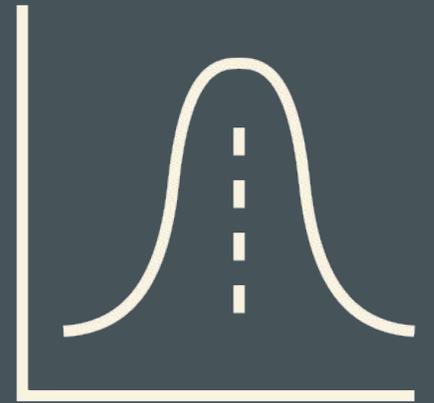
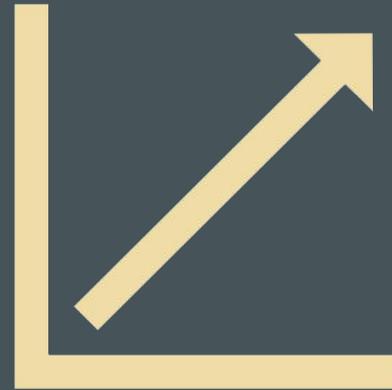
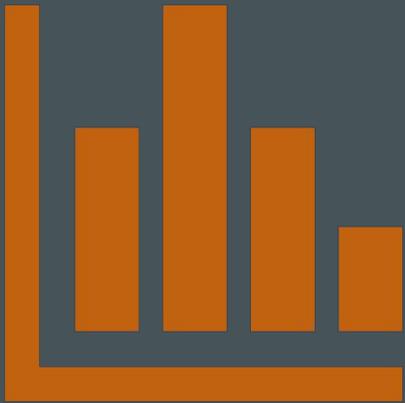

