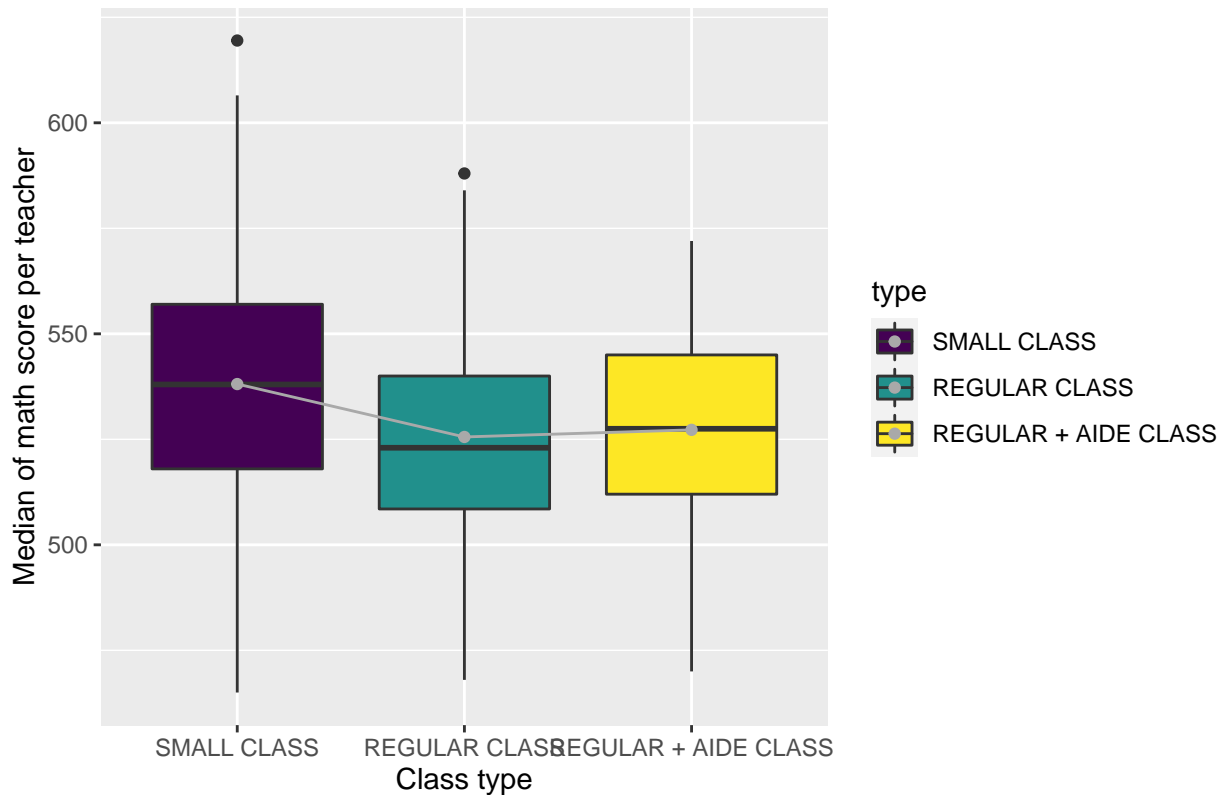Figure1: Boxplot of math for three class types

