

Efficacy of Small-class Size in Early Education

The STAR (Student-Teacher Achievement Ratio) Project is a four year *longitudinal study* examining the effect of class size in early grade levels on educational performance and personal development.

This exercise is in part based on: Mosteller, Frederick. 1997. "The Tennessee Study of Class Size in the Early School Grades." *Bulletin of the American Academy of Arts and Sciences* 50(7): 14-25.

A longitudinal study is one in which the same participants are followed over time. This particular study lasted from 1985 to 1989 involved 11,601 students. During the four years of the study, students were randomly assigned to small classes, regular-sized classes, or regular-sized classes with an aid. In all, the experiment cost around \$12 million. Even though the program stopped in 1989 after the first kindergarten class in the program finished third grade, collection of various measurements (e.g., performance on tests in eighth grade, overall high school GPA) continued through the end of participants' high school attendance.

We will analyze just a portion of this data to investigate whether the small class sizes improved performance or not. The data file name is `STAR.csv`, which is a CSV data file. The names and descriptions of variables in this data set are:

Name	Description
<code>race</code>	Student's race (White = 1, Black = 2, Asian = 3, Hispanic = 4, Native American = 5, Others = 6)
<code>classtype</code>	Type of middle school class (small = 1, regular = 2, regular with aid = 3)
<code>g4math</code>	Total scaled score for math portion of fourth grade standardized test
<code>g4reading</code>	Total scaled score for reading portion of fourth grade standardized test
<code>yearssmall</code>	Number of years in small classes
<code>hsgrad</code>	High school graduation (did graduate = 1, did not graduate = 0)

Note that there are a fair amount of missing values in this data set. For example, missing values arise because some students left a STAR school before third grade or did not enter a STAR school until first grade.

Question 1

Create a new factor variable called `type` in the data frame. This variable should recode `classtype` by changing integer values to their corresponding informative labels (e.g., change 1 to `small` etc.). Similarly, recode the `race` variable into a factor variable with four levels (`white`, `black`, `hispanic`, `others`) by combining Asians and Native Americans as the `others` category. For the `race` variable, overwrite the original variable in the data frame rather than creating a new one. Recall that `na.rm = TRUE` can be added to functions in order to remove missing data.

Run following scripts to get the data:

```
library(devtools)

## Loading required package: usethis
install_github("kosukeimai/qss-package", build_vignettes = TRUE)

## Downloading GitHub repo kosukeimai/qss-package@HEAD
```

```

## rlang      (1.1.1    -> 1.1.3    ) [CRAN]
## cli       (3.6.1    -> 3.6.2    ) [CRAN]
## sys       (3.4.1    -> 3.4.2    ) [CRAN]
## askpass   (1.1      -> 1.2.0    ) [CRAN]
## openssl   (2.0.6    -> 2.1.1    ) [CRAN]
## jsonlite  (1.8.4    -> 1.8.8    ) [CRAN]
## curl      (5.0.0    -> 5.2.0    ) [CRAN]
## utf8      (1.2.3    -> 1.2.4    ) [CRAN]
## vctrs     (0.6.2    -> 0.6.5    ) [CRAN]
## fansi     (1.0.4    -> 1.0.6    ) [CRAN]
## rprojroot (2.0.3    -> 2.0.4    ) [CRAN]
## pkgbuild  (1.4.0    -> 1.4.3    ) [CRAN]
## fs        (1.6.2    -> 1.6.3    ) [CRAN]
## glue      (1.6.2    -> 1.7.0    ) [CRAN]
## withr     (2.5.0    -> 3.0.0    ) [CRAN]
## waldo     (0.5.1    -> 0.5.2    ) [CRAN]
## ps        (1.7.5    -> 1.7.6    ) [CRAN]
## processx  (3.8.1    -> 3.8.3    ) [CRAN]
## pkgload   (1.3.2    -> 1.3.4    ) [CRAN]
## lifecycle (1.0.3    -> 1.0.4    ) [CRAN]
## evaluate  (0.21     -> 0.23     ) [CRAN]
## digest    (0.6.31   -> 0.6.34   ) [CRAN]
## desc      (1.4.2    -> 1.4.3    ) [CRAN]
## brio      (1.1.3    -> 1.1.4    ) [CRAN]
## stringi   (1.7.12   -> 1.8.3    ) [CRAN]
## yaml      (2.3.7    -> 2.3.8    ) [CRAN]
## xfun      (0.39     -> 0.41     ) [CRAN]
## remotes   (2.4.2    -> 2.4.2.1  ) [CRAN]
## Rcpp      (1.0.10   -> 1.0.12   ) [CRAN]
## BH        (1.81.0-1 -> 1.84.0-0 ) [CRAN]
## xml2      (1.3.4    -> 1.3.6    ) [CRAN]
## RCurl     (1.98-1.13 -> 1.98-1.14) [CRAN]
## httr      (1.4.6    -> 1.4.7    ) [CRAN]
## testthat  (3.1.8    -> 3.2.1    ) [CRAN]
## stringr   (1.5.0    -> 1.5.1    ) [CRAN]
## xmlparsedata (NA      -> 1.0.5    ) [CRAN]
## rex       (NA      -> 1.2.1    ) [CRAN]
## knitr     (1.42     -> 1.45     ) [CRAN]
## cyclocomp (NA      -> 1.1.1    ) [CRAN]
## cpp11     (0.4.3    -> 0.4.7    ) [CRAN]
## magick    (NA      -> 2.8.2    ) [CRAN]
## wordcloud (NA      -> 2.6      ) [CRAN]
## swirl     (NA      -> 2.4.5    ) [CRAN]
## maps      (3.4.1.1  -> 3.4.2    ) [CRAN]
## lintr     (NA      -> 3.1.1    ) [CRAN]
## igraph    (1.5.0    -> 2.0.1.1  ) [CRAN]
## animation (NA      -> 2.7      ) [CRAN]

## Installing 47 packages: rlang, cli, sys, askpass, openssl, jsonlite, curl, utf8, vctrs, fansi, rproj
##
## The downloaded binary packages are in
## /var/folders/70/5dx3642905115g_v7tbdbp0h0000gr/T/RtmpC4Qg1w/downloaded_packages
## -- R CMD build -----
##      checking for file '/private/var/folders/70/5dx3642905115g_v7tbdbp0h0000gr/T/RtmpC4Qg1w/remotes1