Basic Statistical Concepts



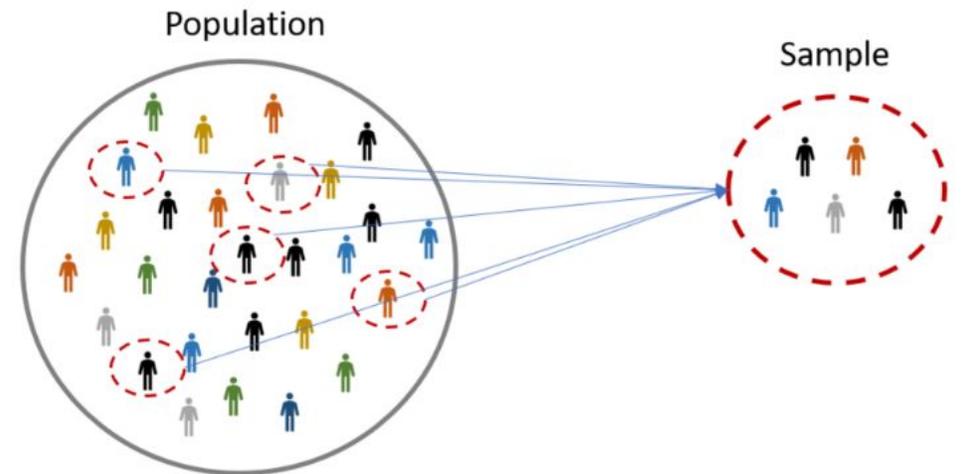
Review of Populations and Samples

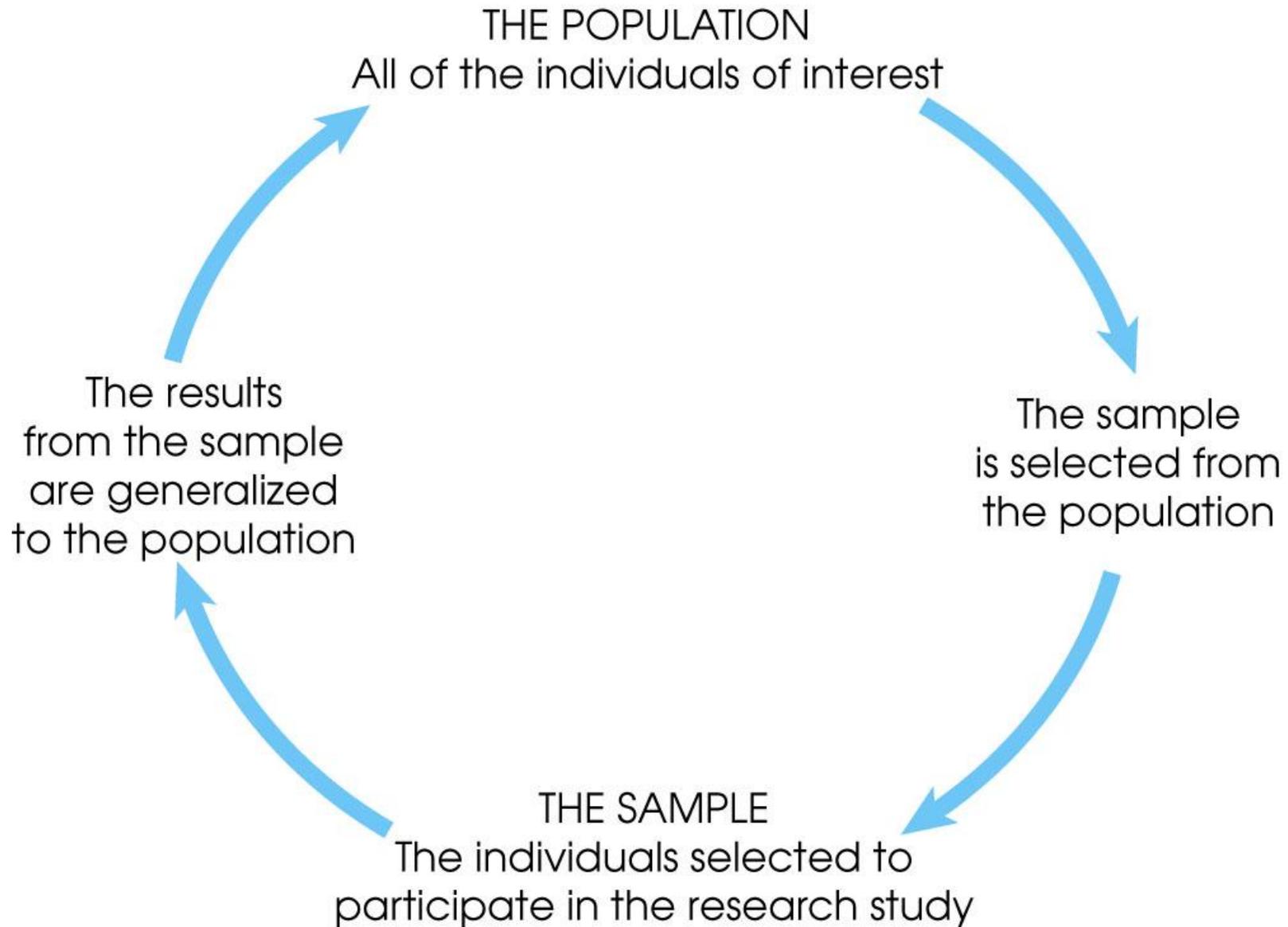
Population

- The entire group of individuals is called the **population**
- For example, a researcher may be interested in the relation between class size (variable 1) and academic performance (variable 2) for college students in the U.S.
- **Parameter** - any summary number, like an average or percentage, that describes the entire population

Why sampling?

- Usually, populations are so large that a researcher cannot examine the entire group
- A **sample** is selected to represent the population in a research study
- Many reasons to choose sampling:
 - Less costs
 - Less field time
 - More accuracy
 - When it's impossible to study the population





Sampling methods

Sampling methods effect out statistics

- Your results are only as good as your sample
 - Non-probability or probability sampling?
 - Random or not?

Non-probability samples

- Leads to convenience sampling
 - Often based on who it's easy for the researchers to contact
- Many strive to control for potential sources of bias
 - Voluntary sampling
 - Snowball sampling
 - Purposive (judgment) sampling
 - Quota sampling

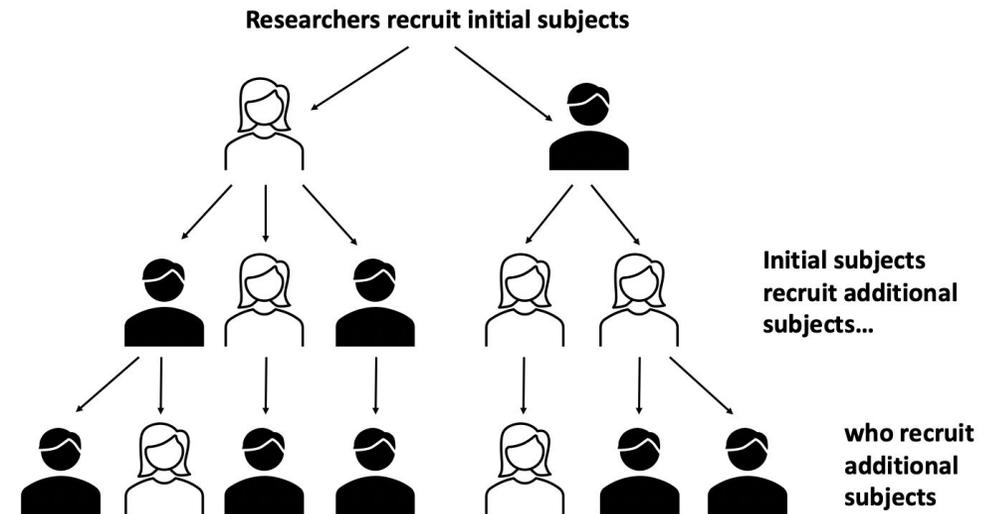
Voluntary sampling

- Occurs when researchers seek volunteers to participate in studies
 - Public postings
 - SONA systems
- Pros - Relatively fast and affordable
- Cons - Subject to bias



Snowball sampling

- Participants assist researchers in identifying other potential subjects
- Pros
 - Ability to reach small or stigmatized groups
- Cons
 - Hard to generalize
 - Slow



