The Effect of Class Types on Math Scores of 1st Grade Students

Wun-Syuan Wu

1/20/2022

1 Introduction

The Student/Teacher Achievement Ration (STAR) was a class-size study funded by the Tennessee General Assembly and conducted by the State Department of Education beginning in 1985. The STAR projected prolong for four years with 7000 students in 79 schools being randomly assigned into three different class types : small class (13 - 17 students per teacher), regular class(22-25 students per teacher) and regular class with aide. In each class, one random teacher was assigned to teach. Students as the interventions were who entered school in kindergarten and continued through third grade.

The initial motivation of this experiment is because that The legislature and the educational community of Tennessee were mindful of a promising study of the benefits of small class size carried out in nearby Indiana, but were also aware of the costs associated with additional classrooms and teachers (F.Mosteller, 1995). To improve school systems, as well minimize the additional funds from government, the study of class size which might have potential impact on the education systems was conducted.

Here, under the STAR experiment, the primary question of interest of this project is whether different class sizes(class types) have any impact on math scaled score of first grade students. If there is significant evidence to conclude that the primary question of interest is true, then our secondary question of interest would be in which class type, first grade students could get the highest math scaled scores.

2 Background

The dataset we used in this project is from Hardvard dataverse. In the original data, 11601 observaitons were obtained with 379 attributes. Within 379 attributes, they primarily include following information from targeted student in kindergarten to grade 3 (grade 4 to grade 8 were taken as additional data):

- the basic information of students (gender, ethnicity, birthday)
- the basic information of teachers (ethnicity, highest degree, teaching experience)
- the identifiers to students, teachers and schools
- the experimental condition (classtype)
- the scale score of sections in SAT (reading, math, listening, word study scores)

