WHY STATISTICAL DATA

The need to gather information arises in almost every conceivable sphere of human activity. Many of the questions that are subject to common conversation and controversy require numerical data for their resolution. Data resulting from the physical, chemical, and biological experiments in the form of observations are used to test different theories and hypotheses. Various social and economic investigations are carried out through the use and analysis of relevant data. The data collected and analyzed in an objective manner and presented suitably serve as basis for taking policy decisions in different fields of daily life. The important users of statistical data, among others, include **government, industry, business, research institutions, public organizations, and international agencies and organizations**. To discharge its various responsibilities, the government needs variety of information regarding **different sectors of economy, trade, industrial production, health and mortality, population, livestock, agriculture, forestry, environment, meteorology, and available resources**. The inferences drawn from the data help in determining future needs of the nation and also in tackling social and economic problems of people. For instance, the information on cost of living for different categories of people, living in various parts of the country, is of importance in shaping its policies in respect of wages and price levels. Data on health, mortality, and population could be used for formulating policies for checking population growth. Similarly, information on forestry and environment is needed to plan strategies for a cleaner and healthier life. Agricultural production data are of immense use to the state for planning to feed the nation. In case of industry and business, the information is to be collected on labor, cost and quality of production, stock, and demand and supply positions for proper planning of production levels and sales campaigns. The research institutions need data to verify the earlier findings or to draw new inferences. The data are used by public organizations to assess the state policies, and to point it out to the administration if these are not up to the expectations of the people. The international organizations collect data to present comparative positions of different countries in respect of economy, education, health, culture, etc. Besides, they also use it to frame their policies at the international level for the welfare of people.

TYPES OF DATA

The collection of required information depends on the nature, object, and scope of study on the one hand and availability of financial resources, time, and man power on the other. The statistical data are of two types: (l) primary data, and (2) secondary data.

Definition 1: The data collected by the investigator from the original source are called primary data.

Definition 2: If the required data had already been collected by some agencies or individuals and are now available in the published or unpublished records, these are known as secondary data.

Technical Terms of Sampling

Element: An element is a unit for which information is sought.

Population: The population or universe is an aggregate of elements, about which the inference is to be made.

Sampling units: Sampling units are non-overlapping collections of elements of the population.

Sampling frame: A list of all the units in the population to be sampled is termed frame or sampling frame.

Sample: A subset of population selected from a frame to draw inferences about a population characteristic is called a sample.

NEED FOR A SAMPLE

Collection of information on every unit in the population for the characteristics of interest is known as complete enumeration or census. The money, manpower, and time required for carrying out a census will generally be large, and there are many situations where with limited means complete enumeration is not possible. There are also instances where it is not feasible to enumerate all units due to their perishable nature. In all such cases, the investigator has no alternative except resorting to a sample survey. The number of units (not necessarily distinct) included in the sample is known as the sample size and is usually denoted by n, whereas the number of units in the population is called population size and is denoted by N. The ratio n/N is termed as sampling fraction. There are certain advantages of a sample survey over complete enumeration. These are given below:

Greater Speed

The time taken for collecting and analyzing the data for a sample is much less than that for a complete enumeration. Often, we come across situations where the information is to be collected within a specified period. In such cases, where time available is short or the population is large, sampling is the only alternative.

Greater Accuracy

A census usually involves a huge and unwieldy organization and, therefore, many types of errors may creep in. Sometimes, it may not be possible to control these errors adequately. In sample surveys, the volume of work is considerably reduced. On account of this, the services of better trained and efficient staff can be obtained without much difficulty. This will help in producing more accurate results than those for complete enumeration.

More Detailed Information

As the number of units in a sample are much less than those in census, it is, therefore, possible to observe/interview each and every sample unit intensively. Also, the information can be obtained on more number of variables. However, in complete enumeration such an effort becomes comparatively difficult.

Reduced Cost

Because of lesser number of units in the sample in comparison to the population, considerable time, money, and energy are saved in observing the sample units in relation to the situation where all units in the population are to be covered. From the above discussion, it is seen that the sample survey is more economical, provides more accurate information, and has greater scope in subject coverage as compared to a complete enumeration. It may, however, be pointed out here that sampling errors are present in the results of the sample surveys. This is due to the fact that only a part of the whole population is surveyed. On the other hand, non-sampling errors are likely to be more in case of a census study than these are in a sample survey. Merits and demerits of sample surveys have been discussed in detail by Zarkovich (1961) and Lahiri (1963).

