

---

# Types of Sampling and Sampling Techniques

## Overview

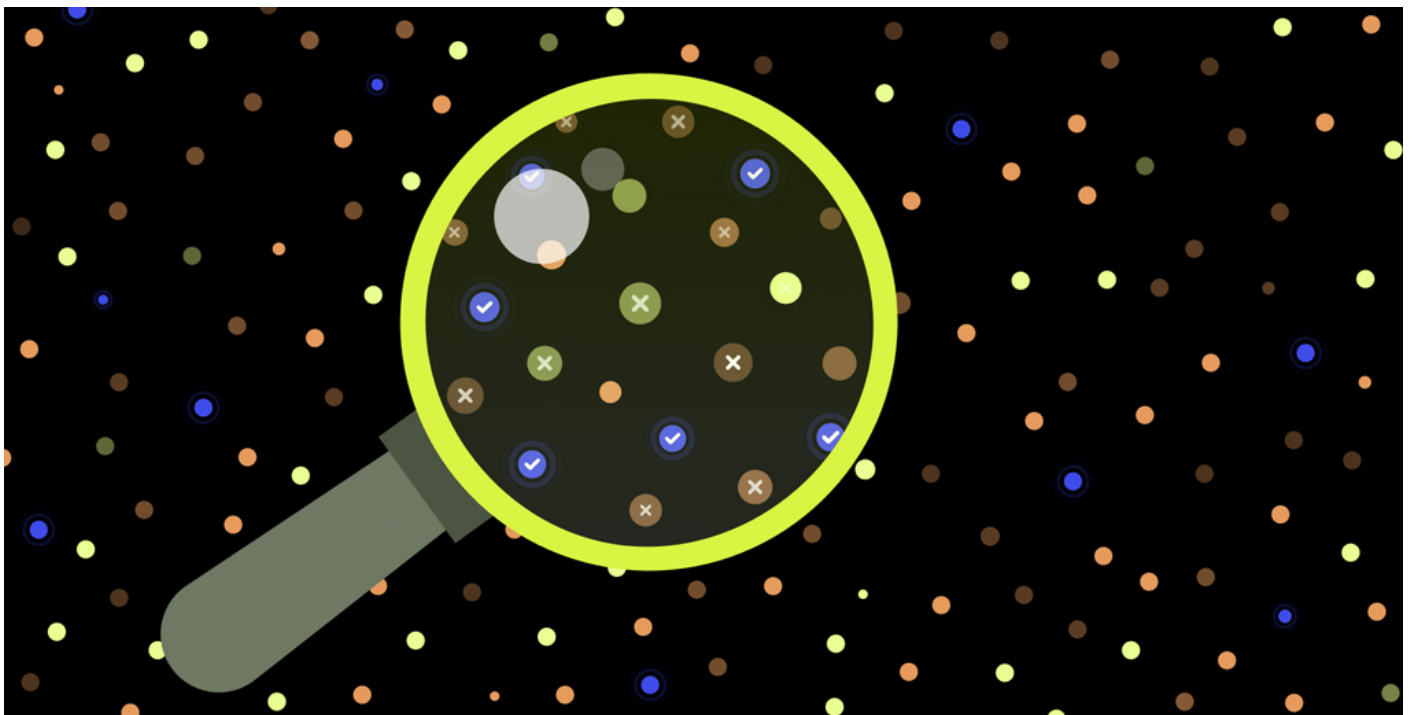
- Sampling is a popular statistical concept – learn how it works in this article
- We will also talk about eight different types of sampling techniques using plenty of examples

## Introduction

Here's a scenario I'm sure you are familiar with. You download a relatively big dataset and are excited to get started with analyzing it and building your machine learning model. And snap – your machine gives an “out of memory” error while trying to load the dataset. In such cases, employing efficient data loading techniques and utilizing types of sampling can help manage memory constraints and facilitate analysis.

