Problem Set 1

Your Name - Net ID - Section Number

Due Oct 7th, 2022

This homework must be turned in on Brightspace by Oct 7th 2022. It must be your own work, and your own work only – you must not copy anyone's work, or allow anyone to copy yours. This extends to writing code. You may consult with others, but when you write up, you must do so alone.

Your homework submission must be written and submitted using Rmarkdown. No handwritten solutions will be accepted. You should submit:

1. A compiled PDF file named yourNetID_solutions.pdf containing your solutions to the problems.
2. A .Rmd file containing the code and text used to produce your compiled pdf named yourNetID_solutions.Rmd.

Note that math can be typeset in Rmarkdown in the same way as Latex. Please make sure your answers are clearly structured in the Rmarkdown file:

1. Label each question part
2. Do not include written answers as code comments.
3. The code used to obtain the answer for each question part should accompany the written answer. Comment your code!

Definitions and Examples (20 points)

Answer the following questions. Be as specific and detailed as possible. Give examples.

1. What is the fundamental problem of causal inference?
2. Why are experiments important?
3. What does ignorability mean?
4. What is SUTVA?

Application (Coding) (20 points)

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 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 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
`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 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.

1. Create a new factor variable called `kinder` in the data frame. 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 Asians and Native Americans as the `others` category. For the `race` variable, overwrite the original variable in the data frame rather than creating a new one. Recall that `na.rm = TRUE` can be added to functions in order to remove missing data.

```
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag

## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
## read in data
STAR <- read.csv("STAR.csv")
# Things to do in this chunk :

# Create a new factor variable called `kinder` in the data
#   frame.  This variable should recode `classtype` by changing
#   integer values to their corresponding informative labels (e.g.,
#   change 1 to `small` etc.)
STAR = STAR %>% mutate(kinder = case_when(classtype == '1' ~ "small",classtype == '2' ~ "regular",classt

# recode the
#   `race` variable into a factor variable with four levels
#   (`white`, `black`, `hispanic`, `others`) by
#   combining Asians and Native Americans as the `others`
#   category. ----(For the `race` variable, overwrite the original
  # variable in the data frame rather than creating a new one.  Recall
  # that `na.rm = TRUE` can be added to functions in order to
  # remove missing data)---
STAR = STAR %>% mutate(race = case_when(race == '1' ~ "white",race == '2' ~ "black",race == '4' ~ "hispa
head(STAR,5)
```

```
##    race classtype yearssmall hsgrad g4math g4reading           kinder na.rm
## 1 white         3          0     NA     NA        NA regular with aid FALSE
## 2 black         3          0     NA    706       661 regular with aid FALSE
## 3 white         3          0      1    711       750 regular with aid FALSE
## 4 black         1          4     NA    672       659            small FALSE
## 5 white         2          0     NA     NA        NA          regular FALSE
```

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? Use means to make this comparison while removing missing values. Give a brief substantive interpretation of the results. To understand the size of the estimated effects, compare them with the standard deviation of the test scores.

```
grad4_small = STAR %>% filter(kinder=='small') %>% filter(!is.na(g4math) & !is.na(g4reading))

# mean and standard deviation for math scores on small class
grad4_small_math_mean = mean(grad4_small$g4math)
grad4_small_math_std = sd(grad4_small$g4math)

grad4_small_math_mean
```

```
## [1] 709.3213
```

```r
grad4_small_math_std
```

```
## [1] 43.70478
```

```r
# mean and standard deviation for reading scores on small class
grad4_small_read_mean = mean(grad4_small$g4reading)
grad4_small_read_std = sd(grad4_small$g4reading)

grad4_small_read_mean
```

```
## [1] 723.518
```

```r
grad4_small_read_std
```

```
## [1] 51.36679
```

```r
grad4_regular = STAR %>% filter(kinder=='regular') %>% filter(!is.na(g4math) & !is.na(g4reading))

# mean and standard deviation for math scores on regular class
grad4_regular_math_mean = mean(grad4_regular$g4math)
grad4_regular_math_std = sd(grad4_regular$g4math)

grad4_regular_math_mean
```

```
## [1] 709.4022
```

```r
grad4_regular_math_std
```

```
## [1] 41.0294
```

```r
# mean and standard deviation for reading scores on regular class
grad4_regular_read_mean = mean(grad4_regular$g4reading)
grad4_regular_read_std = sd(grad4_regular$g4reading)

grad4_regular_read_mean
```

```
## [1] 719.7275
```

```r
grad4_regular_read_std
```

```
## [1] 53.10735
```

Answer: For math scores, there are no significant difference in the means. The standard deviations for the regular size class is smaller which might indicate that the performance on the regular size class was similar amongst their peers. For the reading scores, the small sized class performed slightly better by comparing the mean score and the standard deviation for that class was smaller, which might indicated that the students perfomed closely on that type of test.