```

```
## - preparing 'qss': (343ms)
## checking DESCRIPTION meta-information ... v checking DESCRIPTION meta-information
## - installing the package to build vignettes
## creating vignettes ... v creating vignettes (17s)
## - checking for LF line-endings in source and make files and shell scripts
## - checking for empty or unneeded directories
## - looking to see if a 'data/datalist' file should be added
## - building 'qss_0.1.0.9000.tar.gz' (1.6s)
##
##
```

```
library(qss)
data(STAR, package="qss")
```

Answer 1

```
## create a variable labeling class type
STAR$type<- ifelse(STAR$classtype == 1, "small",
                  ifelse(STAR$classtype == 2, "regular",
                        "reg w/aid"))
STAR$type<- as.factor(STAR$type)

## create race label variable
race <- ifelse(STAR$race == 1, "white",
              ifelse(STAR$race == 2, "black",
                    ifelse(STAR$race == 4, "hispanic", "others")))
race <- as.factor(race)
## overwrite the original variable
STAR$race <- race
```

Question 2

How does performance on fourth grade reading and math tests for those students assigned to a small class in kindergarten compare with those assigned to a regular-sized class (only for those without aid)? Do students in the smaller classes perform better? Use means to make this comparison while removing missing values. Make sure you call them `dif_read` and `dif_math`

Answer 2

```
## diff in means for reading
dif_read <- mean(STAR$g4reading[STAR$type=="small"], na.rm = TRUE) -
            mean(STAR$g4reading[STAR$type=="regular"], na.rm = TRUE)
## diff in means for math
dif_math <-
mean(STAR$g4math[STAR$type=="small"], na.rm = TRUE) -
mean(STAR$g4math[STAR$type=="regular"], na.rm = TRUE)
```

Question 2.1

Give a brief substantive interpretation of the results. To understand the size of the estimated effects, compare them with the standard deviation of the test scores.

Answer 2.1

This output shows that students in small classrooms in kindergarten scored, on average, 3.5 points higher on reading tests in the fourth grade. However, they scored about 0.34 points lower on average in math compared to students in regular classrooms. Comparing these estimated effects with the standard deviation of test scores, we observe that the sizes of the estimated effects are quite small.

```
## standard deviations of reading and math scores
sd(STAR$g4reading, na.rm = TRUE)
```

```
## [1] 52.42592
```

```
sd(STAR$g4math, na.rm = TRUE)
```

```
## [1] 43.09217
```

We conclude that on average small class size in middle school did not substantially increase reading and math test scores in the fourth grade.