3 Descriptive Analysis

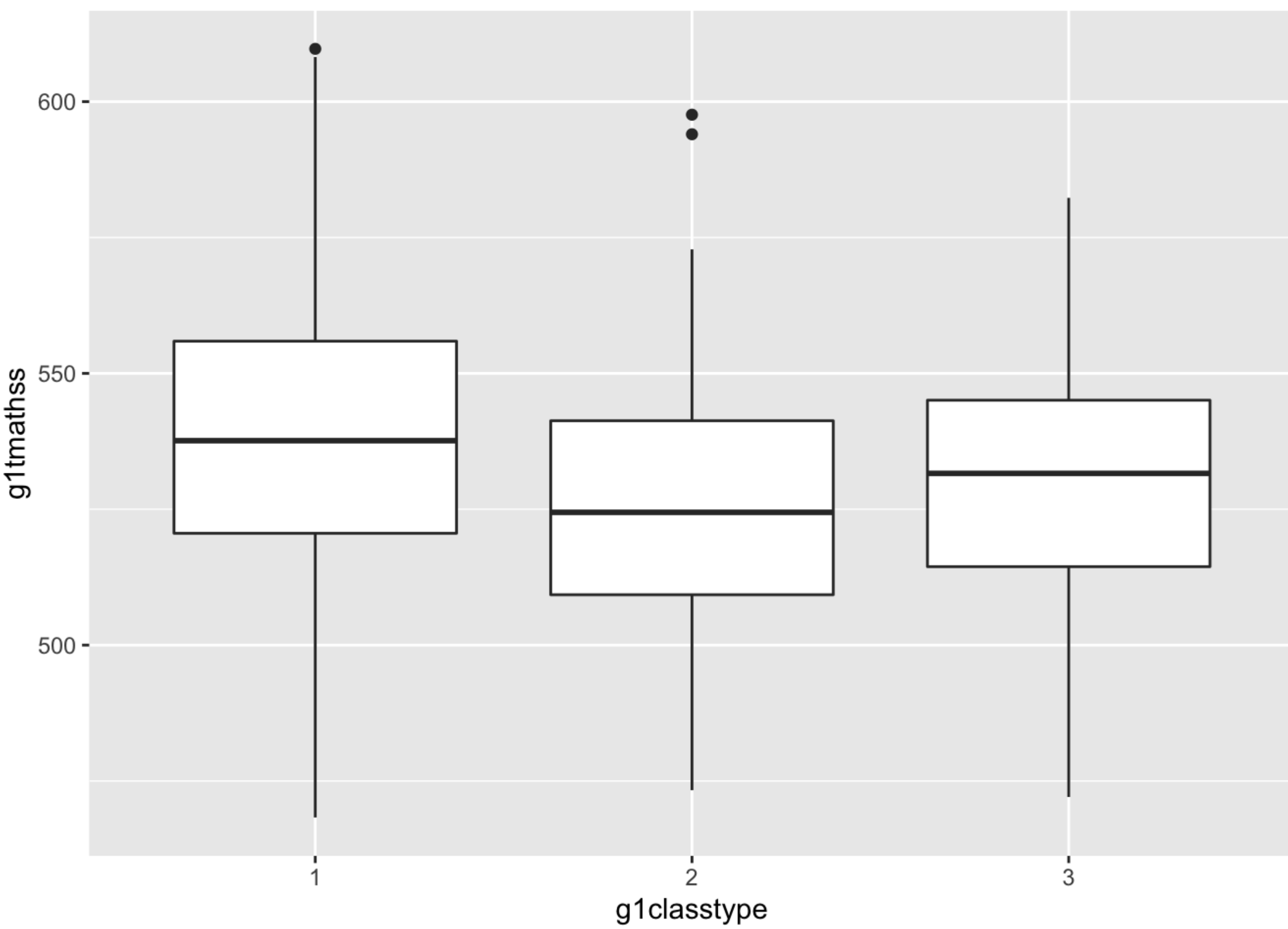
Based on the main question of interest, we selected the data only related to 1st grade students in order to make inferences on the effect on math score by three class types. After removing all the rows containing missing value, we obtained 339 observations of 4 variables with information about the identifier of 76 schools and 339 teachers, 3 class types and math scale scores. There is no obvious outliers in the selected data. In this study, teachers were randomly assigned to a class in original data, and we are more interested in the average student math score in each class, thus, we summarize the measure of math score by using mean instead of median. The distribution of mean scaled math scores in different class types are closely symmetric. However, by examining the amount of teachers in each school, the result shows that schools contains a various amount of teachers. Some of them have 12 teachers(classes) while some others only have 3 teachers(classes). Due to this unbalance of the data, it might cause variance not be constant which needed further examination in the following analysis.

```
##     g1tchid           g1tmathss
##  Min.   :11203804   Min.   :468.3
##  1st Qu.:17180356   1st Qu.:514.8
##  Median :21553308   Median :532.4
##  Mean   :20993925   Mean   :531.6
##  3rd Qu.:24475512   3rd Qu.:548.5
##  Max.   :26494510   Max.   :609.7
```

```
##     g1tchid           g1tmathss
##  Min.   :11203804   Min.   :465.0
##  1st Qu.:17180356   1st Qu.:512.0
##  Median :21553308   Median :532.0
##  Mean   :20993925   Mean   :530.6
##  3rd Qu.:24475512   3rd Qu.:549.0
##  Max.   :26494510   Max.   :619.5
```

```
##     g1tchid
##  Min.   :11203804
##  1st Qu.:17180356
##  Median :21553308
##  Mean   :20993925
##  3rd Qu.:24475512
##  Max.   :26494510
##     g1tmathss.0%       g1tmathss.25%       g1tmathss.50%       g1tmathss.75%       g1
tmathss.100%
##  Min.   :404.000     Min.   :450.2500    Min.   :465.0000    Min.   :475.7500    Min.
:523.0000
##  1st Qu.:446.000     1st Qu.:491.5000    1st Qu.:512.0000    1st Qu.:532.0000    1st
Qu.:578.0000
##  Median :468.000     Median :508.7500    Median :532.0000    Median :553.0000    Medi
an :601.0000
##  Mean   :468.059     Mean   :509.4417    Mean   :530.6431    Mean   :553.2891    Mean
:599.5516
##  3rd Qu.:486.000     3rd Qu.:526.0000    3rd Qu.:549.0000    3rd Qu.:572.0000    3rd
Qu.:619.5000
##  Max.   :562.000     Max.   :578.0000    Max.   :619.5000    Max.   :653.0000    Max.
:676.0000
```

```
## $g1schid
## [1] "numeric"
##
## $g1tchid
## [1] "numeric"
##
## $g1classtype
## [1] "haven_labelled" "vctrs_vctr"     "double"
##
## $g1classsize
## [1] "numeric"
##
## $g1tmathss
## [1] "numeric"
```



