1. Introduction

Welcome to the series of E-learning modules on Sample. In this module, we are going to cover the basic definition of Sample, advantages, essentials required, biased samples, sample size, relation between a sample and a population and role of samples in Sample Theory.

At the end of this session, you will be able to:

- Explain a sample in a sampling theory
- Explain the need of a sample
- Explain objectives and essentials of a sample
- Explain the Biased sample
- Explain the sample size
- Discuss the role of a sample in the Sampling Theory

Shoppers often sample a small piece of cheese before purchasing any. They decide from a small piece what the larger chunk will taste like.
A chemist does the same thing when he takes a sample of alcohol from a still determines that it is ninety proof. If the chemist test all the alchohol or the shoppers tastes all the cheese there will be none to sell. Testing all the products often destroys it and is unnecessary. To determine the characteristic of the whole we have to extract information of only a portion.

“In Statistics, population is the aggregate of objects, animate or inanimate, under study in any statistical investigation.” In sampling theory, the population means a larger group from which the samples are drawn. Based on sample study we can predict and generalize the behaviour of the mass of phenomena. This is possible because there is no statistical population whose elements would vary from each other without limit.

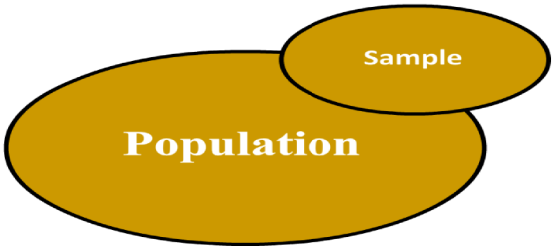
Let us discuss the Population with Sample. Now, observe the figure.
A finite subset of the population selected from it with the objective of investigating its properties is called a sample. However, the sample selected should represent the population. Hence, a sample is a representative portion of the population.

Although the term representativeness has no precise scientific meaning, it carries a common sense that makes it useful in sampling theory. Sample is a representative of the population from which it is selected if the aggregate characteristic of the sample closely approximate those same aggregate characteristics in the population.

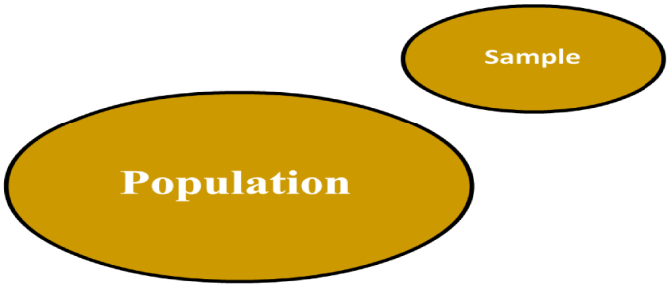
For example, the population contains fifty percent women and the sample must contain “close to” fifty percent women to be representative. Note that the samples need not be representative in all respects. Representativeness is limited to those characteristics that are relevant to the substantive interest of the study. However, we may not know in advance which characteristics are relevant. Following laws guarantees the representativeness of the samples based on its size and randomness.

Observe the figure that represents a bad sample from the population because the sample

selected does not reflect the characteristics of the population.



Look at this figure.



This figure represents a very bad sample from the population because the units that form a sample are not at all a part of the population.

The main objectives for the selection of the samples are:

1) To obtain the optimum results i.e. the maximum information about the characteristics of the population with the available sources at our disposal in terms of time, manpower and money by studying the sample values only.
2) To obtain the best possible estimates of the population parameters with limited resources.

Sampling is an Art of cutting a small portion of material from a large lot and transferring it to the analyzer.

The Theory of Sampling is concerned with estimating the properties of the population from those of the sample and with gauging the precision of the estimate. This sort of movement from the particular (sample) towards general (population) is known as statistical induction or Statistical Inference. In clear terms from the sample, we attempt to draw inference concerning the universe.