Question 3

Some students were in small classes for all four years that the STAR program ran. Others were assigned to small classes for only one year and had either regular classes or regular classes with an aid for the rest. How many such students of each type are in the data set? Create a contingency table of proportions using the `type` and `yearsmall` variables. Does participation in more years of small classes make a greater difference in test scores? Compare the average and median reading and math test scores across students who spent different numbers of years in small classes.

Answer 3

```
## contingency table of proportions
prop.table(table(STAR$type, STAR$yearsmall))
```

```
##
##           0           1           2           3           4
## reg w/aid 0.315573123 0.015335968 0.009486166 0.012332016 0.000000000
## regular   0.310039526 0.015019763 0.009169960 0.012648221 0.000000000
## small     0.000000000 0.091067194 0.043003953 0.030830040 0.135494071
```

```
## mean/median reading score across years in small classes
tapply(STAR$g4reading, STAR$yearsmall, mean, na.rm = TRUE)
```

```
##           0           1           2           3           4
## 719.8754 723.1471 717.8681 719.8986 724.6651
```

```
tapply(STAR$g4reading, STAR$yearsmall, median, na.rm = TRUE)
```

```
##           0           1           2           3           4
## 722.0 724.5 720.0 721.0 726.0
```

```
## mean/median math score
tapply(STAR$g4math, STAR$yearsmall, mean, na.rm = TRUE)
```

```
##           0           1           2           3           4
## 707.9793 707.5524 711.9140 709.6170 710.0519
```

```
tapply(STAR$g4math, STAR$yearsmall, median, na.rm = TRUE)
```

```
##           0           1           2           3           4
## 710 709 714 712 711
```

The contingency table shows that 63% of students were never in the small classes whereas 14% of them were in the small classes for all four years. The analysis suggests that in general, spending all four years in the small classes increases both reading and math test scores. The effect is not huge but is of reasonable size. The analyses based on the mean and median yield similar results.

Question 4

We examine whether the STAR program reduced the achievement gaps across different racial groups. Begin by comparing the average reading and math test scores between white and minority students (i.e., Blacks and hispanics) among those students who were assigned to regular classes with no aid. Conduct the same comparison among those students who were assigned to small classes.

Your output should be four new objects: `race_gap_read_reg`, `race_gap_read_small`, `race_gap_math_reg` and `race_gap_math_small`

Answer 4

```
## subset the data
white <- subset(STAR, subset = (race == "white"))
minority <- subset(STAR, subset = ((race == "black") |
                                  (race == "hispanic")))

## racial gap for reading among students in regular classes
race_gap_read_reg <- mean(white$g4reading[white$type=="regular"], na.rm = TRUE) -
  mean(minority$g4reading[minority$type=="regular"],
        na.rm = TRUE)

## racial gap for reading among students in small classes
race_gap_read_small <- mean(white$g4reading[white$type=="small"], na.rm = TRUE) -
  mean(minority$g4reading[minority$type=="small"],
        na.rm = TRUE)

## repeat the same analysis for math score
race_gap_math_reg <- mean(white$g4math[white$type=="regular"], na.rm = TRUE) -
  mean(minority$g4math[minority$type=="regular"],
        na.rm = TRUE)
race_gap_math_small <- mean(white$g4math[white$type=="small"], na.rm = TRUE) -
  mean(minority$g4math[minority$type=="small"],
        na.rm = TRUE)
```

Question 4.1

Give a brief substantive interpretation of the results of your analysis above.

Answer 4.1

Our analysis shows that there is a substantial racial gap. On average, white students tend to perform better in both reading and math scores than minority students, regardless of class sizes. However, in terms of reading test scores, this achievement gap is reduced when students are assigned to small classes. This suggests that minority students benefited more from small classes than white students. The same conclusion, however, does not apply to math scores. The racial achievement gap is approximately the same size in both regular and small classes.

Question 5

We consider the long term effects of middle school class size. Compare high school graduation rates across students assigned to different class types. Also, examine whether graduation rates differ by the number of

years spent in small classes. Finally, as done in the previous question, investigate whether the STAR program has reduced the racial gap between white and minority students' graduation rates. Briefly discuss the results.

