

Sampling

Sampling

A **population** is the entire group of individuals or objects of interest to us. In practice, it is difficult or impossible to study every individual or object in the population.

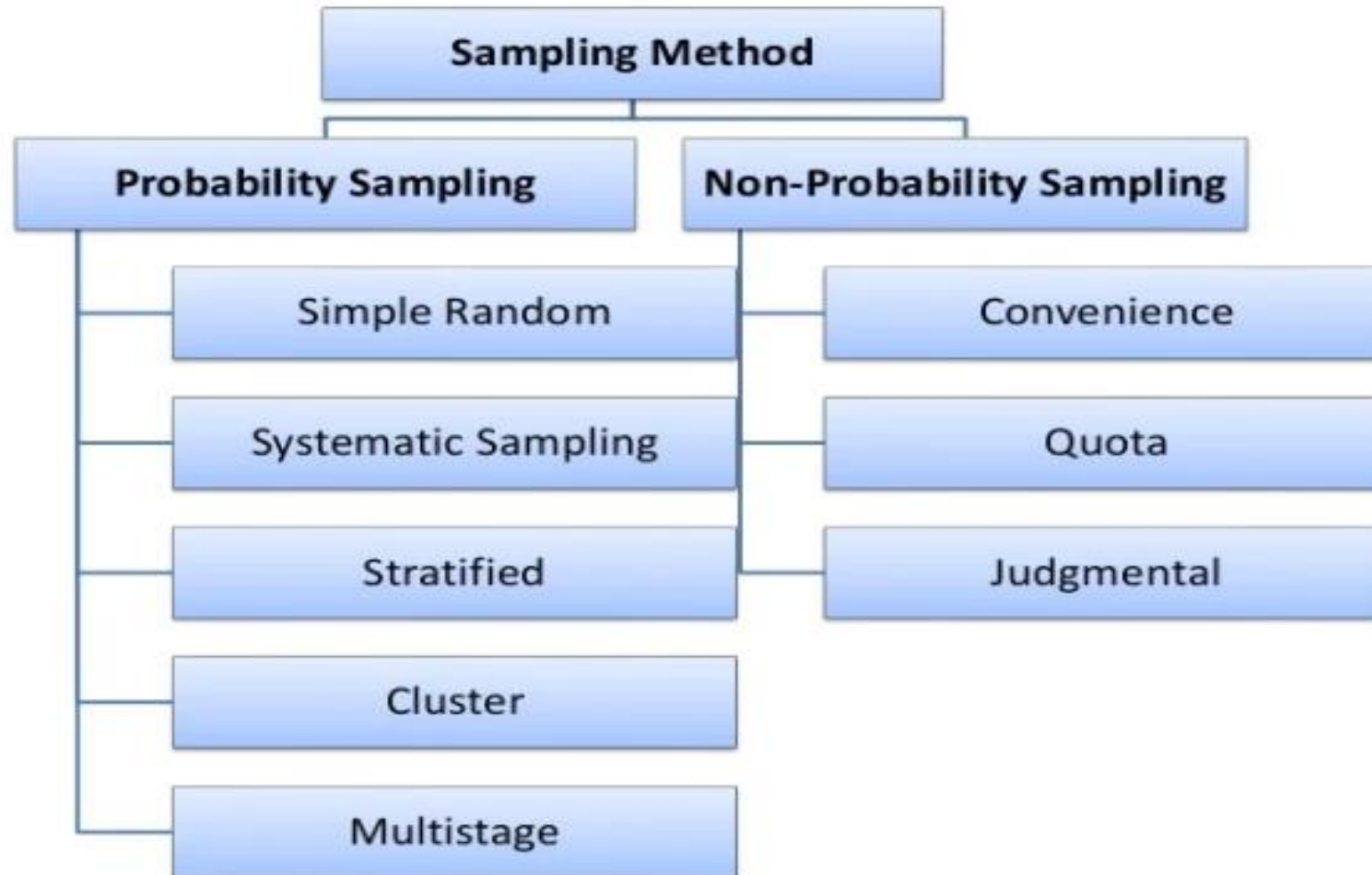
A **sample** is a subset of the population that we can study by collecting or gathering data.

