BLOCK 3
POPULATION AND SAMPLING

BLOCK 3 : POPULATION AND SAMPLING

Introduction

The third block of this course is of *two units*. In the *first unit*, you will be introduced with the concept of population and sampling. The unit will also deal with the meaning of sample size and the different types of bias involved in sampling. The concept and importance of placebo effect will also be explained.

The *second unit* will help you to understand the various techniques of sampling. You will also come to know, which technique of sampling is suitable to be used in what type of research.

UNIT 6 POPULATION AND SAMPLING *

Structure

6.0 OBJECTIVES

After reading this unit, you will be able to:

- Explain the concept of Population and Sampling;
- Differentiate between Population and sample;
- Discuss the concept of sample size;
- Explain the different types of sampling errors; and
- Describe the different kinds of experiments.

* Dr. Sarika Boora, Psychologist, Sambharti Foundations , New Delhi.

6.1 INTRODUCTION

After hearing the word 'population', the first thing we think of is a large group of individuals. Similarly, for statistical purpose a population signifies a large group of elements that have at least one common attribute. The term population is often contrasted with sample, which refers to nothing but a subgroup of the population that is chosen such that it represents the entire group. In other words, population represents the entirety of persons, objects, units and anything that can be thought of, having certain properties. On the contrary, a sample is a finite subset of that population which is chosen by a systematic process, to explore the characteristics of the parent set.

6.2 MEANING OF POPULATION IN CONTEXT OF SAMPLING

Simply put, population refers to the aggregate of all elements under a study having one or more common characteristics, for example, all human beings living in India make up its population. Population is not limited to just people and can also include animals, buildings, objects, events, etc. There is no limit to its size, and the total members or elements in a population is known as population size, i.e. if we consider that India has 125 million people then the size of the population (N) will be 125 million.

A population considered for research is normally a large group of people, objects or individuals who are the primary focus of a scientific examination. Research is conducted to benefit this group, however, due to its large size; researchers find it difficult to test each individual of the population, as it takes a lot of time and money. For this reason researchers rely on sampling techniques. Another definition of research population can be 'a well-defined collection of objects or individuals with similar characteristics'. All individuals or objects within a certain population more often have a prevalent binding attribute or trait.

The different population types are mentioned below:

1) *Finite Population:* We can call the population finite when the number of elements in the population is fixed and thus making it possible to count them it in totality.

2) *Infinite Population:* If we cannot count the number of units in a given population, thus making it impossible to observe all the items of the universe, then it is termed as infinite population.

3) *Existent Population:* It is the population which consists of elements that exist in reality.

4) *Hypothetical Population:* It is the population which is assumed or is imaginary or exists hypothetically.

Examples of population:

All the employees of an automobile factory;

All cars made by a vehicle manufacturer;

The count of bees in a city will be their population;

The income tax payers in the country.

6.2.1 Relationship between Sample and Population in Research

A sample refers to a subgroup or subset of the given population. Because of the researcher's inability to survey all individuals or subjects in the total population. It should represent the population. The researcher selects a sample from that population. It should represent the population from which it is drawn and it should have ample size to conduct statistical analysis. The most important function of a sample is to give researchers the ability to conduct a study on individuals from the population in order to make the results of their study relevant in deriving conclusions that are applicable to the entire population.

6.3 SAMPLE

A sample signifies a part of the population that is chosen at random to participate in a study. The chosen sample should represent all the characteristics of the population, while being free from bias, thereby producing a miniature cross-section, because the generalisations about the population are inferred from the observation made from the sample.

In simple terms, individuals selected from population make up a 'sample', and the process that selects these individuals is termed as 'sampling'. Each individual under observation is known as a sampling unit, and the numbers of individuals in the sample make the sample size.

Hence, we can define a sample group as a subset of a population. The population is the total population for which the information is needed. Ideally, it is the population exposed to the risk. This ''study population'' is the population from which we intend to draw the sample. Generally, due to its large size and for a research study, observing the entire population is often neither practical nor possible. Hence, by a sample unit, researchers get a manageable representation of the population.

During statistical testing, a researcher uses sample when the sample size is very large and is not able include all individuals of the population under study.

So let us understand the difference between a sample and population through an example:

Imagine that there is a box of 1000 pens of 4 different colours. All these Pens together are called the population. Now it will require a lot of time and effort to evaluate each pen for the research, so for measuring we need parameters in descriptive statistics- here, we have yellow, red, black and blue pens. So, by selecting 100 pens from this box- a subset of the population – we get a sample of 100 pens.

- *Population:* The set of all possible conditions of a random variable refers to its population. Its size can be either finite or infinite.
- *Sample:* It can be defined as a finite subset of the population.

The process of taking a subset of participants that represents the entire population is called population sampling. For a statistical analysis, the sample must be of sufficient size. A researcher does sampling because it is nearly impossible to test every subject of the population, the other benefits being reducing the time, effort and money involved. Still, each researcher must take into consideration that the best scenario is to test all the subjects to get valid, reliable and accurate results. Hence sampling techniques are used only when it is impossible to test all individuals.

Thus, to avoid misleading and inaccurate time line we must perform population sampling correctly.

Therefore, the related terms can be briefly explained as:

- *Sampling:* The process of choosing a group that represents the population being studied.
- *Target Population:* The complete group of individuals from which the sample is drawn.
- *Sample:* It refers to the group of individuals (called "participants') who take part in a study.
- *Generalisability:* The extent to which the results of the research can be applied on the target population.

6.3.1 The Purpose of Sampling

The focus of psychological research is to learn about a large group of people with something in common. This group that interests us for the study is our 'target population'. For some research this target population can be as big as all humans, but for others it can just be a smaller group like preteens, teenagers or children or people who are alcoholic, etc.

