Bitchin Practice analictics Report on Project Star

Jared Schultz

2/4/2023

Abstract

This report studies a portion of the STAR Dataset. This dataset studies how class size may effect learning out comes of children form kindergartner to 3rd grade. This report will only use data involving first graders. The data is centered on three types of class: small, regular and regular with an aide. The study used a stratified or block design at the school level. This report cleans the data by removing rows with missing values. To obtain our response variable the data is merged so we will have each unit be a teacher. Exploratory data analysis technique show that backing to the claim that there exists a difference and that the small class type seems to have the highest mean. This report tries to answer two question being is there any differences in math scaled scores in 1st grade across class type and is there a class type associated with the highest score. This report uses a two way anova test and Tukey test to answer these main question. the results fins that there are differences among class types and the small class type is associated with the the highest 1st grade scaled math scores. The sensitivity result of our model finds that we break our normality assumption with all other holding. This report along with others find that small class room sizes produce better results among children and we recommend switching classroom style to teach more efficiently.

Introduction

Questions of interest:

1. Are there any differences in math scaled scores in 1st grade across class types?
2. Which class type is associated with the highest math scaled scores in 1st grade?

Questions of further research studied in ref(Achilles,2012):

- What are the effects of small class size in early grades [short term - long term]?
- What are the effects of teachers education/experience in relation to academic success of students?

Motivation of analysis: Allow schools to achieve better results with teaching styles and sizes.This may also allow for cost effective teaching methods by making classroom smaller it cost the school less money with obtaining high results. How a teacher might preform may or may not be an indicator of education level in which case schools could either look for less or more qualified candidates.

Potential impacts of results: To see how class size effects the students academic outcome. If shown successful then schools could focus changes in policy around class size. This would lead to an optimal pupil-teacher ratio that gives the schools the best results. This may save the school money by reducing the number of students that it can take for the upcoming years and hire the correct amount of teachers for desired lass size.

Background

Explain source of the data, target population, sampling mech. and other variables in data set:

The source of this data comes from the AER R-library of which it was obtained from the project Star website. The Target population of this study would be policy makers on school boards or at a high level.

The description of the data from the AER library is,

"Project STAR (Student/Teacher Achievement Ratio) was a four-year longitudinal class-size study funded by the Tennessee General Assembly and conducted in the late 1980s by the State Department of Education. Over 7,000 students in 79 schools were randomly assigned into one of three interventions: small class (13 to 17 students per teacher), regular class (22 to 25 students per teacher), and regular-with-aide class (22 to 25 students with a full-time teacher's aide). Classroom teachers were also randomly assigned to the classes they would teach. The interventions were initiated as the students entered school in kindergarten and continued through third grade." (Kleiber, 2008)

Further information provided from Achilles, 2012 include:

" STAR school had at least one of each class type (small, regular, and regular with aide) in the robust and parsimonious within-school design. The class arrangement was maintained throughout the day, all school year long. There was no intervention other than class size and a full-time teacher aide provided to assist classes. The large sample size and random assignment overcame threats to validity... Cognitive outcomes were measured by norm-referenced tests and criterion-referenced tests aligned to state standards. Non-cognitive outcomes were also assessed."

Descriptive analysis

To start working with our data it should first be noted that for this project we will only be looking at 6 variables:

math1 : total math scaled score in 1st grade.

school1 : factor indicating school type in 1st grade: "inner-city", "suburban", "rural" or "urban".

degree1 : factor indicating highest degree of 1st grade teacher: "bachelor", "master", "specialist", or "phd".

experience1: years of teacher's total teaching experience in 1st grade.

tethnicity1: factor indicating teacher's ethnicity in 1st grade with levels "cauc" (Caucasian) or "afam" (African-American).

schoolid1 : factor indicating school ID in 1st grade.

star1 : factor indicating small class (13 to 17 students per teacher), regular class (22 to 25 students per teacher), and regular-with-aide class (22 to 25 students with a full-time teacher's aide)

We will clean the data and delete all rows containing NA values. After cleaning this reduces the dimension from 11,598 x 6 to 6558 x 6, omitting 5,040 rows of data. To make sure that this does not alter the our findings we will compare the summary statistics of the original data of math1 scores to our reduced set. The tables below will show the summary statistics of the math1 variable of the original data set and cleaned.

