Efficacy of Small-class Size in Early Education

Introduction

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 aide. 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
race	Student's race (White = 1, Black = 2, Asian = 3, Hispanic = 4, Native American = 5, Others = 6)
classtype	Type of kindergarten class (small = 1, regular = 2, regular with aide = 3)
g4math	Total scaled score for math portion of fourth grade standardized test
g4reading	Total scaled score for reading portion of fourth grade standardized test
yearssmall	Number of years in small classes
hsgrad	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: Reading data into R

Before we can get started working with data, we first need to load the data into R. Load the `tidyverse` package, read the data using the `read_csv()` function and save it as `STAR`.

How many students are there in `STAR`?

Answer 1

```
#load tidyverse
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ----------------------- tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.0     v tibble    3.2.1
## v lubridate 1.9.3     v tidyr     1.3.1
## v purrr     1.0.2
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```r
#read the data in as a tibble
STAR <- read_csv("data/STAR.csv")
```

```
## Rows: 6325 Columns: 6
## -- Column specification -------------------------------------------------------
## Delimiter: ","
## dbl (6): race, classtype, yearssmall, hsgrad, g4math, g4reading
##
## i Use `spec()` to retrieve the full column specification for this data.
## i Specify the column types or set `show_col_types = FALSE` to quiet this message.
```

```r
#check the number of rows in the data
nrow(STAR)
```

```
## [1] 6325
```

Question 2: Creating and changing variables

Before analyzing the data we're going to modify it to make it more easily interpretable. First create a new variable called `kinder` in the data frame. This variable should recode `classtype` by changing integer values to their corresponding informative string labels (e.g., change 1 to `small` etc.). Second, Similarly, recode the `race` variable into a character variable (`white`, `black`, `asian`, `hispanic`, `native american`, `others`). For the `race` variable, overwrite the original variable in the data frame rather than creating a new one.

Hint: Use the `mutate()` function in conjunction with `case_when()`.

Answer 2

```r
#create the new variable and modify the existing one
STAR <- STAR |>
  mutate(kinder = case_when(classtype == 1 ~ "small",
                            classtype == 2 ~ "regular",
                            classtype == 3 ~ "regular with aide"),
         race = case_when(race == 1 ~ "white",
                          race == 2 ~ "black",
                          race == 3 ~ "asian",
                          race == 4 ~ "hispanic",
                          race == 5 ~ "native american",
                          race == 6 ~ "others"))

#check that the changes have been made
head(STAR)
```

```
## # A tibble: 6 x 7
##    race  classtype yearssmall hsgrad g4math g4reading kinder
```

```
##    <chr>      <dbl>     <dbl> <dbl> <dbl>      <dbl> <chr>
## 1 white          3         0    NA    NA         NA regular with aide
## 2 black          3         0    NA   706        661 regular with aide
## 3 white          3         0     1   711        750 regular with aide
## 4 black          1         4    NA   672        659 small
## 5 white          2         0    NA    NA         NA regular
## 6 white          3         0    NA    NA         NA regular with aide
```

Question 3: Covariate balance tables

Let's first check to make sure our covariates are balanced. In this case we have only the one covariate of
`race`. Create a table that shows the proportion of students in each treatment category that fall under each of
your coded values of `race`.

Hint: Remember that taking the mean of booleans (that is `TRUE` or `FALSE`) returns the proportion that are
`TRUE`

Answer 3

```
#using group_by and summarize
STAR |>
  group_by(kinder) |>
  summarize(black = mean(race=="black", na.rm = TRUE),
            white = mean(race=="white", na.rm = TRUE),
            asian = mean(race=="asian", na.rm = TRUE),
            native_american = mean(race=="native american", na.rm = TRUE),
            hispanic = mean(race=="hispanic", na.rm = TRUE),
            others = mean(race=="others")) |>
  knitr::kable(digits=3)
```

kinder	black	white	asian	native_american	hispanic	others
regular	0.324	0.672	0.004	0.000	0.000	NA
regular with aide	0.338	0.658	0.001	0.000	0.000	0.002
small	0.312	0.681	0.002	0.001	0.002	NA

```
#can more easily use table to get count
table(STAR$race, STAR$kinder)
```

```
##
##                  regular regular with aide small
##   asian                8                 3     3
##   black              710               755   593
##   hispanic             0                 1     4
##   native american      1                 0     1
##   others               1                 4     4
##   white             1472              1468  1294
```

```
#and prop table to get proportion: "2" for columnwise %, "1" for rowwise %, and blank for % of total
prop.table(table(STAR$race, STAR$kinder), 2)
```

```
##
##                   regular regular with aide       small
##   asian         0.0036496350      0.0013446885 0.0015797788
##   black         0.3239051095      0.3384132676 0.3122696156
```

```
##    hispanic          0.0000000000    0.0004482295 0.0021063718
##    native american 0.0004562044    0.0000000000 0.0005265929
##    others            0.0004562044    0.0017929180 0.0021063718
##    white             0.6715328467    0.6580008965 0.6814112691
```

```
# Option 2 - coding with lists
# race_categories = c("white", "black", "asian", "hispanic","native american", "others")
# Only works for converting numerical labels from 1..N to factor

# STAR$race_labels = race_categories[race]
```

There do not appear to be any significant, systemic differences in race across the treatments.

Question 4: Difference in means

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? Do students in the smaller classes perform better? Use means to make this comparison while removing missing values. Give a brief substantive interpretation of the results.

Hint: Recall that `na.rm = TRUE` can be added to functions in order to remove missing data.