Example: Math anxiety and community college students

A large community college has about 25,000 students. In a study of 85 students from college, it was determined that about 60 of the students have moderate or high math anxiety.

In this study, the population is all the students at this college. The sample is the 85 students whose math anxiety was measured

Sampling

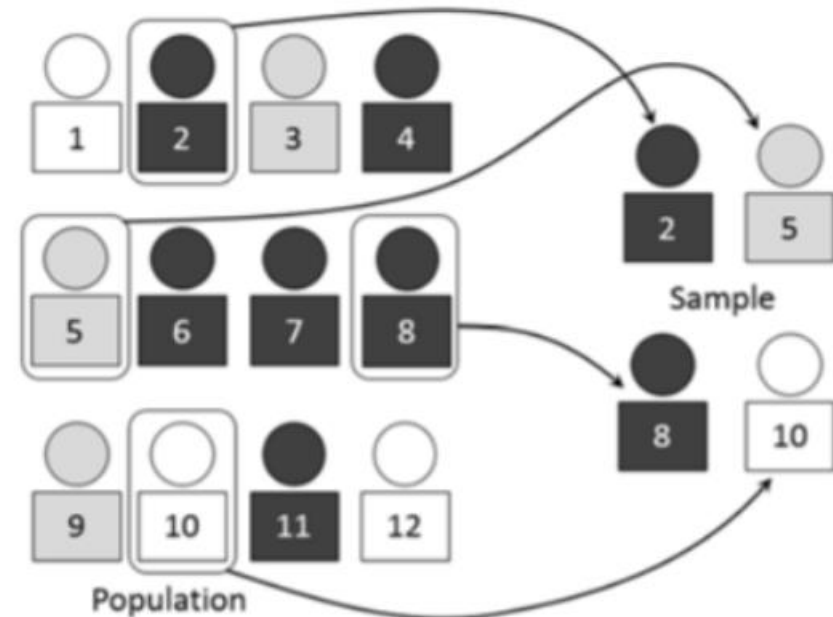


Simple Random Sampling

A **simple random sample** is a subset of a population in which all members of the population have the same chance of being chosen and are mutually independent of each other.

A raffle or lottery in which all names are put in a bowl and then some names are randomly selected.

Random samples in practice are almost impossible to obtain as it is difficult to list every member of the population.



Simple Random Sampling

Advantages of Simple Random Sampling:

- no possibility of bias in the sampling method
- no knowledge of population demographics needed
- easy to measure precision

Disadvantages of Simple Random Sampling:

- often impossible to conduct due to difficulty of cataloguing population
- high expense
- often less precise than a stratified sample

Simple Random Sampling

Example: Custom control searching

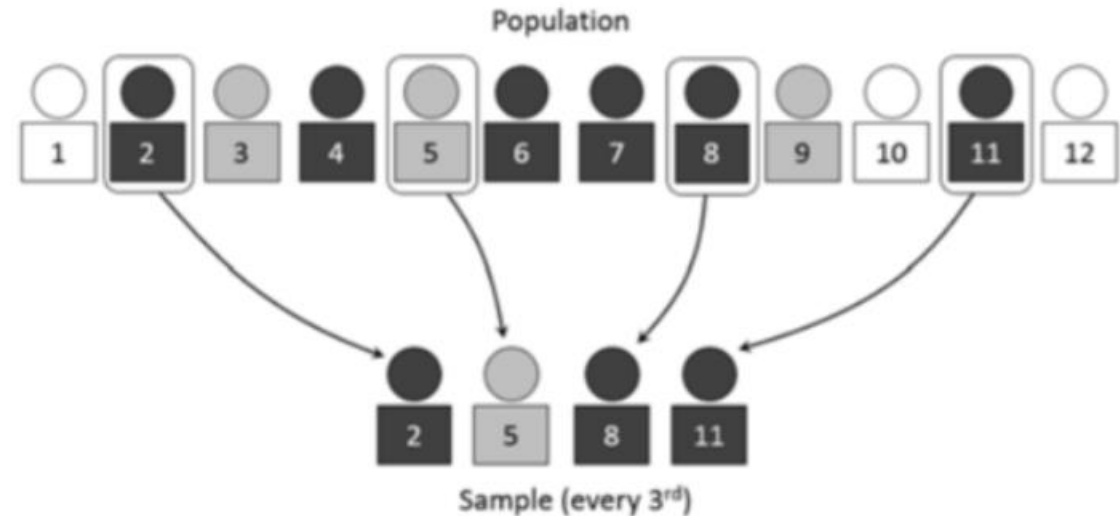
Before leaving customs at several international airports, all passengers must push a button. If the button is red, you will be required to go through an intensive search. If the button is green, you will not be searched.⁴⁷ The button is totally random and has a 20% chance of being red. Passengers who are subject to the intensive search are a true simple random sample of the entire population of arriving passengers.