SAMPLING PROCEDURES

The method which is used to select the sample from a population is known as sampling procedure. These procedures can be put into two categories: probability sampling and non-probability sampling. These two types of surveys are not distinguished by the questionnaire and instructions to be followed, but by the methods of selecting the sample for obtaining the estimates of the population characteristics of interest and their precision.

WHAT IS PROBABILITY

Probability is a measure of the likelihood of an event to occur. Many events cannot be predicted with total certainty. We can predict only the chance of an event to occur i.e. how likely they are to happen, using it. Probability can range in from 0 to 1, where 0 means the event to be an impossible one and 1 indicates a certain event. Probability for Class 10 is an important topic for the students which explain all the basic concepts of this topic. The probability of all the events in a sample space adds up to 1.

Probability Sampling

Definition: If the units in the sample are selected using some probability mechanism, such a procedure is called probability sampling.

This type of survey assigns to each unit in the population a definite probability of being selected in the sample. Alternatively, it enables us to define a set of distinct samples which the procedure is capable of selecting if applied to a specific population. The sampling procedure assigns to each possible sample a known probability of being selected. One can build suitable estimators for different population characteristics for probability samples. For any sampling procedure of this

type, one is in a position to develop theory by using probability apparatus. It is also possible to obtain frequency distribution of the estimator values it generates if repeatedly applied to the same population. The measure of the sampling variation can also be obtained for such procedures, and the proportion of estimates that will fall in a specified interval around the true value can be worked out. The procedures such as these will only be considered in this book.

Non-probability Sampling

Definition: The procedure of selecting a sample without using any probability mechanism is termed as nonprobability sampling.

The convenience sampling and the purposive sampling belong to this category. In convenience sampling, the sample is restricted to a part of the population that is readily accessible. For example, a sample of coal from an open wagon may be taken from the depth of up to 50 cm from the top. In studies where the process of taking observations is inconvenient, unpleasant, or troublesome to the selected person, only the volunteers may constitute the sample. Purposive sampling (also termed judgment sampling) is common when special skills are required to form a representative subset of population. For instance, auditors often use judgment samples to select items for study to determine whether a complete audit of items may be necessary. Sometimes, quotas are fixed for different categories of population based on considerations relevant to the study being conducted, and selections within the categories are based on personal judgment. This type of sampling procedure is also termed quota sampling. Obviously, these methods are subject to human bias. In appropriate conditions, these methods can provide useful results. They are, however, not amenable to the development of relevant theory and statistical analysis. In such methods, the sampling error can not be objectively determined. Hence, they are not comparable with the available probability sampling methods.

WITH AND WITHOUT REPLACEMENT SAMPLING

In with replacement (WR) sampling, the units are drawn one by one from the population, replacing the unit selected at any particular draw before executing the next draw.

As the constitution of population remains same at each draw, some units in with replacement sample may get selected more than once. This procedure gives rise to N^n possible samples when order of selection of units in the sample is taken into account, where N and n denote the population and sample sizes respectively.

PLANNING AND EXECUTION OF SAMPLE SURVEYS

Sample survey techniques are used widely as an organized and fact finding instrument. The quality of the inferences drawn about the population characteristics from the sample data is related to, how well, the sample represents the population. It requires selecting a suitable sampling plan, and implementing it in a way that ensures the sample to be a good representative

of the population under study. It is, therefore, essential to describe briefly the steps involved in the planning and execution of a survey. Surveys vary greatly in their scope and complexity. Problems that are baffling in one survey, may be trivial or nonexistent in another. Some ofthe important aspects requiring attention at the planning stage are grouped under the following heads:

Objectives

The first task is to lay down, in concrete terms, the objectives of the survey. The investigator should ensure that these objectives are commensurate with available resources in terms of money, manpower, and the time limit specified for the survey.