3. Instead of comparing just average scores of reading and math tests between those students assigned to small classes and those assigned to regular-sized classes, look at the entire range of possible scores. To do so, compare a high score, defined as the 66th percentile, and a low score (the 33rd percentile) for small classes with the corresponding score for regular classes. These are examples of *quantile treatment effects*. Does this analysis add anything to the analysis based on mean in the previous question? (Hint: You will use the quantile() function in r.)

```
#taking the quantiles for the small class
quantile(grad4_small$g4math,na.rm = T,probs = c(0.33,0.66))
```

```
## 33% 66%
## 695 726
```

```
quantile(grad4_small$g4reading,na.rm = T,probs = c(0.33,0.66))
```

```
## 33% 66%
## 705 741
```

```
#taking the quantile for the regular size class
quantile(grad4_regular$g4math,na.rm = T,probs = c(0.33,0.66))
```

```
## 33% 66%
## 696 724
```

```
quantile(grad4_regular$g4reading,na.rm = T,probs = c(0.33,0.66))
```

```
##    33%    66%
## 704.56 740.00
```

Conclusions: In general the scores for the small class upper quantile were higher. But the difference is so small that just looking at the means might not be enough to conclude the actual impact this treatment had. Also, the groups might contain individuals who would be going to get a good score regardless of treatment received.

4. We examine whether the STAR program reduced the achievement gaps across different racial groups. Begin by comparing the average reading and math test scores between white and minority students (i.e., Blacks and hispanics) among those students who were assigned to regular classes with no aid. Conduct the same comparison among those students who were assigned to small classes. Give a brief substantive interpretation of the results of your analysis.

```
# this chunk will contain only code for the regular class.

# white group in the regular class
regularClass_white = STAR %>% filter(race=='white' & kinder =='regular' ) %>% filter(!is.na(g4math) & !i

# computing stats
regularClass_white  = regularClass_white %>% group_by(race)  %>%
                    summarise(mean_math = mean(g4math),
                              mean_read = mean(g4reading),
                              .groups = 'drop')

regularClass_white
```

```
## # A tibble: 1 x 3
##   race  mean_math mean_read
##   <chr>     <dbl>     <dbl>
## 1 white      711.      725.
```

```
regularClass_minority = STAR %>% filter((race=='hispanic'| race=='black' | race=='others') & kinder =='

# computing stats for the minority group
regularClass_minority = regularClass_minority %>% group_by(kinder)  %>%
                    summarise(mean_math = mean(g4math),
                              mean_read = mean(g4reading),
                              .groups = 'drop')
regularClass_minority
```

```
## # A tibble: 1 x 3
##   kinder  mean_math mean_read
##   <chr>       <dbl>     <dbl>
## 1 regular      699.      691.
```

```
# showing the difference in the means of the groups
dif_mean_math = (regularClass_white$mean_math - regularClass_minority$mean_math)
cat("difference on math scores:",dif_mean_math)
```

```
## difference on math scores: 12.30924
```

```
dif_mean_read = (regularClass_white$mean_read - regularClass_minority$mean_read)
cat("\ndifference on reading scores:",dif_mean_read)
```

```
##
## difference on reading scores: 33.96172
```

```
# this chunk will contain only code for the small class.

# white group in the small class
smallClass_white = STAR %>% filter(race=='white' & kinder =='small' ) %>% filter(!is.na(g4math) & !is.na

# computing stats
smallClass_white  = smallClass_white %>% group_by(race)  %>%
                    summarise(mean_math = mean(g4math),
                              mean_read = mean(g4reading),
                              .groups = 'drop')

smallClass_white
```

```
## # A tibble: 1 x 3
##   race  mean_math mean_read
##   <chr>     <dbl>     <dbl>
## 1 white      711.      728.
```

```r
# there are only 5 hispanic kids on the dataset, lol?
smallClass_minority = STAR %>% filter((race=='hispanic'| race=='black' | race=='others') & kinder =='sma

# computing stats for the minority group
smallClass_minority = smallClass_minority %>% group_by(kinder)  %>%
                    summarise(mean_math = mean(g4math),
                              mean_read = mean(g4reading),
                              .groups = 'drop')
smallClass_minority
```

```
## # A tibble: 1 x 3
##   kinder mean_math mean_read
##   <chr>      <dbl>     <dbl>
## 1 small       699.      700.
```

```r
# showing the difference in the means of the groups
dif_mean_math = (smallClass_white$mean_math - smallClass_minority$mean_math)
cat("difference on math scores:",dif_mean_math)
```

```
## difference on math scores: 12.50095
```

```r
dif_mean_read = (smallClass_white$mean_read - smallClass_minority$mean_read)
cat("\ndifference on reading scores:",dif_mean_read)
```

```
##
## difference on reading scores: 28.29051
```