4 Inferential Analysis

Based on the primary question of interest, we chose two-way ANOVA to test whether the categorical variables(class type("classtype") and school id("g1schid")) has an effect on the quantitative variable(scaled math scores("g1tmathss")). In the two-way ANOVA model, we take 3 class types and 76 schools as two factos and we set it as following:

$$Y_{ijk} = \mu_{\cdot\cdot} + \alpha_i + \beta_j + \epsilon_{ijk},\ i = 1, \ldots, 3,\ j = 1, \ldots, 76,\ k = 1, \ldots, n_{ij}$$

where $\mu_{..}$ represents the overall average of scaled math score,α_i represents the factor effect of class type and β_j represents the factor effect of different schools. Also, ϵ_{ijk} are i.i.d $N(0, \sigma^2)$ as our assumptions. The reason we did not include interaction of class type and schools in the model is according to the result to the hypothesis testing, that is, there is no obvious evidence to show that adding the interaction term can make any difference on effecting the scaled math scores under significance level 0.05.

In the fitting results, it shows that class type 1(small class) was taken as reference. The second and third type of class tend to have lower average of scaled math scores than the small class type. And then based on the ANOVA table we obtained here, we conducted F test on both factors respectively. The result conclude that under the significance level $\alpha = 0.05$, we can reject the null hypothesis that it is statistically significant that both class type and school factor have unequal mean value, that is, they will affect student's math score significantly.