Population to be studied

The population to be covered by the survey should be clearly defined. An exact description should be given of the geographical region and the categories of the material to be covered by the survey. For instance, in a survey of human population, it is necessary to specify whether such categories as hotel residents, institutions, military personnel, etc., were to be included or not. Population to be sampled should coincide with the target population about which inferences are to be drawn. However, sometimes impracticability and inconvenience may result in the leaving out of certain segments of the population from the scope of the survey. If so, the conclusions drawn will apply only to the population actually sampled. Any supplementary information gathered for the omitted sectors, which can throw some light on the subject matter of the survey, will be useful.

Sampling Unit

The population should be capable of being divided into sampling units, and these should be properly defined. For example, a human population can be considered to be built up of villages, localities, households, persons, etc. The division of population into sampling units should be unambiguous. Every element of the population should correspond to just one and only one sampling unit. The border line cases can be handled by framing some appropriate rules.

The Sampling Frame In surveys, as already discussed, it is always desired that the sampled and the target population should coincide. It should, therefore, be ensured that all the sampling units of the population under study are included in the frame. The frame should be up to date and free from errors of omission and overlapping.

Sample Selection

The size of the sample and manner of selecting the sample should receive careful attention. After taking various technical, operational, and risk factors into consideration, an optimum size of the sample and sampling procedure need to be decided upon. While doing so, the aim of achieving either a given degree of precision with a minimum cost, or the maximum precision with a fixed

cost, should be kept in mind. It should also be ensured that the sample is representative of the population.

Methods of Collecting Information

After a careful examination of the frame, the method of sample selection, available resources, and the objectives of survey, one should decide about the type of data to be used, that means, whether to collect primary data or to use secondary data. In case the primary data are to be collected, the investigator should decide whether data are to be collected by personal face-to-face interview, by mail, through enumerators, or by telephone interview.

Handling of Nonresponse

Procedures should be devised to deal with the respondents, who do not give information by choice, or are not found at home. The reason for nonresponse should also be ascertained. This helps in assessing the effect of refusals and random nonresponse on the conclusions to be drawn.

Pilot Survey

Where some prior information about the nature of population under study, and the operational and cost aspects of data collection and analysis, is not available from past surveys, it is desirable to design and carry out a pilot survey. It will be useful for : (1) discovering shortcomings in the questionnaire/schedule, (2) evolving suitable strategies for field and analysis work, and (3) training the staff for the purpose.

Organization of Field Work

Different aspects of field work such as recruitment and training of investigators, and inspection and supervision of field staff should be given due consideration in the light of the prevailing operational conditions. The personnel engaged in the survey must receive training, not only in the purpose of the survey and in the methods of measurement to be employed, but also in the art of eliciting acceptable responses. The investigators should be able to withstand long and arduous travel, sometimes in inhospitable conditions. The work must be adequately supervised, as it is important for the investigator to adhere to procedures and tact in answering the questions raised by respondents. Besides, it will help in resolving unusual or unforeseen problems in the field. A quality check and editing need to be instituted to make careful review of questionnaires received. It will be valuable in amending the recording errors and deleting the data that are erroneous and superfluous.

Analysis of Data and Preparation of Report

The stage of analysis of collected data and drawing inferences from a sample is a vital issue, as the results of survey are the backbone of the policies to be framed. The errors creeping in the tabulation and statistical analysis of data should be kept under control. Last, but not the least,

comes the report writing. While writing the report, the objectives, the scope, and the subject coverage must be mentioned. It is also essential to clarify the method of data collection, estimation procedure including tabulation and analysis, and cost structure in the report. A brief description of the organizations sponsoring and conducting the survey should also be included. Relevant published papers and reports should be cited for reference. The report should conclude with a summary of findings and suggestions for possible action to be taken. For a report on an actual sample survey, the reader may refer to Des Raj (1968).

Many interesting examples showing the range of applications of the sampling methods in business have been given by Deming (1960) and Slonim (1960).

WITH AND WITHOUT REPLACEMENT SAMPLING

In with replacement (WR) sampling, the units are drawn one by one from the population, replacing the unit selected at any particular draw before executing the next draw.

As the constitution of population remains same at each draw, some units in with replacement sample may get selected more than once. This procedure gives rise to N^n possible samples when order of selection of units in the sample is taken into account, where N and n denote the population and sample sizes respectively.