It is highly unlikely and almost impossible to study each individual in a target population, so psychologists choose a representative subset or subgroup of the target population.

Sampling is very important in order to generalize the result from the sample to the entire population. The higher the representation in the sample the greater is the confidence of the researcher to generalise the result to the target population.

6.3.2 Types of Sampling

6.3.2.1 Non-Probability Sampling

It is the sampling technique where each member of the population does not have an equal chance to get selected. Because of this we can assume that the sample does not completely represent the target population. There is also a possibility that the individuals were chosen deliberately by the researcher to take part in the study.

This method of population sampling is generally applied for qualitative research, case studies, pilot studies and for developing hypothesis.

Non probability sampling method is usually used in studies that do not have an interest in the parameters of the total population. This technique is preferred by some researchers because it is easy, quick and less costly.

6.3.2.2 Probability Sampling

As the name suggests, in this sampling technique each individual from the target population will have an equal probability or equal chance of getting selected for research. This method provides a guarantee that the process is without bias and completely randomized.

The simplest example of this would be to list the names of all individuals in the target population on separate chits of paper and then to draw the papers one by one from the total, till the sample size is met.

Due to this randomized selection, the results from the statistical methods are highly accurate. Also, by eliminating the sample bias and by being representative this method is also used to estimate the population parameters.

Details of the types of sampling will be discussed in the next unit.

6.3.3 Sample Size

A sample size used in a statistical analysis is the count of the total individual samples or the number of observation that are used in a survey or experiment. The sample size is always a positive integer and is denoted by the variable 'n'. There is no exact sample size and it can vary as per the research settings. Though if all conditions remain the same, larger the sample size, higher is the precision in estimating the properties of the target population.

6.3.3.1 Determining the Sample Size

To conduct any research, it is very important to ascertain the sample size. For example, if a researcher is conducting a survey to gather information about the existence of ear problems in all school going children, s/he must answer the important question "How many participants should I choose for the survey"? However, s/he cannot arrive to any conclusion without considering the circumstances and objectives of his/her research. To determine a sample size, one must use both non-statistical and statistical variables. Some of non-statistical variables can be manpower, budget, availability of resources, sampling frame and ethics. For statistical considerations we can include the precision that is desired from the estimate of existance and the expected existence or prevalence of ear problems in school children.

6.3.3.2 Criteria to Infer the Appropriate Sample Size

We need to consider certain factors while selecting an appropriate sample size for own research, few of them are as follows:

- **Level of Precision**

The level of precision or sampling error is the estimated range in which the true value of the population lies and it is denoted in percentage points. Hence if a researcher reaches a result that 25% of school going children have eye problems, with the precision rate being 3%, then s/he can infer that between 22-28% of school going children have ear problems.

- **Confidence Level**

Confidence interval (also called risk level) is the statistical measure that signifies results from an action that are supposed to fall inside a pre-mentioned range. That is, if you take a confidence level of 90% it would signify that results from an action will meet expectation 90% of the time. This basic idea is explained in the Central Limit Theorem as "when a population is repeatedly sampled, the average value of an attribute obtained is equal to the true population value. In other words, if a confidence interval is 95%, it means 95 out of 100 samples will have the true population value within range of precision".

- **Degree of Variability**

The degree of variability can change a lot on the basis of target population and the attributes that are being considered. For a heterogeneous population, a large sample size is needed in order to reach to a high level of precision. One must understand that a percentage of 52% would mean a higher variability than 5% or 90% as in both the cases, a major chunk of majority either possesses or does not possess the attribute considered in the survey. We can follow various approaches to get a sample size which include using published tables, using a census where the population size is small, and using formulae to determine a suitable sample size.

Self Assessment Questions I

Fill in the following blanks

1) The level of precision or ..... error is the estimated range in which the true value of the population lies and it is denoted in percentage points.

2) ..... sampling method is usually used in studies that do not have an interest in the parameters of the total population.

3) A ..... used in a statistical analysis is the count of the total individual samples or the number of observations that are used in a survey or experiment.

4) ..... (also called risk level) is the statistical measure that signifies results from an action that are supposed to fall inside a pre-mentioned range.

6.4 SAMPLING ERROR

Sampling error refers to the deviation of the chosen sample from the original traits, characteristics, qualities, behaviors or figures of the total population.

6.4.1 Why Does This Error Occur?

This error occurs in the case where the researcher takes participants with individual differences for the survey, even though they are from the same population. Sample is just a subset of the total population and hence there might be a difference between the sample and the population whereby the most common cause for this is sampling bias. It should be the goal of every researcher to get a sample that is free from bias and represents the whole population in order to minimize or eliminate sampling error.

6.4.2 Way to Eliminate Sampling Error

The only way to eliminate sampling error is to forego the sample and to test the entire population which is practically not possible in most cases, though through unbiased probability sampling and taking a large sample size, this error can be minimized.

6.5 SAMPLING CRITERIA

It refers to the major characteristics of the subjects that need to be ascertained beforehand to determine their eligibility to be a part of the sample. A researcher determines these on the basis of his/her research problem or the purpose of research. In simple terms, we need to first ask what are we studying, and then conclude the characteristics of the elements or subjects required by us to be able to look at the problem. For example, these characteristics may include (considering human participants only):

- Age ;
- Marital status;
- Gender (male/female);
- Education Level
- Ethnicity;
- Whether suffering from a disease
- type of treatment;
- Languages known; etc.