Figure2: Boxplot of math for all school in increading order
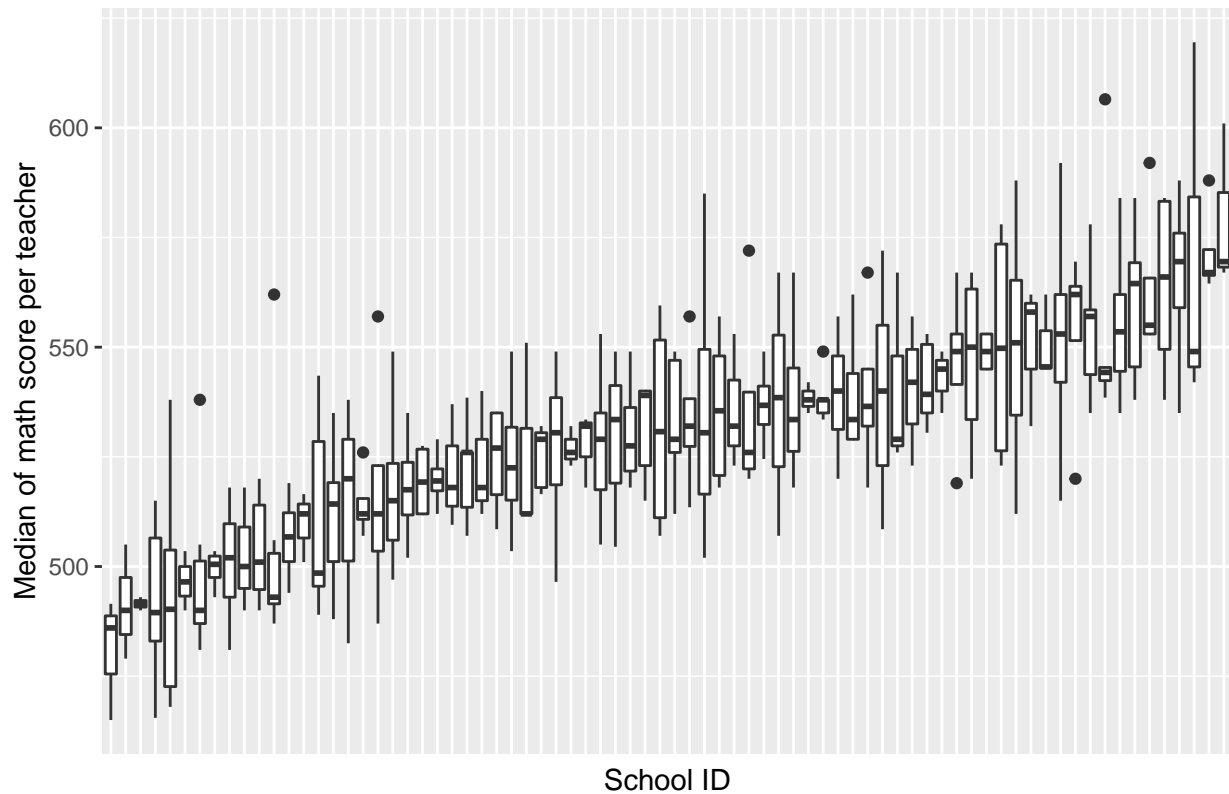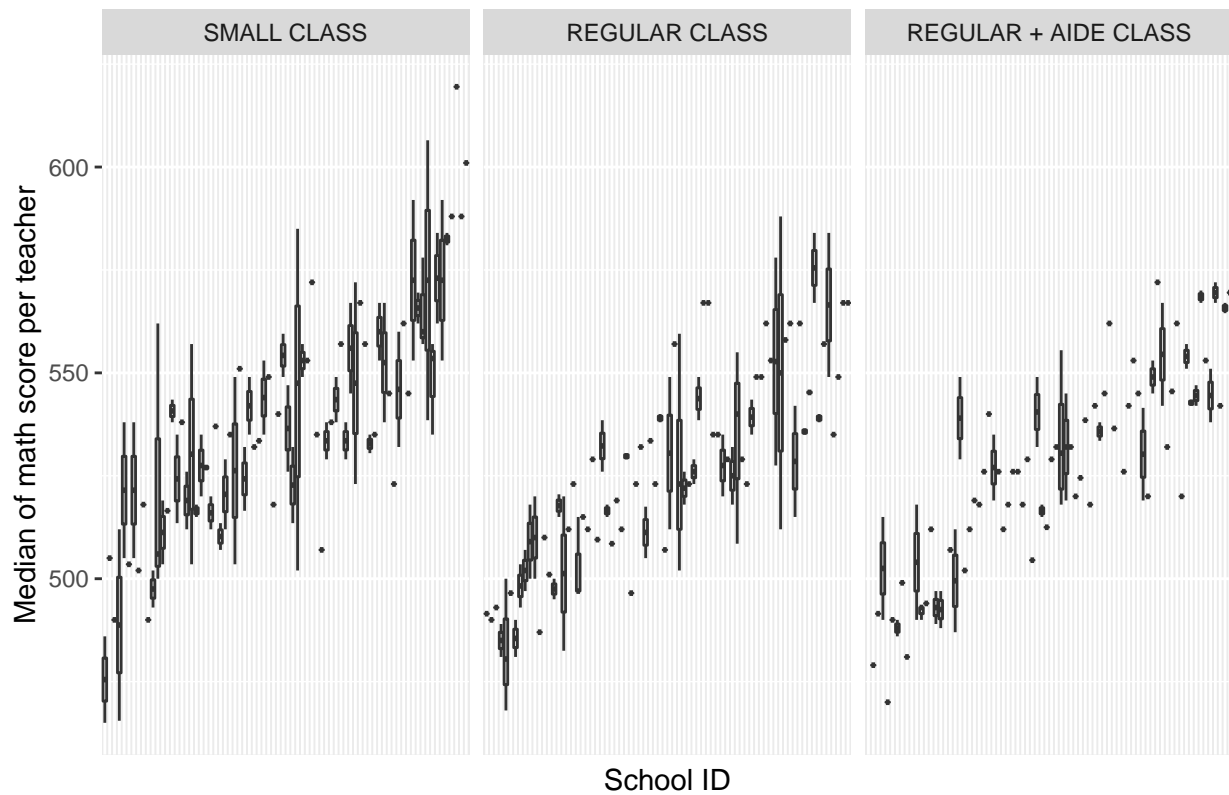

Figure3: Boxplot of math for all school by class type

As for different school, from figure 2 we can see math scores varies among schools, lower ones have math value around 500 while higher ones more than 550, indicating schools might have an association with math scores. Nevertheless, this dispersion was also observed for all three class types [see figure 3].

