

# Data Sampling

Sampling is an important tool for gathering insights from data. With the use of sampling, you can make informed decisions with less cost and time. Learn the different types of data sampling and how you can apply them in real-world scenarios.

© 2023 Tushar Aggarwal.

[Medium.com/@tushar\\_aggarwal](https://Medium.com/@tushar_aggarwal)

[Github.com/2704](https://Github.com/2704)

# What is Data Sampling?

Data sampling is a technique used to select a representative subset of data from a larger dataset. It involves selecting a small group of observations from the larger population to estimate or infer characteristics.

## When do you use sampling?

Sampling is often used in scenarios where it's impossible to gather data from every single member of a population.

## Why is sampling important?

Sampling helps obtain reliable and accurate data, without the need to analyze an entire population.

# Types of Data Sampling

There are different types of data sampling each with its own unique characteristics. The four most common methods are:

<b>Simple Random Sampling</b>	A method where every individual in the population has the same chance of being selected.
<b>Stratified Sampling</b>	Divides the population into groups based on specific characteristics and then selects participants from each group in proportion to their representation in the population.
<b>Cluster Sampling</b>	Select participants based on naturally occurring clusters or groups; can be useful when studying housing units or schools.
<b>Systematic Sampling</b>	Selects participants via a fixed pattern, such as every 5th person in a population.

# Simple Random Sampling

"Simple random sampling is the purest form of probability sampling and is the easiest to understand."

- Investopedia

The simple random sampling method involves selecting a set of observations randomly, with each observation chosen independently. It gives every member of the population an equal chance of being selected.

## 1 Advantages:

Quick and relatively easy to implement

All members of the population have an equal chance of being selected, thereby reducing bias

## 2 Disadvantages:

Can be time-consuming when dealing with large populations

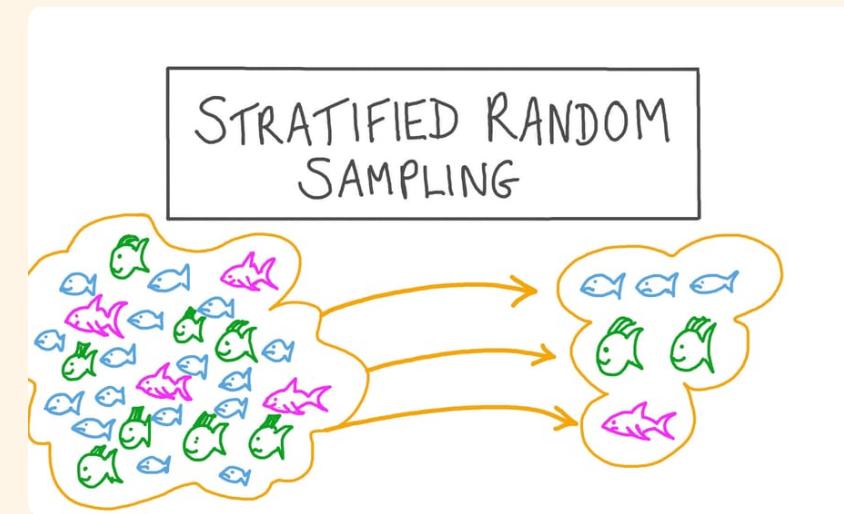
Sampling error could occur because the sample size may not be large enough

# Stratified Sampling

"Stratified sampling is the best sampling method when the population can be divided into different subgroups."

- Survey Anyplace

When different subgroups exist in a population, using stratified sampling may help you better understand it. Stratified sampling allows you to group population members based on shared characteristics and then select a proportion of members from each group.



Identify different strata of a population to understand each subgroup and its characteristics.

Sample a proportion from each stratum.

# Cluster Sampling

"Cluster sampling divides a population into subsets, and the subsets are randomly sampled to obtain the results."

- Investopedia

Cluster sampling is used when it is hard to reach the whole population of interest, but the population can be divided into groups that are similar to each other. These groups are clusters, and a random sample of clusters is selected.

## Advantages

1. Cost-efficient, since individuals can be grouped together to make sampling easier
2. Clustering provides natural groups, which can increase the representation of rare subgroups

# Systematic Sampling

"Systematic sampling is commonly used in research that requires a random sample of a population."

- UNSW

Systematic sampling involves selecting data in a systematic way from an arranged population. You pick the constituent from a list randomly from the selected first constituent.

## Steps in Systematic Sampling

1. Determine the population size
2. Divide the population size by the desired sample size. This gives you the sampling interval
3. Pick a starting point, usually randomly but anyone will do
4. Select every nth person or element on the list. The "n" here refers to the sampling interval
5. The stopping rule: select until you have the number of observations you need from the population

# Real-world Example

Suppose you are a wine producer that receives a large amount of grapes from multiple different regions in California. By using stratified sampling, you can take a sample from each region to determine the characteristics and quality of the grapes. This method would give you a more accurate and unbiased estimate of the grape quality for each region than if you just took one random sample from the entire dataset.



Select a sample of grapes from each region of California



Infer grape quality for the entire population, using the sample obtained from each region.

# Conclusion

Understanding how to sample data is a crucial task for obtaining reliable insights from various populations. Given that sampling is a vast topic, the type of research method that works best will depend on the scenario. From the techniques presented here, choose the sampling method that fits your needs and keeps the standards of accuracy, reliability, and validity.

© 2023 Tushar Aggarwal.

[Medium.com/@tushar\\_aggarwal](https://Medium.com/@tushar_aggarwal)

[Github.com/2704](https://Github.com/2704)