6.5.1 Sampling Bias

Sampling bias occurs when the details of a random variable which are assimilated to estimate its distribution are wrongly selected and are not a reflection of the true distribution due to nonrandom reasons. Let us understand this through an example: If we want to predict the results of an election through an opinion poll, we can go ahead and ask a 1000 voters about their choice and asking them can give a pretty clear picture of the outcome but only if these 1000 voters are selected in an unbiased way then this sample is representative of the population. If the poll just surveyed a particular minority or income class or working class then the prediction might go wrong as this sample is not representing the population.

In case of an unbiased sample, the difference between the individual selected to form a sample from a population and the total population they represent, should be accidental i.e. occur by chance. If the reason for difference is anything other than chance then there exists a bias in the sample (sampling bias). Sampling bias generally comes across because certain variables in a sample are systematically under-represented or over-represented contrasting their true distribution (e.g. our opinion poll). And this bias leads to a systematic distortion of the estimate of the sampled probability distribution. We cannot remove the said error by just increasing the sample size but we need to follow correct sampling technique. In other words, adding 1000 voters will not increase the accuracy of our opinion poll but adding a 1000 voters, who are randomly chosen, would.

6.5.2 Causes of Sampling Bias

The most common reasons for a sampling bias lie in the process of data collection or in the way the study is designed, and either of them may support or reject collection of data from particular individuals or classes or in particular conditions.

It also comes whenever the researcher opts for a sampling strategy that is based on pure convenience or judgment, in which the criterion that is used for selection of sample is associated with the concerned variables. To understand let us go back to our opinion poll, a researcher collecting the data, out of convenience, may choose students living in a nearby hostel for collecting data.

6.5.3 Correcting and Reducing the Sampling Bias

In order to minimize sampling bias the two most critical considerations to be taken into account while designing an experiment are:

i) Convenience sampling or sampling as per own judgment should be avoided

ii) Ensuring that the target population is defined appropriately and that the sample frame is the best possible match.

If sampling the entire population in unviable due to limited resources or efficiency, a researcher must ensure that the population that has been included is not different from the total population when statistically measured. The population representative surveys conducted in social sciences are generally not simple random samples, but they follow sample designs that are much more complex (Cochran 1977). For example, for a normal study of households, the houses are chosen in multiple stages. First might be selection on the basis of the part of the city, second would be selection of the locality in that part of the city and a third stage might involve selecting a certain number of house within that locality. While using these complex sample designs it is important that we make sure that information of the sample frame is properly utilized and that the random selection and probability selection are implemented and documented at each stage of the

sampling process. In complex sample designs the sampling error will always be larger than in the simple random samples (Cochran 1977).

6.5.4 Sampling Bias, Limited Sampling Bias, Bias of Probability Function and Sampling Error

The term "sampling bias" sounds similar to "sampling error", "bias of a probability functional" and "limited sampling bias", but all of them, though related, are distinct and should not be confused. The "sampling error of a functional of the probability distribution" can be understood as the variance of the 'estimate of the probability functional' calculated over the 'sampled distribution' and the true value of the functional calculated over the correct distribution. The bias of a functional of a probability distribution can be defined as the expected value of the sampling error. Sampling bias can cause a bias of a probability functional; although, there is a difference between the concepts.

When one measures a nonlinear function of the probabilities from a small number of samples and because these samples might be picked at random from the total population and thus no sampling bias exists but it can still lead to a bias. This is known as a "limited sampling bias".

6.6 EXPERIMENTS AND CONTROL GROUP

There are various kinds of experiments involved in statistical studies, namely: blind and double -blind tests, control groups, treatment groups and placebos. A study that enforces a treatment (or control), in a controlled environment (e.g. giving them a certain drug or making them stay awake) while taking their responses, to its subjects (participants) is termed as an experiment..

The main reason to conduct an experiment is to ascertain the relationship between two events or factors (like fatty food and heart disease or smoking and cardiovascular function).

Experimental or Treatment-group versus control-group tests

The group that undergoes the treatment procedure or the experiment or a test sample is called the experimental or treatment group. The independent variable in the experiment is tested on this group. The values of the independent variable and its effect on the dependent variable are noted. The experiment can consist of several treatment groups at the same time.

The "treatment" refers to the changes made to the variable under study while the participants under study (human, animals, etc.) make up the "group". For example, a group consisting of humans can be given a new drug or a new counseling technique whereas for plants an experiment can be conducted on the use of organic fertilizer or vertical farming, etc. A "control group" is kept away from the experiment and is not exposed to the variable.

A control group is a group that is kept away from the experiment (i.e. experimental conditions) in such a way that the independent variable used will not be able to effect the results. This is done to isolate the effects of independent variable in order to help negate any alternative explanation of results.

Even though there is an experimental group for all experiments, not all need a control group. Controls prove to be extremely beneficial when the conditions for the experiment tend to be complex and isolation is tough. Such experiments that use control groups are termed as controlled experiments.

The sole difference between the two groups (control and experimental) is the hypothesis being tested. For e.g. in case of testing the effects of a drug on humans the demography of individuals should be the same otherwise the results can vary due to age, gender, better health, economic background, etc.

An example of Controlled Experiment:

To explain a controlled experiment, let us discuss about an experiment it to evaluate whether plants need water for survival or not. We can keep the plants that are not watered in our control group and the experimental group would be plants that are watered. A curious researcher would understand the need of keeping many experiment groups (every one of them being given different amount of water) in order to gauge the effect of amount of water on plants (plants might die in excess water, etc.).

6.6.1 Control Groups and Placebos

The easiest way to have a controlled group is to keep it in normal conditions in order to keep it away from the changing variable. To illustrate this let us consider an experiment to determine how a therapy influencess the modification in behaviour among children. For this we can simply divide the children into an experimental group (the group that will be exposed to therapy) and a control group (the group that will not receive the therapy).