It's happened to the best of us. It's one of the biggest hurdles we face in data science – dealing with massive amounts of data on computationally limited machines (not all of us have Google's resource power!).

So how can we overcome this perennial problem? Is there a way to pick a subset of the data and analyze that – and that can be a good representation of the entire dataset?



Yes! And that method is called types of sampling. I'm sure you've come across this term a lot during your school/university days, and perhaps even in your professional career. Sampling is a great way to pick up a subset of the data and analyze that. But then – should we just pick up any subset randomly?

## Table of contents

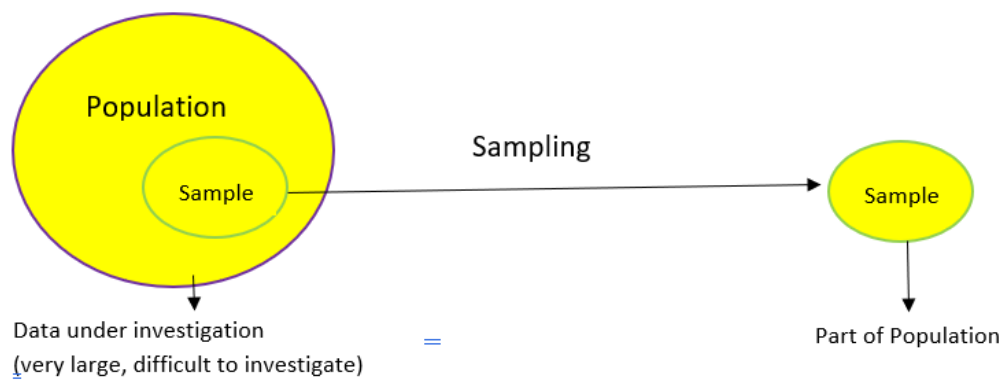
- [What is Sampling?](#)
- [Why do we need Sampling?](#)
- [Steps involved in Sampling](#)
- [Different Types of Sampling](#)
- [Types of Probability Sampling](#)
  - [Simple Random Sampling](#)
  - [Systematic Sampling](#)
  - [Stratified Sampling](#)
  - [Cluster Sampling](#)
- [Types of Non-Probability Sampling](#)
  - [Convenience Sampling](#)
  - [Quota Sampling](#)
  - [Judgment Sampling](#)
  - [Snowball Sampling](#)
- [Conclusion](#)
- [FAQs](#)

*Note: You can also check out our comprehensive collection of articles on statistics for data science [here](#).*

## What is Sampling?

Let's start by formally defining what [sampling](#) is.

Sampling is a method that allows us to get information about the population based on the statistics from a subset of the population (sample), without having to investigate every individual.



The above diagram perfectly illustrates what sampling is. Let's understand this at a more intuitive level through an example.

We want to find the average height of all adult males in Delhi. The population of Delhi is around 3 crore and males would be roughly around 1.5 crores (these are general assumptions for this example so don't take them at face value!). As you can imagine, it is nearly impossible to find the average height of all males in Delhi.

It's also not possible to reach every male so we can't really analyze the entire population. So what *can* we do instead? We can take multiple samples and calculate the average height of individuals in the selected samples.

But then we arrive at another question – how can we take a sample? Should we take a random sample? Or do we have to ask the experts?

Let's say we go to a basketball court and take the average height of all the professional basketball players as our sample. This will not be considered a good sample because generally, a basketball player is taller than an average male and it will give us a bad estimate of the average male's height.

Here's a potential solution – find random people in random situations where our sample would not be skewed based on heights.

## Why do we need Sampling?

I'm sure you have a solid intuition at this point regarding the question.

Sampling is done to draw conclusions about populations from samples, and it enables us to determine a population's characteristics by directly observing only a portion (or sample) of the population.

- Selecting a sample requires less time than selecting every item in a population
- Sample selection is a cost-efficient method

- Analysis of the sample is less cumbersome and more practical than an analysis of the entire population

## Steps involved in Sampling

I firmly believe visualizing a concept is a great way to ingrain it in your mind. So here's a step-by-step [process](#) of how sampling is typically done, in flowchart form!