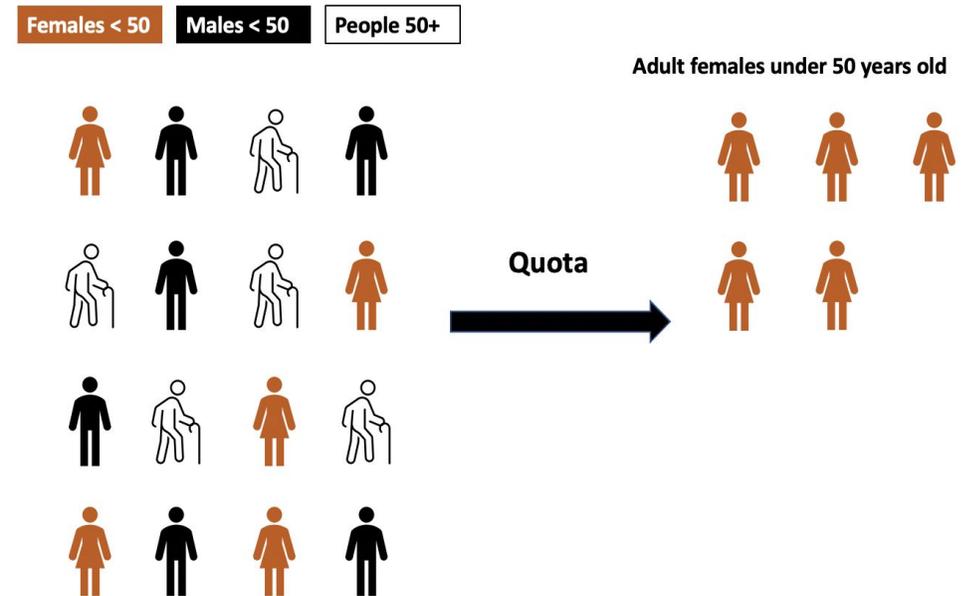
Judgment sampling

- Researcher uses their own judgment to select participants from the population of interest
- Pros
 - Saves time and money when compared to broader sampling strategies
- Cons
 - Vulnerable to errors in judgment by the researcher, leading to bias (it's unsystematic)



Quota sampling

- Researchers identify subsets of the population that are important to represent and then sample participants within each subset
- Pros
 - Ensures representation of important groups within the population
 - Helps eliminate potential confounds
- Cons
 - Potential for bias



Summary of non-probability samples

Pros

- Cheaper
- Faster

Cons

- Probability of being chosen is unknown
- Unable to generalize
- Potential for sampling bias and self-selection bias

Learning Check I



I. To get reactions about a particular new car, readers of a car magazine were asked to mail in their answers to a survey. Classify this sampling method.

- A) Stratified
- B) Convenience
- C) Systematic
- D) Voluntary response

Learning Check I - Answer



1. To get reactions about a particular new car, readers of a car magazine were asked to mail in their answers to a survey. Classify this sampling method.

- A) Stratified
- B) Convenience
- C) Systematic
- D) **Voluntary response**

Learning Check 2



2. Imagine you wanted to study Americans' beliefs about economic mobility, and you needed to make sure your sample included people who make a lot of money, people who make a moderate amount of money, and some people who make a little bit of money. Which sampling method would you choose?

- A) Judgment sampling
- B) Convenience sampling
- C) Quota sampling
- D) Snowball sampling

Learning Check 2 - Answer



2. Imagine you wanted to study Americans' beliefs about economic mobility, and you needed to make sure your sample included people who make a lot of money, people who make a moderate amount of money, and some people who make a little bit of money. Which sampling method would you choose?

- A) Judgment sampling
- B) Convenience sampling
- C) Quota sampling**
- D) Snowball sampling

Probability samples

- Lead researchers to gather representative samples
- Allows application of statistical sampling theory to results to:
 - Generalize
 - Test hypotheses

Probability samples

- Can be collected using several techniques
 - Simple random sampling
 - Systematic sampling
 - Cluster sampling
 - Multi-stage sampling
 - Stratified sampling

Simple random sampling

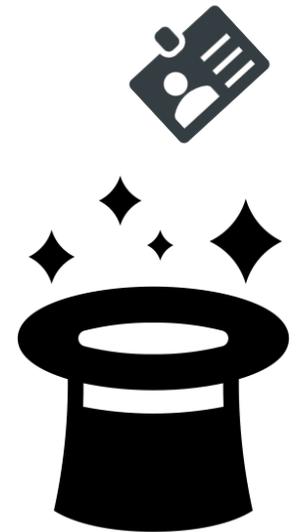
- Every member of the population has an equal chance of being selected into the study