First grade math scores using Original Data

	regular (N = 2584)	small (N = 1925)	regular+aide (N = 2320)
First Grade Math Scores			
min	408	425	404
max	676	676	676
mean	525.27	538.68	529.63
sd	41.65	44.1	42.87

First grade math scores using cleaned data

	regular (N = 2492)	small (N = 1851)	regular+aide (N = 2215)
First Grade Math Scores			
min	408	425	404
max	676	676	676
mean	525.35	538.6	529.68
sd	41.68	44.21	42.82

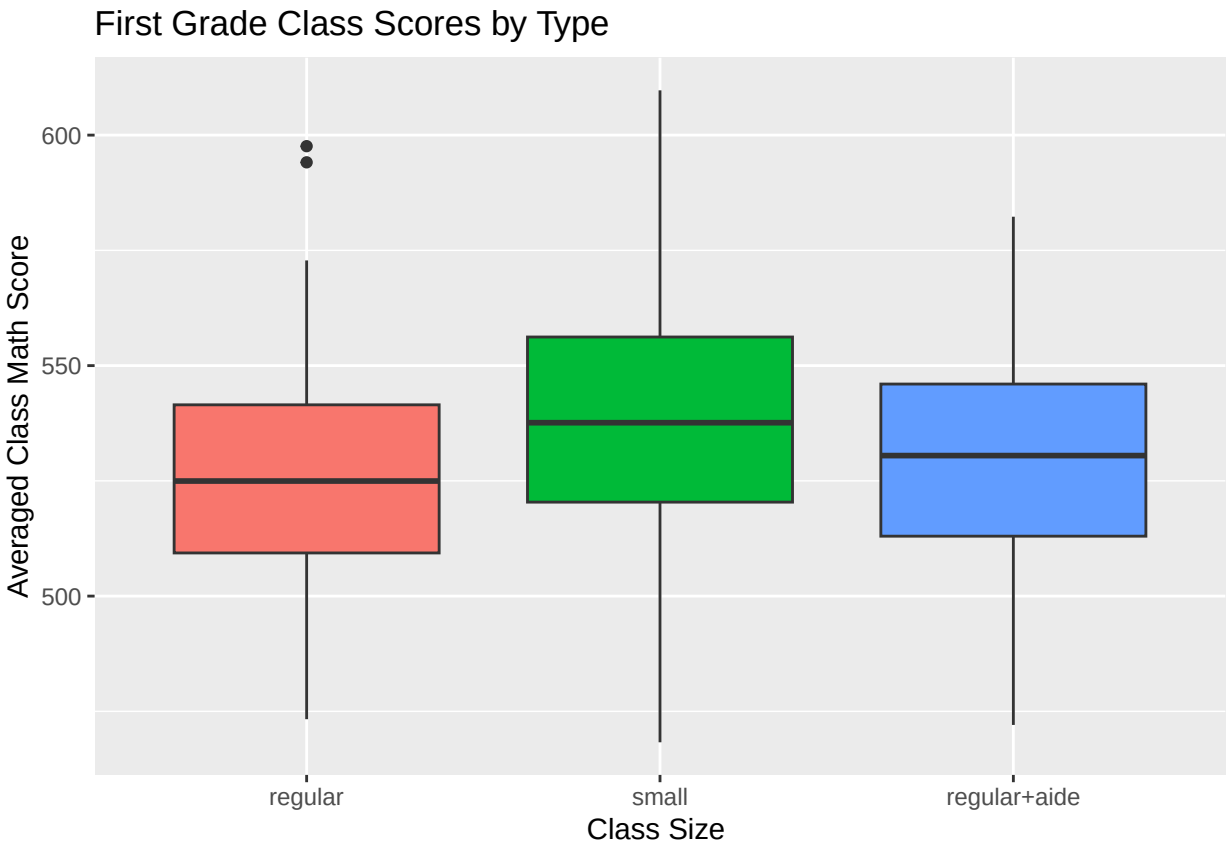
We can see from the tables above that the cleaning procedure did not change the the distribution of math1 scores. This can be seen clearly when looking at the range of each as well as the mean and standard deviation where they differ by less than 0.0025%. The benifit of this dataset is it will allow for a balenced design later in the study.

