

Lab 4

Insert Name

Math 141, Week 4

Due: Before your Week 5 lab meeting

Goals of this lab

- Identify variables and their roles in the study.
- Identify forms of randomness and their implications on the conclusions of the study.
- Identify important data collection features.

Problems

Problem 1

This is problem we will explore sampling strategies for estimating the prevalence of the SARS-CoV-2 infection in Portland, OR.

- a. Feel free to work on this problem with up to 3 other people. List who you worked with here. (If you worked alone, you can leave it blank.)

- b. Current CDC Guidelines recommend the following people get tested: “People who have symptoms of COVID-19, people who have had close contact (within 6 feet of an infected person for at least 15 minutes) with someone with confirmed COVID-19, People who have been asked or referred to get testing by their healthcare provider, local or state health department.” If we are interested in estimating the proportion of people in Portland with COVID-19, will this sample of people be representative? Why or why not?

- No, the sample would leave out individuals who have not met the CDC guidelines. It would count only 2 of the 4 possible groups:
 - **People referred and positive:** Individuals who have some known risk factor, are tested, and are Covid POSITIVE
 - **People referred and negative:** Individuals who have some known risk factor, are tested, and are Covid NEGATIVE
 - **NOT referred and positive:** Individuals who are not tested, and are Covid POSITIVE
 - **NOT referred and negative:** Individuals who are not tested, and are Covid NEGATIVE
- Because the sample would cover only the first two groups (which in all likelihood are the larger groups with less cases).
- We haven’t gone over it in class, but this would be a measure of precision as opposed to percent infected (accuracy)

-
- c. Suppose you have 1000 tests and want to estimate the proportion of Portland residents with COVID. Provide a detailed plan of how you will collect the data that is **representative** and **efficient**. You might consider simple random sampling, stratified sampling, cluster sampling, a hybrid, or another random technique all together. Your plan should be explicit enough that someone else could follow in your plan.
-

- This is one of the most unethically scientific thing imaginable. But this counts for my grade, soooooooooo. . . .
 - **Controlling for those that have not recieved a test**
 - The issue is also that we only want those that are currently infected. So we will count only those that have been tested in the past two weeks from the. And we will estimate those that have not been tested. A form of stratified sampling.
 - With the 1,000 tests we have we will need to control for location.
 - Test individuals that are evenly distibuted among population density (not land area). Only given to those who have never been tested.
 - (percent population tested outside of the study multiplied by the percent positive within the past two weeks) PLUS (one minus percent population tested outside of the study multiplied by the percent positive tests from the study sample
-

- d. Justify why your sampling method is representative and efficient.
-

- Takes advatage of those that have already been tested
 - Accounts for geographical differences
-

Problem 2

A 2006 Washington Post article stated that for the essay portion of the 2005-2006 SAT, the students who wrote in cursive scored significantly higher on the essay, on average, than those who wrote in printed block letters. Researchers wanted to know whether simply writing in cursive would be a way to increase scores.

- a. Identify the explanatory and response variables.
-

- Explanatory: Writting style
 - Response: SAT score
-

- b. Can we conclude that using cursive causes a student to receive a better score? If so, explain why. If not, identify a potential confounding variable and explain how it provides an alternative explanation for why the cursive group had a higher average essay score.
-

- No, we can only say there is a causal effect with students that write in cursive and recieve higher SAT (post 2005) scores.
- Confounding variables
 - Wealth: Students that come from more affluent backgrounds tend to be taught additional “premium topics” In addition they have access to further test prep material.

- Reviewer: Individuals reviewing the test might be biased to cursive, or they might specialize in cursive.
- Smaller sample: Smaller samples skew results. Block might skew higher and have outliers with lower scores
- Essay topic: Perhaps individuals are writing more elaborate essays in cursive.

c. The article also mentioned a study in which the same one essay was given to all graders but half were randomly given a cursive version and the other a printed block letter version. In this study, the cursive style scored significantly higher on average. Explain a key difference between this study and the first study.

-
- The researchers controlled for the essay topic and material, reducing confounding variables.

d. For the second study, can we now conclude a causal link? If so, explain why. If not, identify a potential confounding variable and explain how it provides an alternative explanation for why the cursive group had a higher average essay score.