Pros

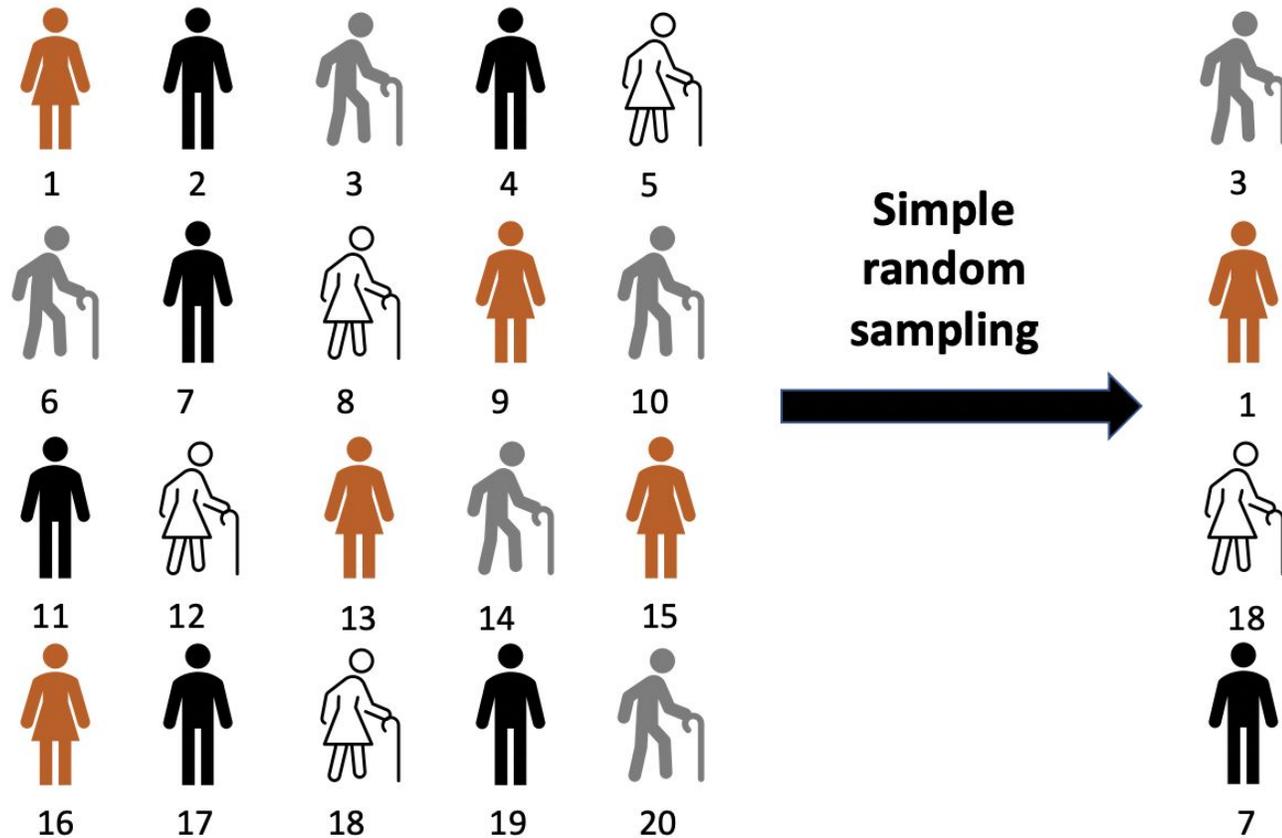
- Generalizable (strong external validity)
- Faster and more efficient than collecting data from every member of the population

Cons

- Expensive
- Time consuming
- Not always possible
 - Researchers may wish to study a group for which there is no organized list to randomly sample from



Simple random sampling – Visualized



Systematic sampling

- Version of random sampling in which every member of the population being studied is given a number
- Researchers randomly select a number from the list as the first participant
- Researchers choose an interval, say 10, and sample every tenth person on the list

