Causality Recap Assignment

Dr. Soliman

Make sure you answer each question (and subsequestions!) to receive full credit, and use your lecture notes and any other resources to ensure these answers are complete.

Load the *STAR* data from here and assign it to an object called `star_df`. Read the help for the data here to understand what the variables correspond to. (Note: the data has been *reshaped* so don't mind the "k", "1", etc. in the variable names in the help.)

1. Filter the STAR dataset to only keep first graders and the small class and regular class groups. Check the data, as there still may be missing information... Call this object `star_df_clean`. Hint: If you wanted just grade 2, you would use a pipe with `filter(grade == "2")`. In this case, you want to keep first graders in either small or regular class groups, and therefore you would need to change and extend the previous `filter()`.

```
star_df <- read.csv("https://www.dropbox.com/s/bf1fog8yasw3wjj/star_data.csv?dl=1")

# one way to filter
star_df_clean <- star_df %>% filter(complete.cases(.)) %>%
    filter(star %in% c("small","regular") & grade == "1")

# alternative way, but if you don't filter complete cases,
# then you need to add something to the mean function below
star_df_clean <- star_df %>%
  filter(grade == "1") %>% filter(star == "small" | star == "regular")
```

2. Compute the average math score for both groups, and the difference between the two. (Use base `R`.)

```
# one way
mean_small = mean(star_df_clean$math[star_df_clean$star == "small"], na.rm = TRUE)
# alternative way
mean_small = mean((star_df_clean %>% filter(star == "small"))$math, na.rm = TRUE)
mean_small
```

```
## [1] 538.6777
```

```
mean_regular = mean((star_df_clean %>% filter(star == "regular"))$math, na.rm = TRUE)
mean_regular
```

```
## [1] 525.2744
```

```
ATE = mean_small - mean_regular
ATE
```

```
## [1] 13.4033
```

3. Create a dummy variable `treatment` equal to `TRUE` if student is in treatment group (i.e. small class size) and `FALSE` if in control group (i.e. regular class size). *Hint:* you can create the dummy variable

with `treatment = (star == "small")`.

```
star_df_clean <- star_df_clean %>%
    mutate(treatment = (star == "small"))
table(star_df_clean$treatment)
```

```
##
## FALSE  TRUE
##  2584  1925
```

4. Regress math score on the treatment dummy variable. Are the results in line with question 2?

```
lm(math ~ treatment, star_df_clean)
```

```
##
## Call:
## lm(formula = math ~ treatment, data = star_df_clean)
##
## Coefficients:
##   (Intercept)  treatmentTRUE
##         525.3           13.4
```

Yes, the coefficient on treatmentTRUE, the treatment assignment indicator for being in a small class, is identical to what we found in question 2.

5. How do you interpret these coefficients? Is this a causal estimate? Why or why not?

The intercept corresponds to the expected math score for first graders in the control group, defined as a regular-sized class. In other words the expected math score of first grades in the control group is 526,44. You can see this by yourself by comparing the coefficient to the same average computed in question 2. The slope coefficient corresponds the difference in expected math scores between first graders in the treatment and in the control group. In other words, first graders in small classes are expected to score 12.65 points higher than students in regular-sized classes. Again, you can compare this coefficient with the difference in averages computed in question 2. This is a causal estimate because the data is generated by a randomized experiment.