

Project 2: Project STAR II

Note: The report might be shown to the class as examples if it is done well very or very badly, with your identity redacted.

This Document is the project 2 of Team 7 in STA 207, Winter 2020.

Group Partners:

1. Bingdao Chen bdchen@ucdavis.edu contribution: explore data and test causal effect
2. Yahui Li yhuli@ucdavis.edu contribution: write down and hold one-way ANOVA model
3. Zihan Wang zihwang@ucdavis.edu contribution: test on difference
4. Jian Shi jnshi@ucdavis.edu contribution: model diagnose

The repository in Github is on <https://github.com/yhli097/STA207Project2.git>

Introduction

Background

Tennessee Student/Teacher Achievement Ratio study (Project STAR) was conducted in the late 1980s to evaluate the effect of class size on test scores. The study randomly assigned students to small classes, regular classes, and regular classes with a teacher's aide. In order to randomize properly, schools were enrolled only if they had enough studybody to have at least one class of each type. Once the schools were enrolled, students were randomly assigned to the three types of classes, and one teacher was randomly assigned to one class.

The primary scientific question of interest is whether there is a treatment effect of class types on math scaled scores across teachers in 1st grade, with the school indicator as the other factor. In this study, we implement exploratory data analysis, two-way anova model, model diagnostics, hypothesis testing. In the end, we conclude our findings and discuss any causal statements that could possibly be made based on our analysis and assumptions, and compare the results between Project 2 and Project 1.

Statistical questions of interest

To answer the primary scientific question of interest, we propose to fit a two-factors anova model with the quantitative test scores grouped by teacher as the response variable and the class type as well as school as the predictor variable. We will then run our model diagnostics to see if the assumptions of the model hold and test whether or not there is a treatment effect on the test scores.

Analysis Plan

Population and study design

According to the description of the dataset, over 7,000 students in 76 schools were randomly assigned into one of three different treatments: small class (13 to 17 students per teacher), regular class (22 to 25 students per teacher), and regular-with-aid class (22 to 25 students with a full-time teacher's aide). Classroom teachers were also randomly assigned to the classes they would teach. We will only focus on the data about the 1st grade students as study targets[1].

Descriptive Analysis

First of all, we will pick up all valuable variables we are interested in. To deal with missing value, we will draw plots to show the percentage of missing value for these variables, then decide whether drop the missing value or not.

We will then explore the math scores grouped by teacher ID. To choose the summary information on math scores in each unit, we will plot different distributions to find a statistic which is the most informative as well as normal-like.

In the end, we study the treatment of class type as well as school ID. Based on different class type as well as school ID, we will confirm whether there is any difference on math scores across teachers not.

Two-way Anova Model

In order to investigate whether there are effects of class types on math scaled scores in 1st grade and minimize the effects of systematic error, randomized block design is used, and we will create blocks by different schools. Within blocks, students and teachers are randomly assigned to class in different types. Therefore, it is possible to assess the effect from different levels of class types without worry about the influence from different schools. Although there are two factors in this experiment, we are not interested in the effect from different class type as well as different school. As a result, interaction term in this model should be excluded.

The two-way ANOVA model for a randomized block design is shown as below:

$$Y_{ijk} = \mu + \tau_i + \beta_j + \epsilon_{ijk}, \quad \text{with} \quad \epsilon_{ijk} \stackrel{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2).$$

By notation,

- The index i denotes main factor level. There are 3 levels about class type: **small**, **regular** and **regular+aid**.
- The index j denotes blocking factor level. There are 76 levels about different school ID.
- The index k denotes experimental unit, about mean score of k -th teachers in 1st grade. n_{ij} is the number of experimental units in treatment (i, j) from both factors.
- Y_{ijk} refers to the mean math score for k -th teacher ID in i -th type of class and j -th school ID.
- μ denotes the overall mean value of the mean math scores in 1st grade for all treatments.
- τ_i denotes main effect of level i from the class type factor.
- β_j denotes blocking effect of level j from the school ID factor.
- ϵ_{ijk} denotes random errors.

Table 1: Table 1: basic information of variables

Name	Type	The.number.of.levels
g1classtype	qualitative variable	3
g1schid	qualitative variable	76
g1tchid	qualitative variable	339
g1tmathss	quantitive variable	NA

Moreover, there is a “sum-to-zero” constraint shown in the following :

$$\sum_i \tau_i = \sum_j \beta_j = 0$$

Assumptions of two-way ANOVA model specified as follows:

- the mean of the response variable is influenced additively (if not interaction term) and linearly by the factors;
- The random errors are assumed to be identically and independently distributed from a normal distribution with mean 0 and variance σ^2 .