Inferential analysis

Our question of interest is the relationship between math score and class type. A one-way ANOVA model would be able to answer the question. However, from figure 2, we noticed that schools also influence math scores. Thus, a two-way ANOVA model, which considers schools, would be more suitable than a one-way ANOVA model.

Two-way ANOVA model

$Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + \epsilon_{ijk}$, where $\sum_i \alpha_i = \sum_j \beta_j = 0$, $\epsilon \stackrel{iid}{\sim} N(0, \sigma^2)$.

Class type index $i = 1, 2, 3$, school id indicator $j = 1, 2, \ldots, 76$. α_i denotes the effect of class type and β_j denotes the effect of school, and k denotes teacher/class in each school of each class type. Notice: k is unequal in different cells.

Hypothesis testing

- Full model $Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$ vs. Reduced model $Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + \epsilon_{ijk}$.

Based on the Likelihood ratio test (p-value=0.3), the reduced model is more suitable, thus the interaction term was not included in the model. In other words, the interaction between class type and school id does not significantly affect a student's math score.

- In order to test the primary question of interest, Two hypothesis testing were conducted (at level $\alpha = 0.05$): $H_0 : \alpha_1 = \alpha_2 = \alpha_3 = 0$ vs. H_1 : not all α_i are zero.

(1) Full model $Y_{ijk} = \mu_{..} + \alpha_i + \epsilon_{ijk}$ vs. Reduced model $Y_{ijk} = \mu_{..} + \epsilon_{ijk}$.

F test (p-value=0.0004) rejects the null hypothesis at level $\alpha = 0.05$, which means the class type has a significant association with math scores.

(2) Full model $Y_{ijk} = \mu_{..} + \alpha_i + \beta_j + \epsilon_{ijk}$ vs. Reduced model $Y_{ijk} = \mu_{..} + \beta_j + \epsilon_{ijk}$.

Based on the Likelihood ratio test (p-value=$2.7 * 10^{-9}$), the full model is more appropriate than the reduced model.

- For different class types, the 95% family-wise confidence level Tukey multiple comparisons of means for different class type shows a significant difference between the small class and the other two classes (regular and regular+aid class). See table 3 for details.

table 3: 95% family-wise confidence level Tukey multiple comparisons of means for different class type:

	Difference	Lower	Higher	P-value
Regular - Small	-12.5	-17.9	-7.2	0.0000003*
Regular (aid) - Small	-10.9	-16.4	-5.3	0.0000221*
Regular(aid) - Regular	1.7	-4.0	7.4	0.7650696

As for the difference between different schools, some comparisons are significant (e.g., school id 161183-112038,

p-value=0.0000369). At level $\alpha = 0.05$, math scores are significantly different between some schools.

Results

ANOVA table and estiamted coeffient are shown below:

```
##                Df Sum Sq Mean Sq F value   Pr(>F)
## type            2  11014    5507  17.668 6.36e-08 ***
## sid            75 149370    1992   6.389  < 2e-16 ***
## Residuals     261  81354     312
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## (Intercept)      type.L      type.Q
##  491.333333   -9.254795    4.712475
```

From the ANOVA table, we can see both class type and school id have significant effect on math scores.

Estimates for difference between the means of each class type and small type is -9.3 and 4.7 for regular class type.

Estimates for difference between the means of each school and school 130085 is 58.8.

Sensitivity analysis

Model assumptions diagnostics

- Homogeneity of variance: The plot of residuals vs. Fitted value seems to have an almost similar expansion of residuals. The red line is almost identical to the dashed line, which means the equal variance assumption holds.

- Independence of observations: Based on the experiment's design, teachers were randomly assigned to each class, which indicates that math scores of different classes are mutually independent. Thus, the independence assumption holds.

- Normally-distributed dependent variable: From the QQ plot, we can see a slight deviation from the normal distribution for two tails; thus, it is reasonable to suspect this assumption's slight violation.

