Non-Probability Sampling, Part 2

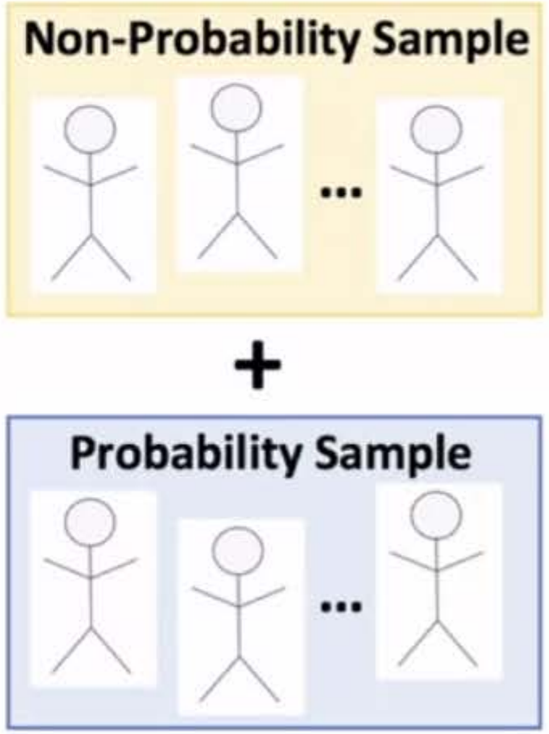
Brady T. West

Research Associate Professor, Survey Research Center,
Institute for Social Research

Population Inference Approaches

"Pseudo-Randomization Approach"

- **Combine non-probability sample with a probability sample**
- **Estimate probability of being included in non-probability sample** as a function of auxiliary information available in both samples
- **Treat estimated probabilities of selection as "known"** for non-probability sample, use probability sampling methods for analysis



Then we use the probability sampling methods that we discussed previously for analysis.

Population Inference Approaches

"Calibration" Approach

– **Compute weights for responding units** in non-probability sample that allow weighted sampled to mirror a known population

Non-probability sample: 70% female, 30% male

Population: 50% female, 50% male

→ **Down-weight females and Up-weight males**

to look more like that known population in terms of a 50,50 distribution.

Population Inference Approaches

"Calibration" Approach

- **Compute weights for responding units** in non-probability sample that allow weighted sampled to mirror a known population

- **Limitation:** if weighting factor not related to variable(s) of interest
  → will not reduce possible sampling bias

the weights that are closely related to the variables that you're actually interested in.

Twitter Example: Non-Probability Sample

API to extract info from several hundred thousand tweets and indicator of support for President Trump computed

- **Probability** of a tweet being selected **cannot be determined**
- **Twitter users not a random sample** of larger population
- **Lots of data**, ***but*** „„
  - high potential for sampling bias
  - lack of representation: may only capture people with strong opinions!

Logo from Twitter

not random sampling, there is high potential for sampling bias.

What prevents a simple random sample of tweets from being representative of the opinions of all adults in the U.S. population? Please select all responses that apply.

- [x] Twitter users decide to become Twitter users, and are not selected at random from the larger U.S. population.

Correct
Simple random samples are probability samples, and any size sample of tweets would still be limited by the same problems of representation and sampling bias.

- [ ] A simple random sample is not a probability sample.

Un-selected is correct

- [x] Twitter users may not necessarily be adults.

Correct
Simple random samples are probability samples, and any size sample of tweets would still be limited by the same problems of representation and sampling bias.

- [x] Twitter users may not necessarily be living in the United States.

Correct
Simple random samples are probability samples, and any size sample of tweets would still be limited by the same problems of representation and sampling bias.

- [ ] We need to know the size of the simple random sample before answering this question.

Un-selected is correct

What's Next?

- **Sampling distributions and sampling variance ~** how to estimate features of these distributions *based on only one probability sample*
- **Examples of making population inferences** *based on type of sample selected*
- Introduce **model-based** approaches to analyzing data