In the table below we explore the aggregated students first grade math score by each unique teacher. Thus the scores can be thought of the mean of the students in first grade taught by that teacher or averaged class math score where each teacher is now a unit. Now we can use this data to answer our questions about class-types.

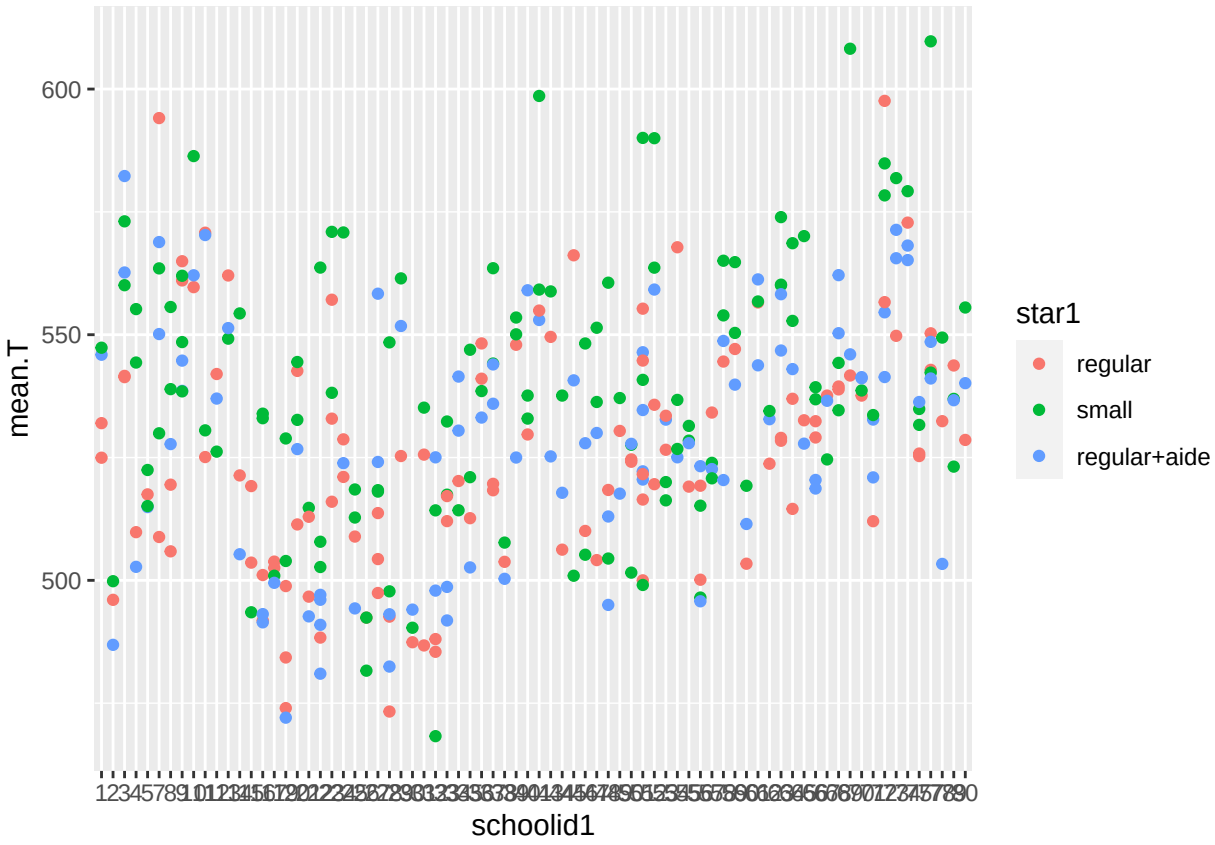
Mean score of students taught by each unique teacher