Answer 5

```
## graduation rate by classsize
tapply(STAR$hsgrad, STAR$type, mean, na.rm = TRUE)

## reg w/aid  regular    small
## 0.8392857 0.8251619 0.8359202

tapply(STAR$hsgrad, STAR$yearssmall, mean, na.rm = TRUE)
```

```
##          0          1          2          3          4
## 0.8286020 0.7910448 0.8131868 0.8324607 0.8775510
```

We observe little difference in graduation rates across students who were assigned to different class types in middle school. However, those who spent all four years in small classes have a higher graduation rate than the others. This result is consistent with the analysis for a previous question where we found spending all four years in small classes increases the average reading score.

```
## racial gap
tapply(white$hsgrad, white$type, mean, na.rm = TRUE) -
  tapply(minority$hsgrad, minority$type, mean, na.rm = TRUE)

## reg w/aid  regular    small
## 0.1440545 0.1173787 0.1227890

tapply(white$hsgrad, white$yearssmall, mean, na.rm = TRUE) -
  tapply(minority$hsgrad, minority$yearssmall, mean, na.rm = TRUE)
```

```
##          0          1          2          3          4
## 0.13933075 0.06692506 0.09706514 0.12804878 0.12046459
```

We observe substantial racial gaps in high school graduation rates regardless of kindergarten class types. These gaps appear to remain even when students were assigned to small classes for all four years. Therefore, the STAR program appears to have little impact in closing racial gaps of high school graduation rates.

Question 6

Now, let's try a simple matching exercise using this data set. And for this question, we are only interested in comparing classsize **without aid** (the regular and small groups). We will see small size as the **treated** group and regular size as the control group. And we also want to get rid of observations with "NA" in their values.

What is the potential problem of getting rid of people with missing values?

Set up a treatment variable where it is TRUE if small class size and FALSE otherwise. For simplicity, change race variable to be an indicator of white = 1, otherwise = 0.

```
library(MatchIt)
```

Answer 6

```
STAR_2 <- STAR[STAR$type != "reg w/aid",]
STAR_2 <- STAR_2[rowSums(is.na(STAR_2)) == 0,]
STAR_2$treatment <- ifelse(STAR_2$type == "small", TRUE, FALSE)
STAR_2$race <- ifelse(STAR_2$race == "white", 1, 0)
```

Question 7

Generate an exact matching set using Mahalanobis distance measure for those with a small class size. Let's ignore "yearssmall" variable for this task - since it complicates the design by making treatment non-binary.

Calculate the average treatment effect on the treated (small class size) and interpret it.

Answer 7

```
m.out <- matchit(treatment ~ race + g4math + g4reading , data = STAR_2,
                 distance = "mahalanobis", replace = TRUE,
                 exact = ~ race )
summary(m.out)

##
## Call:
## matchit(formula = treatment ~ race + g4math + g4reading, data = STAR_2,
##         distance = "mahalanobis", exact = ~race, replace = TRUE)
##
## Summary of Balance for All Data:
##           Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean
## race           0.8649         0.8564         0.0248         .      0.0085
## g4math          713.4844        712.6990         0.0181         1.2763  0.0206
## g4reading        726.4636        723.2664         0.0622         1.1195  0.0192
##           eCDF Max
## race           0.0085
## g4math          0.0641
## g4reading        0.0532
##
## Summary of Balance for Matched Data:
##           Means Treated Means Control Std. Mean Diff. Var. Ratio eCDF Mean
## race           0.8649         0.8649         0.0000         .      0.0000
## g4math          713.4844        713.5800        -0.0022         1.0536  0.0045
## g4reading        726.4636        725.9252         0.0105         1.0061  0.0059
##           eCDF Max Std. Pair Dist.
## race           0.0000         0.0000
## g4math          0.0166         0.0586
## g4reading        0.0208         0.0805
##
## Sample Sizes:
##           Control Treated
## All           578.      481
## Matched (ESS) 215.62    481
## Matched        291.      481
## Unmatched      287.        0
## Discarded       0.        0

m.data <- match.data(m.out)
head(m.data)

##   race classtype yearssmall hsgrad g4math g4reading  type treatment
## 9     1          1           4      1    709      725  small      TRUE
## 10    1          2           0      1    698      692 regular    FALSE
## 27    1          1           4      1    740      836  small      TRUE
## 40    1          1           4      1    771      725  small      TRUE
```

```
## 48 1 1 4 1 709 761 small TRUE
## 70 1 2 0 1 698 778 regular FALSE
## weights
## 9 1.0000000
## 10 0.6049896
## 27 1.0000000
## 40 1.0000000
## 48 1.0000000
## 70 0.6049896
```

```
sum(m.data$hsgrad[m.data$treatment==1])/sum(m.data$treatment==1)- sum(m.data$hsgrad[m.data$treatment==0])/sum(m.data$treatment==0)
## [1] -0.01396718
```

Question 8

Visualize the matching quality and offer some discussion, on the improvement before and after matching.

Answer 8

```
plot(m.out, type = "density", interactive = FALSE,
      which.xs = ~race + g4math + g4reading)
```

