What is 'Random Sampling'

**Definition:** Random sampling is a part of the sampling technique in which each sample has an equal probability of being chosen. A sample chosen randomly is meant to be an unbiased representation of the total population. If for some reasons, the sample does not represent the population, the variation is called a sampling error.  
  
**Description:**Random sampling is one of the simplest forms of collecting data from the total population. Under random sampling, each member of the subset carries an equal opportunity of being chosen as a part of the sampling process. For example, the total workforce in organizations is 300 and to conduct a survey, a sample group of 30 employees is selected to do the survey. In this case, the population is the total number of employees in the company and the sample group of 30 employees is the sample. Each member of the workforce has an equal opportunity of being chosen because all the employees which were chosen to be part of the survey were selected randomly. But, there is always a possibility that the group or the sample does not represent the population as a whole, in that case, any random variation is termed as a sampling error.  
  
An unbiased random sample is important for drawing conclusions. For example when we took out the sample of 30 employees from the total population of 300 employees, there is always a possibility that a researcher might end up picking over 25 men even if the population consists of 200 men and 100 women. Hence, some variations when drawing results can come up, which is known as a sampling error. One of the disadvantages of random sampling is the fact that it requires a complete list of population. For example, if a company wants to carry out a survey and intends to deploy random sampling, in that case, there should be total number of employees and there is a possibility that all the employees are spread across different regions which make the process of survey little difficult.

* Random sampling, also known as probability sampling, is a sampling method that allows for the randomization of sample selection.
* It is essential to keep in mind that samples do not always produce an accurate representation of a population in its entirety; hence, any variations are referred to as sampling errors.
* There are four primary, random (probability) sampling methods – simple random sampling, systematic sampling, stratified sampling, and cluster sampling.

### Types of Random Sampling Methods

There are four primary, random (probability) sampling methods. These methods are:

#### 1. Simple random sampling

Simple random sampling is the randomized selection of a small segment of individuals or members from a whole population. It provides each individual or member of a population with an equal and fair probability of being chosen. The simple random sampling method is one of the most convenient and simple sample selection techniques.

#### 2. Systematic sampling

Systematic sampling is the selection of specific individuals or members from an entire population. The selection often follows a predetermined interval (k). The systematic sampling method is comparable to the simple random sampling method; however, it is less complicated to conduct.

#### 3. Stratified sampling

Stratified sampling, which includes the partitioning of a population into subclasses with notable distinctions and variances. The stratified sampling method is useful, as it allows the researcher to make more reliable and informed conclusions by confirming that each respective subclass has been adequately represented in the selected sample.

#### 4. Cluster sampling

Cluster sampling, which, similar to the [stratified sampling method](https://corporatefinanceinstitute.com/resources/data-science/stratified-random-sampling/), includes dividing a population into subclasses. Each of the subclasses should portray comparable characteristics to the entire selected sample. This method entails the random selection of a whole subclass, as opposed to the sampling of members from each subclass. This method is ideal for studies that involve widely spread populations.

### Practical Example

A company currently employs 850 individuals. The company wishes to conduct a survey to determine employee satisfaction based on a few identified variables. The research team decides to have the sample set at 85 employees. The 85 employees will be part of the survey and will be used as a representation for the total population of 850 employees.

In such a scenario, the sample is the 85 employees, and the population is the entire workforce consisting of 850 individuals. Based on the sample size, any employee from the workforce can be selected for the survey. It goes to say that each employee has an equivalent probability of being randomly selected for the survey.

It is important to keep in mind that samples do not always produce an accurate representation of a population in its entirety; hence, any variations are referred to as [sampling errors](https://corporatefinanceinstitute.com/resources/data-science/sampling-errors/). A sampling error can be defined as the difference between the respective statistics (sample values) and parameters (population values). The sampling error is inevitable when sample data is being used.

### Why an Unbiased Random Sample Matters

Unbiased random sampling results in more reliable and unbiased conclusions.

For example, the [employee satisfaction survey](https://www.surveymonkey.com/mp/employee-satisfaction-surveys/) mentioned above makes use of a sample size of 85 employees. Of these employees, it is possible to have selected more females than males for the study, despite the entire workforce having 450 men and 400 women. It would result in a sampling error, as it causes variations in the results obtained. Ideally, results should be objective and unbiased.

### Probability (Random) Sampling vs. Non-Probability Sampling

Probability – or random sampling – is the random selection of sample participants to derive conclusions and assumptions about an entire population. On the other hand, non-probability sampling is the selection of sample participants based on specified criteria or suitability.

## How to perform simple random sampling

There are 4 key steps to select a simple random sample.

### Step 1: Define the population

Start by deciding on the population that you want to study.

It’s important to ensure that you have access to every individual member of the population, so that you can collect data from all those who are selected for the sample.