Model Diagnostics

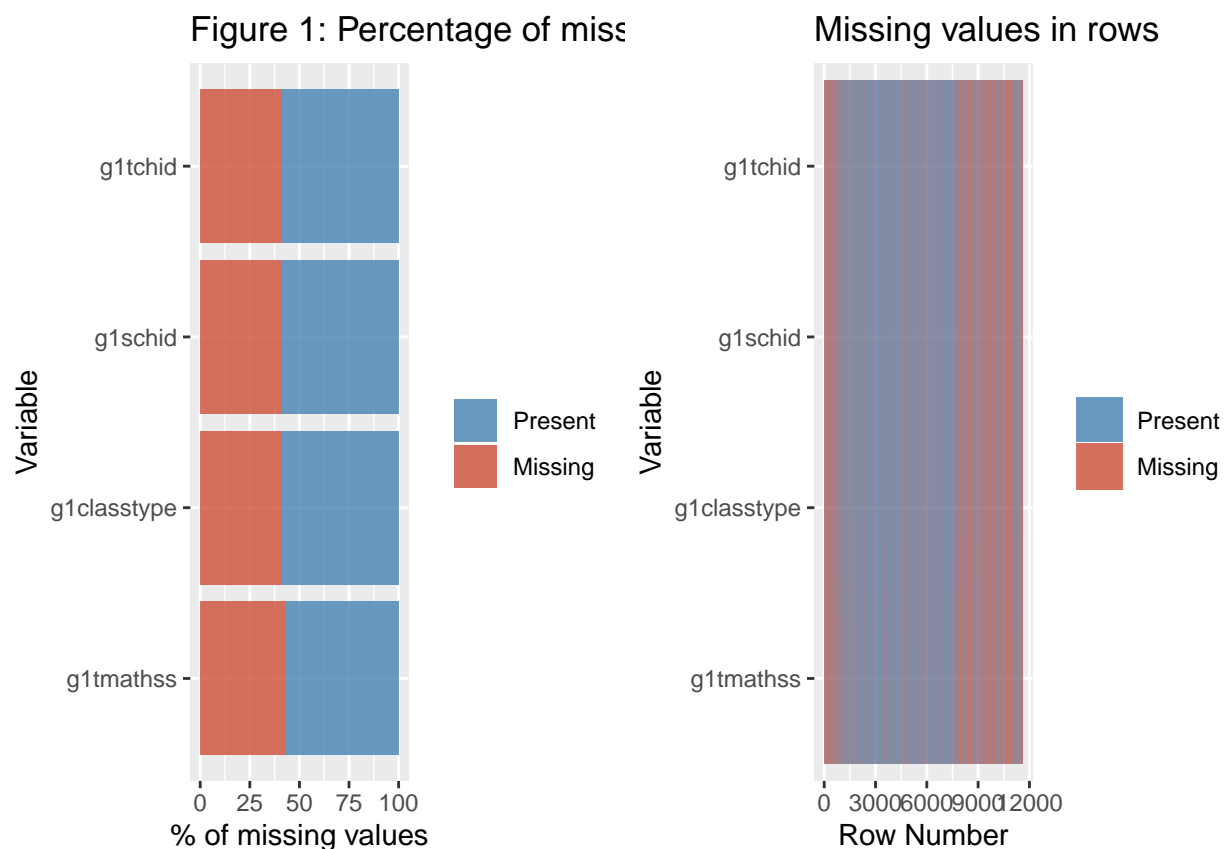
Results

Note: In this template, we present the unpolished tables and figures generated with generic `r` functions, and we skip all interpretation of the analysis results and the sensitivity analysis. You can adapt any functions in this section in your report.

Descriptive Analysis

The dataset we investigate has 11601 rows and 379 cloumns, and we narrow it down to four columns, that is, `g1classtype`, `g1schid`, `g1tchid`, `g1tmathss` which are related with math scaled scores in the 1st grade. `g1classtype` represents the class type of students in 1st grade. `g1schid` represents school’s id. `g1tchid` represents teacher’s id. `g1tmathss` represents math scaled scores of students in 1st grade. Figure1 shows the basic information of four variables.

In Figure 1, the above plot shows the percentage of missing value of four variables are nearly same, and the plot below shows that the location of missing values for three qualitative variables are consistent. It can be explained that because of low grades in kindergarten in STAR project, students quit STAR project and go to another school, or they join STAR project after 1st grade. Because we don’t know how to tackle missing values, considering that we have enough data and believe that it won’t have the significant impact on the results, we drop them directly.



In this project, we are interested in math scaled scores in the 1st grade with teachers as the unit. After verifying each teacher are only employed in specific school and teach one specific class type, we select the mean of math scaled scores of one particular class to judge teachers. Therefore, we have a new dataset which has 339 rows and four columns including g1classtype, g1schid, g1tchid and mean. Figure 2 shows the distribution of the mean of math score by different teacher. It is bell-shaped and roughly closed to normal distribution.

