STAT 107 – Fall 2020

Sampling Methods

Ha Khanh Nguyen (hknguyen)

Sampling

- In order to establish strong external validity, researchers need to make the case that their sample represents the population they are generalizing their findings to.
- Sample size is more of a consideration when deciding how much statistical power we have in our conclusions; larger samples allow us more confident conclusions about the population.
- Some statistical methods we'll discuss later might require a certain sample size to work properly.

Simple Random Sample (SRS)

- Simple random sampling has two requirements:
  - **Random Selection**: From the outset, every member of the population has an equal chance to be chosen for the sample.
  - **Independent Selection**: Every member of the population remains in equal consideration of being chosen as the sampling continues.
- Common biases:
  - **Under-coverage Bias**: Some in the population don't have an equal chance of being selected for the sample.
  - **Volunteer Bias** (Non-response Bias): The sample is composed of subjects who chose to participate and others who chose to ignore or forgot.

land-line → younger people

email → senior citizens

Stratified/Quota Sampling

- When we know ahead of time that we have an under-coverage bias and/or a non-response bias, we might consider certain demographic (or medical) characteristics of our population and attempt to get a representative sample across different key subgroups.
- Just as "blocking" is to sorting experimental groups, "quota sampling" is to gathering a sample.
- Within each bracket, participants may be selected randomly (or quasi-randomly sampled).

Convenience Sampling

- This non-random sampling method is quite common in people-centered research.
- In the case of a survey or poll that is posted in non-representative places, we should expect the sample will be quite biased.
- Be very cautious of using convenience samples to generalize to a larger population.

- **Snowball Sampling**: a common component of convenience sampling that relies on word of mouth (think shares on social media) to get participants. This would be a huge threat to independence in that people participate based on whether they know someone who did. The sample might be an echo chamber.

Practice: Identify which sampling design was used in each of these situations and any limitations to note:

A poll on msn.com asks American users 18 and over whether they plan to vote in the upcoming Midterm elections. After voting, the website encourages people to share the link of the poll with their friends on social media.

convenience sampling (thread to independence)

A clinic is surveying recent customers to assess satisfaction with their recent visits. Using the demographic information they have on their customers, the clinic works to get a certain number of responses from each gender, age bracket, and race/ethnicity.

quota sampling (no-response bias)

A university selects 50 graduating seniors by randomly selecting their email addresses from among those who have applied for graduation. These 50 students are asked to complete an exit interview for $25. 43 of them end up completing the survey.

simple random sampling (money for interview)

To-do:

- Finish Lab 2, commit and push the lab using git commands!
- Complete HW 1 on Prairie Learn!

This lecture notes were adapted from "Course Notes for STAT 100: Statistics" by Kelly Findley & Ha Khanh Nguyen.