### Step 2: Decide on the sample size

Next, you need to decide how large your sample size will be. Although larger samples provide more statistical certainty, they also cost more and require far more work.

There are several potential ways to decide upon the size of your sample, but one of the simplest involves using a formula with your desired confidence interval and confidence level, estimated size of the population you are working with, and the standard deviation of whatever you want to measure in your population.

The most common confidence interval and levels used are 0.05 and 0.95, respectively. Since you may not know the standard deviation of the population you are studying, you should choose a number high enough to account for a variety of possibilities (such as 0.5).

### Step 3: Randomly select your sample

This can be done in one of two ways: the lottery or random number method.

In the**lottery method**, you choose the sample at random by “drawing from a hat” or by using a computer program that will simulate the same action.

In the **random number method**, you assign every individual a number. By using a random number generator or random number tables, you then randomly pick a subset of the population. You can also use the random number function (RAND) in Microsoft Excel to generate random numbers.

### Step 4: Collect data from your sample

Finally, you should collect data from your sample.

To ensure the validity of your findings, you need to make sure every individual selected actually participates in your study. If some drop out or do not participate for reasons associated with the question that you’re studying, this could bias your findings.

For example, if young participants are systematically less likely to participate in your study, your findings might not be valid due to the underrepresentation of this group.

## **How to Conduct a Simple Random Sample**

The simple random sampling process entails size steps. Each step much be performed in sequential order.

### **Step 1: Define the Population**

The origin of statistical analysis is to determine the population base. This is the group in which you wish to learn more about, confirm a [hypothesis](https://www.investopedia.com/terms/h/hypothesistesting.asp), or determine a statistical outcome. This step is to simply identify what that population base is and to ensure that group will adequately cover the outcome you are trying to solve for.

***Example:***I wish to learn how the stocks of the largest companies in the United States have performed over the past 20 years. My population is the largest companies in the United States as determined by the S&P 500.

### **Step 2: Choose Sample Size**

Before picking the units within a population, we need to determine how many units to select This sample size may be constrained based on the amount of time, capital rationing, or other resources available to analyze the sample. However, be mindful to pick a sample size large enough to be truly representative of the population. In the example above, there are constrains in analyzing the performance for every stock in the S&P 500, so we only want to analyze a sub-set of this population.

***Example:*** My sample size will be 20 companies from the S&P 500.

### **Step 3: Determine Population Units**

In our example, the items within the population are easy to determine as they've already been identified for us (i.e. the companies listed within the S&P 500). However, imagine analyzing the students currently enrolled at a university or food products being sold at a grocery store. This steps entails crafting the entire list of all items within your population.

***Example:***Using exchange information, I copy the companies comprising the S&P 500 into an Excel spreadsheet.

### **Step 4: Assign Numerical Values**

The simple random sample process call for every unit within the population receiving an unrelated numerical value. This is often assigned based on how the data may be filtered. For example, I could assign the numbers 1 to 500 to the companies based on market cap, alphabetical, or company formation date. How the values are assigned doesn't entirely matter; all that matters is each value is sequential and each value has an equal chance of being selected.

***Example:***I assign the numbers 1 through 500 to the companies in the S&P 500 based on alphabetical order of the current CEO, with the first company receiving the value '1' and the last company receiving the value '500'.

### **Step 5: Select Random Values**

In step 2, we selected the number of items we wanted to analyze within our population. For the running example, we choose to analyze 20 items. In the fifth step, we randomly select 20 numbers of the values assigned to our variables. In the running example, this is the numbers 1 through 500. There are multiple ways to randomly select these 20 numbers discussed later in this article.

***Example:***Using the random number table, I select the numbers 2, 7, 17, 67, 68, 75, 77, 87, 92, 101, 145, 201, 222, 232, 311, 333, 376, 401, 478, and 489.

### **Step 6: Identify Sample**

The last step of a simple random sample is the bridge step 4 and step 5. Each of the random variables selected in the prior step corresponds to a item within our population. The sample is selected by identifying which random values were chosen and which population items those values match.

***Example:***My sample consists of the 2nd item in the list of companies alphabetically listed by CEO's last name. My sample also consists of company number 7, 17, 67, etc.

## Random Sampling Techniques

There is no single method for determining the random values to be selected (i.e. Step 5 above). The analyst can not simply choose numbers at random as there may not be randomness with numbers. For example, the analyst's wedding anniversary may be the 24th, so they may consciously (or subconsciously) pick the random value 24. Instead, the analyst may choose one of the following methods:

* **Random lottery.**Whether by ping-pong ball or slips of paper, each population number receives an equivalent item that is stored in a box or other indistinguishable container. Then, random numbers are selected by pulling or selecting items without view from the container.
* **Physical Methods.** Simple, early methods of random selection may use dice, flipping coins, or spinning wheels. Each outcome is assigned a value or outcome relating to the population.
* **Random number table.** Many statistics and research books contain sample tables with randomized numbers.
* **Online random number generator.**Many online tools exist where the analyst inputs the population size and sample size to be selected.
* **Random numbers from Excel.** Numbers can be selected in Excel using the =RANDBETWEEN formula. A cell containing =RANDBETWEEN(1,5) will selected a single random number between 1 and 5.