What is probability sampling?

Definition: Probability sampling is defined as a sampling technique in which the researcher chooses samples from a larger population using a method based on the theory of probability. For a participant to be considered as a probability sample, he/she must be selected using a random selection.

The most critical requirement of probability sampling is that everyone in your population has a known and equal chance of getting selected. For example, if you have a population of 100 people, every person would have odds of 1 in 100 for getting selected. Probability sampling gives you the best chance to create a sample that is truly representative of the population.

Probability sampling uses statistical theory to randomly select a small group of people (sample) from an existing large population and then predict that all their responses will match the overall population.

What are the types of probability sampling?

Simple random sampling, as the name suggests, is an entirely random method of selecting the sample. This sampling method is as easy as assigning numbers to the individuals (sample) and then randomly choosing from those numbers through an automated process. Finally, the numbers that are chosen are the members that are included in the sample.

There are two ways in which researchers choose the samples in this method of sampling: The lottery system and using number generating software/ random number table. This sampling technique usually works around a large population and has its fair share of advantages and disadvantages.

Stratified random sampling involves a method where the researcher divides a more extensive population into smaller groups that usually don't overlap but represent the entire population. While sampling, organize these groups and then draw a sample from each group separately.

A standard method is to arrange or classify by sex, age, ethnicity, and similar ways. Splitting subjects into mutually exclusive groups and then using simple random sampling to choose members from groups.

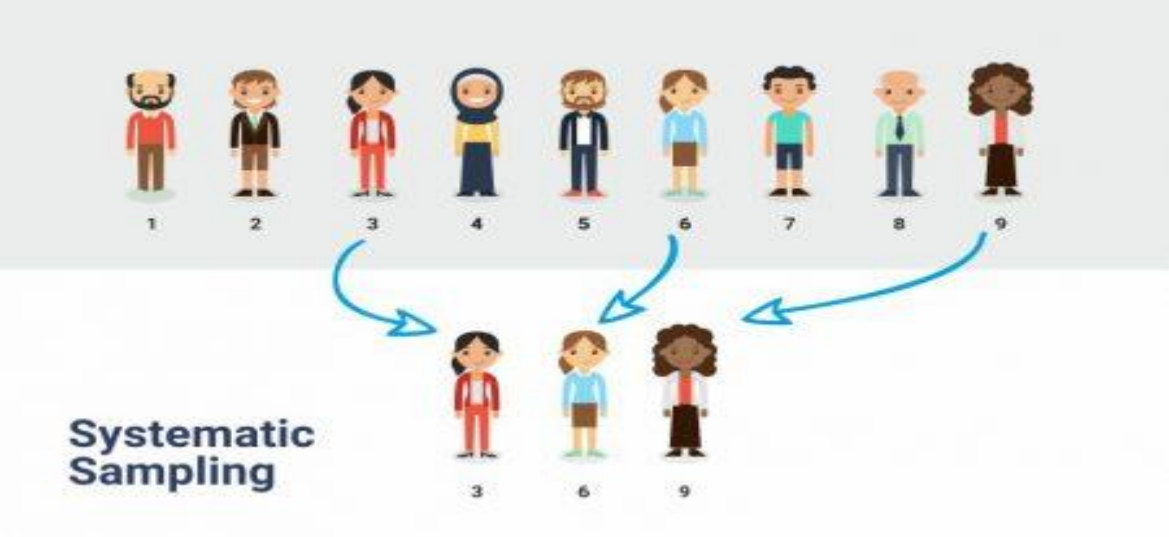
Members of these groups should be distinct so that every member of all groups get equal opportunity to be selected using simple probability. This sampling method is also called "random quota sampling."

Random cluster sampling is a way to select participants randomly that are spread out geographically. For example, if you wanted to choose 100 participants from the entire population of the U.S., it is likely impossible to get a complete list of everyone. Instead, the researcher randomly selects areas (i.e., cities or counties) and randomly selects from within those boundaries.

Cluster sampling usually analyzes a particular population in which the sample consists of more than a few elements, for example, city, family, university, etc. Researchers then select the clusters by dividing the population into various smaller sections.

