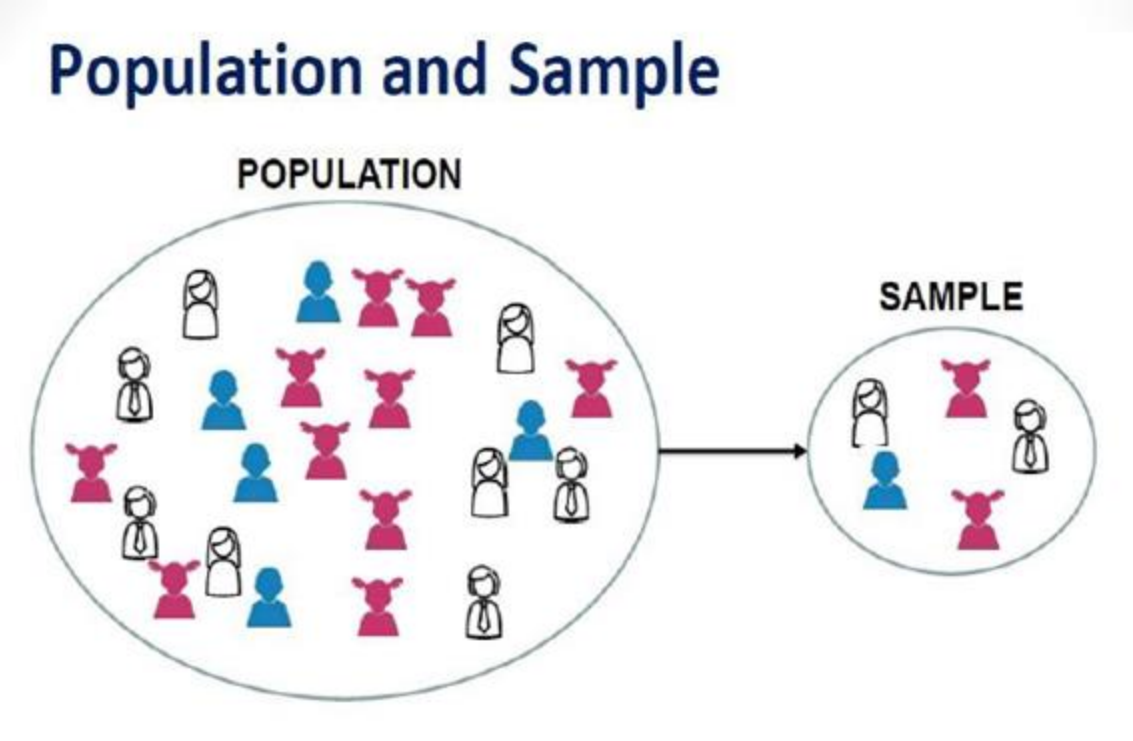
Week 1: Sampling

Professor Kathryn Jacobs

To Do

- What is a sample?
- Representation
- Different methods of sampling
- Sampling bias

Population vs. Sample



Population vs. Sample

A population is EVERYTHING you're interested in

Ex: Americans, cars, coffee-based drinks

A sample is a SPECIFIC group, taken from your population

Ex: 100 americans chosen at random, 30 cars produced in a single day at a factory, all Starbucks brand drinks

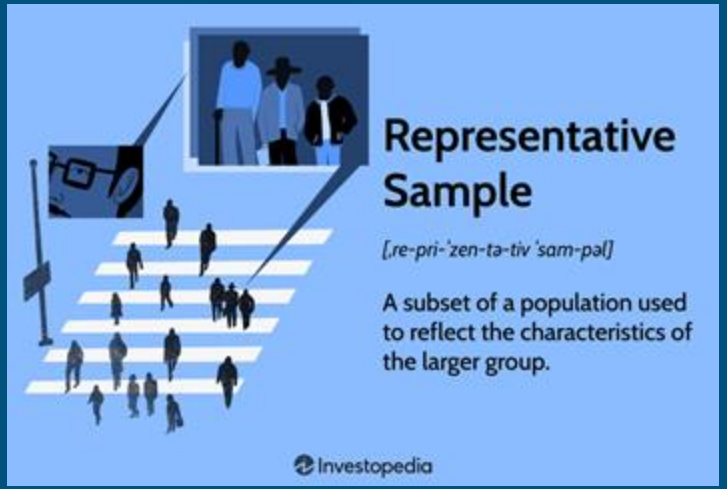
Population vs. Sample

Why not just collect data on the entire population?