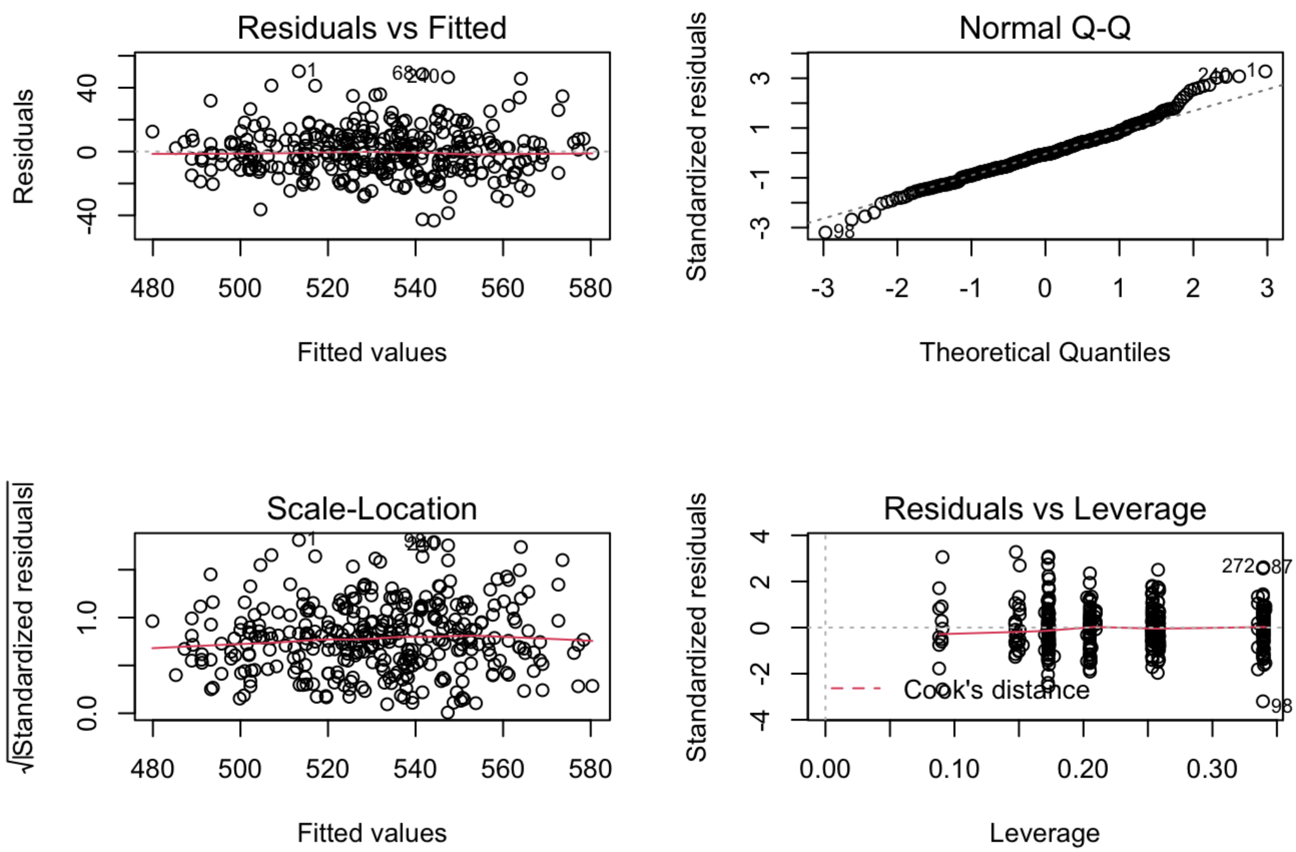
Because the primary question of interest has been satisfied, we conducted Tukey's test to answer our second question of interest: which class type affect math score the most. Under the test result, we could conclude that under significance level 0.05, it is significant that there are differences of average math score between small class with regular class(i=1,2), and small class with regular class with aide(i=1,3). Furthermore, the small class type has the most effect on the math score.

```
## Analysis of Variance Table
##
## Response: g1tmathss
##              Df Sum Sq Mean Sq F value    Pr(>F)
## g1classtype   2  11617  5808.6  20.991 3.522e-09 ***
## g1schid      75 136833  1824.4   6.593 < 2.2e-16 ***
## Residuals   261  72225   276.7
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
##  (Intercept) g1classtype2 g1classtype3
##    502.50070    -13.36987    -11.39667
```

5 Sensitivity Analysis

We have (1)normality,(2) constant variance,(3) independence assumptions of error term. Thus, we obtained normal Q-Q plot and residuals v.s. fitted value plot to see if the assumptions are violated. In the residuals v.s. fitted value plot, there is no obvious pattern of residuals; and in the normal Q-Q plot, we can see that there is heavy tail in the distribution. The normality and constant variance assumptions of error term held both well in conclusion. On top of that, from residuals v.s. leverage plot, it is also showed that there is no extreme outliers or leverage points. The independence of the error term is hold because students and teachers were randomly assigned to classes, these repetitive actions reduced the correlation happened during experiments.



6 Conclusion

In this study, as we analyze above, the average of scale math score will be affected by the class type. Student assigned in the small class tends to have better math score while students in regular class with or without aide has relatively lower math score. However, there are still many possible factors that may have effect on students math score such as the experience of teacher or the urbanicity grade of one school. More analysis and testing are needed to get further information on this topic.