Systematic Sampling

A **systematic sample** is a subset of the population in which the first member of the sample is selected at random and all subsequent members are chosen by a fixed periodic interval.

An example would be having a list of the entire population and then taking every 3rd person on the list.



Systematic Sampling

Advantages of Systematic Sampling:

- easy to design and explain
- more economical than random sampling
- avoids random clustering (several adjacent values)

Disadvantages of Systematic Sampling:

- may be biased if population is patterned or has a periodic trait
- easier for researcher to wrongly influence data
- population size needs to be known in advance

Systematic Sampling

Example: Random drug testing of employees

A shipping company has approximately 20,000 employees. The company decided to administer a random drug test to 5% of the employees, a sample size of 1000. The company has a list of all employees sorted by social security number. A random number is selected between 1 and 20. Starting with that person, every subsequent 20th person is also sampled. For example, if the selected number is 16, then the company would select persons 16, 36, 56, 76, ... , 19996 for drug testing.



Stratified Sampling

- A **stratified sample** is designed by breaking the population into subgroups called strata, and then sampling so the proportion of each subgroup in the sample matches the proportion of each subgroup in the population.
- For example, if a population is known to be 60% female and 40% male, then a sample of 1000 people would have 600 women and 400 men.

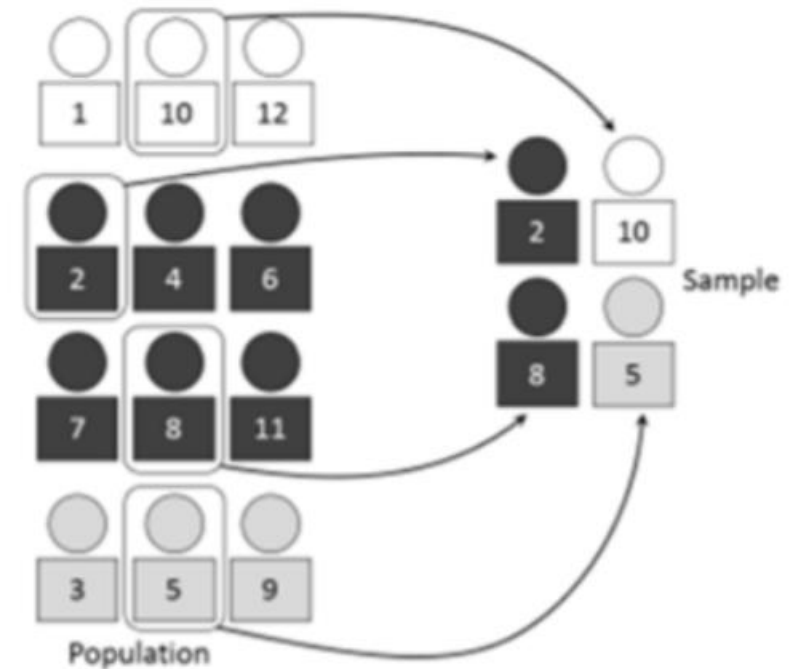
Stratified Sampling

Advantages of Stratified Sampling:

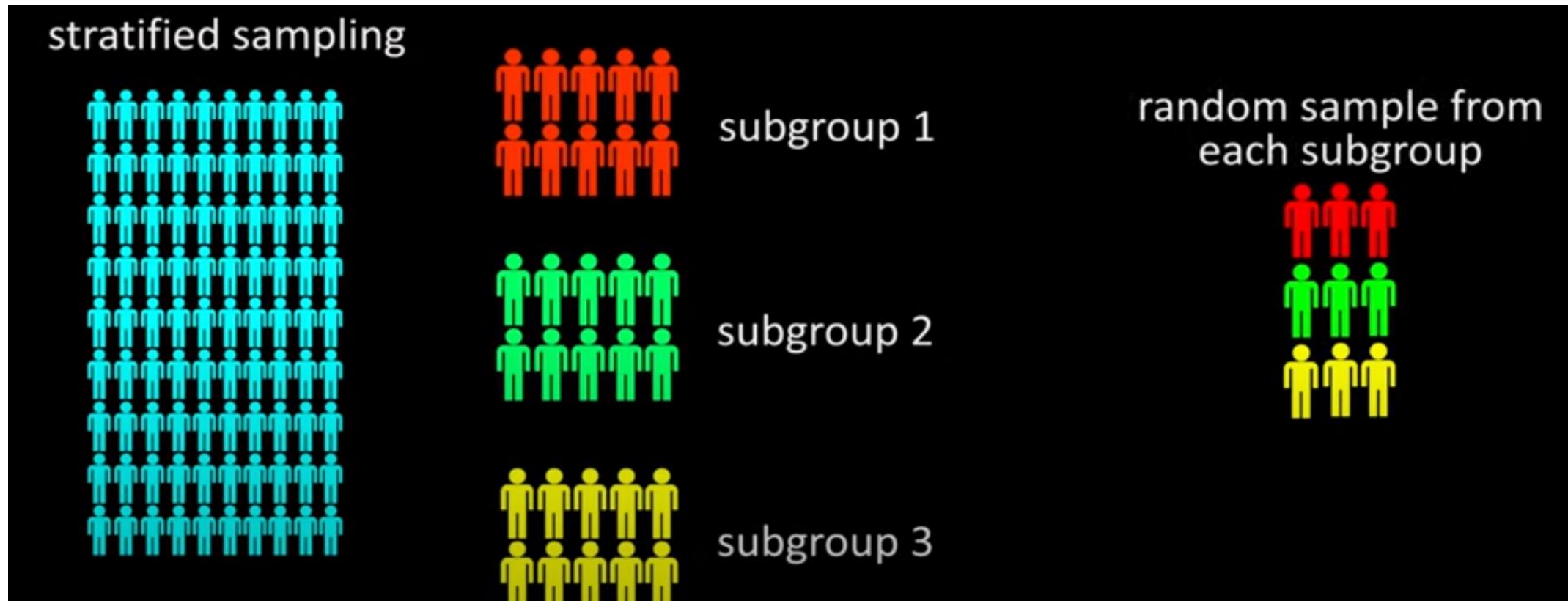
- minimizes selection bias as all strata are fairly represented
- each subgroup receives proper representation
- high precision (low standard deviation) compared to other methods

Disadvantages of Stratified Sampling:

- high knowledge of population demographics needed
- not all populations are easily stratified
- time consuming and expensive