Figure 2: the mean of math score by different teacher

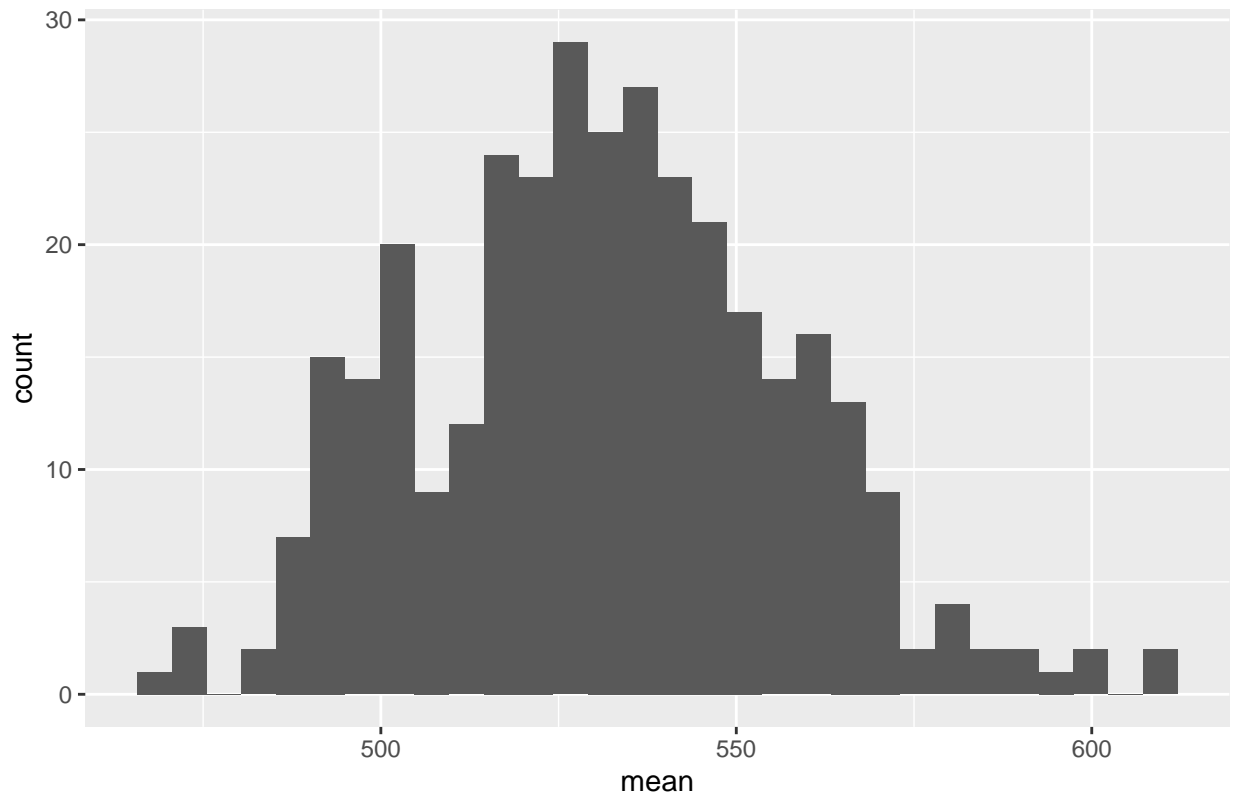


Figure 3 shows the frequency number of class type and school id. Due to too many levels in `glschid`, x axis labels in the plot of `glschid` is omitted for readability. We can find that `g1classtype` data is roughly balanced, and there is one specific school whose frequency number which is much higher than rest of schools. Figure 4 shows that the frequency number of class of different types in each school. Excepts a few extremely cases, data is more balanced than data group by class type.

