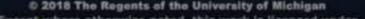


# Sampling from Well Defined Populations

### Brady T. West

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











### Getting Started

- Where data come from?
- Up until early-to-mid 20<sup>th</sup> century, researchers attempted to take a census = measure every unit in a given population
- 1930s: Jerzy Neyman and others enabled researchers to use random sampling



And what that meant is that we didn't have to measure every single unit in



### Target Populations

Important first step =

Define Target Population of interest in concrete terms!

— Who are we measuring?

- Males?
- African-American males?
- Older African-American males?
- What does "Older" mean? Be specific!



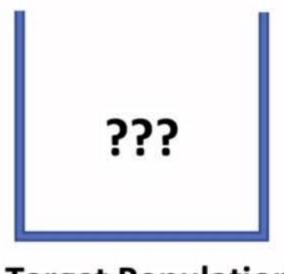
You have to be very specific at this initial stages when





## Target Populations

- Target Population:
  - What time frame are we interested in?
    - 2018? April 2018?
    - The first half of April in 2018?
  - Where is the population located?
    - The Midwest? Michigan?
    - Washtenaw county?
    - The city of Ann Arbor?



**Target Population** 

Or maybe even the city of Ann Arbor.





# Target Populations

The target population should be clearly defined, in a manner that EVERYONE can understand!

> The target population should be clearly defined



A researcher selects a sample of 100 adults ages 18 and older, and asks them a single question about their political party preference. Based on the data collected, the researcher concludes that 40% of adults prefer the Republican party. What is unclear about the researcher's target population? Select all responses that apply. The socio-demographic features of the target population

### Correct

All responses should be selected. We don't know anything about the race/ethnicity distribution, gender distribution, etc. of the target population, where the sample was selected, or when the sample was selected. The target population is ill defined.

The location of the target population

Correct All responses should be selected. We don't know anything about the race/ethnicity distribution, gender distribution, etc. of the target population, where the sample was selected, or when the sample was selected. The target population is ill defined.

The time when the sample was selected

Correct

All responses should be selected. We don't know anything about the race/ethnicity distribution, gender distribution, etc. of the target population, where the sample was selected, or when the sample was selected. The target population is ill defined.



### Given a Target Population, Now What?

- Well-defined target population?
- How can we make inferential statements about it?
  - Option I: Conduct a Census
  - Option 2: Select a Scientific Probability Sample
     from the population, and attempt to measure all units in the sample
  - Option 3: Select a <u>Non-Probability Sample</u>
     from the population, and attempt to measure all units in the sample

And then option 3 is to select what's called a non-probability sample from







### Option I: Conducting a Population Census

- Easier for smaller target populations
- Incredibly expensive for larger populations \$\$\$\$\$
- Requires a careful evaluation of
  - how much it will cost to measure all population units
  - what administrative data sources already available

That is,



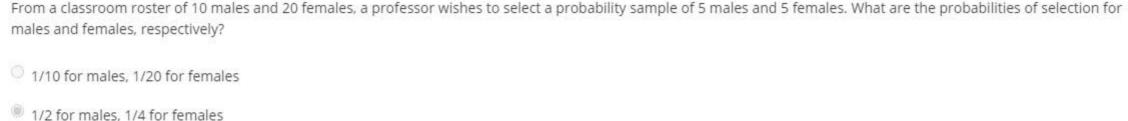




## Option 2: Probability Sampling

- Probability sampling basics (more details later)
  - Construct list of all units in population
     = sampling frame
  - Determine probability of selection for every unit on list (known and non-zero!)
  - Select units from list at random, with sampling rates for different subgroups determined by probabilities of selection
  - Attempt to measure randomly selected units

We might collect data from other sources about those randomly selected units but



### .

### Correct One out of every two males is selected, and one out of every four females is selected. The probabilities of selection are known from the well-defined population list (the classroom roster) and the clearly stated sample sizes.

1/5 for males, 1/5 for females



## Option 3: Non-Probability Sampling

- Generally does not involve random selection
- Probabilities of selection <u>can't be determined</u> for population units

And with non-probability sampling, as the names adjust,







## Option 3: Non-Probability Sampling

### Examples:

- opt-in web surveys
- quota sampling
- snowball sampling
- convenience sampling

For example, ou might go out on the street and





### Option 3: Non-Probability Sampling

 Main Problems: No statistical basis for making inference about the target population; high potential for bias

More on these issues in a later lecture!

probabilities of selection from some larger list,





## Why Probability Sampling?

The known probabilities of selection for all units allow us to make unbiased statements about both population features and the uncertainty in survey estimates

See Introductory Text for Week 4

estimate when we analyze the data, the features of the population





### Why Probability Sampling?

Random selection of population units protects us against bias from the sample selection mechanism,

~ allows us to make population inferences based on sampling distributions.

I talked about some of these sources of bias in the way that we select our sample.





## Why Probability Sampling?

### **Big Idea:**

With careful sample design, probability samples yield representative, realistic, random samples from larger populations; such samples have important statistical properties!

And these types of samples have very important statistical properties.







### What's Next?

- Probability sampling details with lots of examples
- Examples of non-probability samples + potential pitfalls
- Sampling distributions
  - + methods for making population inferences based on analyses of data from different types of samples

So what's next?