Stratified Sampling



Stratified Sampling

Example: Social media conversations about race

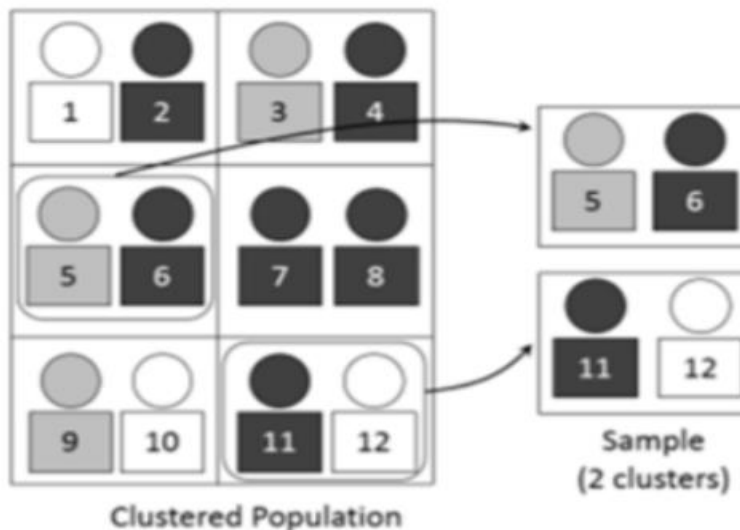
In 2016, Pew Research Center conducted a study to examine how people use social media such as Twitter or Facebook.⁵⁰ The study focused on the content and hash tags used on people's comments about events involving racially motivated attacks by the police and differences in opinions about groups such as Black Lives Matter.



Since the study involved people's opinions about race, it was important that Pew used stratified sampling by race. Particular care was taken to make sure that there was appropriate representation in the sample from traditionally undersampled African American and Latino groups.

Cluster Sampling

- A **cluster sample** is created by first breaking the population into groups called clusters, and then taking a sample of clusters.
- An example of cluster sampling is randomly selected several classes at a college and then sampling all the students in those selected classes.



Cluster Sampling

- **Advantages of Cluster Sampling:**

- most economical form of sampling because only clusters need to be randomized
- study can be completed in less time
- suitable for surveying populations that are broken into natural clusters

- **Disadvantages of Cluster Sampling:**

- sample may not be as diverse as population
- clusters may have a similar bias, causing sample to be biased
- less precision (higher standard deviation)

Cluster Sampling

Example: Student homelessness⁵³

The Bill Wilson Center of Santa Clara County provides social services for children, teens and adults. In 2017, the center conducted a study documenting homeless youth populations, surveying both high school students and community college students.⁵⁴



For community college students, the researchers chose two community colleges from the eight in Santa Clara County and surveyed students from Winter 2017 to Spring 2017. One finding was that a staggering 44% of community college students surveyed at these two colleges reported that they were homeless. (Homeless in this study means living on the street, living in cars, or couch surfing).

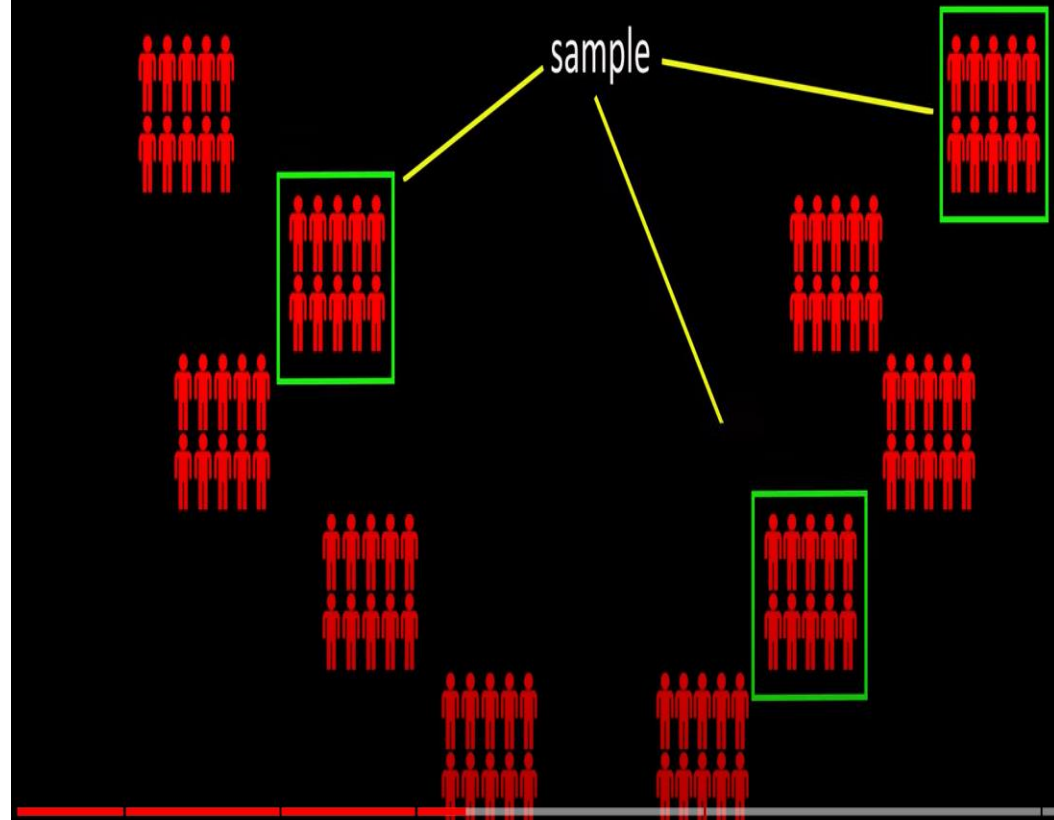
This study is an example of cluster sampling. Out of the eight Santa Clara County community colleges, the researchers chose 2. Although not reported in the study, it would be important that the demographics of the two chosen colleges match the average of all community college students in the county.