Figure 3: The frequency bar plots of class type and school id

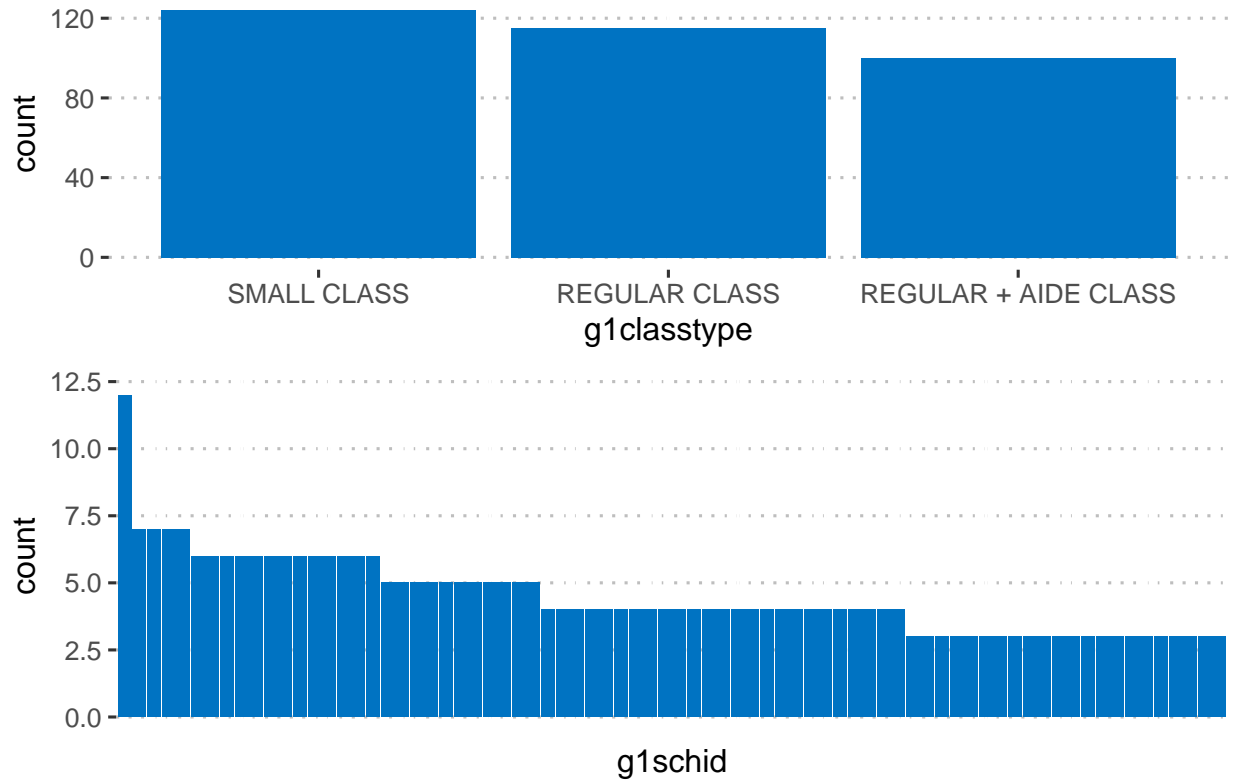