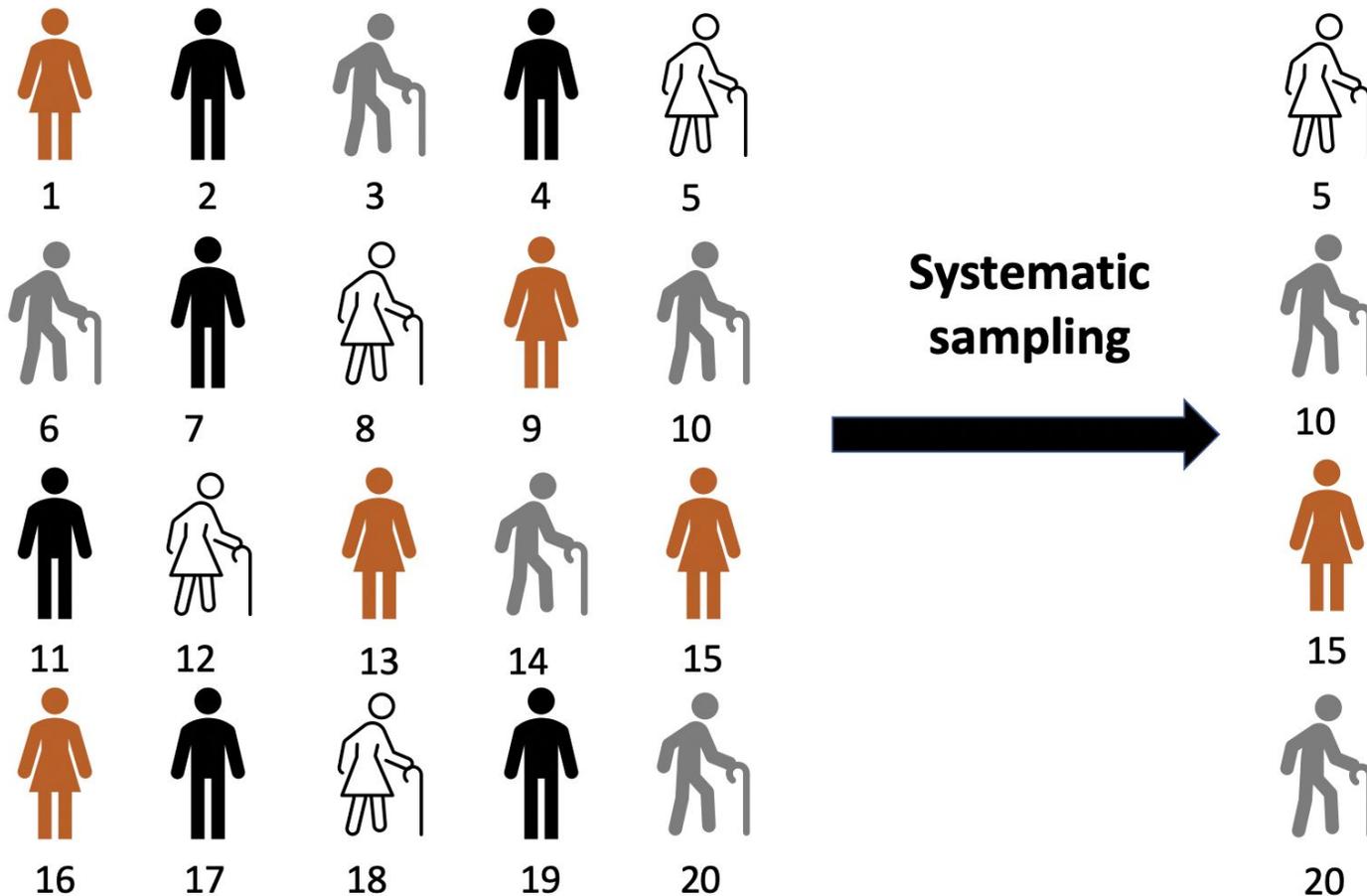
Pros

- Generalizable (high external validity)
- Faster than contacting all members of the population or simple random sampling

Cons

- Not possible without a list of all members of the population (limited feasibility)

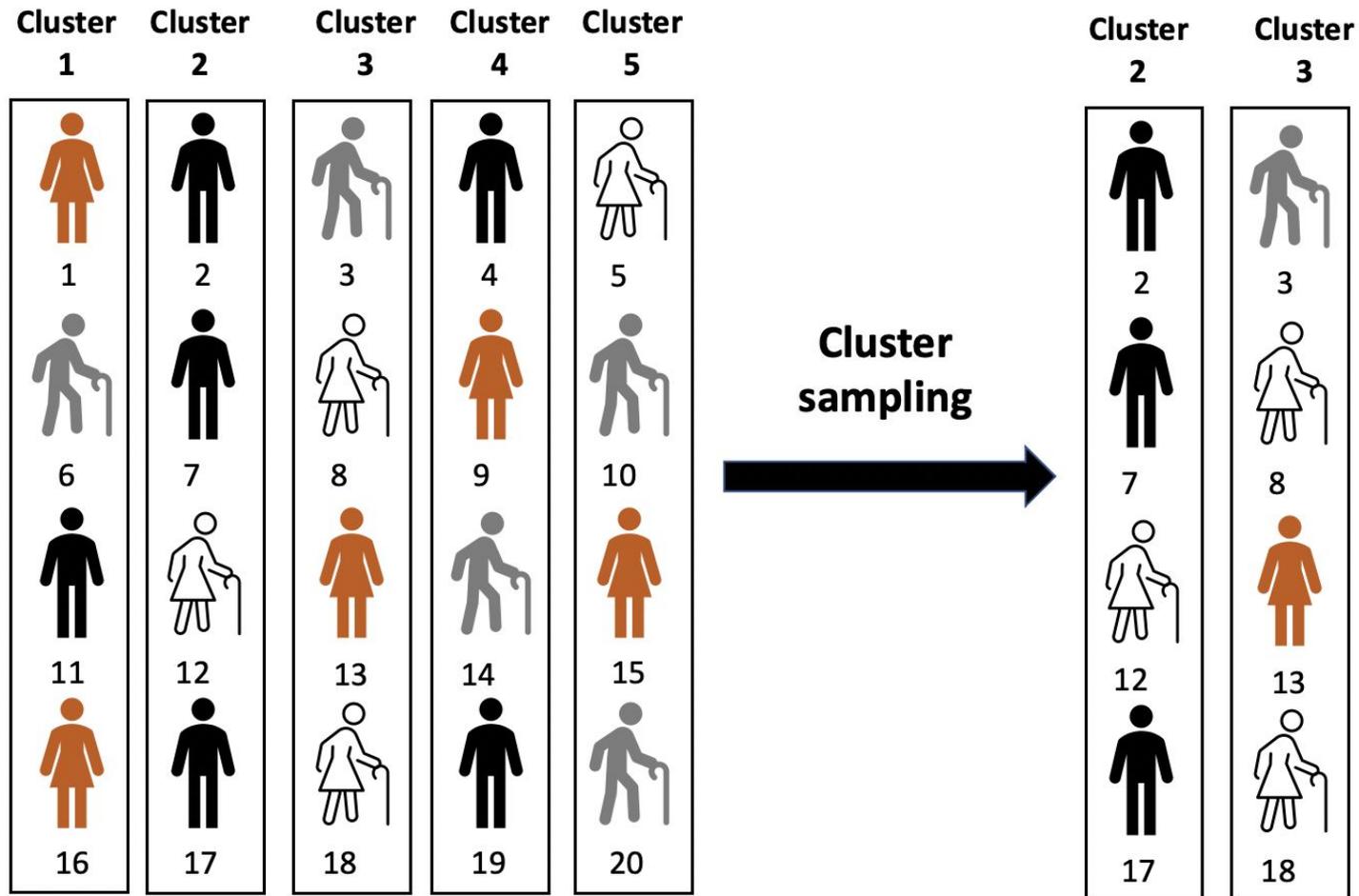
Systematic sampling – Visualized



Cluster sampling

- Occurs when researchers randomly sample people within groups or clusters the people already belong to
- Pros
 - External validity
 - Faster and more efficient than sampling all groups or all people in the population
- Cons
 - Not always possible