-
- Yes, we can conclude a stronger causal link. Confounding variables have been reduced.
 - There is still the question of what is the SATs grading standards
-

Problem 3

Researchers from Harris Interactive were interested in determining what percent of people wash their hands after using the washroom. They collected data by standing in public restrooms and pretending to comb their hair or put on make-up as they observed patrons' behavior. Public restrooms were observed at Turner's Field in Atlanta, Penn Station and Grand Central Station in New York, the Museum of Science and Industry and the Shedd Aquarium in Chicago, and the Ferry Terminal Farmers Market in San Francisco. Of the over 6,000 people whose behavior was observed, 85% washed their hands.

a. What are the cases?

-
- 6,000 observed individuals
-

b. What is the variable of interest?

-
- Did the individual wash their hands?
-

c. Who is included in the sample? Who is missing? To what population are you willing to generalize the results to?

-
- People that go to public bathrooms in NY, Chicago, and SF landmarks
-

- d. In a separate telephone survey of more than 1000 adults, more than 96% said they always wash their hands after using a public restroom. Why do you think there is such a discrepancy in the percent from the telephone survey compared to the percent observed?

-
- People can be dishonest when asked point blank due to embarrassment.
 - Actions speak louder than words.
-

Problem 4

Researchers conducted a double-blind experiment to determine whether taking large amounts of Vitamin E protects against prostate cancer. To study this question, they enrolled 29,133 Finnish men, all smokers, between the ages of 50 and 69. The men were divided into two groups: one group took vitamin E and the other a placebo. The researchers followed all the men for eight years and found that the participants taking vitamin E had lower rates of prostate cancer.

- a. Explain how random assignment was used and why it was used.

-
- There was no bias in who went to which group (placebo or treatment)
-

- b. Explain what “double-blind” means in the context of the study and why it was used.

-
- Neither the scientists nor the individuals that partook in the study know what treatment they were receiving.
-

- c. Who is included in the sample? Who is missing? What population is it reasonable to generalize these results to? Explain your reasoning.

-
- 29,133 Finnish smokers between 50 and 69 who have prostates.
-

- d. For the population you described in c, is it reasonable to conclude that taking vitamin E causes a reduction in the probability of developing prostate cancer? Explain your reasoning.

-
- Yes, but I would like to know more.
 - I would like to know how much higher of a percentage and if there were any additional treatments in the study.
 - That being said, it is a sufficiently large sample where only apples are being compared to apples with placebo accounted for.
 - Measuring vs a placebo.
-

Problem 5

A recent study compared two groups of physicians: one group was reminded of the sacrifices that physicians have to make as part of their training and the other was given no such reminder. All physicians in the study were then asked whether they consider it acceptable for physicians to receive free gifts from industry representatives. In the “sacrifice reminders” group, 47.5% answered that gifts were acceptable compared to 21.7% in the “no reminder” group.

- a. Identify the explanatory and response variables.

- Explanatory variable: Reminded of “sacrifices that physicians have to make as part of their training”
- Response variable: Answers yes to: “[is] it acceptable for physicians to receive free gifts from industry representatives.”

- b. Does this study contain a control group? Explain.

- Yes, no mention of “sacrifices that physicians have to make as part of their training”

- c. Do you believe this study used random assignment? What would that involve?

- Yes, a doctor has a fifty fifty chance of being in either group. That would involve randomly grouping a doctor into the control or treatment group.

- d. Do you believe this study used random sampling? What would that involve?

- Yes and no. Not every doctor was asked to participate in the study, so a sample were selected to take part in the study. I strongly suspect that it was not evenly distributed at every hospital.

Problem 6

Some professional sports teams have better winning percentages when their home games are not sold out than when their home games are sold out.

- a. What is the response variable?

- Game win percentage (win or not win)

- b. What is the explanatory variable?

- Sold out or not sold out

- c. Identify a possible confounding variable. Make sure to explain how the suggested variable is related both to the response variable and the explanatory variable.
-

- Opponent difficulty
 - Fans tend not to go to games which are considered boring which can happen if there is a skew in talent
 - There is an example of this in college football where fans were asked to rank “rival score” (a higher score means fans rank a given college a larger rival).
 - Boston College gave Notre Dame 74.17 points and Notre Dame gave Boston College 1.71 points.
 - Boston College had a worse overall NCAA score than Notre Dame, so Notre Dame fans are less likely to go to home Notre Dame v BC games as the outcome is likely Notre Dame winning.
 - Boston College fans showed up to a game where they tended to lose.
 - <https://knowrivalry.com/media/>
 - <https://knowrivalry.com/category/ncaaf/>
-