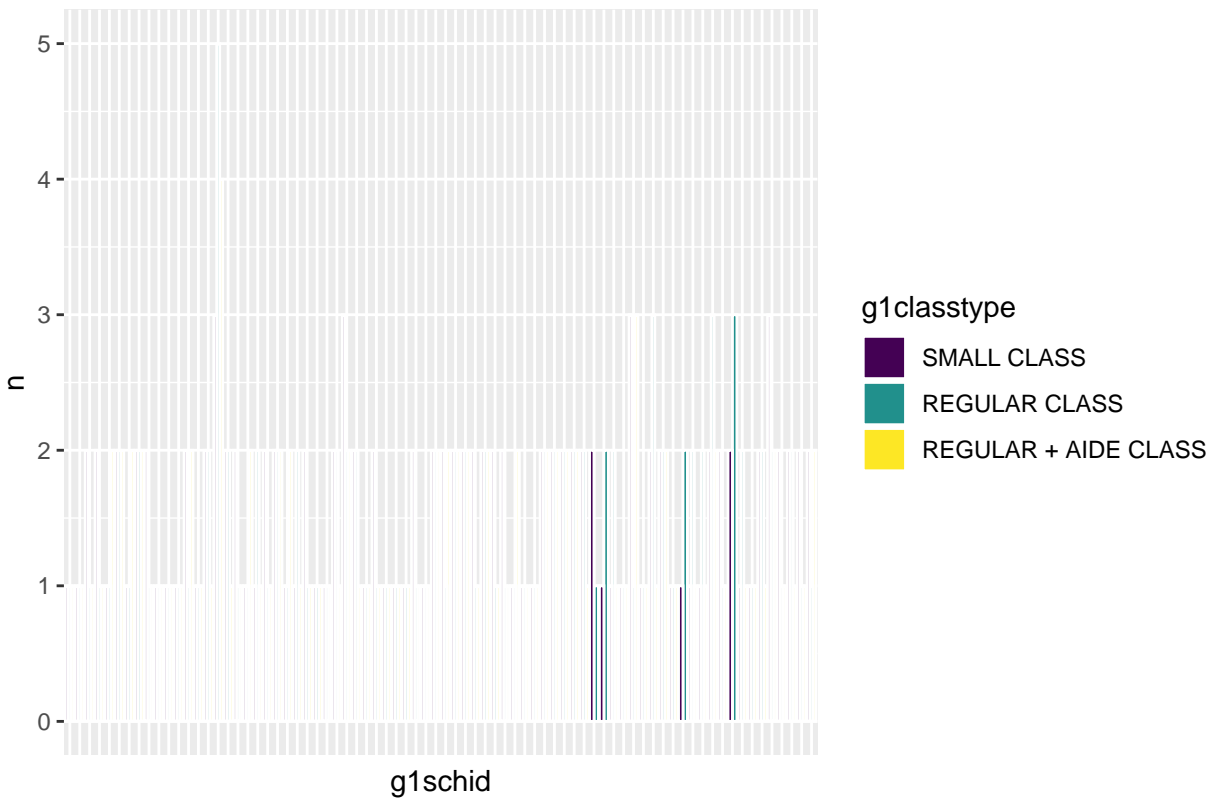
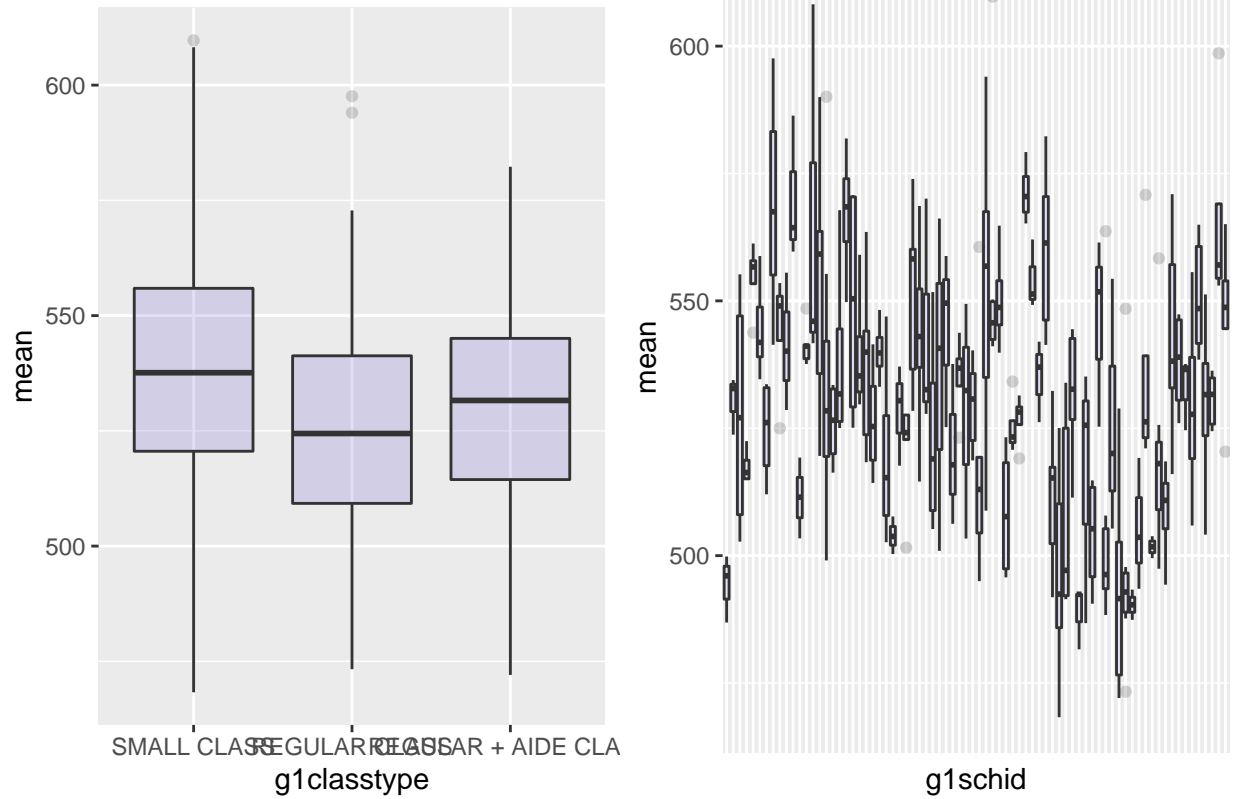