Acknowledgement

Reference

Frederick Mosteller (1995).The Future of Children Vol. 5, No. 2, Critical Issues for Children and Youths (Summer - Autumn, 1995), pp. 113-127. Princeton University

Appendix

R code

Load Data

```
library(haven)
Star <- read_sav("STAR_Students.sav")
```

Section 3

```
### Processing data: selected variables, deal with missing value

#filter data with g1 in the column name
library(dplyr)
starg1 <-Star %>% select(starts_with("g1"))
```

```
#summary the filtered data
summary(starg1)
```

```
#choose the summary measure to be used
summary(aggregate(g1tmathss ~ g1tchid, data = starg1, FUN = mean))
```

```
##     g1tchid           g1tmathss
##  Min.   :11203804   Min.   :468.3
##  1st Qu.:17180356   1st Qu.:514.8
##  Median :21553308   Median :532.4
##  Mean   :20993925   Mean   :531.6
##  3rd Qu.:24475512   3rd Qu.:548.5
##  Max.   :26494510   Max.   :609.7
```

```
summary(aggregate(g1tmathss ~ g1tchid, data = starg1, FUN = median))
```

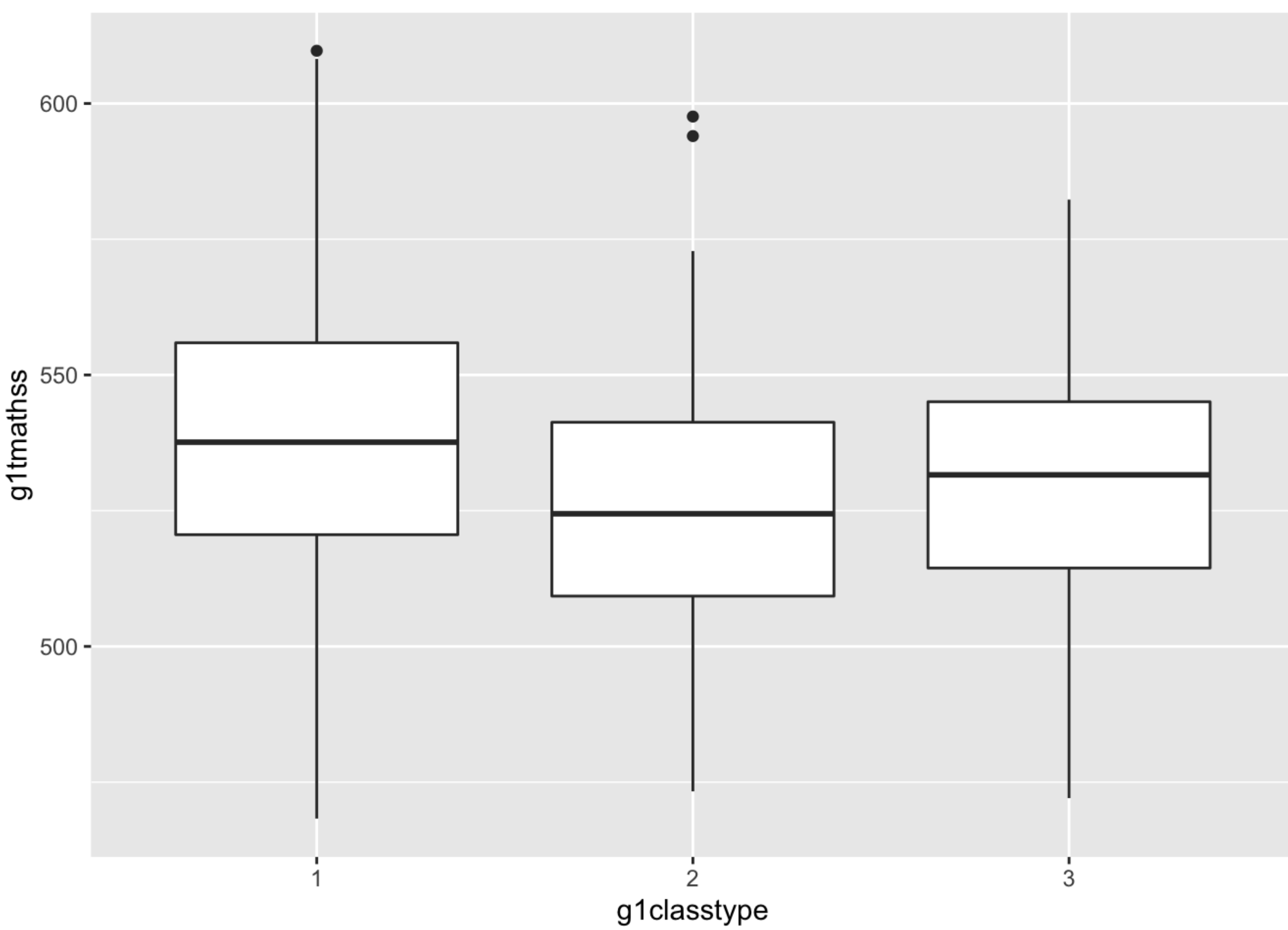
```
##     g1tchid           g1tmathss
##  Min.   :11203804   Min.   :465.0
##  1st Qu.:17180356   1st Qu.:512.0
##  Median :21553308   Median :532.0
##  Mean   :20993925   Mean   :530.6
##  3rd Qu.:24475512   3rd Qu.:549.0
##  Max.   :26494510   Max.   :619.5
```

```
#calculate the summary measure and get the needed variables
starg1_aggr <- aggregate(g1tmathss ~ g1schid + g1tchid + g1classtype + g1classsize , dat
a = starg1, FUN = mean)

#dealing with NA value
starg1_complete <- na.omit(starg1_aggr)
```

```
#summary the selected variables
sapply(starg1_complete, class)
```

```
#convert the data type in starg1_complete table
library(hablar)
starg1_complete <- starg1_complete %>% convert(fct(1,2,3,4))
```

```
### Multivariate descriptive analysis

#boxplot for scale math score v.s. class types
library(ggplot2)
ggplot(starg1_complete, aes(x = g1classtype, y = g1tmathss)) +
  geom_boxplot()
```




```
#outcome v.s. school id
starg1_complete %>% count(g1schid, sort = TRUE)
```

```
## # A tibble: 76 × 2
##    g1schid     n
##    <fct>   <int>
##  1 169229     12
##  2 201449      7
##  3 244755      7
##  4 244806      7
##  5 257905      7
##  6 159171      6
##  7 161183      6
##  8 180344      6
##  9 209510      6
## 10 215533      6
## # … with 66 more rows
```

```
starg1_groupby_schid <- starg1_complete %>%
  group_by(g1schid) %>%
  summarize(mean = mean(g1tmathss))
```