Systematic sampling is when you choose every "nth" individual to be a part of the sample. For example, you can select every 5th person to be in the sample. Systematic sampling is an

extended implementation of the same old probability technique in which each member of the group is selected at regular periods to form a sample. There's an equal opportunity for every member of a population to be selected using this sampling technique.



Example of probability sampling

Let us take an example to understand this sampling technique. The population of the US alone is 330 million. It is practically impossible to send a survey to every individual to gather information. Use probability sampling to collect data, even if you collect it from a smaller population.

For example, an organization has 500,000 employees sitting at different geographic locations. The organization wishes to make certain amendments in its human resource policy, but before they roll out the change, they want to know if the employees will be happy with the change or not. However, it's a tedious task to reach out to all 500,000 employees. This is where probability sampling comes handy. A sample from the larger population i.e., from 500,000 employees, is chosen. This sample will represent the population. Deploy a survey now to the sample.

From the responses received, management will now be able to know whether employees in that organization are happy or not about the amendment.

What are the steps involved in probability sampling?

Follow these steps to conduct probability sampling:

1. Choose your population of interest carefully: Carefully think and choose from the population, people you believe whose opinions should be collected and then include them in the sample.

2. Determine a suitable sample frame: Your frame should consist of a sample from your population of interest and no one from outside to collect accurate data.

3. Select your sample and start your survey: It can sometimes be challenging to find the right sample and determine a suitable sample frame. Even if all factors are in your favor, there still might be unforeseen issues like cost factor, quality of respondents, and quickness to respond. Getting a sample to respond to a probability survey accurately might be difficult but not impossible.

But, in most cases, drawing a probability sample will save you time, money, and a lot of frustration. You probably can't send surveys to everyone, but you can always give everyone a chance to participate, this is what probability sample is all about.

When to use probability sampling?

Use probability sampling in these instances:

1. When you want to reduce the sampling bias: This sampling method is used when the bias has to be minimum. The selection of the sample largely determines the quality of the research's inference. How researchers select their sample largely determines the quality of a researcher's findings. Probability sampling leads to higher quality findings because it provides an unbiased representation of the population.

2. When the population is usually diverse: Researchers use this method extensively as it helps them create samples that fully represent the population. Say we want to find out how many people prefer medical tourism over getting treated in their own country. This sampling method will help pick samples from various socio-economic strata, background, etc. to represent the broader population.

3. To create an accurate sample: Probability sampling help researchers create accurate samples of their population. Researchers use proven statistical methods to draw a precise sample size to obtained well-defined data.

Advantages of probability sampling

Here are the advantages of probability sampling:

1. It's Cost-effective: This process is both cost and time effective, and a larger sample can also be chosen based on numbers assigned to the samples and then choosing random numbers from the more significant sample.

2. It's simple and straightforward: Probability sampling is an easy way of sampling as it does not involve a complicated process. It's quick and saves time. The time saved can thus be used to analyze the data and draw conclusions.

3. It is non-technical: This method of sampling doesn't require any technical knowledge because of its simplicity. It doesn't require intricate expertise and is not at all lengthy.

What is the difference between probability sampling and non-probability sampling?

Here's how you differentiate probability sampling from non-probability sampling,

Probability sampling	**Non-probability sampling**
The samples are randomly selected.	Samples are selected on the basis of the researcher's subjective judgment.
Everyone in the population has an equal chance of getting selected.	Not everyone has an equal chance to participate.
Researchers use this technique when they want to keep a tab on sampling bias.	Sampling bias is not a concern for the researcher.
Useful in an environment having a diverse population.	Useful in an environment that shares similar traits.

Probability sampling	Non-probability sampling
Used when the researcher wants to create accurate samples.	This method does not help in representing the population accurately.
Finding the correct audience is not simple.	Finding an audience is very simple.

What is non-probability sampling?

Definition: Non-probability sampling is defined as a sampling technique in which the researcher selects samples based on the subjective judgment of the researcher rather than random selection. It is a less stringent method. This sampling method depends heavily on the expertise of the researchers. It is carried out by observation, and researchers use it widely for qualitative research.