Figure 4: The frequency bar plots of grouped contingency tables



In Figure 5, it can be seen that different class types have effects on mean scores of math. Therein, teachers in small class have the higher mean scores of math and teachers in regular class have the lower mean scores of math in general. Also, different schools have impacts on mean scores of math. Aiming to investigate the class types effect on mean scores of math, it is a good strategy that using schools to creat blocks for eliminating effects between different schools.

Figure 5: The distribution of mean



Two-way ANOVA Model

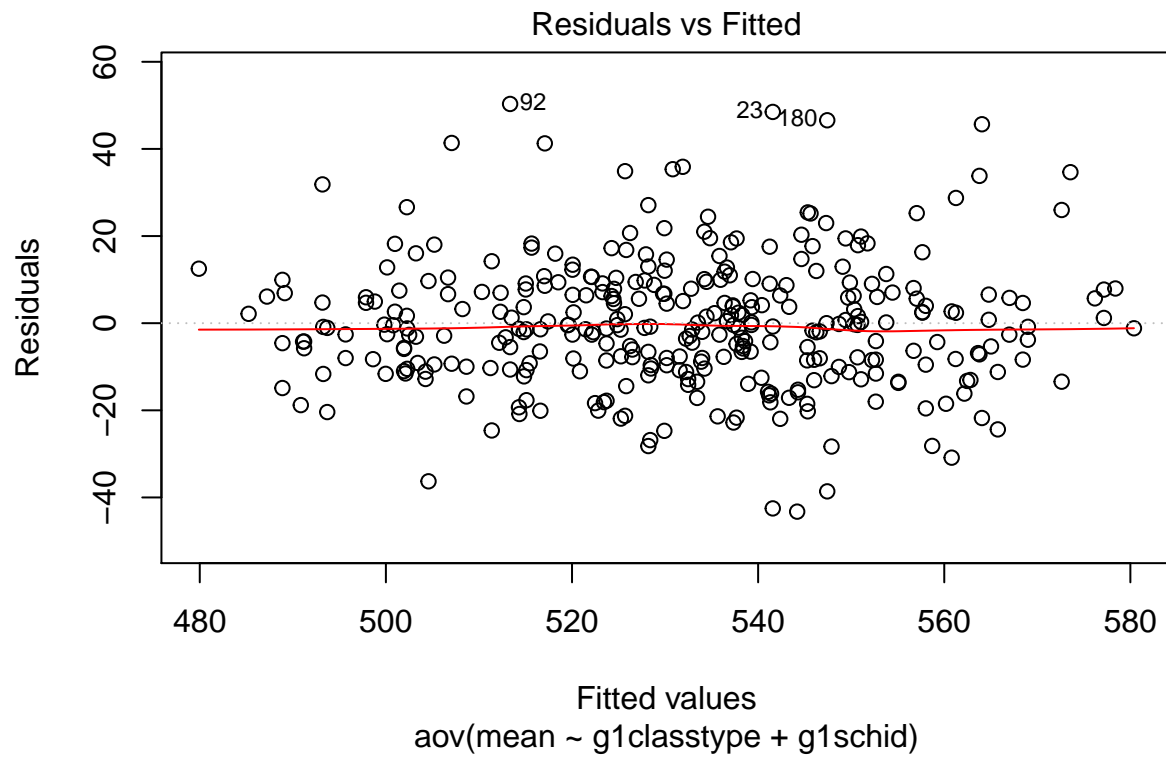
Table . ANOVA Table

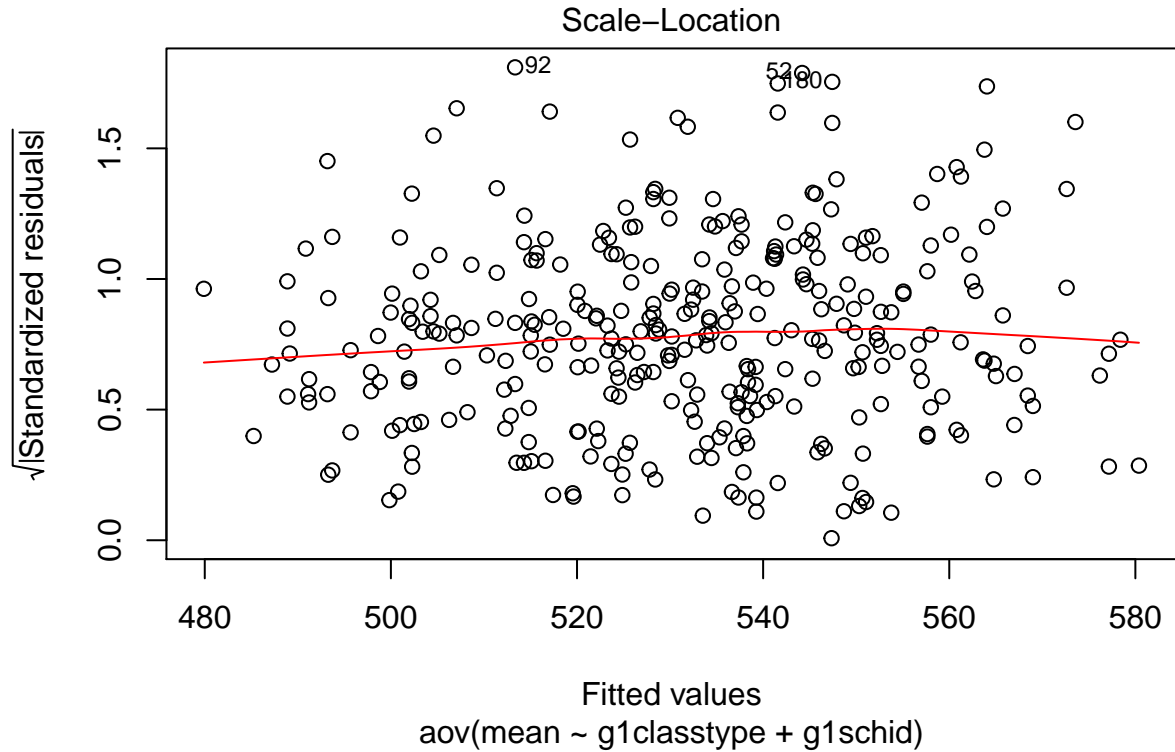
Source of Variation	Degree of Freedom	Sum Square	Mean Square	F value	P value
g1classtype	2	11617	5809	20.991	3.52e-09
g1schid	75	136833	1824	6.593	< 2e-16
Residuals	261	72225	277		