Section 4

```
### Fitting models

#two-way anova without interaction
g1.fit <- aov(g1tmathss ~ g1classtype + g1schid, data = starg1_complete)
summary(g1.fit)
```

```
##              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
```

```
anova(g1.fit)
```

```
## Analysis of Variance Table
## 
## Response: g1tmathss
##              Df Sum Sq Mean Sq F value    Pr(>F)    
## g1classtype   2  11617  5808.6  20.991 3.522e-09 ***
## g1schid      75 136833  1824.4   6.593 < 2.2e-16 ***
## Residuals   261  72225   276.7                      
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
g1.fit$coefficients[1:3]
```

```
##  (Intercept) g1classtype2 g1classtype3
##    502.50070    -13.36987    -11.39667
```

```
#two-way anova with interaction
g1.interaction <- aov(g1tmathss ~ g1classtype*g1schid, data = starg1_complete)
anova(g1.interaction)
```

```
## Analysis of Variance Table
##
## Response: g1tmathss
##                      Df Sum Sq Mean Sq F value    Pr(>F)
## g1classtype           2  11617  5808.6 19.9362 3.687e-08 ***
## g1schid              75 136833  1824.4  6.2617 < 2.2e-16 ***
## g1classtype:g1schid 146  38718   265.2  0.9102    0.7056
## Residuals           115  33507   291.4
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
### Hypothesis Testing
#Test for the main effect of class type

#F statistics = 20.99087
anova(g1.fit)[1,4]
```

```
## [1] 20.99087
```

```
#critical value = 3.030382
qf(1-0.05,2,261)
```

```
## [1] 3.030382
```

```
#Test for the main effect of school indicator

#F statistics = 6.593013
anova(g1.fit)[2,4]
```

```
## [1] 6.593013
```

```
#critical value = 1.338174
qf(1-0.05,75,261)
```

```
## [1] 1.338174
```

```
###Test for interaction

#Test whether we need the interaction
anova(g1.fit, g1.interaction)
```

```
## Analysis of Variance Table
##
## Model 1: g1tmathss ~ g1classtype + g1schid
## Model 2: g1tmathss ~ g1classtype * g1schid
##   Res.Df   RSS  Df Sum of Sq      F Pr(>F)
## 1    261 72225
## 2    115 33507 146     38718 0.9102 0.7056
```

```
#test statistics = 0.9101785
anova(g1.fit, g1.interaction)$F[2]
```

```
## [1] 0.9101785
```

```
#critical value = 1.341821
qf(0.95, 261-115, 115)
```