2. Need for Sampling

Let us now discuss the need for sampling.
Sampling is used in practice for variety of reasons such as

- Sampling can save time and money. A sample study is usually a less expensive than enumeration of the entire population and produces results at a relatively faster speed.
- Sampling may enable measurements that are more accurate for a sample study is generally conducted by trained and experienced investigators.
- Sampling remains the only way when the population consists of infinitely many members.
- Sampling remains the only choice when the test involves the destruction of the item under study.
- Sampling usually enables to estimate sampling errors and thus assists in obtaining information concerning some characteristics of the population.

In most of the research work and surveys, the usual approach happens to be to generalize or to draw the inferences based on samples about the parameters of the population from which the samples are taken.

The researcher quite often selects only a few items from the population for his study purposes. All this is done on the assumption that the sample data will enable him to estimate the population parameters.

The items so selected constitute what is technically called a sample, their selection process or technique is called sample design and the survey conducted on the basis of sample is described as sample survey. Sample should be truly representative of population characteristics without any bias so that it may result in valid and reliable conclusions.

Let us discuss the Biased samples.
Consider a situation where the Congress is debating some “gun control laws”. You are asked to conduct an opinion survey. Because hunters are the ones that are most affected by the gun control laws, you went to a hunting lodge and interviewed the members there. Then, you reported that in a survey done by you ninety seven percent of the respondents were in favour of repealing all gun control laws.

A week later, Congress took up another bill “ Should working pregnant women be given a maternity leave of one year with full pay to take care of new born babies?” Because this issue affects women most, this time you went to all high-rise office complexes in your city and interviewed several working women of child- bearing age. Again, you reported that in a survey done by you about ninety-three percent of the respondents were in favour of the one-year maternity leave with full pay. In both these situations, you picked a biased sample by choosing people who would have very strong feelings on one side of the issue. This way of selection of samples may lead to disastrous results in decision-making process.

3. Essentials of Sampling and Approaches to Sample Size Calculations

Let us now list the essentials of Sampling.
Samples possess a few essentials. They are

- Adequacy: The size of the sample should not be very small. It should be adequate enough to represent the universe completely.
- Independent selection: All the items of the sample must be selected independently of one another and all the items of the universe must have an equal chance of being selected in the sample.
- Homogeneity: There should be no basic difference in the nature of units of the population and that of the sample. If two samples from the same population are taken they should be more or less similar.
- Representative: A sample should be so selected that it truly represents the population otherwise the results obtained may be misleading to ensure this selection of sample must be random.

Size of a sample:
The number of units which constitute a sample is known as a sample size. One crucial aspect of study design is deciding how big your sample should be. If you increase your sample size you increase the precision of your estimates, which means that, for any given estimate/size of effect, the greater the sample size the more “statistically significant” the result will be. The smaller the deference between the population and a sample you regard as important to detect, the greater the sample size required.

Factors such as time, cost, and how many subjects are actually available are the constraints that often have to be taken account when designing a study. However, these factors should not dictate the sample size. There is no point in carrying out a study that is too small, only to come up with results that are inconclusive, since you will then need to carry out another study to confirm or refute your initial results.

There are two approaches to sample size calculations:
• Precision-based: With what precision do you want to estimate the population constants...
• Power-based: How small a deference is it important to detect and with what degree of certainty?

Sample size: How big is enough?
Qualitative Research: Sample until phenomenon is clear
Quantitative Research: Linked to data Analysis and can be determined statistically
Power Analysis: Statistics are available to calculate the sample size
Non probability sampling: Consider the study purpose, design, sampling, data analysis

Approximate Sample Size

- The bigger the population: It is better up to two thousand five hundred. Beyond two thousand five hundred, it doesn't really matter because accuracy increases very slowly after this point
- The smaller the population: The bigger the sampling ratio that is needed.
- For populations of size under thousand: You need sampling ratio of thirty percent (three hundred elements) to be accurate.
- For populations of about ten thousand: Need sampling ratio of about ten percent.