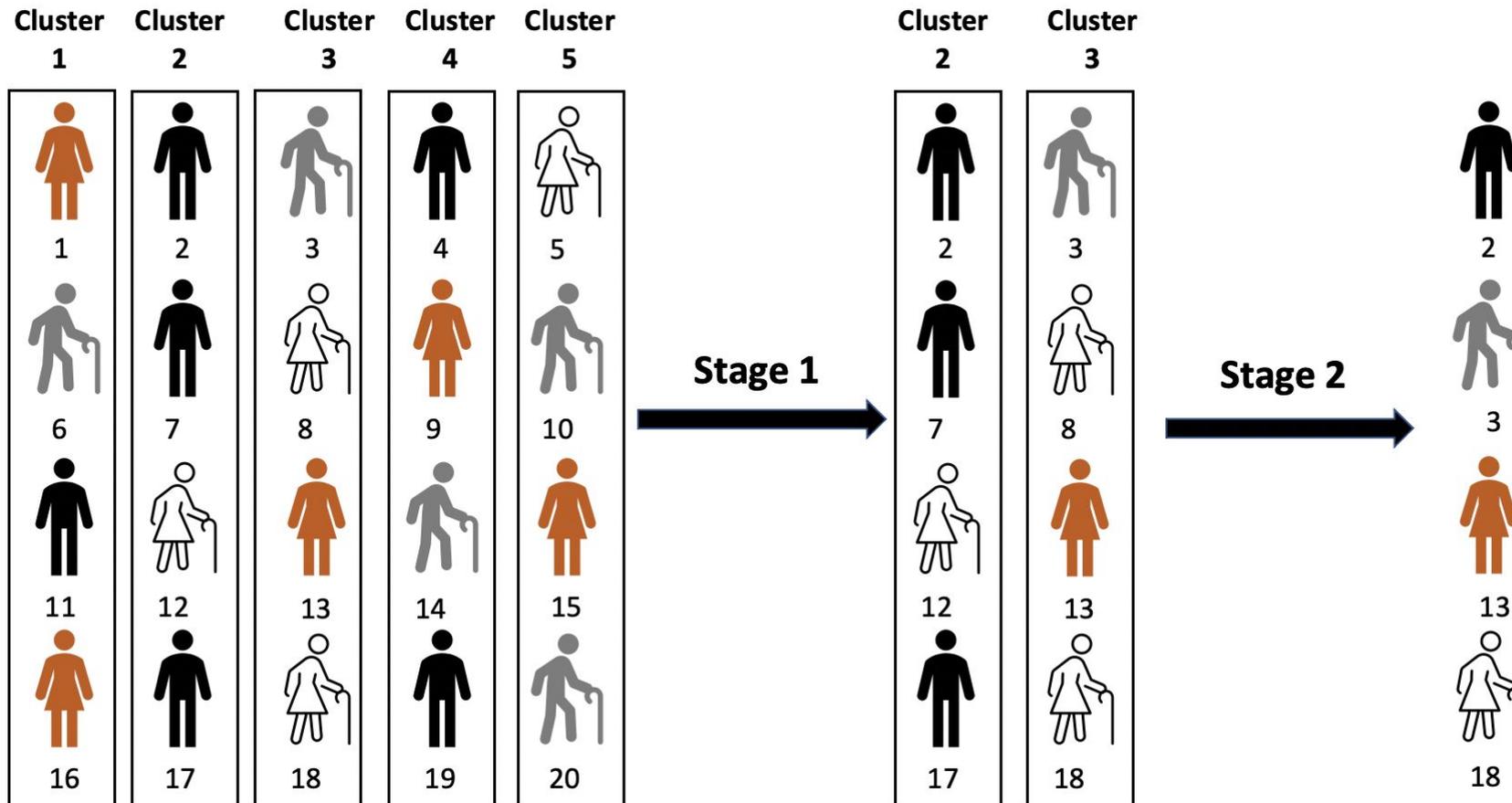
Cluster sampling – Visualized



Multi-stage sampling

- Version of cluster sampling
 - Multistage sampling begins when researchers randomly select a set of clusters or groups from a larger population
 - Then, researchers randomly select people within those clusters, rather than sampling everyone in the cluster
- Pros
 - External validity
 - Relative speed