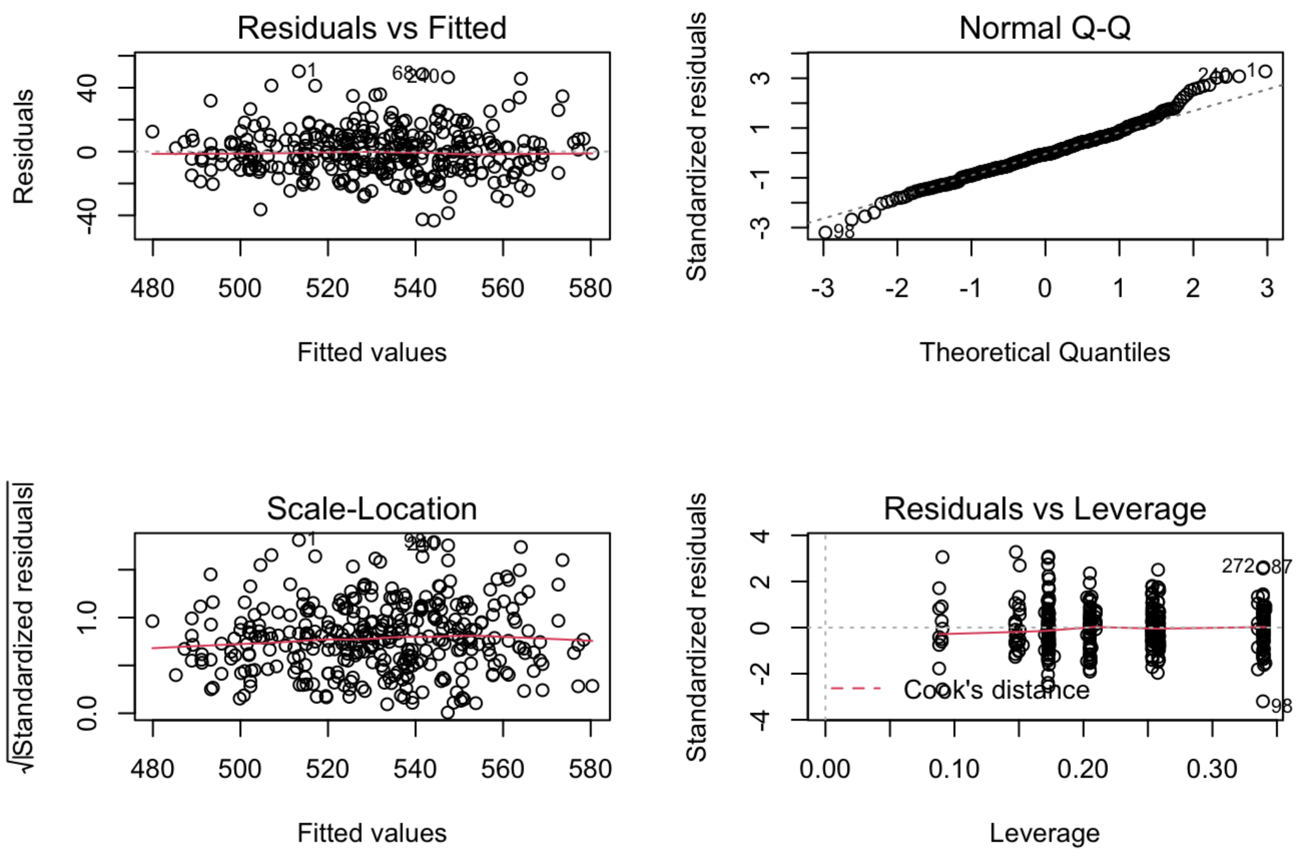
```
## [1] 1.341821
```

```
###Tukey test
TukeyHSD(g1.fit, conf.level = 0.95)$g1classtype
```

```
##           diff        lwr       upr        p adj
## 2-1 -13.580024 -18.656367 -8.503681 3.651884e-09
## 3-1  -9.485401 -14.755575 -4.215226 9.091937e-05
## 3-2   4.094623  -1.266825  9.456071 1.715156e-01
```