	regular (N = 115)	small (N = 123)	regular+aide (N = 101)
Mean First grade math score per teacher			
min	473.3	468.29	472.05
max	597.6	609.71	582.3
mean	525.49	538.9	529.06
sd	23.89	26.68	24.47

Right away looking at the means of each class type it is easy to see that small class type has the largest mean. We can also explore these summary statistics through box plots below.



Now that we have looked at class type we should also look at the averaged class scores vs school id. WE WILL ALSO REPORT SUMMARY STATISTICS ?



Inferential analysis

We will use the additive factor effects two way model with constants:

$$Y_{ijk} = \mu.. + \alpha_i + \beta_j \epsilon_{ijk}, \ k = 1, \ldots, n, \ j = 1, \ldots, b, i = 1, \ldots, a, \tag{1}$$

where $\{\epsilon_{ijk}\}$ are i.i.d. $N(0, \sigma^2)$ and

$$\sum_i \alpha_i = \sum_j \beta_j = 0 \tag{2}$$

It should also be noted that: $\alpha_i = \mu_i. - \mu..$ and $\beta_j = \mu_{.j} - \mu..$ where $\mu..$ is the population mean and $\mu_{.j}$ is the overall mean for factor A taking the i'th level. The same definition holds for $\mu_{.j}$

The response variable in the model is the mean.T variable which is the averaged first grade class score.The other two factor effect terms are Star1 and schoolid1. Star1 has 3 facot levels and schoolid has 80 factor levels. This is also a balenced model since we cleaned our data.

Explain why interaction terms are not in the model: We do not need interaction terms since our two factor variables are independent given how the test was conducted. The students were randomized into different classes at the class level and since our factor is a school indicator class size and school will be independent. I believe this is called stratification of the design experiment in regards to school CHECK THIS

To answer our first question we will consider a test on α_i which in our case is the difference of the mean of class size i (i=1,2,3) to the population mean. We will test at a type one error rate of $\alpha = 0.01$.

$H_0 : \alpha_i = 0 \quad \forall i = 1, 2, 3 \quad v.s \quad H_a$: not all α_i equal zero

$F^* = 20.57$ and p-value $=5.05e^{-16}$

Since our found p-value is less than $\alpha \quad = \quad 0.01$ we will reject the null hypothesis that all alphas are equal. In other words we would reject the claim that the mean of first grade scores are the same per each class type.