Model Diagnostics

In this experiment, researchers assigned students to small classes, regular classed and regular classes with aide randomly. Besides, teachers are also randomly assigned to classed in each school. By adding the randomized block design of schools in the experiment, there are adepuate reasons to believe that the outcomes do not depend on each other. In other words, the model satisfies the independence assumption.

From the Residuals vs Fitted Values scatterpoint, these points are uniformly distributed on both sides of x-axis, which means that our model does not violate the equal variance assumption.

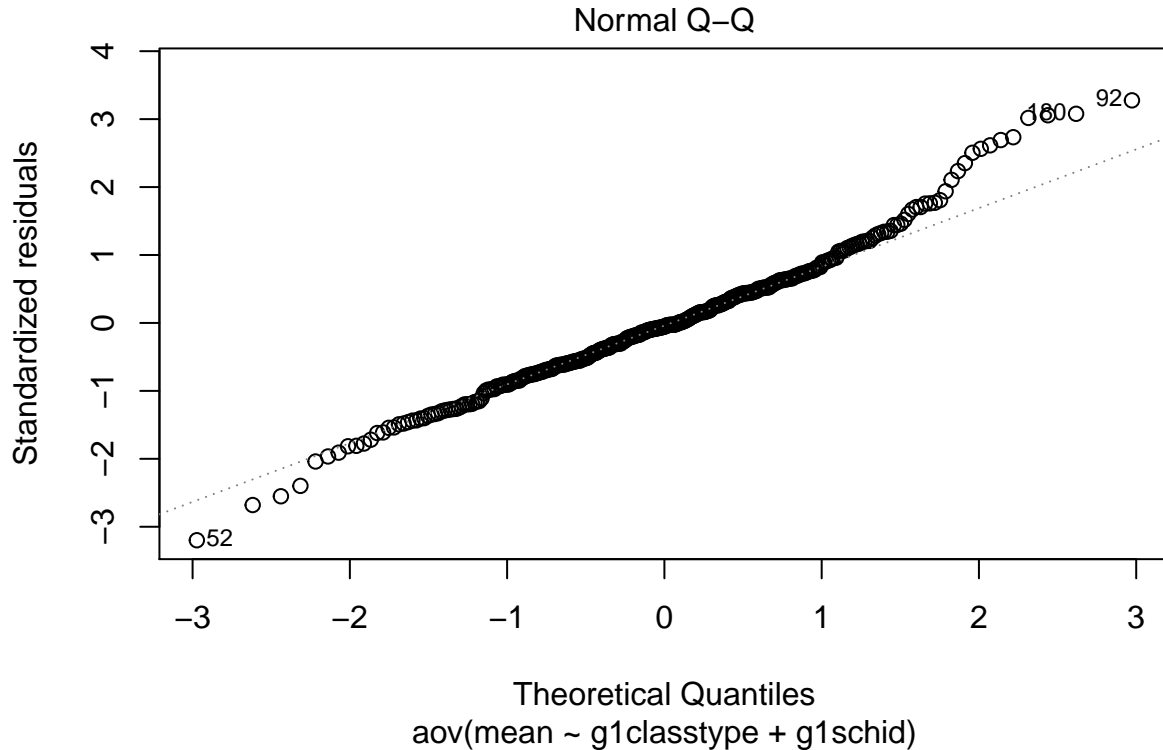




```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## g1classtype  2  11617    5809  20.991 3.52e-09 ***
## g1schid      75 136833    1824   6.593 < 2e-16 ***
## Residuals   261  72225     277
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

## Levene's Test for Homogeneity of Variance (center = median)
##           Df F value    Pr(>F)
## group 223  2.1025 7.149e-06 ***
##          115
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

The Normal Q-Q plot illustrates that the residuals are heavy tailed compared with the normal distribution. In order to check the normality assumption more precisely, we conduct the Shapiro-Wilk Normality Test, whose null hypothesis is H_0 : the population is normally distributed. The result shows that $W = 0.9802$ and the p -value is 0.0001, which is much smaller than the significant level 0.01. Therefore, we are 99% confident to reject the null hypothesis H_0 : the residuals are normally distributed and our model does not follow the normality assumption.