### Simple Random Sampling

Advantages

* Each item within a population has an equal chance of being selected
* There is less of a chance of sampling bias as every item is randomly selected
* This sampling method is easy and convenient for data sets already listed or digitally stored

Disadvantages

* Incomplete population demographics may exclude certain groups from being sampled
* Random selection means the sample may not be truly representative of the population
* Depending on the data set size and format, random sampling may be a time-intensive process

**Characteristics of Sampling**

Various characteristics of sampling are discussed in points given below: –

1. **Goal-oriented:**Design of sampling should be goal oriented. It must align clearly with the objectives of research being conducted and should be in accordance with conditions of survey.
2. **Proper universe representation:**Sample chosen should adequately represent the characteristics of whole population from which it is taken. It should fairly represent details about all units without any biasness. There are different methods of choosing a sample and it need to be chosen with utmost care as improper sampling would lead to error in survey.
3. **Proportional:**Size of sample should be proportional with the size of population. It should be large enough for representing the whole universe and must provide statistical reliability. Sample must ensure proper accuracy for carrying out the particular research study.
4. **Economical:**Process of sampling should be economical requiring minimum cost and efforts for attaining the objectives of survey.
5. **Random selection:**Sample units should be selected on a random basis under which every unit has an equal chance of being chosen. It will ensure that sample is a fair representative of whole population.
6. **Practical:**Design of sample should be simple and practical. It must be capable of easily understood and applicable in fieldwork.

## **Advantages of Sampling**

Various advantages of sampling are as discussed below: –

1. **Lower sampling cost:**Sampling reduces the overall cost involved in doing research. The cost for collecting data about entire population is quite high. Sampling reduces the population into small manageable units. Acquiring data about sample of population involves lower cost which is one of the major advantage.
2. **Less time consuming:**Sampling reduces the overall time by reducing the size of population. Data is not collected about every member in population but only related to sample is gathered. It is less time-consuming in comparison to census technique.
3. **Higher accuracy of data:**A sample represents the whole population from which it is drawn. It is used for calculation of desired descriptive statistics and a stability of derived sample value can be easily determined. Samples permit a high level of accuracy because of limited area of operations. It enables in proper execution of field work and results of studies conducted on the basis of theses sample units turn out to be accurate.
4. **Higher scope of sampling:**Sampling enables investigators to easily arrive at generalizations about set of data. It would be totally impractical to study whole population as it is too large for measuring characteristics of all individual members. Process of sampling by analyzing variables within small proportion of population ease in arriving at generalizations.
5. **Intensive and exhaustive data:**In studies based on sample units, observations are made of a limited number. Therefore, exhaustive and intensive data are collected.
6. **Suitable in case of limited resources:**Sampling is very effective technique of collecting information in presence of limited resources with organization. Studying the whole population requires large amount of resources both in term of money and time. Sampling makes it possible to cover whole population satisfactorily even by employing limited resources.
7. **Better rapport:**Good rapport in between the researcher and respondents is must for carrying out an effective research study. In presence of large population, various issues of rapport arise.

## **Disadvantages of Sampling**

Accuracy of sample is dependent upon appropriateness of sample method used. Theory of sampling focuses on improving the efficiency of sampling. Major difficulties are pose at the time of estimation, selection and administration of samples. Various disadvantages of sampling process are discussed in points given below: –

1. **Chance of Bias:**Major limitation that arises with sampling is chance of biasness in choosing sample units. Selection of samples is a judgmental task as it is based on mindset of individual choosing them. These biased selection does not truly represent the whole population and may lead to faulty conclusions by researcher.
2. **Difficulty in choosing a truly representative sample:** Choosing an adequate and reliable sample that is a truly representative of population remains a difficult task. In case the phenomena under study is of complex nature involving heterogeneous data, it becomes difficult to select proper samples.
3. **Lack of adequate subject knowledge:**Application of sampling process requires proper knowledge regarding sampling technique by individual selecting sample units. This process requires computation of probable error and statistical analysis. There are chances of serious mistakes being committed by researcher in case if he lacks specialized knowledge about sampling. Consequently, overall results of research study conducted will be misleading.
4. **Impossibility of sampling:**Process of sampling is not applicable in cases where universe is too small consisting of heterogeneous set of data. It is difficult to derive a representative sample in such cases. Census study is the only alternative for doing study for such phenomena. Also, sampling is inadequate for studies that needs a high degree of accuracy. There are always chance of errors in sampling even if sample units are chosen with utmost care.