Problem Set 1. Definitions

Relevant material will be covered by **Aug 24**. Problem set is due **Aug 31**.

Welcome to the problem set! This homework will practice conceptual and notation ideas for descriptive and causal inference.

To complete the problem set, **Download the .Rmd** and complete the homework. Omit your name so we can have anonymous peer feedback. Compile to a PDF and submit the PDF on Canvas.

1. Practice with potential outcomes

Jose says that coming to Cornell caused him to discover statistics, and it became his major! He says that if he had gone to NYU, he would have stuck with biology.

1.1 (7 points)

Answer. In Jose's claim, what is the treatment?

Jose's treatment was choosing to attend Cornell University

1.2 (7 points)

Using the mathematical notation we discussed in class, define the two potential outcomes to which Jose is referring **Answer.**

If Jose chooses to attend Cornell than his major would be Statistics

$$Y_{Jose}^{Cornell} = Statisitcs$$

If Jose chooses to attend NYU then his major would be Biology

$$Y_{Jose}^{NYU} = Biology$$

1.3 (7 points)

In a sentence or two, say how the Fundamental Problem of Causal Inference applies to Jose's claim.

Answer. The fundamental Problem of Causal Inference when applied to Jose's claim is that we have him going to Cornell and then picking up statistics, we don't have the data of Jose if he had stayed in NYU to see if he would continue only doing Biology or if he would have eventually discovered statistics. Its impossible to get both pieces of data to test.

1.4 (7 points)

Using conditional expectations or probabilities, write the following in math: the probability of majoring in statistics is higher among students who attend Cornell than among students who attend NYU.

Answer.

$$P(Degree = Statisitcs|College = Cornell) > P(Degree = Statisitcs|College = NYU)$$

1.5 (7 points)

Give one reason why the average causal effect of attending Cornell versus NYU (the quantity from 1.2, averaged over all students) might be different from the average descriptive difference (from 1.4) in rates of majoring in statistics.

Answer. The reason why the causal effect of attending Cornell versus NYU will be different from the average descriptive difference is because one can be obtained just from looking at numbers while the other can't really be tested properly as a second option can not be tested. It is important to remember that when looking at causal effects we can only look at one outcome, which is the problem of causal inference, while with descriptions it is just simply observations, so it makes easier to get both sides. To boil it down both values will be different because one doesn't give you the full story.

2. A sailing class

You are looking into a sailing class through Cornell Wellness! For each claim below, tell us whether the claim is causal or descriptive.

2.1 (5 points)

Last year, there was a survey of students who did and did not take the class. The proportion reporting that they felt prepared to sail on Cayuga Lake was higher among those who took the class.

Answer. This would be considered a descriptive claim, as they are only stating that they students felt more prepared after taking the class. There was no claim that stated if they didn't take the class that they wouldn't be prepared.

2.2 (5 points)

Last year, there was a survey of students before and after the class. The proportion reporting that they felt prepared to sail on Cayuga Lake was higher in the survey taken after the class.

Answer. This would be considered a descriptive claim, as again they are only stating that they students felt more prepared after taking the class. There is still no claim that stated if they didn't take the class that they wouldn't be prepared.

2.3 (5 points)

On average, the students in this class emerged more prepared to sail than they would have been without the class.

Answer. This would be considered a casual claim, as they claim that the students where more prepared because they took the class, implying that if they didn't, they would not be prepared.

Session info

The chunk below will record information about your R session, which is useful for debugging issues in homework assignments that contain code.

```
sessionInfo()
```

```
## R version 4.2.2 (2022-10-31 ucrt)
## Platform: x86_64-w64-mingw32/x64 (64-bit)
## Running under: Windows 10 x64 (build 22621)
##
## Matrix products: default
##
## locale:
## [1] LC_COLLATE=English_United States.utf8
## [2] LC_CTYPE=English_United States.utf8
## [3] LC_MONETARY=English_United States.utf8
## [4] LC_NUMERIC=C
## [5] LC_TIME=English_United States.utf8
##
## attached base packages:
## [1] stats     graphics  grDevices utils     datasets  methods   base
##
## loaded via a namespace (and not attached):
##  [1] compiler_4.2.2    fastmap_1.1.1     cli_3.6.1         tools_4.2.2
##  [5] htmltools_0.5.6   rstudioapi_0.15.0 yaml_2.3.7        rmarkdown_2.24
##  [9] knitr_1.43        xfun_0.40         digest_0.6.33     rlang_1.1.1
## [13] evaluate_0.21
```