Sometimes we do! That’s called a census

But most of the time….

- Money
- Time
- Logistics



Census: Why she so complicated

- Some people are hard to find

- Populations are constantly changing

- Harder than sampling!

Representation

Why do we want our samples to represent the population?

Trying to avoid the “representativeness heuristic” - what I see is what exists everywhere

Ex: What percentage of Americans have Scandinavian heritage?

A. 15.1%
B. 2.8%
C. 7.6%
D. 23.2%

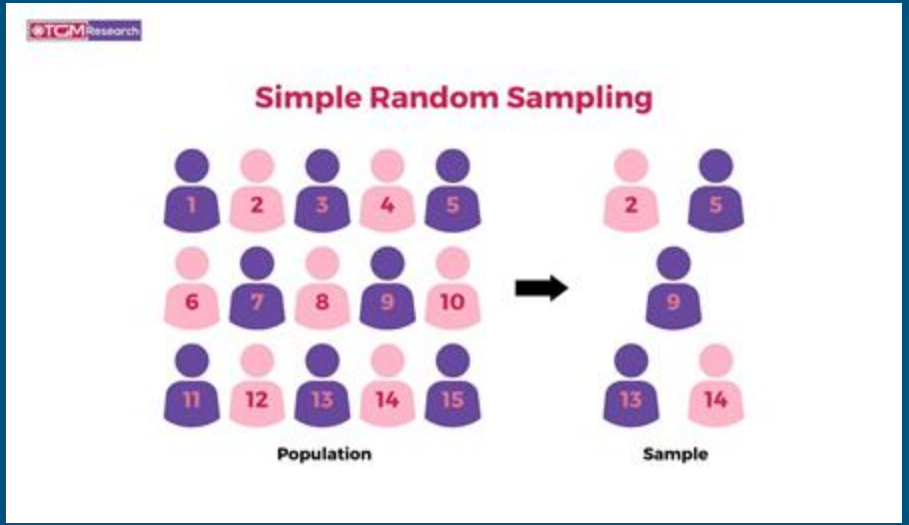
Types of Sampling

- Simple random
- Stratified
- Cluster
- Multistage
- Systematic

Types of Sampling: Simple Random

Each unit (let's just say person) has the same chance of being chosen out of the population

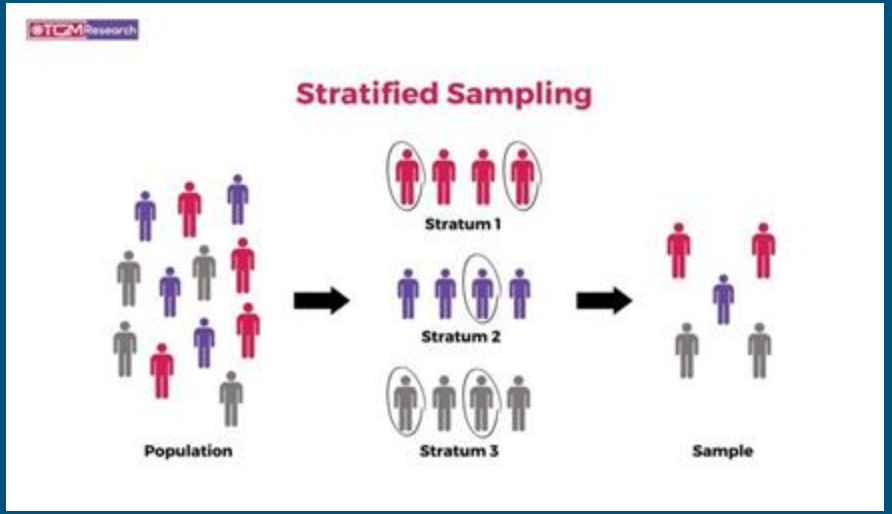
Ex: Randomly shuffle SSNs, pull random numbers