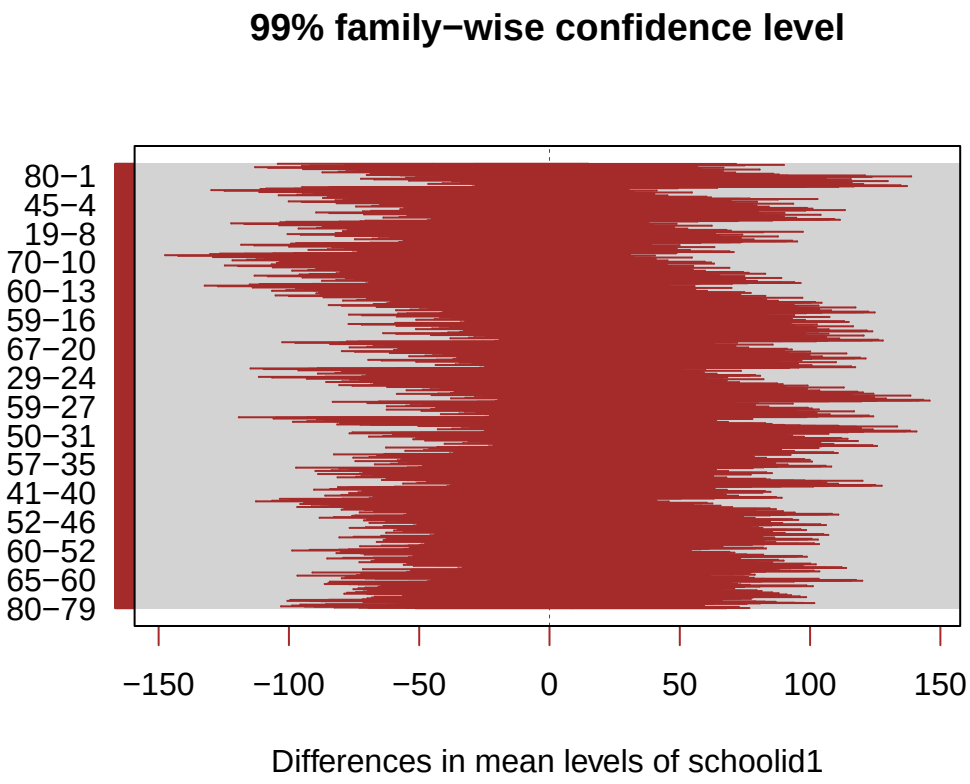
Doing the same test for schoolid1 would end in the results that we would reject the claim that mean of first grade scores are the same per school. We can see these results and information from testing from the table below with our fitted model summary statistics.

```
##              Df Sum Sq Mean Sq F value   Pr(>F)
## star1         2  11484    5742  20.574 5.05e-09 ***
## schoolid1    75 138921    1852   6.637  < 2e-16 ***
## Residuals   261  72844     279
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Now to answer our secondary question of interest: which class type is associated with the highest math scaled scores in 1st grade. To answer this question we will preform simultaneous inference using a Turkey test at a type one error rate of 1 percent. Our model hypothesis test will be of the following form:

H_0 : For all i there exists a j such that $\mu_i \leq \mu_j \quad v.s \quad H_a$: There exists an i such that $\mu_i > \mu_j \quad \forall i \neq j$

The simultaneous testing for both factors are plotted graphically below:



