

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
<code>race</code>	Student’s race (White = 1, Black = 2, Asian = 3, Hispanic = 4, Native American = 5, Others = 6)
<code>classtype</code>	Type of kindergarten class (small = 1, regular = 2, regular with aide = 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: 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`?

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()`.

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`.

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.

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.

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.

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.