Who is being surveyed?

Department of Political Science and Government
Aarhus University

September 15, 2014

1 Review of Last Week

2 New Material to Cover
- Total Survey Error
- Populations
- Representativeness
- Sampling Frames
- Sampling without a Frame
- Simple Random Sampling

3 Preview of Next Week

Paul Lazarsfeld (1901–1976)

- Originally from Austria; spent career at Columbia University
- Pioneered the study of mass media (Princeton Radio Project)
  - *The War of the Worlds* (1938)
- Created the survey panel to study radio impact
- First ever election surveys: *The People's Choice* (1940) and *Voting* (1948)
- Two-step flow of influence: *Personal Influence*

Criteria for Causal Inference

1. Relationship
2. Temporal precedence
3. Nonconfounding
4. Mechanism
5. Level of analysis

Assignment for this week

1. Form groups of 3 (or so)
2. Present your research question idea(s)
3. Give feedback to your peers on the idea
4. Share some with the whole class

2 New Material to Cover

- Total Survey Error
- Populations
- Representativeness
- Sampling Frames
- Sampling without a Frame
- Simple Random Sampling

Total Survey Error

- Envision the perfect survey!
- Errors introduced in design, implementation, and analysis
- Late 20th-century survey research focused on minimizing particular sources of error
- “Total Survey Error” approach is about trade-offs between all sources of error, costs, and time

2 New Material to Cover

- Total Survey Error
- Populations
- Representativeness
- Sampling Frames
- Sampling without a Frame
- Simple Random Sampling

Inference Population

- We want to speak to a population
- But what population is it?

Inference Population

- We want to speak to a population
- But what population is it?
- Example: “The Danish population”

Population Census

- All population units are in study

Population Census

- All population units are in study
- History of national censuses
  - Denmark 1769–1970 (sporadic)
  - U.S. 1790 (decennial)
  - India 1871 (decennial)

Population Census

- All population units are in study
- History of national censuses
  - Denmark 1769–1970 (sporadic)
  - U.S. 1790 (decennial)
  - India 1871 (decennial)
- Other kinds of census
  - Citizen registry
  - Commercial, medical, government records
  - “Big data”

Advantages and Disadvantages

- Advantages
- Disadvantages

Advantages and Disadvantages

- Advantages
  - Perfectly representative
  - Sample statistics are population parameters
- Disadvantages

Advantages and Disadvantages

- Advantages
  - Perfectly representative
  - Sample statistics are population parameters
- Disadvantages
  - Costs
  - Feasibility
  - Need

2 New Material to Cover

- Total Survey Error
- Populations
- Representativeness
- Sampling Frames
- Sampling without a Frame
- Simple Random Sampling

Representativeness

- What does it mean for a sample to be representative?

Obtaining Representativeness

- Quota sampling (common prior to the 1940s)

Obtaining Representativeness

- Quota sampling (common prior to the 1940s)
- Simple random sampling

Obtaining Representativeness

- Quota sampling (common prior to the 1940s)
- Simple random sampling
- Advanced survey designs (discuss next week)

Convenience Samples

- What is a convenience sample?

Convenience Samples

- What is a convenience sample?
- Different types:
  - Passive/opt-in
  - Sample of convenience (not a sample per se)
  - Sample matching
  - Online panels

Convenience Samples

- What is a convenience sample?
- Different types:
  - Passive/opt-in
  - Sample of convenience (not a sample per se)
  - Sample matching
  - Online panels
- “Purposive” samples (common in qualitative studies)

Questions about convenience samples?

2 New Material to Cover

- Total Survey Error
- Populations
- Representativeness
- Sampling Frames
- Sampling without a Frame
- Simple Random Sampling

Sampling Frames

- Enumeration (listing) of all units eligible for sample selection
- Two flavors:
  - Random sample from an ordered list
  - Systematic sampling from a randomized list
- Building a sampling frame
  - Combine existing lists
  - Canvass/enumerate from scratch

Big considerations

- Coverage!
  - Undercoverage
  - Overcoverage

Big considerations

- Coverage!
  - Undercoverage
  - Overcoverage
- What is a unit?

Big considerations

- Coverage!
  - Undercoverage
  - Overcoverage
- What is a unit?
- Clustering

Big considerations

- Coverage!
  - Undercoverage
  - Overcoverage
- What is a unit?
- Clustering
- Overlap between units

Big considerations

- Coverage!
  - Undercoverage
  - Overcoverage
- What is a unit?
- Clustering
- Overlap between units
- List maintenance

Multi-frame Designs

- Construct one sample from multiple sampling frames
- E.g., “Dual-frame” (landline and mobile)
- Analytically complicated
  - Overlap of frames
  - Sample probabilities in each frame

2 New Material to Cover

- Total Survey Error
- Populations
- Representativeness
- Sampling Frames
- Sampling without a Frame
- Simple Random Sampling

Sampling without a Sampling Frame

- Sometimes we have a population that can be sampled but not (easily) enumerated in full

Sampling without a Sampling Frame

- Sometimes we have a population that can be sampled but not (easily) enumerated in full
- Examples
  - Protest attendees

Sampling without a Sampling Frame

- Sometimes we have a population that can be sampled but not (easily) enumerated in full
- Examples
  - Protest attendees
  - Streams (e.g., people buying groceries)

Sampling without a Sampling Frame

- Sometimes we have a population that can be sampled but not (easily) enumerated in full
- Examples
  - Protest attendees
  - Streams (e.g., people buying groceries)
  - Points in time

Sampling without a Sampling Frame

- Sometimes we have a population that can be sampled but not (easily) enumerated in full
- Examples
  - Protest attendees
  - Streams (e.g., people buying groceries)
  - Points in time
- Population is the sampling frame

Rare or “hidden” populations

- Big concern: coverage!