Non-probability sampling is a sampling method in which not all members of the population have an equal chance of participating in the study, unlike probability sampling. Each member of the population has a known chance of being selected. Non-probability sampling is most useful for exploratory studies like a pilot survey (deploying a survey to a smaller sample compared to pre-determined sample size). Researchers use this method in studies where it is impossible to draw random probability sampling due to time or cost considerations.

Select your respondents

Types of non-probability sampling

Here are the types of non-probability sampling methods:

Convenience sampling:

Convenience sampling is a non-probability sampling technique where samples are selected from the population only because they are conveniently available to the researcher. Researchers choose these samples just because they are easy to recruit, and the researcher did not consider selecting a sample that represents the entire population.

Ideally, in research, it is good to test a sample that represents the population. But, in some research, the population is too large to examine and consider the entire population. It is one of the reasons why researchers rely on convenience sampling, which is the most common non-probability sampling method, because of its speed, cost-effectiveness, and ease of availability of the sample.

Consecutive sampling:

This non-probability sampling method is very similar to convenience sampling, with a slight variation. Here, the researcher picks a single person or a group of a sample, conducts research over a period, analyzes the results, and then moves on to another subject or group if needed. Consecutive sampling technique gives the researcher a chance to work with many topics and fine-tune his/her research by collecting results that have vital insights.

Quota sampling:

Hypothetically consider, a researcher wants to study the career goals of male and female employees in an organization. There are 500 employees in the organization, also known as the population. To understand better about a population, the researcher will need only a sample, not the entire population. Further, the researcher is interested in particular strata within the population. Here is where quota sampling helps in dividing the population into strata or groups.

Judgmental or Purposive sampling:

In the judgmental sampling method, researchers select the samples based purely on the researcher's knowledge and credibility. In other words, researchers choose only those people who they deem fit to participate in the research study. Judgmental or purposive sampling is not a scientific method of sampling, and the downside to this sampling technique is that the preconceived notions of a researcher can influence the results. Thus, this research technique involves a high amount of ambiguity.

Snowball sampling:

Snowball sampling helps researchers find a sample when they are difficult to locate. Researchers use this technique when the sample size is small and not easily available. This sampling system works like the referral program. Once the researchers find suitable subjects, he asks them for assistance to seek similar subjects to form a considerably good size sample.

Non-probability sampling examples

Here are three simple examples of non-probability sampling to understand the subject better.

1. An example of convenience sampling would be using student volunteers known to the researcher. Researchers can send the survey to students belonging to a particular school, college, or university, and act as a sample.
2. In an organization, for studying the career goals of 500 employees, technically, the sample selected should have proportionate numbers of males and females. Which means there should be 250 males and 250 females. Since this is unlikely, the researcher selects the groups or strata using quota sampling.
3. Researchers also use this type of sampling to conduct research involving a particular illness in patients or a rare disease. Researchers can seek help from subjects to refer to other subjects suffering from the same ailment to form a subjective sample to carry out the study.

When to use non-probability sampling?

- Use this type of sampling to indicate if a particular trait or characteristic exists in a population.
- Researchers widely use the non-probability sampling method when they aim at conducting qualitative research, pilot studies, or exploratory research.
- Researchers use it when they have limited time to conduct research or have budget constraints.
- When the researcher needs to observe whether a particular issue needs in-depth analysis, he applies this method.
- Use it when you do not intend to generate results that will generalize the entire population.

Advantages of non-probability sampling

Here are the advantages of using the non-probability technique

- Non-probability sampling techniques are a more conducive and practical method for researchers deploying surveys in the real world. Although statisticians prefer probability sampling because it yields data in the form of numbers, however, if done correctly, it can produce similar if not the same quality of results.
- Getting responses using non-probability sampling is faster and more cost-effective than probability sampling because the sample is known to the researcher. The respondents respond quickly as compared to people randomly selected as they have a high motivation level to participate.