Types of Sampling: Stratified

1. Sort people into strata (groups)
    a. Ex: Age groups, economic levels, education levels
    b. All people in a subgroup are “the same” in some way

1. Randomly sample from within those groups

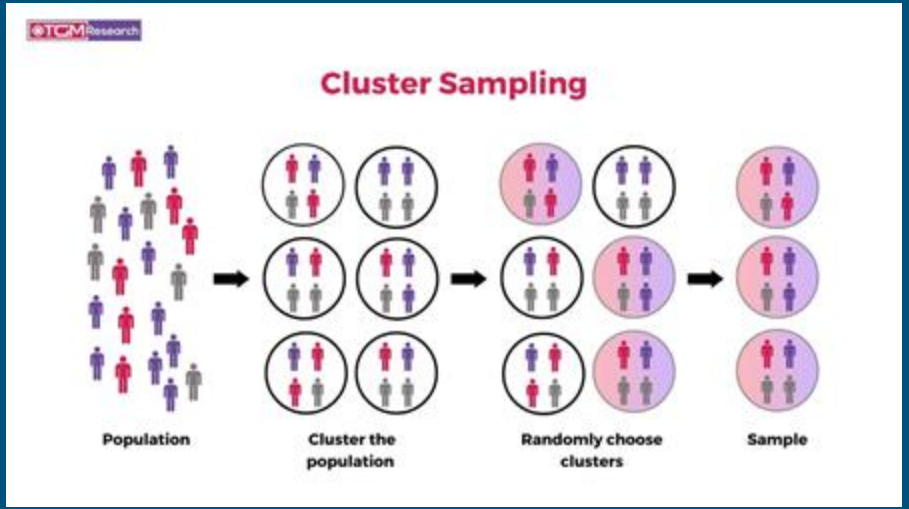


Can be faster/cheaper than true random sampling

Types of Sampling: Cluster

1. Sort people into clusters (groups)
    a. Ex: Neighborhoods of a city, regions of a country
    b. People within groups are all different from one another
    c. Each cluster should be representative of the population

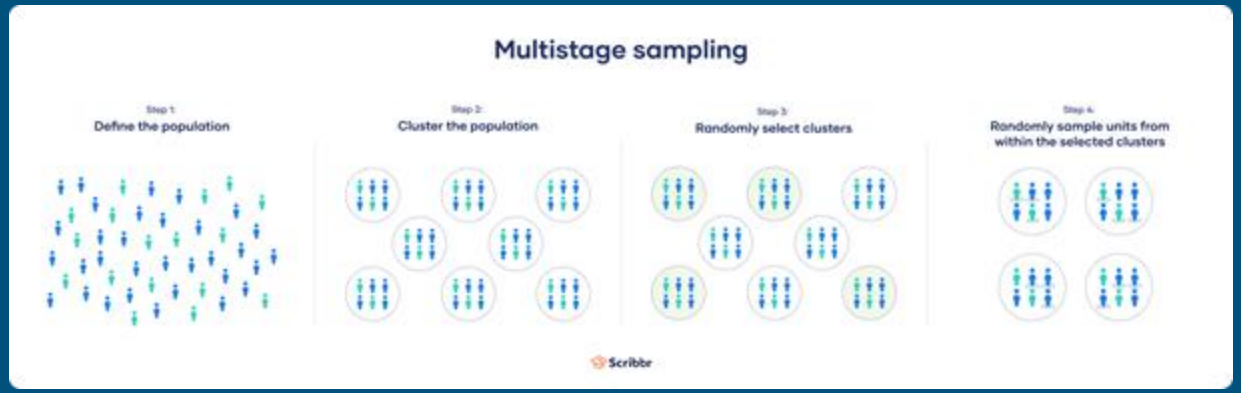
1. Randomly sample *entire clusters*



Types of Sampling: Multistage

1. Cluster sample (get a random sample of clusters)
2. Simple random sample from *within* those clusters
3. Geez this is complicated