The complexity of the experiment increases with the involvement of human subjects. In order to test the efficacy of a drug, the subjects in the controlled group might expect that they will not be affected (since they are not exposed) and this might skew the result. To prevent this we administer them a placebo (it looks similar to the drug but does not contain the chemicals of the drug, e.g. can be plain salt). This would ensure that the participants of the control group are unaware of whether they are being given the drug or not and hence making them same expectations as the experimental group.

Therefore we must also take into consideration the placebo effect. This is the opposite of earlier example, here the individual in the controlled group feels that there is an improvement or effect of the drug even though she/he has been given any experimental treatment. Also there is a concern that it is difficult to make a placebo that is free from any active ingredient; a salt pill might also have an effect on the outcome of the study.

These Placebo-controlled studies are done to test a medical therapy where apart from giving the subject the targeted treatment a separate group gets a placebo that is intended to have zero effect. In most of these trials the subjects are unaware that they have received a placebo and all of them believe that they are getting the real drug or treatment. Most often, the researcher also maintains a natural history group that is kept away from any sort of treatment or exposure to the drug.

These are control groups that can be explained as:

- **Positive Control Groups:** It refers to the control groups in which the conditions produce a positive outcome. These control groups are important to show that the experiment is being performed as per the plan.
- **Negative Control Groups:** They can be defined as those control groups in which the conditions generate a negative result. These are helpful in identifying the external influences that might be existing but were not accounted for, like contaminants.

Key Points: Control vs. Experimental Group

During an experiment the control group and an experimental group are contrasted against each other. The difference amongst the groups is that the independent variable remains constant in the control group whereas it is changed in the experimental group.

Multiple experimental group can be a part of a single experiment and they can all be contrasted against the control group.

A control group is formed in order to rule out outside influences or factors that may be affecting the experiment and such experiments are termed as "controlled experiments".

Sometimes a placebo is added to the experiment. It is not a replacement of the control group as individuals receiving a placebo might be affected by the information of being tested.

Self Assessment Questions II

State whether the following are 'True' or 'False':

1) A control group is included within the experimental conditions in such a way that the independent variable used will be able to affect the results. ..... .

2) The "sampling error of a functional of the probability distribution" can be understood as the variance of the 'estimate of the probability functional' calculated over the 'sampled distribution' and the true value of the functional calculated over the correct distribution. ..... .

3) In case of an unbiased sample, the difference between the individual selected to form a sample from a population and the total population they represent, should be accidental i.e. occur by chance ..... .

4) Population refers to the deviation of the chosen sample from the original traits characteristics, qualities, behaviors or figures of the total population. ..... .

6.7 LET US SUM UP

The above unit started with an introduction to the concept of population and sample. The population was described as a set of all possible conditions of a random variable. Its size can be either finite or infinite. Further, Sample was defined as a finite subset of the population. The above unit also discussed about the concept of sampling and sampling bias. The unit discussed about the ways to reduce the bias. At the end, you were introduced about the concepts of experimental group, control group and placebo effect.

6.8 ANSWERS TO SELF ASSESSMENT QUESTIONS

Self Assessment Questions I

1) Sampling

2) Non probability

3) sample size

4) Confidence interval

Self Assessment Questions II

1) False

2) True

3) True

4) False

6.9 UNIT END QUESTIONS

1) Define Population. Discuss the relationship between population and sampling.

2) Explain the types of sampling.

3) Describe sampling bias.

4) Discuss the ways of reducing the sampling bias.

5) Enumerate controlled group and placebos.

6.10 GLOSSARY

Sampling : The process of choosing a group that represents the population being studied.

Target Population : The complete group of individuals from which the sample is drawn.

Sample : It refers to the group of individuals (called "participants') who take part in a study.

Generalisability : The extent to which the results of the research can be applied on the target population is termed as generalizability.

Level of Precision : The level of precision or sampling error is the estimated range in which the true value of the population lies and it is denoted in percentage points.

Confidence Level : Also called risk level, is the statistical measure that signifies results from an action that are supposed to fall inside a pre-mentioned range.

Sampling Error : Sampling error refers to the deviation of the chosen sample from the original traits characteristics, qualities, behaviors or figures of the total population.

Sampling Criteria : It refers to the major characteristics of the subjects that need to be ascertained beforehand to determine their eligibility to be a part of the sample.

Experimental Group : The group that undergoes the treatment procedure or the experiment or a test sample .

Control group : A group that is kept away from the experiment (i.e. experimental conditions) in such a way that the independent variable used will not be able to affect the results.

6.11 SUGGESTED READINGS AND REFERENCES

Bailey, R. A. (2008). Design of Comparative Experiments. Cambridge University Press.

Bailey, R. A. (2008). Design of Comparative Experiments. Cambridge University Press. ISBN 978-0-521-68357-9.

Bradburn, N.M., & Sudman, S. (1988). Polls and surveys: Understanding what they tell us. San Francisco: Jossey-Bass.

Chaplin, S. (2006). "The placebo response: an important part of treatment". Prescriber: 16–22.

Chaplin, S. (2006). "The placebo response: an important part of treatment". Prescriber: 16–22.

Cochran W.G. (1977), "Sampling Techniques", Wiley

Fowler, F. J. Jr. (1984). Survey research methods. Thousand Oaks, CA: Sage.

Groves R. and Cooper M. (1998) "Nonresponse in Household Interview Surveys", New York: John Wiley and Sons.

Hinkelmann, Klaus; Kempthorne, Oscar (2008). Design and Analysis of Experiments, Volume I: Introduction to Experimental Design (2nd ed.). Wiley.