- Outliers: Based on the Cook's distance, it is still safe to say we do not have outliers.
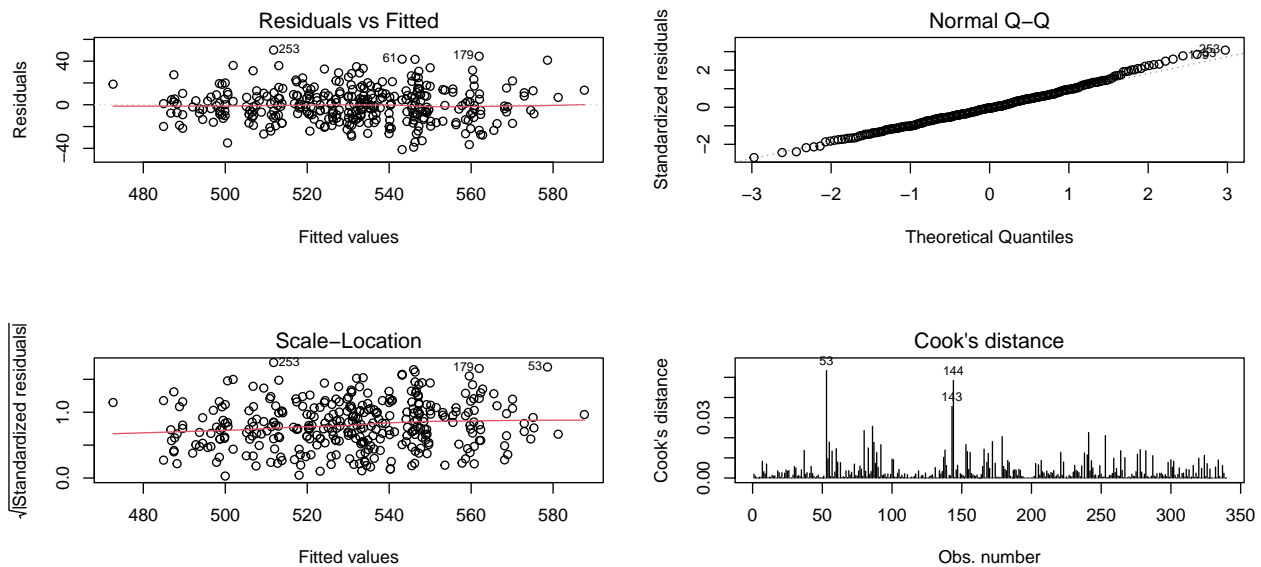
Alternative methods

When using mean math scores of students in a class as a measure score for each teacher, though the magnitude of estimated effects is slightly different from using the median, the conclusion is the same: class type has a significant effect on math score.

```
##               Df Sum Sq Mean Sq F value   Pr(>F)
## type           2  11617    5809  20.991 3.52e-09 ***
## sid           75 136833    1824   6.593  < 2e-16 ***
## Residuals    261  72225     277
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## (Intercept)      type.L      type.Q
##   494.245189   -8.058664    6.263783
```

Causal interpretation

- Ignorability of Treatment Assignment: Since students and teachers were randomly assigned to each class, this assumption holds.

- Stable Unit Treatment Value(SUTVA): Though teachers were randomly assigned to each class, it is unknown whether a school has decided the number of each type before randomization. In other words, the class type assignment for one teacher might affect another teacher's potential outcomes. For example, if a school only has one small class and one regular class, as long as one teacher is assigned to one class, the other teacher's class type would be decided.) Thus, assumption SUTVA might not hold.

- one version of each treatment possible for each unit: Since once teachers were assigned to a class, the class type would not change, it seems this assumption holds. However, sometimes a student would change his/her class type during the experiment, which might affect the outcome of interest (median math score in a class). Hence, this assumption might not hold.

In a nutshell, we can only draw an association between class type and math scores and cannot draw a causal conclusion between them (ref [1]).

Discussion

We analyzed first-grade students' math scores in different class types while treating one class/teacher as a unit. After descriptive data analysis and taking the impact of the school on math score into consideration, a two-way ANOVA model was utilized to investigate the association between class type and math score. Model assumptions were examined, and no apparent violation was noticed.

Based on the two-way ANOVA model results, we found a significant association between class type and math score. Notably, on average, students in small classes have higher math scores than regular classes (with or without aid).

As an education reform goal, class size reduction (CSR) has shown a variety of benefits for students and teachers. Our findings verify the significant effect of small classes on math studying for first-grade students in the short term.

However, since causal inference assumptions are violated, our conclusion fails to show a causal relationship between class type and math scores.

Hence, I call for more rigorous experiments to investigate whether small class sizes cause students' better performance.

Besides, for the data analysis section, in the further, I would like to see one use a mixed-effect model rather than an ANOVA model to investigate the question of interest.

Reference

[1] Lam, Patrick Kenneth. 2013. Estimating Individual Causal Effects. Doctoral dissertation, Harvard University.