Multi-stage sampling – Visualized

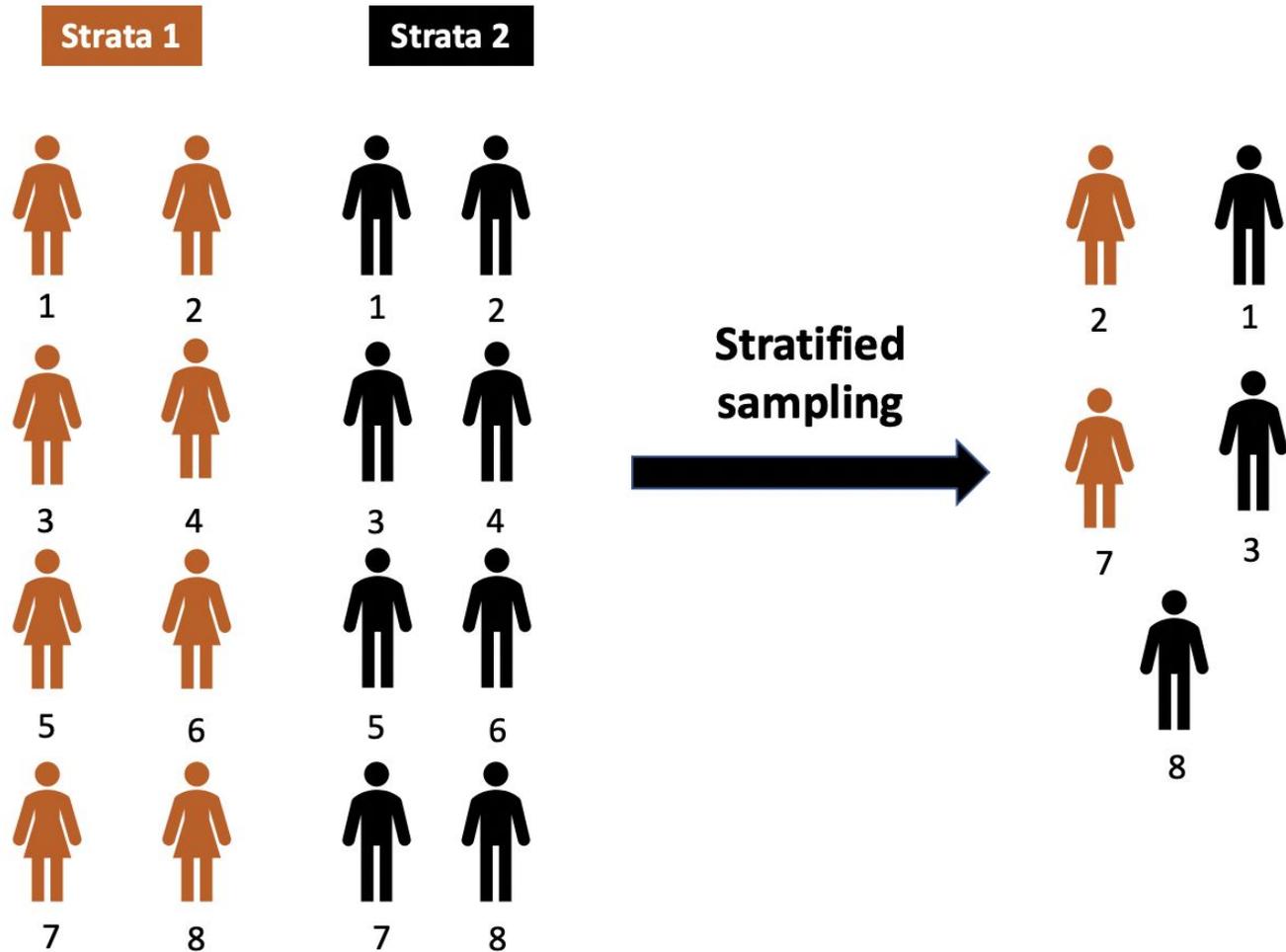


Stratified sampling

- Version of multistage sampling, in which a researcher selects specific demographic categories, or strata, that are important to represent within the final sample
- Once these categories are selected, the researcher randomly samples people within each category

- Pros
 - External validity
 - Relative speed

Stratified sampling – Visualized



Summary probability sampling

- There is a possibility that the statistics we obtain do not accurately reflect the population

Potential sources of error

in estimating a population distribution using a sample

Sampling error

Because the sample is not the whole population

Non-sampling error

Poor sampling method

Questionnaire or measurement error

Behavioural effects

Learning Check 3



1. When a random starting point is chosen, followed by every n th individual, this sampling method is

- A) Simple sampling
- B) Cluster sampling
- C) Systematic sampling
- D) Stratified sampling

Learning Check 3 - Answer



1. When a random starting point is chosen, followed by every n th individual, this sampling method is

- A) Simple sampling
- B) Cluster sampling
- C) Systematic sampling**
- D) Stratified sampling

Learning Check 4



2. A university polled 500 of its students, randomly selecting them proportional to the number of students enrolled in each degree program. Classify the sampling method.