Conclusions:

5. We consider the long term effects of kindergarden 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 classses. Finally, as done in the previous question, investigate whether the STAR program has reduced the racial gap between white and minority students' graduation rates. Briefly discuss the results.

Bed Nets and Malaria (20 points)

Article: Free Distribution or Cost-Sharing? Evidence from a Randomized Malaria Prevention Experiment by Jessica Cohen and Pascaline Dupas

Some economists have argued that "cost-sharing" makes it more likely that a product will be used (versus giving it away for free). Cohen and Dupas partnered with 20 Kenyan prenatal clinics to distribute subsidized anti-malarial bed nets. For each clinic, they varied the extent of the subsidy: either full (free bed-nets, $D_i = 1$) or partial (90% cheaper bed-nets, $D_i = 0$). They measure (among other things) whether women who received bed nets used them (Y_i).

1. What is $\mathbb{E}[Y_i|D_i = 0]$?
2. What is $\mathbb{E}[Y_i(1)]$?
3. What is $\mathbb{E}[Y_i(1)|D_i = 0]$?
4. Cohen and Dupas randomized treatment at the level of the clinic, but the outcomes of interest are at the individual level. Is there a violation of consistency/SUTVA? Why or why not? Argue your case.

Let's Help a Small Business! (20 points)

1. Imagine you are a consultant working with a restaurant to increase their number of customers. In the past, the owner spent a lot of money sending a postcard to every individual in the city. The owner says that every time they send out postcards, there's an increase in their business. Is the relationship between sending postcards and increased business for the restaurant causal? What do you need to know to say that this is a causal relationship? Be specific and give examples.

2. Imagine you are designing an experiment for this small business. You want to know whether advertising (sending a postcard) will increase a person's probability of going to the restaurant. How would you do it? (Detail what kind of experiment you would run. Explain how and why the experiment will provide an average treatment effect. What assumptions do you need to be true for the experiment to be causal.)

ATE (20 points)

A study of 10 dancers took place last year. The researchers wanted to know whether new shoes improved the dancers performance. We denote T = 1 to mean that the dancers received new shoes and T = 0 to mean the dancers used old shoes.

The table below contains the data. For each dancer, we know whether they were assigned to the treatment (T = 1) or control group (T = 0). We also see their observed outcome or quality of performance on a 0 to 100 scale. Lastly, we have information about whether the individual is currently a dance trainer (instructor) denoted by DT={yes, no}. Note: The empty cells indicate the counterfactual outcomes that are not possible to observe.

Dancer	Trainer(DT)	T	Y(1)	Y(0)
1	Yes	1	60	
2	Yes	1	75	
3	Yes	1	53	
4	Yes	1	69	
5	Yes	0		63
6	No	1	50	
7	No	0		42
8	No	0		50
9	No	0		58
10	No	0		59

1. Estimate the ATE using the data in the table.

$$ATE_1 = E[Y|T = 1] - E[Y|T = 0]$$

2. As we can see from the data, the treatment was given mostly to dance trainers. Now compute the dance-trainer-specific effect using the data in the table.

$$ATE_2 = E_{P(DT)}[E[Y|T = 1, DT = Yes] - E[Y|T = 0, DT = Yes]]$$

3. Which one of ATE_1 and ATE_2 estimates the ATE better? Briefly explain your choice.

Design Your Experiment (20 points)

Design your own experiment from start to finish. Choose an *interesting* question. Explain why observational data may give you the wrong answer. Detail the potential outcomes and a well-defined treatment. Explain the type of experiment (completely random, cluster-design, block/stratified). Will your design ensure a causal treatment effect? (Remember: Be as specific as possible and give examples.)