Problem Set #2

Econ224

Due Date: Sunday, September 9th by 11:59pm

Instructions

Submit your code and solutions to the following questions as an RMarkdown document. In particular, upload both the `.Rmd` file you used to generate your report and the resulting `.html` output to *canvas* by the due date listed above. You may discuss this problem set with your classmates, provided that you adhere to the *empty hands* policy: after any such discussion, all parties must leave the room *empty-handed* i.e. without code files or written notes or any kind. In other words, the final code and write-up that you produce must be entirely your own work. If you discuss the problem set with any other students be sure to list their names at the top of your problem set. You are likewise welcome to consult printed or internet resources provided that you cite them. Violations of this policy constitute cheating and will be reported to the Office of Student Conduct.

Data Description

This exercise is partly based on the following paper: 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.

The STAR (Student-Teacher Achievement Ratio) Project was a four year *longitudinal study* examining the effect of class size in early grade levels on educational performance and personal development. 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. The data are contained in a CSV file called `STAR.csv`. Details on how to download and read this file appear below under "Questions." Here is a description of the variables in `STAR.csv`:

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 aid = 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 many missing observations in this data set. These arise for various reasons. For example, some students left a STAR school before third grade or did not enter a STAR school until first grade. Recall that `na.rm = TRUE` can be added to functions such as `mean` to remove missing observations.

Questions

Before beginning, download `STAR.csv` from http://ditraglia.com/econ224/STAR.csv and read it into a tibble called `star` using the function `read_csv` from the `readr` package. (Recall that `readr` is part of `tidyverse` so you don't have to install or load it separately.)

1. Add a new variable called `kinder` to `start`. This variable should recode `classtype` by changing integer values to their corresponding informative labels (e.g. change 1 to `small`). Similarly, recode the `race` variable so that it takes on the values `white`, `black`, `hispanic`, `others` by combining Asians and Native Americans to create the `others` category. Overwrite the original `race` variable in `star` rather than creating a new one.
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? Do students in the smaller classes perform better on average? Interpret your results. How large are the differences of means compared to the standard deviations of test scores? Can you reject the null hypothesis that there is no difference in average test scores against the two-sided alternative? Interpret your results.
3. Suppose that we wanted to compare *quantiles* of reading and math test scores rather than means. How do the 33th and 66th percentiles of each test score differ between small and regular class sizes? (This kind of comparison is called a *quantile treatment effect*.) Compare with your results from Question 2.
4. 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 students of each type are in the data set? Compare the average and reading and math test scores across students who spent different numbers of years in small classes. Then form two groups: students who spent all four years in regular classrooms versus students who spent all four years in small classrooms. Using these groups, repeat your statistical tests from Question 2. Briefly discuss your findings.
5. Did the STAR program reduce achievement gaps between different racial groups? To answer this question, first compute average reading and math scores separately for minority (defined as black or hispanic) and white students, grouped by `kinder`. Explain your results.
6. Fourth grade test scores are a short-term outcome. But the STAR dataset also contains information on a longer-term outcome: high school graduation. Repeat Questions 2, 4, and 5 using `hsgrad` in place of test scores. Discuss your findings.