Introduction

Sampling is a critical element of the research process. It is generally not feasible to study the entire population we are interested in, so instead we study a sample drawn from that population. As long as the sample is representative of the larger population, we can assume that the findings for the sample hold for the larger population as well. The quality of the sample determines the “generalizability” or “external validity” of the results. There are now well-established ways of ensuring a high quality sample.

Questions

1. Let’s think about our study of biking in Davis. Let’s say we’re planning to do a survey that will let you quantify levels of biking and characteristics that might influence biking. For the study, individuals are the unit of observation and adult Davis residents are the target population. We have lots of different options for sampling.

a. A good starting point is to think about possible sampling frames. Remember that the sampling frame is the list of elements for your population of interest, in this case, individuals living in Davis. Here are a couple of classic sampling frames you might use:

- Phone listings: What systematic biases might this sampling frame introduce into the study? What might you do to reduce this bias?
- Marketing listings (available for a small fee from commercial providers): What systematic biases might this sampling frame introduce into the study? What might you do to reduce this bias?
- Addresses (available through the Census bureau and local government sources): What is the unit of analysis for this sampling frame? If we still want to study individuals, how might we choose them, using this sampling frame? How can we ensure that we end up with a random sample of individuals?

b. Let’s say you use the address list as your sampling frame. What approach might you use for simple random sampling? What approach might you use for systematic random sampling?

c. Researchers use stratified sampling for three general purposes: when there isn’t one sampling frame for the entire target population but there are separate sampling frames for different subsets of the population (usually proportionate), to create more homogenous subsets of the target population and thus tighten the confidence intervals (usually proportionate), or to ensure sufficient numbers of respondents for groups that are a small share of the population (usually disproportionate).

What characteristic might you want to stratify the population by? Why?
What sampling frame would you use for each strata? Think carefully: is there a list of the elements of strata that you can get access to?
For proportionate sampling, how do you determine what share of the sample falls into each strata?
For disproportionate sampling, how do you determine what share of the sample falls into each strata?
Which one – proportionate or disproportionate - would you want choose for this project? Why?
d. Researchers use cluster sampling for convenience, usually when the elements of interest (Davis residents, in our case) are clustered into geographic or social groups and when the groups are easier to find than the elements themselves. Remember that you can use the same characteristic to define clusters as strata, but the process is different.

What is one type of cluster you could use for this study?
What is the sampling frame for the clusters? Once again: you need to be able to divide your sampling frame into clusters, or you must be able to find a separate sampling frame for each cluster.
How do you pick the clusters?
How would you pick the elements from within the clusters?

e. How big do you think the overall sample should be?

f. Do you think you could generalize the results for Davis to other populations? Why or why not?

2. Let’s say you decide to do a more exploratory study, one that uses the qualitative research methods we’ll be talking about in class. In this kind of study, we are not so concerned about generalizability, rather we want to understand more about the range of factors that influence whether someone bikes or not. That means we can use nonprobability sampling methods. Again, we have lots of choices.

a. What approach might you use for availability sampling?

b. What characteristic of the population might you be particularly interested in studying? What approach might you use for quota sampling for this characteristic?

c. What kind of people might be useful as “key informants” for this study? What approach might you use for sampling these informants, a technique called purposive sampling?

d. What hard-to-reach subset of the Davis population might you want to focus on? What approach might you use for snowball sampling for this population subset?

e. How large a sample do you think you need for this exploratory study? How do you know when you’ve got a big enough sample?

3. Let’s say you want to test the effect of bicycle infrastructure on levels of biking. Maybe the presence of bike paths and bike lanes in Davis encourage people to bike more here. What approach to sampling would you use to ensure that your bicycle infrastructure variable actually varies?

Don’t forget to have your recorder turn in your notes!

Readings for Thursday: Chapter 6 (in both books) on Causation and Research Design