Further

- Heterogeneity: If the population is heterogeneous, need larger sample to study population that is more diverse.
- Desired precision: If we are aiming at particular level of precision then we need larger sample to get smaller error
- Sampling design: Based on various sampling designs smaller sample size may be used for stratified and larger for cluster sampling
- Nature of analysis: Complex multivariate statistics need larger samples
- Accuracy of sample depends upon the sample size not on n/N

4. Statistical Laws Related to the Samples

Let us now state the statistical laws related to the samples.
The characteristics of the sample will provide a fairly good idea about the population characteristics if it is governed by the following laws:
Law of statistical Regularity:
In words of L. R. Conner “The Law of statistical Regularity lays down that a group of objects Chosen at random from a large group are almost sure on the average to possess the characteristics of the large group”.

According to King, “The Law of statistical Regularity lays down that a moderately large number of items chosen at random from a large group are almost sure on the average to possess the characteristic of their large group”.

The law of statistical regularity impresses upon the following two points:

1. Large sample size: It seems that as the sample size increases, the sample is more likely to reveal the true characteristics of the population and thus provides the better estimates of the population constants.
2. Random selection: Samples should be selected at random from the population. That is we have to select units for the sample such that each and every unit in the population has an equal chance of being selected in the sample.

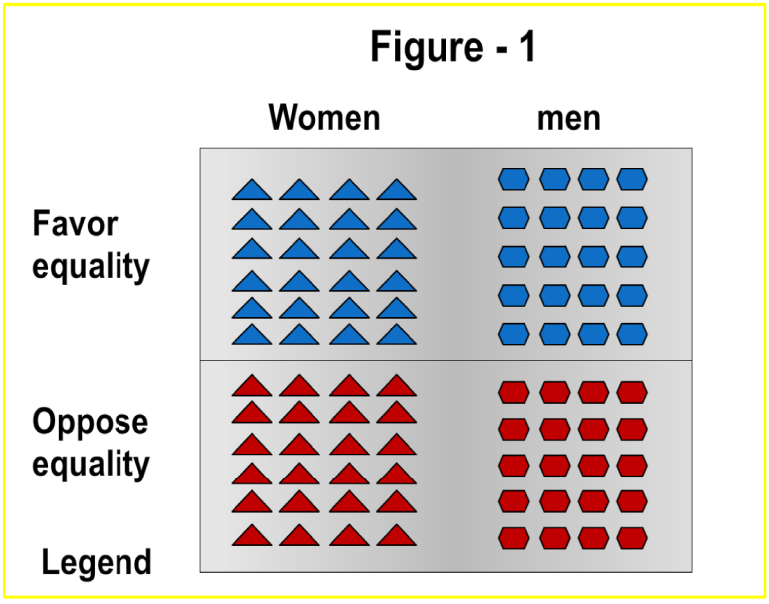
For example, if we are interested in studying the average height of students in Pune University then it is not desirable to resort to hundred percent enumerations of the students in the university. A fairly adequate sample of students from each constituent college may be selected at random and the average height of the students in the selected sample may be computed. Since the sample is random it would be representative of the population and the average so obtained will not differ much from the actual average height of the students of Pune University.

The Logic of statistical significance in the selection of the sample.
We can illustrate the logic of statistical significance best in a series of diagrams representing the selection of samples from a population.
Table shows the distribution of men and women in a hypothetical population to study whether there is any relationship between gender and feelings about equality for men and women.

