SSI1005 Lab 8: Experiments

Dr Hannah Bunting

20/11/2023

EFFICACY OF SMALL CLASS SIZE IN EARLY EDUCATION

The STAR (Student-Teacher Achievement Ratio) Project is a 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 and 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, the collection of various measurements (e.g., performance on tests in eighth grade, overall high-school GPA) continued through to the end of participants' high-school attendance. We will analyse just a portion of this data to investigate whether the small class sizes improved educational performance or not. The data file name is STAR.csv, which is in csv format. The names and descriptions of variables in this data set are displayed in the table below.

Variable	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 the math portion of the fourth-grade standardized test
g4reading	total scaled score for the reading portion of the fourth-grade standardized test
yearssmall	number of years in small classes
hsgrad	high-school graduation (did graduate = 1, did not graduate = 0)

Load the data into RStudio and load the `tidyverse` package into your RStudio session.

QUESTION 1

In the experiment, kindergarten pupils were randomly assigned into one of three groups. These are identified in the `classtype` variable. We want to look at just the differences between small classrooms and regular sized classrooms, i.e. Treatment group 1 and the control group.

Create a new dataframe that contains just these two groups. The new dataframe should be called `star_2`.

QUESTION 2

Remember from the lecture that one problem which arises in experiments is that subjects do not comply with their randomly allocated status. The number of years spent in a small classroom is recorded in the `yearssmall` variable. Some pupils may have spent time in a small class after their initial assignment to either the treatment or control group. We want those in the treatment category to have only been in regular classes.

Remove any pupils who are in the regular category and have been in a small class.

QUESTION 3

It's important to make sure you interpret any results from your analyses correctly. Create a new factor variable called `classtype_fac` that gives each group a label. This will help us identify the treatment and control groups accurately.

QUESTION 4

The experiment also recorded the ethnicity of each pupil using the `race` variable. We want to check if there are any systematic differences between the two groups in their ethnicities. Most pupils in the study identified as either white or black.

Create a new character variable that writes the categories in text and combines the 'Asian', 'Hispanic' and 'Native American' categories with the 'others' category, meaning the new variable has a total of 3 categories. Name the new variable `race_3cats`.

QUESTION 5

Create a crosstab of proportions showing each new race category variable by the classroom types factor variable.

What proportion of pupils are black in the treatment group?

What proportion of pupils are black in the control group?

QUESTION 6

Recall from the lecture that attrition can be a problem in all research studies, including experiments. The 'hsgrad' variable shows whether the pupil graduated high school many years after the class size experiment was conducted. Missing values indicate the study could not record whether they graduated because the pupil had dropped out of the study.

Quantify the attrition rate of the study in two ways:

1. What number of pupils dropped out of the study before their graduation year?

2. What proportion of pupils remained in the study until graduation?

QUESTION 7

We can now investigate whether the classroom size had any causal effect on improving pupil outcomes. Find the average treatment effect of being in a small classroom on Grade 4 reading scores, recorded in the `g4reading` variable.

How would you interpret this finding?

QUESTION 8

In addition to quantifying them, we can also visualise the differences across groups to communicate our findings.

Create a boxplot to visualise the differences in reading scores across the treatment and control groups.

QUESTION 9

Find the average treatment effect of being in a small classroom on Grade 4 maths scores, recorded in the `g4math` variable.

QUESTION 10

Create a boxplot to visualise the differences in maths scores across the treatment and control groups.

End