Hinkelmann, Klaus; Kempthorne, Oscar (2008). Design and Analysis of Experiments, Volume I: Introduction to Experimental Design (2nd ed.). Wiley. ISBN 978-0-471-72756-9.

Kalton, G. (1983). Introduction to survey sampling. Beverly Hills: Sage Publications.

Kish, L. (1965). Survey sampling. New York: Wiley.

Panzeri S., Senatore R., Montemurro M.A., Petersen R.S. (2007), "Correcting for the Sampling Bias Problem in Spike Train Information Measures", J Neurophysiol, 98: 1064– 1072.

Schuman, H., & Kalton, G. (1985). Survey methods. In G. Lindzey & E. Aronson (Eds.),. Handbook of Social Psychology, Volume 1 (pp.635-697). New York: Random House.

Smith, T. W. (1983). The hidden 25 percent: An analysis of nonresponse on the 1980 General Social Survey. Public Opinion Quarterly, 47, 386-404.

Sudman, S. (1976). Applied sampling. New York: Academic Press.

UNIT 7 SAMPLING TECHNIQUES*

Structure

7.0 OBJECTIVES

With the help of this unit, you will be able to:

- Explain the meaning and characteristics of sampling techniques;
- Identify the qualities of an ideal sample;
- Describe the uses of sampling techniques; and
- Discuss the different methods or techniques of sampling.

7.1 INTRODUCTION

In this unit, you will basically be introduced about the meaning and characteristics of sampling techniques. Sampling technique should be based on the nature of research. So, after going through this unit, you will be in a position to understand

* Dr. Sarika Boora, Psychologist, Sambharti Foundations, New Delhi.

the uses of appropriate sampling technique with reference to research. The present unit will also deal with different techniques of sampling methods.

7.2 SAMPLING AND SAMPLING TECHNIQUES

Sampling is very useful in research. It helps in determining the accuracy of the research or survey. If there is a mistake in the sample then it has a direct implication on the result. We can use various techniques to help us gather a sample based on the situation and need of the researcher.

For the purpose of quantified research, sampling techniques are used extensively and their importance cannot be downplayed in any field of research. It is of great importance for our day to day activities. For example, when we buy fruit or vegetables we do not look at each individual piece, we just pick and examine a few items and form an idea about the whole lot. As said by famous researcher (Snedecor) "A cart load of coal is accepted or rejected on the evidence gained from testing only a few pounds. The physicians make inference about a patient's blood through examination of a single drop. Samples are devices for learning about large masses by observing a few individuals". Sampling is also widely used in the field of education and the census being hardly used like the case of population count.

For investigating a sample, we select a small number of units from the entire domain and think of them as being representative. Then we study them in depth and reach to the conclusions that are generalized for the entire domain. A sample investigation does not study all units like a census investigation, it only chooses a few unit on certain predefined basis. To illustrate this let us suppose a researcher want to find out the percentage of expense on food by university students. Now, it is not feasible for him/her to survey the entire thousands of students, rather she/he may choose a representative sample of 500 students and collect the figure from them. This figure can then be generalized for the entire university. If the data is collected diligently and the selection of students is representative then the reliability of the entire set will be very high.

7.2.1 Characteristics of Sampling Technique

The significance and value of the sampling techniques are brought forward by the following characteristics

1) Cost & Time: It is highly cost effective and efficient thereby taking less time than census technique. (Technique of Measuring the variables of whole population).

2) Reliability: If we choose the sample units carefully and the matter surveyed is not diverse in character then the results derived from the sample should have the same reliability as the census survey.

3) In-depth Study: Due to the fact that the number of units in a sample are relatively small, they can be studied in depth and observed from various angles and viewpoints.

4) Scientific Approach: The sampling technique follows and scientific approach and the results garnered from a particular set of units can be verified and also applied to other set of units. The deviation from the norm can be determined by taking random samples.

5) Suitability: Maximum number of researches and surveys use a sampling technique because whenever the matter is not diverse in nature; the study of a small number of unit is enough.

This is applicable to most situations.

7.2.2 Errors of Sampling Technique

Few of the errors of sampling technique can be discussed as follows:

- **Accuracy:** While comparing it to a census, the observations from a sample can have more faults, hence it can be less accurate in comparison to a census technique.
- **Changeability of Units:** In the field of survey the units can change and if these are not in harmony then the sampling technique can prove to be wrong. It will no longer be scientific to extend the results from one set to other.
- **False conclusions:** If we do not take enough care while selecting samples the results from the survey can be wrong if extended to all the units. If we go back to our earlier example of food expense and choose only well off students then the result for the food expense will be misleading if extended to the entire university.
- **Requirement of Specific Knowledge or Expertise:** The success of sample technique depends on the skills of the researcher; any researcher with mediocre skill level can jeopordise the entire selection process.
- **Impossible Sampling:** There are certain circumstances that make it difficult to apply the sample technique. If we need 100% accuracy or the time is too less to make a selection then this technique cannot be applied. It might not be used when the material is diverse in nature.

7.2.3 When and Where to Use Sampling Technique

Despite of the shortcomings or errors that can occur in sampling techniques, yet they are used significantly. Below mentioned are few of the points to keep in mind before applying this technique:

- **Large or Infinite Data Size:** If the total size of the data is very large then sampling technique is used as it reduces the cost and effort and time involved in the research.
- **If full accuracy is not required:** If the researcher does not require absolute accuracy then he/she can go for a sampling technique.
- **Where Census is not possible:** Suppose one wants to know the salt content in oceans, she cannot measure it in all the water of the oceans, she/he will have to take a sample.
- **Homogeneity:** If there is no difference between the units of a domain then sampling technique is most suitable to use.