```
##
##  Shapiro-Wilk normality test
##
## data:  anova.fit$residuals
## W = 0.98016, p-value = 0.0001262
```

The model diagnostic part will confirm whether the model assumptions of the two-way ANOVA model so that it is appropriate for the problem setting. For the residuals, we will use the residuals vs fitted values plot to check equal variance assumption. Then, Q-Q plot is drawn to see whether the residuals are normally distributed roughly. To be more precise, the Shapiro-Wilk Normality Test is conducted.

Causal Analysis

From [Diane Whitmore Schanzenbach Brookings Papers on Education Policy No. 9 (2006/2007), pp. 205-228] we can know, project STAR is a randomized experiment since it randomly assigns students to different class types and teachers are also randomly assigned to class types. Therefore, we can make some causal inference if it follows two assumptions:

Stable unit treatment value assumption: The potential outcomes for any unit do not vary with the treatments assigned to other units, and, for each unit, there are no different forms or versions of each treatments level, which lead to different potential outcomes.

Ignorability: Treatment assignment is independent of the outcomes.

For the first assumption, it means that we could exclude the probability that a teacher was assigned or not assigned to small class does not effect the mean of math grade in other class types. Also, same version of treatment need to be ensured, that means teachers in the class of the one specific type give consistent

education of that class type. Both of two points are hold based on the setting of the experiment. For the second assumptions, because of randomized experiment, other factors don't effect our treatments and we don't have the selection bias, the average difference in outcomes between the three groups can only be attributable to treatment.

After verifying two assumptions, we use avarage treatment effect to measure the causal effect. Since there are three class types, we do pairwise comparisons to investigate it. We compare factor level means between each pair. From the above discussion, we can make a conclusion that, small class can cause higher mean scores by different teachers.

Discussion

Note: Discuss the limitations of the presented projects, and comment on how this project enlightens future research or analysis.

Appendix I. Reference

[1] C.M. Achilles; Helen Pate Bain; Fred Bellott; Jayne Boyd-Zaharias; Jeremy Finn; John Folger; John Johnston; Elizabeth Word, 2008, "Tennessee's Student Teacher Achievement Ratio (STAR) project", <https://doi.org/10.7910/DVN/SIWH9F>, Harvard Dataverse, V1, UNF:3:Ji2Q+9HCCZAbw3csOdMNdA== [fileUNF]

Appendix II. Session information

```
## R version 3.6.2 (2019-12-12)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 17763)
##
## Matrix products: default
##
## attached base packages:
## [1] stats      graphics  grDevices utils      datasets  methods   base
##
## other attached packages:
## [1] car_3.0-6      carData_3.0-3  ggpubr_0.2.4   magrittr_1.5
## [5] gridExtra_2.3  kableExtra_1.1.0 forcats_0.4.0  stringr_1.4.0
## [9] dplyr_0.8.3    purrr_0.3.3    readr_1.3.1    tidyr_1.0.0
## [13] tibble_2.1.3   ggplot2_3.2.1  tidyverse_1.3.0
##
## loaded via a namespace (and not attached):
## [1] Rcpp_1.0.3      lubridate_1.7.4  lattice_0.20-38  assertthat_0.2.1
## [5] zeallot_0.1.0   digest_0.6.23    R6_2.4.1         cellranger_1.1.0
## [9] backports_1.1.5  reprex_0.3.0     evaluate_0.14    httr_1.4.1
## [13] pillar_1.4.3    rlang_0.4.2      curl_4.3         lazyeval_0.2.2
## [17] readxl_1.3.1     data.table_1.12.8 rstudioapi_0.10  rmarkdown_2.0
## [21] labeling_0.3     webshot_0.5.2    foreign_0.8-72   munsell_0.5.0
## [25] broom_0.5.3      compiler_3.6.2    modelr_0.1.5     xfun_0.11
## [29] pkgconfig_2.0.3  htmltools_0.4.0  tidyselect_0.2.5 rio_0.5.16
## [33] fansi_0.4.0      viridisLite_0.3.0 crayon_1.3.4     dbplyr_1.4.2
## [37] withr_2.1.2      grid_3.6.2        nlme_3.1-142     jsonlite_1.6
## [41] gtable_0.3.0     lifecycle_0.1.0  DBI_1.1.0        scales_1.1.0
```

## [45]	zip_2.0.4	cli_2.0.1	stringi_1.4.3	farver_2.0.1
## [49]	ggsignif_0.6.0	fs_1.3.1	xml2_1.2.2	ellipsis_0.3.0
## [53]	generics_0.0.2	vctrs_0.2.1	openxlsx_4.1.4	tools_3.6.2
## [57]	glue_1.3.1	hms_0.5.3	abind_1.4-5	yaml_2.2.0
## [61]	colorspace_1.4-1	rvest_0.3.5	knitr_1.26	haven_2.2.0