Table 1

	Women	Men
Favor Equality	50%	50%
Oppose Equality	50%	50%

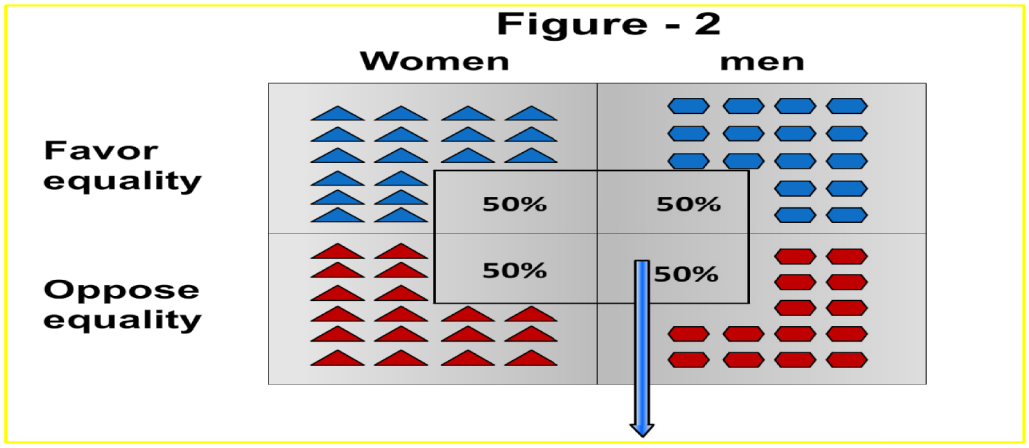
Figure one represents a hypothetical population of people, half are women and half are men. The diagram also indicates how each person feels about seeing women as equal to men. In the diagram, women favouring equality will have a symbol of triangle with blue colours, and men have a “hexagon” with blue colours, those opposing it have the same symbols filled with red colour.



The question we will be investigating whether there is any relationship between gender and feelings about equality for men and women. More specifically, we will see if women are more likely to favour equality than men are, because women would presumably benefit more from it. Take the moment to look at this figure to understand the population.

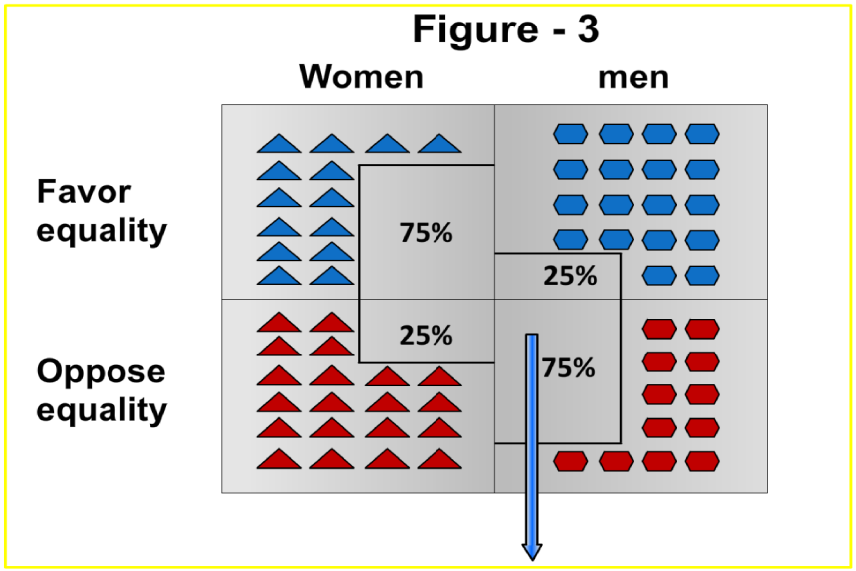
Figure two represents the selection of one-fourth sample from the hypothetical population. In terms of graphic illustration, a “square” selection from the centre of the population provides a representative sample. Notice that our sample contains three lines of each type of person. Half are men, half are women, half of each gender favor equality, and the other half opposes

it. The sample selected in this figure would allow us to draw accurate conclusions about the relationship between gender and feelings about equality in larger population.



Of course, real life samples are seldom such perfect reflections of the population from which they are drawn. It would not be unusual for us to have selected say one or two extra men who opposed equality and a couple of extra women who favoured it. Such minor variations are part and parcel of sampling procedures.

Figure three represents the sample that fall far short of the mark in reflecting the larger population. Notice it includes too many supportive women and opposing men. Figure shows three-fourths of the women in the sample support equality, but only one-fourth of the men to do so. If we select this sample from a population, we would be surely misled by the sample.



Now observe this table.