7.3 QUALITIES OF AN IDEAL SAMPLE

An ideal sample contains the following essentials:

- **Representation:** A perfect sample is where the whole data can be represented adequately. It must be ensured that units with same characteristics are selected as represented in the data.

- **Independent:** The other aspect of a sample is it being independent which means the units should be interchangeable. There should be a freedom for every unit has a chance to be included in the sample.
- **Adequate:** The size of the sample should be adequate in order to infer any results and apply the conclusions to the entire data.
- **Homogenous:** The sampling units must have the same characteristics as other units or else the sample will be rendered unscientific.

Self Assessment Questions I

Answer the following questions in one or two lines:

1) Explain any one error of sampling.

.....

.....

2) What is the use of sampling?

.....

.....

3) Give any one characteristics of sampling

.....

.....

4) Describe quality of an ideal sample.

.....

.....

7.4 METHODS/TECHNIQUES USED IN SAMPLING

There can be many methods to derive a sample from a given data. The problem, its scope and nature will determine the method used by the researcher but there must be a careful consideration given to the choice.

As mentioned in the previous unit that the sampling technique can be broadly categorized into two:

- Probability Sampling
- Non- Probability Sampling

The fact that differentiates both is whether the sample is randomly selected or not. In case of a random sample each element or unit has an equal chance to be chosen for the study.

7.4.1 Probability Sampling

As mentioned, it is a sampling technique in which each element of the population has an equal probability of selection and this is because of randomization and hence it is also known as random sampling.

It has the following types:

7.4.1.1 Simple Random Sampling

Here each unit or element or individual from the population gets an equal opportunity to get selected. It is applied when the researcher does not have knowledge about the target population beforehand.

To illustrate: If we were to choose 25 students from a class of 60 then each student has an equal probability of 1/60 of getting selected.

From all the ways of choosing the samples, random sampling technique is used the most; and is widely considered as the best for sample selection. This technique gives every unit an equal probability of getting selected and this selection is free from any kind of personal bias or preference of the researcher. No unit is chosen on the basis of personal likes or whims of the researcher and the entire process is based on chance giving equal opportunity to each unit. Random sampling is done through the following methods:

i) **Lottery Method:** It is the easiest way of choosing the sample. Each unit is assigned a particular number and these numbers are then written on a piece of paper and put in a box. Then a neutral person, who is blindfolded, is made to pullout the required number of units for the sample from the box. Here the sample is being chosen by simple chance and there is no favor or partiality involved. It is also important to ensure that the sheets of paper that are used should be of equal size and quality.

ii) **Using the Rotating Drum:** This process is similar to the lottery method but with a slight modification. Here the units are itemized into lists and divided into categories from say 0 to 5. Then the same categories 0-5 are printed on pieces of wood or tin etc.(of same size) and placed in a drum. This drum is then rotated and the required number of the pieces is drawn. Now if we draw 5 zeroes 10 fives and 20 twos then we pick 5 units from the zero list 10 units from five list and 20 units from the twos list respectively.

iii) **Selection based on a Sequential List:** This process involves maintaining the units in alphabetical, numerical or geographical sequence. In this procedure units are broken up into numerical, alphabetical or geographical Sequence. So for a numerical selection one can choose units that fall in multiples of 3, for alphabetical selection we can choose all the names that begin with vowels, etc.

Precautions to be considered while choosing sample:

While conducting a random sample, the researcher must consider that the sample represents the entire population and the number of units selected are sufficient. Following should be considered while choosing a random sample:

- The researcher should be aware of the entire variation of population from which she/he wants to select his/her then sample. She/he should be aware of the main features of the field and its scope with the number of units in the population, so as to make a fair selection.
- The various items of a population must be similar and have common characteristics (homogeneous). If they differ too much then the sample will not be representative.
- The units of the field should not be dependent or linked otherwise it will not be possible for a random selection to take place.

7.4.1.2 Stratified Sampling

When the population is divided into subgroups called strata which refers to a way in which the units within the group share common characteristics with each other but are heterogeneous with other subgroups, after which units are randomly chosen from each stratum, it is called stratified sampling. The researcher needs to have knowledge of the population beforehand so that she/he can form subgroups.

This technique uses both deliberate and random sampling. Firstly, we divide the entire population into subsets based on their homogeneous character, then by using random sampling we choose elements from these subsets. Thus, this is a mixed sampling technique. This technique can be applied when the population can be divided into subgroups based on common characteristics. This stratification is made in accordance to the special qualities of a group and from these strata, units are randomly selected. For example, if a researcher wants to collect the data of the distribution of expenses of all residents in a city, she/he can divide the population on the basis of their profession and then choose some individuals or units randomly from these subgroups.

Process of Stratifying:

The stratification of the total population or division of data must be done with utmost care as the success of this technique is based on successful stratification. Below mentioned points should be considered for the same:

i) The researcher should have in-depth information about all units in the data and should be able to identify the common characteristics to bunch them into subgroups that are different from each other.

ii) Each stratum should be large enough to conduct random sampling.

iii) While dividing them into subgroups it must be considered that each stratum is related to the domain in a similar way though being homogeneous themselves.

iv) The number of units chosen from each subgroup for the survey, through random sampling, must be in the same proportion as the subgroup is to the entire population. For e.g if the total population is 10,000 out of which the businessmen are 30% then the number of units chosen from the subgroup of businessmen should be 30% of all the units picked by random sampling.

If the researcher observes the above mentioned precautions then she/he can achieve goals from this method as it has the qualities of both deliberate and random sampling.