Answer 4

```
#difference in means for reading
STAR |>
  filter(kinder %in% c("small", "regular")) |>
  group_by(kinder) |>
  summarize(mean_reading = mean(g4reading, na.rm = TRUE))  |>
  pivot_wider(names_from = kinder,
              values_from = mean_reading) |>
  mutate(reading_diff = small - regular)
```

```
## # A tibble: 1 x 3
##    regular small reading_diff
##      <dbl> <dbl>        <dbl>
## 1    720.  723.         3.50
```

```
#difference in means for math
STAR |>
  filter(kinder %in% c("small", "regular")) |>
  group_by(kinder) |>
  summarize(mean_math = mean(g4math, na.rm = TRUE)) |>
  pivot_wider(names_from = kinder,
              values_from = mean_math) |>
  mutate(math_diff = small - regular)
```

```
## # A tibble: 1 x 3
##    regular small math_diff
##      <dbl> <dbl>     <dbl>
## 1    710.  709.    -0.336
```

This is evidence that, on average, small class size in kindergarten did not substantially increase reading and math test scores in the fourth grade.

Question 5: Comparing multiple treatments

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 aide for the rest. How many such students of each type are in the data set? 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 5

```r
#can do contingency table of counts with table:
table(STAR$kinder, STAR$yearssmall)
```

```
##
##                       0    1    2    3    4
##   regular          1961   95   58   80    0
##   regular with aide 1996   97   60   78    0
##   small               0  576  272  195  857
```

```r
#or using group_by and count()
STAR |>
  select(kinder, yearssmall) |>
  group_by(kinder, yearssmall) |>
  count() |>
  pivot_wider(names_from = yearssmall,
              values_from = n)
```

```
## # A tibble: 3 x 6
## # Groups:   kinder [3]
##   kinder              `0`   `1`   `2`   `3`   `4`
##   <chr>             <int> <int> <int> <int> <int>
## 1 regular            1961    95    58    80    NA
## 2 regular with aide  1996    97    60    78    NA
## 3 small                NA   576   272   195   857
```

```r
#differences in math and reading scores based on years in a small classroom
STAR |>
  group_by(yearssmall) |>
  summarize(reading_avg = mean(g4reading, na.rm = TRUE),
            reading_median = median(g4reading, na.rm = TRUE),
            math_avg = mean(g4math, na.rm = TRUE),
            math_median = median(g4math, na.rm = TRUE)) |>
  knitr::kable(digits=3)
```

yearssmall	reading_avg	reading_median	math_avg	math_median
0	719.875	722.0	707.979	710
1	723.147	724.5	707.552	709
2	717.868	720.0	711.914	714
3	719.899	721.0	709.617	712
4	724.665	726.0	710.052	711

The analysis suggests that in general, spending all four years in the small classes marginally increases both reading and math test scores. However the effect is not very big at all. The analyses based on the mean and median yield similar results.

5

Question 6: Subgroup effects

We next examine whether the STAR program had differential effects among different groups of students. Evaluate the treatment effect between small classes and regular classes with no aide among white students and black students respectively. Give a brief substantive interpretation of the results of your analysis.

Answer 6

```r
#differences in reading
STAR |>
  filter(race %in% c("white", "black"),
         kinder %in% c("small", "regular")) |>
  group_by(race, kinder) |>
  summarize(reading_avg = mean(g4reading, na.rm = TRUE)) |>
  pivot_wider(names_from = kinder,
              values_from = reading_avg) |>
  mutate(reading_score_diffs = small - regular)
```

```
## # A tibble: 2 x 4
## # Groups:   race [2]
##   race  regular small reading_score_diffs
##   <chr>   <dbl> <dbl>               <dbl>
## 1 black    689.  699.                9.26
## 2 white    725.  728.                2.72
```

```r
#differences in math
STAR |>
  filter(race %in% c("white", "black"),
         kinder %in% c("small", "regular")) |>
  group_by(race, kinder) |>
  summarize(math_avg = mean(g4math, na.rm = TRUE)) |>
  pivot_wider(names_from = kinder,
              values_from = math_avg) |>
  mutate(math_score_diffs = small - regular)
```

```
## # A tibble: 2 x 4
## # Groups:   race [2]
##   race  regular small math_score_diffs
##   <chr>   <dbl> <dbl>            <dbl>
## 1 black    699.  698.           -1.03
## 2 white    711.  711.           -0.220
```

In terms of reading test scores black students benefited more from the smaller classrooms, scoring roughly 9 points higher on reading than those in regular classes while white students only scored 2.7 points higher. The same conclusion, however, does not apply to math scores where both groups performed slightly woorse in the small classrooms with the decrease among black students a full point where it is only a fraction of a point for white students.

Question 7: Long term effects

We next consider the long term effects of kindergarten 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. Briefly discuss the results.

Answer 7

```r
## graduation rate by classsize
STAR |>
  group_by(kinder) |>
  summarize(avg_hsgrad = mean(hsgrad, na.rm = TRUE))
```

```
## # A tibble: 3 x 2
##   kinder            avg_hsgrad
##   <chr>                  <dbl>
## 1 regular                0.825
## 2 regular with aide      0.839
## 3 small                  0.836
```

```r
## By years in the small class
STAR |>
  group_by(yearssmall) |>
  summarize(avg_hsgrad = mean(hsgrad, na.rm = TRUE))
```

```
## # A tibble: 5 x 2
##   yearssmall avg_hsgrad
##        <dbl>      <dbl>
## 1          0      0.829
## 2          1      0.791
## 3          2      0.813
## 4          3      0.832
## 5          4      0.878
```

We observe little difference in graduation rates across students who were assigned to different class types in kindergarten. 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.