Rare or “hidden” populations

- Big concern: coverage!
- Solutions?

Rare or “hidden” populations

- Big concern: coverage!
- Solutions?
  - Snowball sampling
  - Informant sampling
  - Targeted sampling
  - Respondent-driven sampling
- How does RDS work?

Questions?

Activity!

- Work in pairs
- Develop two sampling frames/sampling strategies for a population
- Share with class and discuss

2 New Material to Cover

- Total Survey Error
- Populations
- Representativeness
- Sampling Frames
- Sampling without a Frame
- Simple Random Sampling

Simple Random Sampling (SRS)

- Advantages
  - Simplicity of sampling
  - Simplicity of analysis
- Disadvantages
  - Need sampling frame and units without any structure
  - Possibly expensive

Sample Estimates from an SRS

- Each unit in frame has equal probability of selection
- Sample statistics are unweighted
- Variances are easy to calculate
- Easy to calculate sample size need for a particular variance

Sample mean

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \tag{1}$$

where y_i = value for a unit, and
n = sample size

$$SE_{\bar{y}} = \sqrt{(1 - f)\frac{s^2}{n}} \tag{2}$$

where f = proportion of population sampled,
s^2 = sample variance, and
n = sample size

Sample proportion

$$\bar{y} = \frac{1}{n} \sum_{i=1}^{n} y_i \tag{3}$$

where y_i = value for a unit, and
n = sample size

$$SE_{\bar{y}} = \sqrt{\frac{(1-f)}{(n-1)} p(1-p)} \tag{4}$$

where f = proportion of population sampled,
p = sample proportion, and
n = sample size

Estimating sample size

- Imagine we want to conduct a political poll
- We want to know what percentage of the public will vote for which coalition/party
- How big of a sample do we need to make a relatively precise estimate of voter support?

Estimating sample size

$$Var(p) = (1 - f)\frac{p(1-p)}{n-1} \tag{5}$$

Given the large population:

$$Var(p) = \frac{p(1-p)}{n-1} \tag{6}$$

Need to solve the above for n.

(7)

Estimating sample size

$$Var(p) = (1 - f)\frac{p(1-p)}{n-1} \tag{5}$$

Given the large population:

$$Var(p) = \frac{p(1-p)}{n-1} \tag{6}$$

Need to solve the above for n.

$$n = \frac{p(1-p)}{v(p)} = \frac{p(1-p)}{SE^2} \tag{7}$$

Estimating sample size

Determining sample size requires:

- A possible value of p
- A desired precision (SE)

If support for each coalition is evenly matched ($p = 0.5$):

$$n = \frac{0.5(1 - 0.5)}{SE^2} = \frac{0.25}{SE^2} \tag{8}$$

Estimating sample size

What precision (margin of error) do we want?

- +/- 2 percentage points: $SE = 0.01$

$$n = \frac{0.25}{0.01^2} = \frac{0.25}{0.0001} = 2500 \tag{9}$$

Estimating sample size

What precision (margin of error) do we want?

- +/- 2 percentage points: $SE = 0.01$

$$n = \frac{0.25}{0.01^2} = \frac{0.25}{0.0001} = 2500 \tag{9}$$

- +/- 5 percentage points: $SE = 0.025$

$$n = \frac{0.25}{0.000625} = 400 \tag{10}$$

Estimating sample size

What precision (margin of error) do we want?

- +/- 2 percentage points: $SE = 0.01$

$$n = \frac{0.25}{0.01^2} = \frac{0.25}{0.0001} = 2500 \tag{9}$$

- +/- 5 percentage points: $SE = 0.025$

$$n = \frac{0.25}{0.000625} = 400 \tag{10}$$

- +/- 0.5 percentage points: $SE = 0.0025$

$$n = \frac{0.25}{0.00000625} = 40,000 \tag{11}$$

Important considerations

- Required sample size depends on p and SE

Important considerations

- Required sample size depends on p and SE
- In large populations, population size is irrelevant

Important considerations

- Required sample size depends on p and SE
- In large populations, population size is irrelevant
- In small populations, precision is influenced by the proportion of population sampled

Important considerations

- Required sample size depends on p and SE
- In large populations, population size is irrelevant
- In small populations, precision is influenced by the proportion of population sampled
- In anything other than an SRS, sample size calculation is more difficult

Important considerations

- Required sample size depends on p and SE
- In large populations, population size is irrelevant
- In small populations, precision is influenced by the proportion of population sampled
- In anything other than an SRS, sample size calculation is more difficult
- Much political science research assumes SRS even though a more complex design is actually used

Questions about SRS?

Next week's agenda

- Stratified sampling
- Cluster sampling
- Estimates, variances, and design effects
- Discuss sampling schemes in published research

Presentations?

- Burnham et al.: Mortality in Iraq
- Reinisch et al.: Registry data study
- Walker and Enticott: Surveying public managers

Assignment for next week

- Find a real survey or published study based on a survey
- Figure out its population, sampling frame, and sample
- Write up 0.5-1.0 pages discussing and evaluating its sampling approach
- We will discuss these in class next week