Cluster Sampling

cluster sampling



cluster sampling



Multistage Sampling

- **Multistage sampling** is a method of obtaining a sample from a population by splitting a population into smaller and smaller groups and taking samples of individuals from the smallest resulting groups.
- For example, suppose we're interested in estimating the average household income in the U.S. For simplicity, let's assume there are 100 million households. This represents the entire population we're interested in.

Multistage Sampling

- However, it would be too expensive and time-consuming to collect income data on each household, so instead we may take a simple random sample of 15 states.
- Then, within each state we may take a simple random sample of 10 counties.
- Then, within each county we may take a simple random sample of 100 households.
- Our resulting sample would contain **15,000** total households:
- $\text{Sample} = 15 \text{ states} * 10 \text{ counties} * 100 \text{ households} = 15,000$ households.

Non-probability sampling

Non-probability sampling is a sampling method that uses non-random criteria like the availability, geographical proximity, or expert knowledge of the individuals you want to research in order to answer a research question.

Convenience sampling

Convenience sampling is primarily determined by convenience to the researcher.

This can include factors like:

- Ease of access

- Geographical proximity

- Existing contact within the population of interest

Convenience samples are sometimes called “accidental samples,” because participants can be selected for the sample simply because they happen to be nearby when the researcher is conducting the data collection.

Example: Convenience sampling

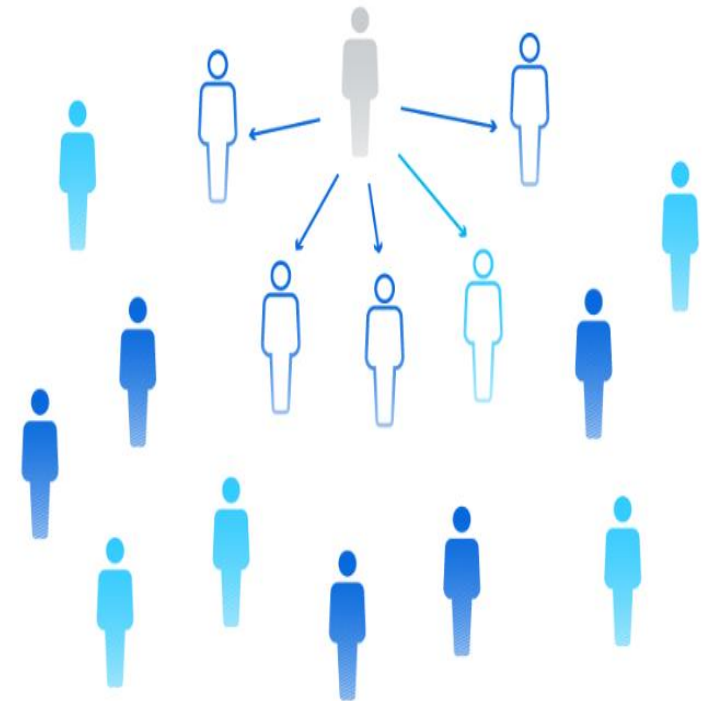
You are investigating the association between daily weather and daily shopping patterns. To collect insight into people’s shopping patterns, you decide to stand outside a major shopping mall in your area for a week, stopping people as they exit and asking them if they are willing to answer a few questions about their purchases.