7.4.1.3 Cluster Sampling

When the total population is bifurcated into clusters or sections and then sections are chosen randomly, it is called cluster sampling. All the units in a given cluster are surveyed. These clusters can be determined on the basis of age, gender, geographic presence, etc. Cluster sampling is performed using following methods:

- Single Stage Cluster Sampling: We choose the whole section for sampling.
- Two Stage Cluster Sampling: Firstly the section is randomly chosen and then the elements are selected randomly for sampling.

7.4.1.4 Systematic Sampling

In this type of sampling, we choose the elements systematically and not randomly with the exception of the first element. There is a regular interval in the population at which these elements are chosen. All these elements are sequenced first wherein each of them gets an equal probability of being chosen.

For sample size denoted as ‘n’, we fragment our population (with size N) into subsets of k elements. Then we choose our first element on random basis from the first subset of k elements. To choose the other elements of the sample, we do it as follows:

We know number of elements in each group is k i.e N/n

So let us keep the first element as n1

Second element would be n1+k i.e n2;

Third element n2+k i.e n3 and so on.

For example of N=100, n=10

No of elements in each of the subgroups is N/n i.e 100/10 =10= k

Now, randomly choose the first element from the first subgroup. If we select n1= 2

n2 = n1+k = 2+10 = 12

n3 = n2+k = 12+10 = 22

7.4.1.5 Multi-Stage Sampling

It combines two or more methods that have been explained above. In this type of sampling, the total population is split up into various clusters and these are further split and put together in multiple subgroups or strata based on common characteristics. Then we can randomly choose a single cluster or multiple clusters from each stratum and this process will go on till the cluster cannot be fragmented any further. To illustrate, we can split the population of a country on the basis of their states, districts, urban vs rural and the homogeneous areas can be joined to form a strata.

The researchers do not favour this sampling technique to a great extent. The elements are selected at random at various stages. For example, if one needs to find the yield per hectare of land in Maharashtra, she/he will first do a random selection of certain number of districts, say 5, and then she/he will randomly choose a certain number of villages in a district, say 10. After that she/he will randomly pick a certain number acres of fields in each village, say 5. Hence she/he will have to examine 250 acres of land, spread over 10 villages in 5 districts spread over entire Maharashtra. This number can vary as per the suggestions given by experts. This can also be explained by the below mentioned table:

Entire Data:	Maharashtra	
	Districts	Stage 1
	Villages	Stage 2
	Acres of land	Stage 3

7.4.2 Non- Probability Sampling

It is the sampling technique that does not depend on randomization. It banks upon the ability of the researcher to choose the elements of a sample. The result from this technique can have a bias thereby making it difficult for all elements or units of population to have an equal representation. It is also known as non-random sampling. Below mentioned are its types:

7.4.2.1 Purposive Sampling

It is also called deliberate sampling. This method allows the investigator to choose the sample as per his/her whims with complete freedom. Some items are chosen from the entire data and then studied. The investigator is absolutely free to choose any item that she/he feels or as per his/her judgment will be able to represent the complete data or total population.

This simple technique of sampling is used by the investigator when the data is not diverse and she/he has knowledge of various facets of the problem. This method does not point to a random choice even though it is deliberate; on the contrary it signifies that only the elements that represent the entire population will be selected. It is determined by the purpose or intent of the study and selection is done on the suitability of elements for the purpose of study.

Criticism of Purposive Sampling

The major criticism of this technique is the fact that the units are selected by the investigator with his/her preconceived opinion and these preferences might prevent him/her from being objective. In this case, in theory, there is no way to prevent the selection of sample from being unfair. For example, if a researcher wants to study the bifurcation of total expenses of boys staying in a hostel and she/he has a preconceived notion that most of the money is spent by them in going to expensive restaurants to dine out (there may be only a small percentage of students who do that), she/he may choose only those students that she/he feels dine out at expensive restaurants. This may lead to the result being biased and wrong in terms of depicting the expenses made by hostel students as the result from this sample will always show that students spent a major part of their money on dining out.

Hence, this technique is considered majorly flawed by most of the statisticians and not many recommend it. As per Parten, "Statisticians as a class, have nothing to say in favour of purposive selection."

This technique can still be useful, when done carefully by an expert statistician else it is not appropriate or suitable because there are no inbuilt checks.

7.4.2.2 Quota Sampling

This sampling method is seldom used. In this method the data is divided equally amongst the investigators and each of them is required to pick a certain number of items from his/her subgroup to form the sample. This method is dependent on the honesty and proficiency of the investigator for its success. Since the proficiency of the investigators is subjective and varied, the results of the study tend to be erraneous, thus this method is seldom used.

There is a need of a predefined standard to conduct this sampling technique. The sample thus selected represents the population and the proportion of traits should be similar to that of the population.

7.4.2.3 Convenience Sampling

In this sampling technique the elements or items of the study are selected as per the convenience of the investigator. There is no endeavor to collect information as per a plan or process. This is a technique which is frequently used by tourists while travelling. They interact with a few people whom they meet, transact with them and based on this they generate a perception and generalize this to the entire population of the country or state. This approach has zero scientific credibility and is termed as a "hit or miss method". This is why some of the travelogues written by such travellers are extreme.

Convenience selection chooses its samples on the basis of availability and used when this availability turns out to be costly and rare.

7.4.2.4 Referral /Snowball Sampling

This method of sampling is used by the researcher when he/she is completely unaware about the population, she/he then picks one element or individual and asks him/her for reference or recommendation of other individuals whose characteristics she/he may describe to suit the need of the sample. This is called as a snowball technique because the size of the sample constantly grows as new referrals keep pouring in.

7.4.3 Advantages and Disadvantages of Probability Sampling

Advantages

- Is easier to undertake;
- Lesser degree of preconceived notions and judgement involved;
- The research finding are highly reliable;
- Sampling error can be estimated accurately;
- Is easily performed even on non-technical persons;
- There is no sampling and systematic bias.