Let's take an interesting case study and apply these steps to perform sampling. We recently conducted General Elections in India a few months back. You must have seen the public opinion polls every news channel was running at the time:

Were these results concluded by considering the views of all 900 million voters of the country or a fraction of these voters? Let us see how it was done.

## Step 1

The first stage in the sampling process is to clearly define the target population.

So, to carry out opinion polls, polling agencies consider only the people who are above 18 years of age and are eligible to vote in the population.

## Step 2

**Sampling Frame** – It is a list of items or people forming a population from which the sample is taken.

So, the sampling frame would be the list of all the people whose names appear on the voter list of a constituency.

## Step 3

Generally, probability sampling methods are used because every vote has equal value and any person can be included in the sample irrespective of his caste, community, or religion. Different samples are taken from different regions all over the country.

## Step 4

**Sample Size** – It is the number of individuals or items to be taken in a sample that would be enough to make inferences about the population with the desired level of accuracy and precision.

Larger the sample size, more accurate our inference about the population would be.

For the polls, agencies try to get as many people as possible of diverse backgrounds to be included in the sample as it would help in predicting the number of seats a political party can win.

## Step 5

Once the target population, sampling frame, sampling technique, and sample size have been established, the next step is to **collect data from the sample**.

In opinion polls, agencies generally put questions to the people, like which political party are they going to vote for or has the previous party done any work, etc.

Based on the answers, agencies try to interpret who the people of a constituency are going to vote for and approximately how many seats is a political party going to win. Pretty exciting work, right?!

## Different Types of Sampling

Here comes another diagrammatic illustration! This one talks about the different types of sampling techniques available to us:

- **Probability Sampling:** In probability sampling, every element of the population has an equal chance of being selected. Probability sampling gives us the best chance to create a sample that is truly representative of the population
- **Non-Probability Sampling:** In non-probability sampling, all elements do not have an equal chance of being selected. Consequently, there is a significant risk of ending up with a non-representative sample which does not produce generalizable results

For example, let's say our population consists of 20 individuals. Each individual is numbered from 1 to 20 and is represented by a specific color (red, blue, green, or yellow). Each person would have odds of 1 out of 20 of being chosen in probability sampling.

With non-probability sampling, these odds are not equal. A person might have a better chance of being chosen than others. So now that we have an idea of these two sampling types, let's dive into each and understand the different types of sampling under each section.

## Types of Probability Sampling

### Simple Random Sampling

This is a type of sampling technique you must have come across at some point. Here, every individual is chosen entirely by chance and each member of the population has an equal chance of being selected.

**Simple random sampling reduces selection bias.**

One big advantage of this technique is that it is the most direct method of probability sampling. But it comes with a caveat – it may not select enough individuals with our characteristics of interest. *Monte Carlo methods use repeated random sampling for the estimation of unknown parameters.*

## **Systematic Sampling**

In this type of sampling, the first individual is selected randomly and others are selected using a fixed 'sampling interval'. Let's take a simple example to understand this.

Say our population size is  $x$  and we have to select a sample size of  $n$ . Then, the next individual that we will select would be  $x/n$ th intervals away from the first individual. We can select the rest in the same way.

Suppose, we began with person number 3, and we want a sample size of 5. So, the next individual that we will select would be at an interval of  $(20/5) = 4$  from the 3rd person, i.e. 7 ( $3+4$ ), and so on.

3,  $3+4=7$ ,  $7+4=11$ ,  $11+4=15$ ,  $15+4=19$  = **3, 7, 11, 15, 19**

Systematic sampling is more convenient than simple random sampling. However, it might also lead to bias if there is an underlying pattern in which we are selecting items from the population (though the chances of that happening are quite rare).