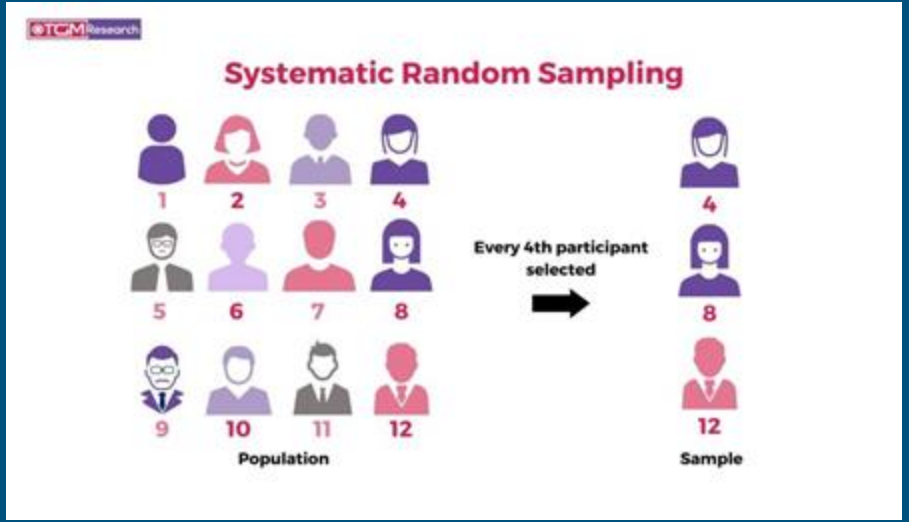
Ex: Cluster St Paul into neighborhoods. Choose random neighborhoods to include in study. In those neighborhoods, use random number program to choose only some houses to survey



Types of Sampling: Systematic

Every [blank]th person in an ordered list.

Ex: List all SSN’s in order, survey every 10th person



Sampling bias: General

Not always on purpose!

In fact, usually accidental

4 main types: Non-response, voluntary response, convenience sample, undercoverage

Sampling bias: Non-response

Whenever you try to sample a population, some people chosen will not respond

Those people may differ from everyone else in some way that will cause bias in your findings

Ex: A researcher wants to know people's opinions on an upcoming election. She chooses to text surveys to every 10th cell phone number in her area code. Why might this produce bias?

Sampling bias: THINK ~1 min

Whenever you try to sample a population, some people chosen will not respond

Those people may differ from everyone else in some way that will cause bias in your findings

Ex: A researcher wants to know people's opinions on an upcoming election. She chooses to text surveys to every 10th cell phone number in her area code. Why might this produce bias?

Sampling bias: PAIR ~2 min

Whenever you try to sample a population, some people chosen will not respond

Those people may differ from everyone else in some way that will cause bias in your findings

Ex: A researcher wants to know people's opinions on an upcoming election. She chooses to text surveys to every 10th cell phone number in her area code. Why might this produce bias?

Sampling bias: Share ~5 min

Whenever you try to sample a population, some people chosen will not respond

Those people may differ from everyone else in some way that will cause bias in your findings

Ex: A researcher wants to know people's opinions on an upcoming election. She chooses to text surveys to every 10th cell phone number in her area code. Why might this produce bias?

Sampling bias: Voluntary response

Amazon reviews

“How’s my driving? Call 1-800-”

Mail-in surveys sent to an entire city

Only very negative or very positive responses tend to get recorded

Sampling bias: Convenience sample

What it sounds like

Ex:

Survey all of your followers on Instagram

Ask your friends and family

Convenience samples are ALWAYS biased. Depending on what you're asking though, it might not matter too much.

For example, asking everyone at a Planet Fitness how likely they are to hire a personal trainer, versus asking them how strongly they believe in rinsing dishes before loading them into the dishwasher

Sampling bias: Undercoverage

- Door to door surveys will miss people who work during the day
- Evening phone surveys will miss restaurant and entertainment workers
- Surveys in general will pretty much always miss people without homes, people in prison, and children

Often, no good answer to this one. We just have to do our best

Sampling bias: Other

There are lots of ways to get bias, intentional or unintentional

How you word questions can accidentally or very intentionally bias responses.

Ex: Asking people “Do you agree with raising taxes?” leads to a much lower agreement rate vs. “Do you agree with increased public funding for education?”

Also, people can lie! And sometimes do