Inference from plot 1 : Plot 1 shows the comparison of difference of the three factor types. When our confidence interval contains zero it tells us that we can not reject our null hypotheses at that level combination. That is to say that we cannot reject that the means are different at our stated type one error level.

small - regular : Confidence interval does not contain 0, thus reject the null hypothesis that means of small and regular class size are the same. Note CI is positive

regular+aide - regular : Confidence interval does contain 0, thus do not reject the null hypothesis that means of regular+aide and regular class size are the same.

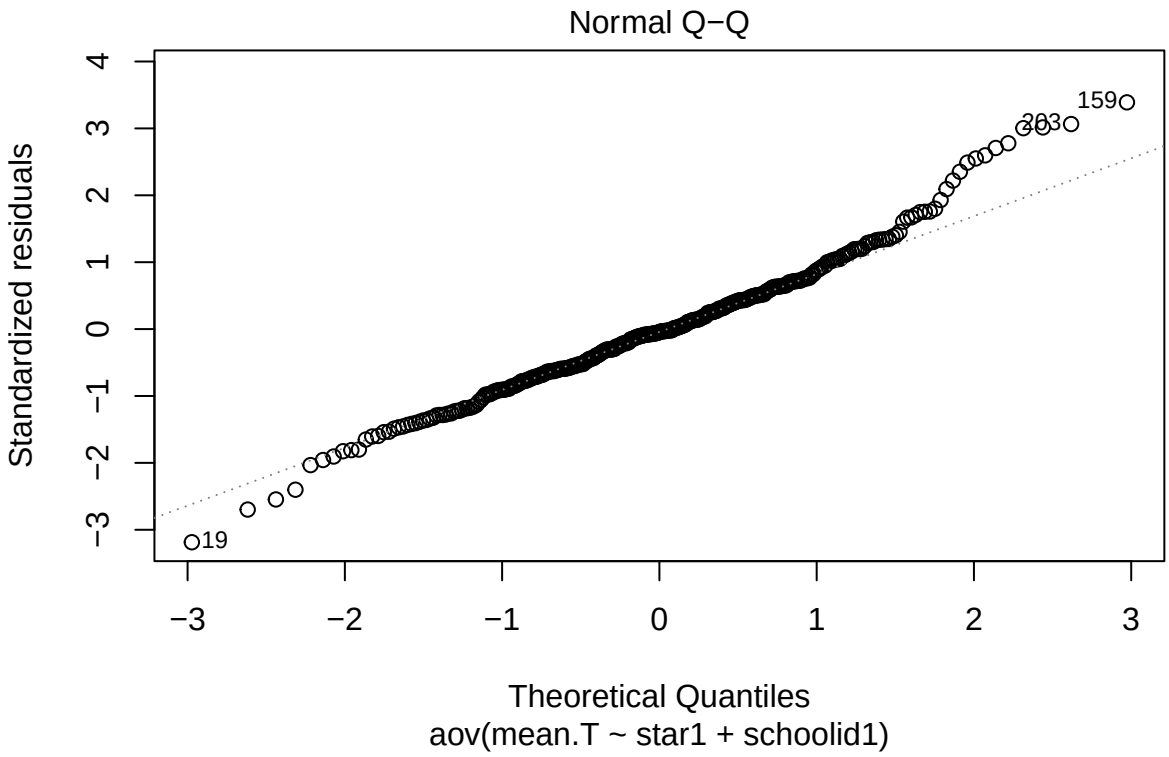
regular+aide - small : Confidence interval does not contain 0, thus reject the null hypothesis that means of small and regular+aide class size are the same. Note CI is negative

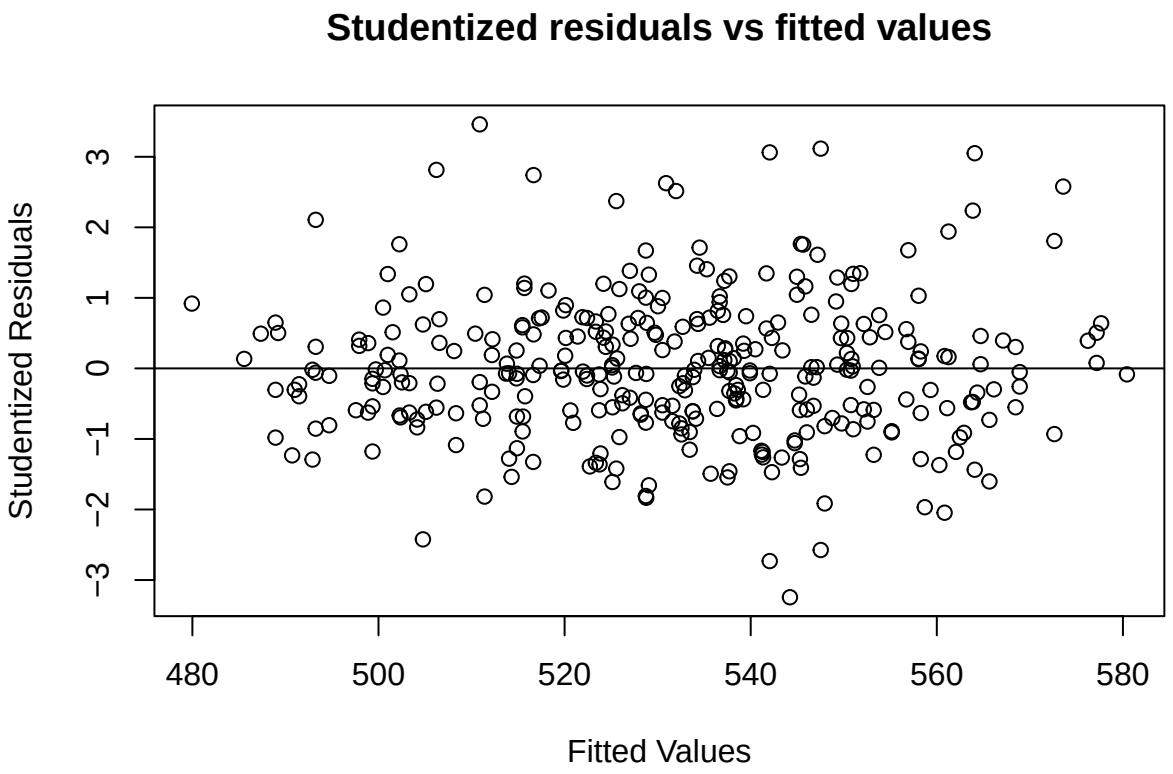
Results : Since our test for small - regular and regular+aid - small both showed differences and the confidences intervals were positive and negative respectively. Thus we can conclude that the small class size has the highest first grade mean scores

Inference from plot 2 : inference from plot 2 is not possible since there are so many confidence intervals. There are 2850 confidence intervals for this factor and it can be found that only 181 Confidence intervals do not contain 0 at a 1% type one error level. Thus a mass for a majority there exists no difference of means between schools.

Sensitivity analysis

Residual analysis : From our model above we make several assumptions on our error terms that can be verified through residual analysis. In the plots below we will explore if our assumptions hold up.


Normal Q−Q
Standardized residuals
Theoretical Quantiles
aov(mean.T ~ star1 + schoolid1)
19
159
203



Assumptions on Error terms **Normal Error Distribution :** From the Normal Q-Q plot we can see that our errors have right heavy tailed distribution. Thus the normal error distribution assumption does not hold.

Non-Constant Variance : From the Studentized residuals vs fitted values there is very slight evidence of non-constant variance thus from the plot we will not assume this assumption holds. We will further verify it with a proper test at $\alpha = 0.05$. The test we will use will be levene's test.

First allow d_{ij} to be the absolute value of model1's residuals. The testing claims will be

H_0 : All $E[d_{ij}.]$ are the same $\quad v.s \quad$ H_a : Not all $E[d_{ij}.]$ are the same