Disadvantages

- Monotonous;
- Chances that only a specific class may be chosen;
- Degree of complexity is higher;
- Is generally more time consuming and costly.

7.4.4 Advantages and Disadvantages of Non-Probability Sampling

Advantages

- In case an individual is not willing to participate, s/he can be easily replaced by someone else;
- Less expensive;
- Highly time effective;
- Sampling methods are easy to execute.

Disadvantages

- Only accessible individuals are chosen and this reduces the chance of assimilating important data;
- No way to judge the accurate representation of population;
- Too much dependence on researchers' judgement;
- No way to estimate margin of error;
- There can be a bias while selecting units;
- Does not focus on effectiveness, instead thrives on simplicity.

Self Assessment Questions II

Fill in the following blanks:

1) ..... method of sampling is used by the researcher when he/she is completely unaware about the population.

2) ..... sampling technique does not depend on randomization.

3) In ..... method, each unit is assigned a particular number and these numbers are then written on a piece of paper and put in a box.

4) In ..... sampling, we choose the elements systematically and not randomly with the exception of the first element.

7.5 LET US SUM UP

It can be summed up from the above discussion that, sampling technique is very useful in research. It helps in determining the accuracy of the research or survey. If there is a mistake in the sample then it has a direct implication on the result. We can use various techniques to help us gather a sample based on the situation and need of the researcher. Later part of the unit discussed about the characteristics and meaning of sampling techniques. The unit also discussed about the different techniques of sampling and their advantages as well as disadvantages.

7.6 UNIT END QUESTIONS

1) Describe the relevance of sampling techniques in research.

2) Describe types of sampling techniques.

3) Discuss the advantages and disadvantages of sampling techniques.

7.7 ANSWERS TO SELF ASSESSMENT QUESTIONS

Self Assessment Questions I

1) Changeability of Units: In the field of survey the units can change and if these are not in harmony then the sampling technique can prove to be wrong. It will no longer be scientific to extend the results from one set to other.

2) It helps in determining the accuracy of the research or survey.

3) Due to the fact that the number of units in a sample are relatively small, they can be studied in depth and observed from various angles and viewpoints.

4) Representation: A perfect sample is where the whole data is represented adequately. It must be ensured that units with same characteristics are selected as represented in the data.

Self Assessment Questions II

1) Snowball sampling

2) Non probability

3) Lottery method

4) Systematic sampling

7.8 GLOSSARY

Probability Sampling: It is a sampling technique in which each element of the population has an equal probability of selection and this is because of randomization and hence it is also known as random sampling.

Simple Random Sampling: Here each unit or element or individual from the population gets an equal opportunity to get selected.

Stratified Sampling: A strata formed in a way that the units within the group share common characteristics with each other but are heterogeneous with other subgroups, after which units are randomly chosen from each stratum.

Cluster Sampling: When the total population is bifurcated into clusters or sections and then sections are chosen randomly, it is called cluster sampling.

Systematic Sampling: In this type of sampling, we choose the elements systematically and not randomly with the exception of the first element.

Multi-Stage Sampling: It combines two or more methods of sampling.

Referral /Snowball Sampling: When he/she is completely unaware about the population, he then picks one element or individual and asks him for reference or recommendation of other individuals whose characteristics he may describe to suit the need of the sample.

Non- Probability Sampling: It is the sampling technique that does not depend on randomization. It banks upon the ability of the researcher choose the elements of a sample.

Purposive Sampling: It signifies that only the elements that represent the entire population will be selected.

Quota Sampling: In this method the data is divided equally amongst the investigators and each of them is required to pick a certain number of items from his subgroup to form the sample.

Convenience Sampling: In this sampling technique the elements or items of the study are selected as per the convenience of the investigator.

7.9 SUGGESTED READINGS AND REFERENCES

Andale. (2015). Probability Sampling: Definition,Types, Advantages and Disadvantages. Statistics How To. Retrieved from http://www.statisticshowto.com/probability-sampling/.

Babbie, E. R. (n.d.). The Logic of Sampling. The Basics of Social Research. Wadsworth Cengage Learning, 208.

Barreiro, P. L., & Albandoz, J. P. (2001). Population and sample. Sampling techniques. MaMaEuSch† (Management Mathematics for European School). Retrieved from http:// optimierung.mathematik.unikl.de/mamaeusch/ veroeffentlichungen/ ver_texte/ sampling_en.pdf.

Chaudhuri, A., & Stenger, H. (2005). Survey Sampling: Theory and Methods - 2nd ed.

Chapman & Hall/CRC.

Daniel, J. (2012). Sampling Essentials: Practical Guidelines for Making Sampling Choices. Sage Publications, 103.

Doherty, M. (1994) Probability versus Non-Probability Sampling in Sample Surveys, The New Zealand Statistics Review March, 21-28.

Fink, A. (2003) How to Sample in Surveys. 2nd Edition. Thousand Oaks: Sage.

Henry, G.T. (n.d.). Practical Sampling. Sage Publications, 23.

King, R. M. Types of sampling. Advanced Research Methods. Retreived from http:// www.psyking.net/HTMLobj-3829/Types_of_Sampling.pdf.

Lynchi, G. (n.d.). Sampling. Retreived from https://www.kent.ac.uk/ religionmethods/ documents/Sampling.pdf.

Wiid, J., & Diggines, C. (n.d.). Marketing Research. Juta, pp 200. 12. Zikmund, W., &Babin, B. (n.d.). Sampling Designs and Sampling Procedures. Exploring Marketing Research.

Thomson South –Western, 411.