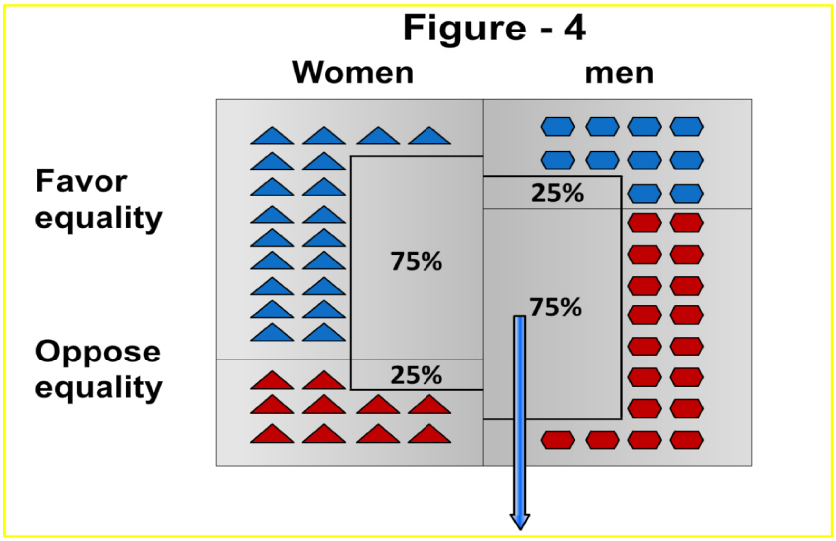
Table 2

	Women	Men
Favor Equality	75%	25%
Oppose Equality	25%	75%

Table shows the distribution of men and women in the same hypothetical population but change in the number of units in the respective groups. Seventy five percent of women and twenty-five percent of men favour equality. The rest of the men and women oppose it.

Figure four represents the most likely situation where we have selected perfectly representative sample. In this figure, women are more likely to support equality than men are, that is the case in the population and the sample reflects it. In practice, of course we never know what is for the total population that is why we select samples. Hence, we should try to select such a sample, which reflects the relationship within the population units.

Figure - 4



5. Testing of Reliability of Samples

The reliability of samples can be tested in the following ways.

- More samples of the same size should be taken from the same population and results be compared. If results are similar, the sample will be reliable.
- If the measurements of the universe are known then they should be compared with the measurements of the sample. In case of similarity of measurements, the sample will be reliable.
- Sub-sample should be taken from the samples and studied. If the results of the sample and sub sample show similarity, then sample should be considered reliable.

Suppose that a personnel director of a large bank need to write a report describing all the employees who have voluntarily left the company in the last 10 years. We would have a difficult task locating all these thousands of people. They are not easily accessible as group- Many have died, moved from community, left the country or acquired a new name by marriage. How do we write the report? The best idea is to locate a representative sample and interview them in order to generalize about the entire group.
Time is also a factor when managers need information quickly in order to adjust an operation or change a policy.

Consider an automatic machine that sorts thousands of pieces of mails daily. Why wait for an entire day's output to check whether the machine is working accurately. Instead, samples can be taken at specific intervals and if necessary, the machine can be adjusted right away. We use sampling when it is not possible to count or measure every item in the population.

In the above cases, the population is all the employees of the large bank who voluntarily left in the last ten years, and all mails sorted by the automatic machine. The samples are selected employees of the large bank who voluntarily left in the last ten years and selected mails sorted by the automatic machine at a specific intervals of time.

It is obvious that the most important aim of drawing a sample from a population is to obtain maximum information about the population under study with the least sacrifice of time, money and energy. If the sample study has been made in such a manner that we can obtain a large variety of information about the phenomena to which the sample relates, it would be easy for us to have an idea about similar information related to the population.

For example, the sample study related to expenditure of selected students in Indian Universities has been done properly; they would give us an idea about the distribution of expenditure of all the students in Indian Universities. Thus, the aim of selection of representative portion of the population is to obtain the best possible values of the parameters of the population.

Here's a summary of our learning in this session:

- Statistical definition of sample
- Advantages of a sample
- Objectives and essentials of samples
- Laws governing the samples

- Biased samples
    - Size of the sample and its reliability
    - Importance of samples in sampling theory