Intro to Surveys and Sampling

Math 255, St. Clair

day 1

What is a survey?

- Study designed to get a snapshot of a population at a particular period of time.
- Results used by government policy planners, politicians, marketing departments, media, quality control managers, etc....
  - Demographic studies (migration, mortality, fertility, gender,...)
    - US Census Bureau
  - Natural resource management
    - MN Department of Natural Resources
  - Public health monitoring
    - National Health And Nutrition Examination Survey
  - Public opinion polls
    - Gallup, Roper, Harris, Pew Research Center
  - Marketing surveys
    - study consumer buying practices

Sampling Terminology

- **Element** An object on which a measurement is taken.
- **Population** A collection of all elements at a particular period of time.
  - **target population**: who we want to study
  - **sampled population**: who we actually sampled from

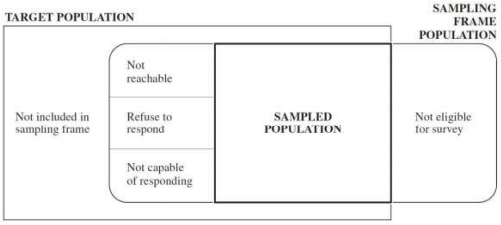

Figure 1: Target population and sampling frame (Lohr, 2010, p. 4)

Sampling Terminology

- **Sampling Unit** An element, or group of elements, which we actually sample. Sampling units are nonoverlapping and must cover the entire population.
- **Sampling Frame** A list of sampling units in a population (sampled population).
- **Sample** A collection of units drawn from a sampling frame.
- **Census** An (attempted) 100% sample from the population.
- **Sampling Design** A plan for selecting sampling units
- **Parameter** A particular population characteristic, usually a number, that we want to estimate in a survey.
- **Statistic** A number computed from a sample.
  - Statistics are used to estimate population parameters.
  - The formula used for an estimate depends on (1) the parameter of interest and (2) the sampling design used to collect the data.

Sampling vs. Non-sampling Errors

- **Sampling Error** Sample-to-sample variation in a statistic
  - Sampling error is quantifiable if a probability sampling design is used.
  - Sampling error is determined by the sampling design, the sample size, and the population variability.
- **Non-sampling Error** Systematic variation between sample data and a population data that can't be explained by sampling error. Can introduce *bias* into results.
  - selection bias
  - measurement error

Selection bias

- **Undercoverage** Occurs when the sampling frame does not cover the targeted population.
  - units in the undercoverage part of the population have no chance of being selected for the sample.
  - effect not easily measured or corrected.
- **Nonresponse** Failure to collect responses from all sampling units selected for a survey.
  - can't contact (no substitutions!!)
  - can't respond (e.g. illness, lack of knowledge)
  - refuse to answer (e.g. privacy, fear, embarrassment)
  - effect not easily measured but nonresponse rates can be reduced by careful planning (callbacks, mailings, incentives, shorten interview length)

Measurement Error

- **Measurement Instrument Errors** Occur due to inaccurate measurement tools/machines, poorly worded questions, inappropriate questions/questionnaire design, etc.
- **Interviewer bias** Occurs because people tend to give different answers to different interviewers.
- **Response Errors** People differ in motivation and ability to respond to survey questions.

Principle Steps in a Survey Sample

Taken from *Sampling Techniques*, 3rd ed. by W.G. Cochran (1977)

1. Why? Start by clearly stating the survey objectives.

2. Who? Carefully define the target population and elements. Construct a sampling frame that closely matches the target population.

Principle Steps in a Survey Sample

3. What to measure? Use objectives to determine the population parameters of interest (e.g. proportions, means, etc). Define relevant variables to measure and, if appropriate, design questionnaire. Identify which variables are categorical and which are quantitative.

4. How to measure? Decide how to gather data: direct observation, mail, web, telephone, in-person interview, or some combination. Determine data entry method and the form of the database.

Principle Steps in a Survey Sample

5. How to sample? Determine the amount of error that can be tolerated when estimating the population parameters of interest, then choose a sampling design and a sample size. Design type, time and budget constraints, and precision help to determine a sample size. Try out survey on a small scale (a pretest) to determine if any problems exist in questionnaire design, survey design or cost.

6. What to do with data? Summarize data and estimate population parameters of interest with standard errors. Check for mistakes in data entry. Learn from mistakes and use survey results to improve designs of future samples.

Questionnaire Design - some topics

Wording of Questions

- Keep questions simple but specific.
- Make sure all terms or measurement amounts are well-defined.
- Don't use leading or loaded questions or words.
- Don't use **double-barreled** questions that deal with 2 or more issues.
    - 2019 Trump tweet: "Wow! A Suffolk/USA Today Poll, just out, states, “50% of Americans AGREE that Robert Mueller’s investigation is a Witch Hunt.”
    - but here is the question asked: "President Trump has called the Special Counsel’s investigation a ‘witch hunt’ and said he’s been subjected to more investigations than previous presidents because of politics. Do you agree?"
    - https://www.factcheck.org/2019/03/trump-touts-questionable-survey-results/

Questionnaire Design - some topics

Question Ordering

- General vs. Specific: usually best to ask general questions prior to specific questions
- Order of possible responses
    - verbal questions: response more likely to be the last option in the list
    - written questions: response more likely to be the first option in the list
- Solutions: vary ordering, repeat question

Questionnaire Design - some topics

Open vs. Closed Questions

- Open question - respondents not prompted with categories for response. Good for exploratory surveys or sensitive questions but can be hard to analyze.
- Closed question - respondents given a list of possible responses to choose from.

Questionnaire Design - some topics

Response Options for Closed Questions

- Can't list all possible responses but pretesting of questions should narrow the list to the most common responses.
- Have other" category when appropriate.
- Detailed/narrow topic questions: Provide a screening question to assess an individuals knowledge of a specific issue.