Section 5

```
###Diagnostic plots
par(mfrow=c(2,2))
plot(g1.fit)
```



Session Info

```
sessionInfo()
```

```
## R version 4.1.1 (2021-08-10)
## Platform: aarch64-apple-darwin20 (64-bit)
## Running under: macOS Monterey 12.2.1
##
## Matrix products: default
## BLAS:   /Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/lib/libRblas.0.d
ylib
## LAPACK: /Library/Frameworks/R.framework/Versions/4.1-arm64/Resources/lib/libRlapack.d
ylib
##
## locale:
## [1] en_US.UTF-8/en_US.UTF-8/en_US.UTF-8/C/en_US.UTF-8/en_US.UTF-8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## other attached packages:
## [1] ggplot2_3.3.5 hablar_0.3.0  dplyr_1.0.7   haven_2.4.3
##
## loaded via a namespace (and not attached):
##  [1] highr_0.9        pillar_1.6.2     bslib_0.3.1      compiler_4.1.1
##  [5] jquerylib_0.1.4  forcats_0.5.1    tools_4.1.1      digest_0.6.28
##  [9] gtable_0.3.0     jsonlite_1.7.2   evaluate_0.14    lifecycle_1.0.0
## [13] tibble_3.1.4     pkgconfig_2.0.3  rlang_0.4.11     rstudioapi_0.13
## [17] cli_3.1.1        DBI_1.1.1        yaml_2.2.1       xfun_0.26
## [21] fastmap_1.1.0    withr_2.4.3      stringr_1.4.0    knitr_1.34
## [25] generics_0.1.0   vctrs_0.3.8      sass_0.4.0       hms_1.1.1
## [29] grid_4.1.1       tidyselect_1.1.1 glue_1.4.2       R6_2.5.1
## [33] fansi_0.5.0      rmarkdown_2.11   farver_2.1.0     readr_2.0.2
## [37] tzdb_0.1.2       purrr_0.3.4      magrittr_2.0.1   scales_1.1.1
## [41] ellipsis_0.3.2   htmltools_0.5.2  assertthat_0.2.1 colorspace_2.0-2
## [45] labeling_0.4.2   utf8_1.2.2       stringi_1.7.6    munsell_0.5.0
## [49] crayon_1.4.1
```