- A) Simple random
- B) Stratified
- C) Systematic
- D) Convenience

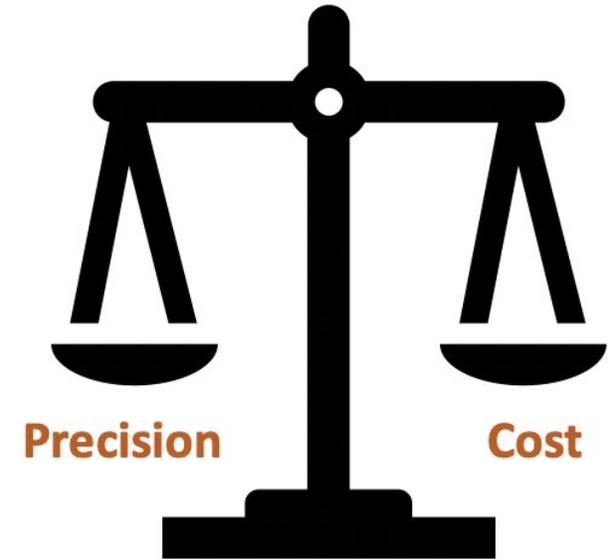
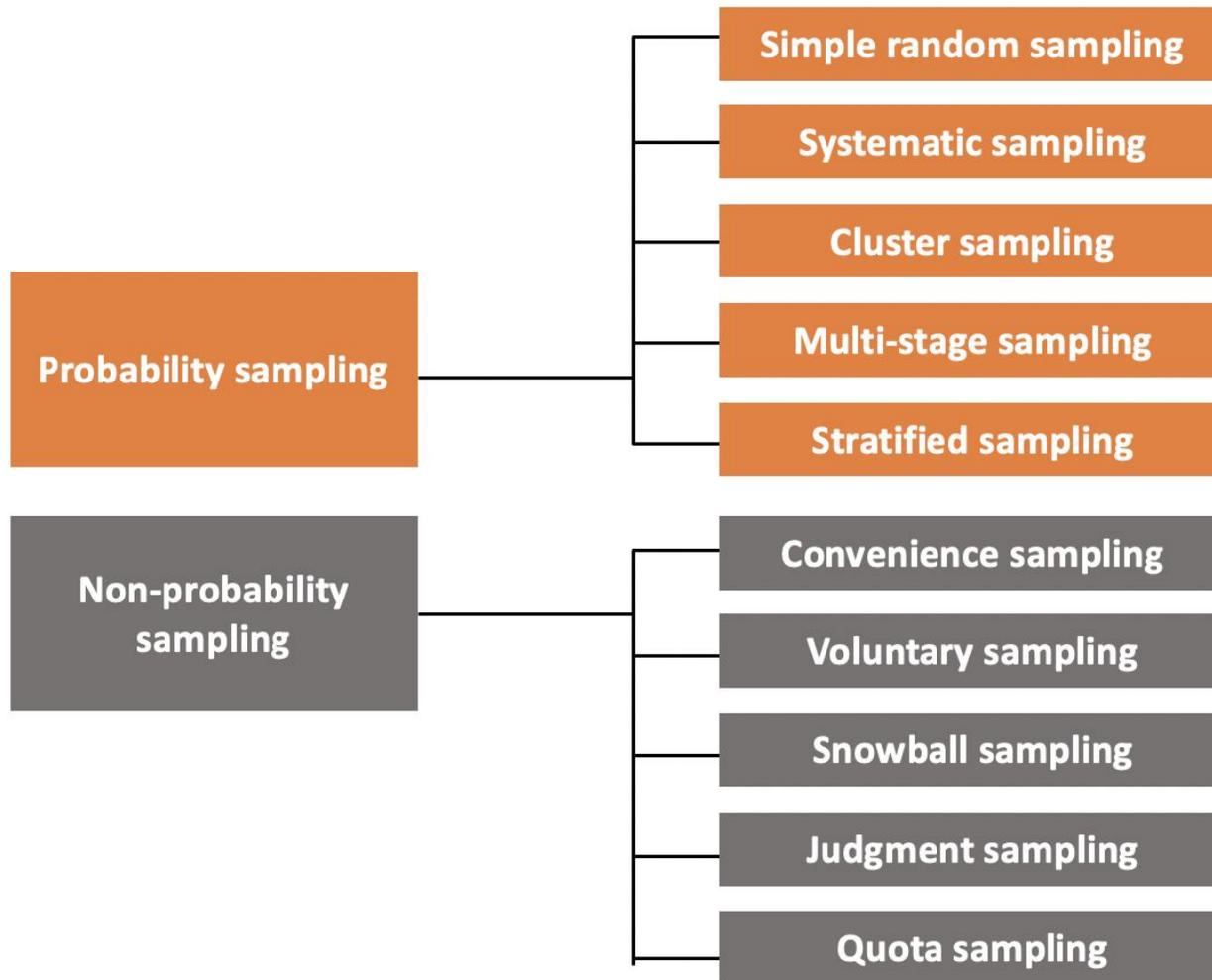
Learning Check 4 - Answer



2. A university polled 500 of its students, randomly selecting them proportional to the number of students enrolled in each degree program. Classify the sampling method.

- A) Simple random
- B) Stratified**
- C) Systematic
- D) Convenience

Sampling method techniques



You have to weigh these out when deciding what sampling method to choose

Does size matter?

YES!

- The more heterogeneous a population is, the larger the sample needs to be
- Probability sampling
 - The larger the sample size, the better
- Nonprobability sampling
 - Your results are not generalizable regardless...but sample size helps reduce variability (small samples increase your margin of error)



Sampling error

- Sample statistics are not perfect estimates of their corresponding population parameters
- Random variability in every sample
 - Mean number of depressive symptoms for adults can be 8.73 in one sample and 6.45 in another
- Random variability is called **sampling error**

Learning objectives

By the end of this lecture, you should be able to:

- Identify probability and non-probability sampling methods