Example of Consecutive sampling

Here is an easy to understand example of consecutive sampling

- One of the most common examples of a consecutive sample is when companies/ brands stop people in a mall or crowded areas and hand them promotional leaflets to purchase a luxury car.
- In this example, the people walking in the mall are the samples, and let us consider them as representative of a population.
- Now, the researcher hands these people an advertisement or a promotional leaflet. A few of them agree to stay back and respond to the questions asked by the promotion executive (we can consider him/her as a researcher).
- The responses are collected and analyzed, but there is no conclusive result that people would want to buy that car based on the features described in the leaflet.
- The promotion executive now asks questions to another group of people who analyze the details of the car and its features and show a keen interest in buying the luxury car. Thus, this group of people has provided conclusive results for purchasing the vehicle.

However, there is a downside to this sampling method. You cannot consider the sample to be representative of the entire population. In this example, not all people who have taken this leaflet were interested in buying the car.

Sampling Distribution

What is a Sampling Distribution?

A sampling distribution is a probability distribution of a statistic obtained from a larger number of samples drawn from a specific population. The sampling distribution of a given population is

the distribution of frequencies of a range of different outcomes that could possibly occur for a statistic of a population.

In statistics, a population is the entire pool from which a statistical sample is drawn. A population may refer to an entire group of people, objects, events, hospital visits, or measurements. A population can thus be said to be an aggregate observation of subjects grouped together by a common feature.

- A sampling distribution is a statistic that is arrived out through repeated sampling from a larger population.
- It describes a range of possible outcomes that of a statistic, such as the mean or mode of some variable, as it truly exists a population.
- The majority of data analyzed by researchers are actually drawn from samples, and not populations.

Understanding Sampling Distribution

A lot of data drawn and used by academicians, statisticians, researchers, marketers, analysts, etc. are actually samples, not populations. A sample is a subset of a population. For example, a medical researcher that wanted to compare the average weight of all babies born in North America from 1995 to 2005 to those born in South America within the same time period cannot within a reasonable amount of time draw the data for the entire population of over a million childbirths that occurred over the ten-year time frame. He will instead only use the weight of, say, 100 babies, in each continent to make a conclusion. The weight of 200 babies used is the sample and the average weight calculated is the sample mean.

Now suppose that instead of taking just one sample of 200 newborn weights from each continent, the medical researcher takes repeated random samples from the general population, and computes the sample mean for each sample group. So, for North America, he pulls up data for 200 newborn weights recorded in the US, Canada and Mexico as follows: four 200 samples from select hospitals in the US, five 70 samples from Canada and three 150 records from Mexico, for a total of 1200 weights of newborn babies grouped in 12 sets. He also collects a sample data of 100 birth weights from each of the 12 countries in South America.

Each sample has its own sample mean and the distribution of the sample means is known as the sample distribution. Sampling distribution is conducted from sample averages rather than sample units.

The average weight computed for each sample set is the sampling distribution of the mean. Not just the mean can be calculated from a sample. Other statistics, such as the standard deviation, variance, proportion, and range can be calculated from sample data. The standard deviation, range and variance measure the variability of the sampling distribution. In averages, we include mean, median, mode, quartiles, quintiles, harmonic mean, geometric mean etc.

The number of observations in a population, the number of observations in a sample and the procedure used to draw the sample sets determine the variability of a sampling distribution. The

standard deviation of a sampling distribution is called the **standard error**. While the mean of a sampling distribution is equal to the mean of the population, the standard error depends on the standard deviation of the population, the size of the population and the size of the sample.

Objective of sampling distribution: Knowing how spread apart the mean of each of the sample sets are from each other and from the population mean will give an indication of how close the sample mean is to the population mean. The standard error of the sampling distribution decreases as the sample size increases.

Special Considerations

A population or one sample set of numbers will have a normal distribution. However, because a sampling distribution includes multiple sets of observations, it will not necessarily have a bell-curved shape.

Following our example, the population average weight of babies in North America and in South America has a normal distribution because some babies will be underweight (below the mean) or overweight (above the mean), with most babies falling in between (around the mean). If the average weight of newborns in North America is seven pounds, the sample mean weight in each of the 12 sets of sample observations recorded for North America will be close to seven pounds as well.

However, if you graph each of the averages calculated in each of the 1,200 sample groups, the resulting shape may result in a uniform distribution, but it is difficult to predict with certainty what the actual shape will turn out to be. The more samples the researcher uses from the population of over a million weight figures, the more the graph will start forming a normal distribution.