## **Stratified Sampling**



In this type of sampling, we divide the population into subgroups (called strata) based on different traits like gender, category, etc. And then we select the sample(s) from these subgroups:

Here, we first divided our population into subgroups based on different colors of red, yellow, green and blue. Then, from each color, we selected an individual in the proportion of their numbers in the population.

**We use this type of sampling when we want representation from all the subgroups of the population.** However, stratified sampling requires proper knowledge of the characteristics of the population.

## **Cluster Sampling**

In a clustered sample, we use the subgroups of the population as the sampling unit rather than individuals. The population is divided into subgroups, known as clusters, and a whole cluster is randomly selected to be included in the study:

In the above example, we have divided our population into 5 clusters. Each cluster consists of 4 individuals and we have taken the 4th cluster in our sample. We can include more clusters as per our sample size.

**This type of sampling is used when we focus on a specific region or area.**

## **Types of Non-Probability Sampling**

### **Convenience Sampling**

This is perhaps the easiest method of sampling because individuals are selected based on their availability and willingness to take part.

Here, let's say individuals numbered 4, 7, 12, 15 and 20 want to be part of our sample, and hence, we will include them in the sample.

Convenience sampling is prone to significant bias, because the sample may not be the representation of the specific characteristics such as religion or, say the gender, of the population.

## **Quota Sampling**

In this type of [sampling](#), we choose items based on predetermined characteristics of the population. Consider that we have to select individuals having a number in multiples of four for our sample:

Therefore, the individuals numbered 4, 8, 12, 16, and 20 are already reserved for our sample.

In quota sampling, the chosen sample might not be the best representation of the characteristics of the population that weren't considered.

## **Judgment Sampling**

It is also known as selective sampling. It depends on the judgment of the experts when choosing whom to ask to participate.

Suppose, our experts believe that people numbered 1, 7, 10, 15, and 19 should be considered for our sample as they may help us to infer the population in a better way. As you can imagine, quota sampling is also prone to bias by the experts and may not necessarily be representative.

## **Snowball Sampling**

I quite like this sampling technique. **Existing people are asked to nominate further people known to them so that the sample increases in size like a rolling snowball.** This method of sampling is effective when a sampling frame is difficult to identify.

Here, we had randomly chosen person 1 for our sample, and then he/she recommended person 6, and person 6 recommended person 11, and so on.

**1->6->11->14->19**

There is a significant risk of selection bias in snowball sampling, as the referenced individuals will share common traits with the person who recommends them.

## Conclusion

In this article, we learned about the concept of sampling, steps involved in sampling, and the different [types of sampling](#) methods. Sampling has wide applications in the statistical world as well as the real world.

Are there any other types of [sampling techniques](#) you feel the community should know? Let me know in the comments section below and we'll discuss!

And as I mentioned earlier, if you're new to data science and statistics, you should really check out the below courses:

- [Introduction to Data Science](#)
- [Applied Machine Learning: Beginner to Professional](#)

## FAQs

### Q1. What are the 5 steps to sampling?

1. Define the target population (who/what to learn about).
2. Select the sampling frame (list of all target population members).
3. Choose a sampling technique (random selection method).
4. Determine the sample size (how many members to include).
5. Collect data from samples (surveys, interviews, or observations).

### Q2. What is 95-5 sampling?

95 5 sampling is a type of stratified sampling in which the target population is divided into two strata: the majority group (95%) and the minority group (5%). Randomly sample each stratum.. This technique is often used to ensure that the minority group is adequately represented in the sample.

### Q3. What is Sampling techniques in research?

Sampling techniques in research are used to select a subset of a population to study. We often cannot study the entire population because it is impractical or impossible. Sampling techniques can be used in all types of research, including quantitative research, qualitative research, and mixed-methods research.

### Q4. What are Sampling techniques in statistics?

Sampling techniques in statistics are used to collect data from a sample of a population in order to make inferences about the entire population. Sampling techniques are an important part of statistical analysis, and they can be used to design a variety of different types of studies.

---



**Ronak Gangwal**