

Lab 3: Small Classes and Early Education

David Sichinava

October 18, 2019

Topics

- Read external files
- Recoding variables
- Simple arithmetic operations

Instruction:

Follow the assignment step-by-step. Name your *.rmd* file in a following format: *your_surname_lab3.rmd*. For example,

```
sichinava_lab3.Rmd
```

Introduction:

This assignment is based on the study by [Frederick Mosteller](#) published in the *The Future of Children* in 1995. Mosteller wanted to identify how does class size affect students' academic achievement and whether the effect is long-term.

Data Manipulation and Descriptive Statistics

First of all, create a new notebook and read the data to R. Do not forget to indicate working directory. As you may notice, we name the newly created table *star*. You can read data files from the web, or alternatively, you may want to download them to your computer and read offline.

```
### link to download data file: https://goo.gl/pnNug3

star <- read.csv("STAR.csv")

## Warning in file(file, "rt"): cannot open file 'STAR.csv': No such file or directory
## Error in file(file, "rt"): cannot open the connection
```

It is a rule of thumb that you do a descriptive statistics before you start analyzing your data. First let's check the variables included in the *star* table:

```
names(star)

## Error in eval(expr, envir, enclos): object 'star' not found
```

Display the dimensions of *star* table and the descriptive statistics (mean, maximum, minimum and median values). You can find the corresponding syntax [here](#) and [here](#).

There are seven variables in the *star* table which take the following values:

Description	
Variable and Variable Value	
race	Race (1-white, 2-black, 3-asian, 4-hispanic, 5-native American, 6-Other)
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)

Create a new variable *kinder* out of *classType* variable. 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 the Asian and Native American categories with the others category. For the *race* variable, overwrite the original variable in the data frame rather than creating a new one.

```
### Create a new variable

star$kinder<-factor(star$classtype,
  levels = c(1,2,3),
  labels = c("small", "regular", "regular w/aid")
)

## Error in factor(star$classtype, levels = c(1, 2, 3), labels = c("small", : object
'star' not found

table(star$kinder)

## Error in table(star$kinder): object 'star' not found

### Recode race into new categories:

table(star$race)

## Error in table(star$race): object 'star' not found
```

```
star$race[star$race == 4] <- 4
## Error in star$race[star$race == 4] <- 4: object 'star' not found
```

Great. The hardest part is behind. Now let's check what do variables *g4math* and *g4reading* look like. These variables are so called *quantitative variables*, therefore you should calculate their mean, median and standard deviation. The necessary syntax was given during the previous presentation.

Analysis

Mosteller was interested whether the reading and numeric comprehension of fourth-graders differ across classrooms of different sizes. In order to check that we have to calculate, what were *average scores* in math and reading.

The following code will calculate the mean values for small classes. Use the same approach and calculate means for regular and regular with aid classes. Notice that we also indicate *na.rm = TRUE* option. It tells *R* to ignore missing values. Otherwise the code will break.

```
mean(star$g4math[star$kinder=="small"], na.rm=TRUE)
## Error in mean(star$g4math[star$kinder == "small"], na.rm = TRUE): object 'star'
not found
```

Based on these results, write a paragraph summarizing the effect of class size on the academic achievements of the students. Additionally, use *standarddeviation* of scores, which would help you evaluating how dispersed or concentrated the scores are.

Apart from the average, you may also want to assess high and low scores, that is a *quantile effect* of the experiment. To put it in simple terms, the experiment may have no effect on the average, however, it may affect students who scored high or low. Maybe small classes are better for better students? Let's calculate scores for low (33%) and high (66%) quantiles of math and reading scores and write a short interpretation. The example below describes how to calculate these scores for reading in small classes.

```
quantile(star$g4reading[star$kinder=="small"], c(.33, .66), na.rm=TRUE)
## Error in quantile(star$g4reading[star$kinder == "small"], c(0.33, 0.66), : object
'star' not found
```

As we can see, only a portion of kids studied in small classes throughout all four years. Others were put back to the regular classes or were assigned to the regular classes with special aid. Using *kinder* and *yearssmall* variables calculate the proportion of students who spend one, two, three, or four years in small classes. How do you think, does spending long time in small classes affect kids' math and reading comprehension? Hint: you should calculate the mean and the median of scores for each year. Write a short interpretation.

Assess the difference in reading and math comprehension between white and black kids. However, in this case use data for only regular and small classes. Calculate average scores of reading and math comprehension for each group. Write a short paragraph describing the results.

Finally, examine the long-term effect of spending time in small classes. Compare the graduation level for each class type and examine, whether there is a difference in graduation level between kids who spend different time in small classes. Additionally, check whether spending time in small classes yielded different graduation level among white and minority kids. Write a short interpretation of your analysis.

Annotated bibliography

Höfler, M. (2005). [Causal inference based on counterfactuals](#). *BMC medical research methodology*, 5(1), 28.

Holland, P. W. (1986). [Statistics and causal inference](#). *Journal of the American statistical Association*, 81(396), 945-960.

Thanks God, finished! Want to submit my assignment A.S.A.P!

Zip the whole folder for the second lab. Name the file according to the following format: *surname_lab3.zip* like this:

```
sichinava_lab3.zip
```

Upload the file to [this link](#) before the start of our next meeting.