```
## Levene's Test for Homogeneity of Variance (center = "mean")
##        Df F value    Pr(>F)
## group 223  4.3554 2.259e-16 ***
##       115
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The results from the table above show that we would reject the null-hypothesis and thus we would state that the non-constant variance assumption holds.

Independence of error terms: not covered

Outliers : From the Studentized residuals vs fitted values there is no clear evidence of outliers.

Missing variables : not covered

Discussion

In this project we explore a cleaned and manipulated version of the STAR data set. The Study wishes to answer two main questions which are if there exists a difference and if there is a class type that is associated with the highest mean. Preforming EDA using boxplots shows that there appears to be a difference of means within class types and all the variances of each class type appear the same.[ADD SOMETHING ABOUT SCHOOLID].To answer our intiial questions we use a two way factor effect form anova model with out interaction term. We negate this term since the study used a block design so there interaction from different schools and class type.Our findings show that there is a difference in means and that the small class type has the highest averaged class score. The model does lack normality of error terms. From this analysis it is recomded to compare analysis of other grades and use other observation variables to confirm these results. If the results are validated then it is recomended to switch class styles to a small to better education and save money.

Findings from Aftunion, 2014 show that other projects such as project STAR have been studied and found similar results. A 1999/2010 update and follow up on project star's student indicated the following,

"1999 Update Researchers reported that the effects of small class sizes in grades K-3 lasted all the way through high school. Students from small classes are:

- More likely to graduate from high school on schedule and less likely to drop out;
- More likely to have enrolled in honors classes and to graduate in the top 10 percent of their class; and
- More likely to take SAT or ACT exams, indicating that they plan to go on to college. Further, the black-white achievement gap is reduced by 56 percent for black students who began school in small classes.

Researchers also found that students in small classes in grades K-3 were between six and 13 months ahead of their regular-class peers in math, reading, and science in each of grades 4, 6, and 8. Researchers reported that for the benefits to be sustained through later grades, at least three years in a small class are necessary. In addition, the benefits of having been in a small class in the primary years increase from grade to grade.

2010 Update Researchers from Amherst College attempted to differentiate which kinds of schools and students benefit the most from smaller class sizes. Findings:

The researchers found that all students in high-poverty schools benefit from reduced class sizes, with high-achieving students benefiting the most. Researchers also found no evidence of changes in teacher behavior or pedagogical practice when class sizes are reduced, indicating that "student differences account for the positive relationship between achievement and the benefit of smaller classes.""

Ref(Aftunion,2014)

Acknowledgement

Reference

Achilles, C. M. (2012, September 30). Class-size policy: The Star Experiment and related class-size studies. NCPEA policy brief. volume 1, Number 2. NCPEA Publications. https://eric.ed.gov/?id=ED540485

Aftunion. (2014, September 16). Supporting research - class size. American Federation of Teachers. Retrieved February 11, 2023, from https://www.aft.org/education/publications/school-improvement/supporting-research-class-size

Kleiber,C & Zeileis,A. (2008). Star: Project star: Student-teacher achievement ratio. RDocumentation. Springer-Verlag. https://www.rdocumentation.org/packages/AER/versions/1.2-9/topics/STAR

Session info

Report information of your `R` session for reproducibility.

```
sessionInfo()
```

```
## R version 4.2.2 (2022-10-31 ucrt)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 19044)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.utf8
## [2] LC_CTYPE=English_United States.utf8
## [3] LC_MONETARY=English_United States.utf8
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.utf8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
##  [1] gridExtra_2.3  ggplot2_3.4.0  dplyr_1.0.10   qwraps2_0.5.2  AER_1.2-10
##  [6] survival_3.4-0 sandwich_3.0-2 lmtest_0.9-40  zoo_1.8-11     car_3.1-0
## [11] carData_3.0-5
##
## loaded via a namespace (and not attached):
##  [1] tidyselect_1.1.2 xfun_0.33        purrr_0.3.4      splines_4.2.2
##  [5] lattice_0.20-45  colorspace_2.0-3 vctrs_0.5.0      generics_0.1.3
##  [9] htmltools_0.5.3  yaml_2.3.5       utf8_1.2.2       rlang_1.0.6
## [13] pillar_1.8.1     glue_1.6.2       withr_2.5.0      DBI_1.1.3
## [17] lifecycle_1.0.3  stringr_1.4.1    munsell_0.5.0    gtable_0.3.1
## [21] evaluate_0.16    labeling_0.4.2   knitr_1.40       fastmap_1.1.0
## [25] fansi_1.0.3      highr_0.9        Rcpp_1.0.9       scales_1.2.1
## [29] abind_1.4-5      farver_2.1.1     digest_0.6.29    stringi_1.7.8
## [33] grid_4.2.2       cli_3.4.1        tools_4.2.2      magrittr_2.0.3
## [37] tibble_3.1.8     Formula_1.2-4    pkgconfig_2.0.3  ellipsis_0.3.2
## [41] Matrix_1.5-1     assertthat_0.2.1 rmarkdown_2.16   rstudioapi_0.14
## [45] R6_2.5.1         compiler_4.2.2
```