Sampling from Well-Defined Populations

Brady T. West

Getting Started

- **Where data come from?**
- Up until early-to-mid 20th 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

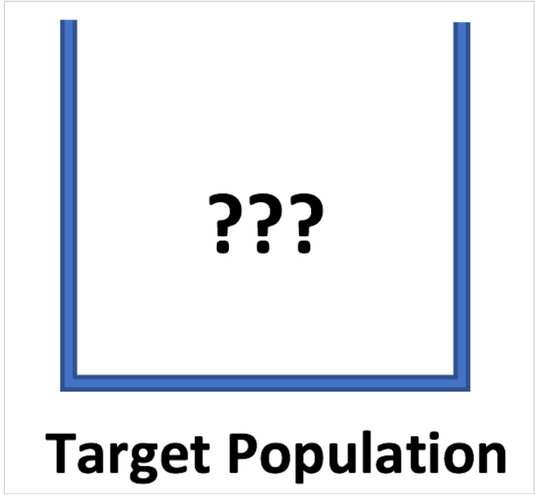
https://commons.wikimedia.org/wiki/File:Jerzy_Neyman.jpg

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!

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 Populations

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

Given a Target Population, Now What?

- Well-defined target population? ✓
- **How can we make inferential statements about it?**
  - **Option 1:** 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 **Non-Probability Sample** from the population, and attempt to measure all units in the sample

Option 1: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

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

Option 3: Non-Probability Sampling

- Generally does **not** involve random selection
- Probabilities of selection can't be determined for population units

Option 3: Non-Probability Sampling

- **Examples:**
  - opt-in web surveys
  - quota sampling
  - snowball sampling
  - **convenience sampling**
  - **“survey on the street”**

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!**

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

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.**

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!

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