Convenience sampling

Example: Convenience sampling

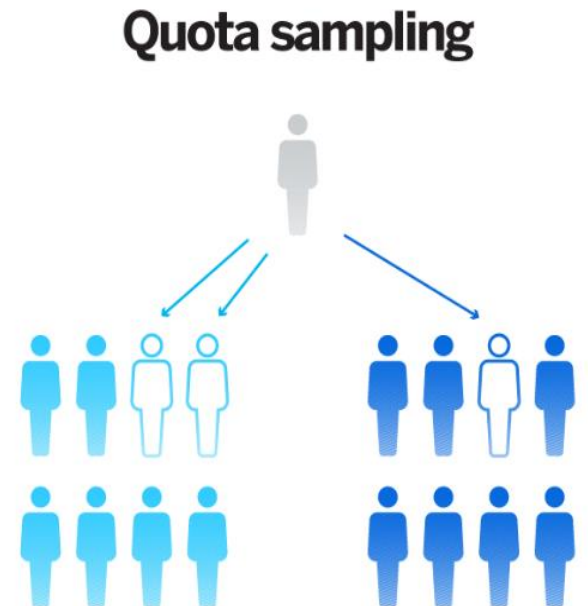
You are investigating the association between daily weather and daily shopping patterns. To collect insight into people's shopping patterns, you decide to stand outside a major shopping mall in your area for a week, stopping people as they exit and asking them if they are willing to answer a few questions about their purchases.

Convenience sample



Quota sampling

- In quota sampling, you select a predetermined number or proportion of units, called a quota.
- Your quota should comprise subgroups with specific characteristics (e.g., individuals, cases, or organizations) and should be selected in a non-random manner.
- Your subgroups, called strata, should be mutually exclusive.



Quota sampling

Your estimation can be based on previous studies or on other existing data, if there are any.

This helps you determine how many units should be chosen from each subgroup.

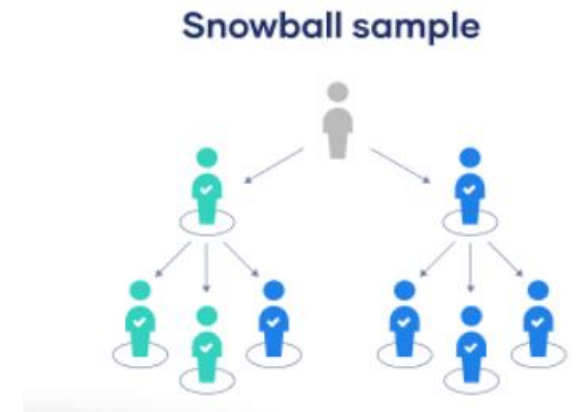
In the data collection phase, you continue to recruit units until you reach your quota.

Example: Non-proportional quota sampling

Let's say you are seeking opinions about the design choices on a website, but do not know how many people use it. You may decide to draw a sample of 100 people, including a quota of 50 people under 40 and a quota of 50 people over 40. This way, you get the perspective of both age groups.

Snowball sampling

- **Snowball sampling** is used when the population you want to research is hard to reach, or there is no existing database or other sampling frame to help you find them.
- Research about socially marginalized groups such as drug addicts, homeless people, or sex workers often uses snowball sampling.
- To conduct a snowball sample, you start by finding one person who is willing to participate in your research. You then ask them to introduce you to others.



Judgmental Sampling

- In judgmental sampling, the researcher selects participants based on their own judgment about who would be the most appropriate for the study.
- For example, a researcher may select participants for a study on the effects of a new cancer drug based on their experience with the disease and their likelihood of benefiting from the treatment.

