UNIT 7 : SAMPLING

UNIT STRUCTURE

7.1 Learning Objectives
7.2 Introduction
7.3 Census and Sample
7.4 Advantages of Sampling
7.5 Parameters and Statistic
7.6 Sample Size
7.7 Sample Design and Sampling Methods
7.8 Sampling and Non-sampling Error
7.9 Let Us Sum Up
7.10 Further Reading
7.11 Answers to Check Your Progress
7.12 Possible Questions

7.1 LEARNING OBJECTIVES

After studying this unit you will be able to:

- discuss the ideas of sample and sampling methods and advantages of sampling
- distinguish various types of sample and appreciate their utility
- describe the various types of errors involved in sampling theory and their role
- know how appropriate sample size is obtained, and
- analyze how a representative sample can be drawn

7.2 INTRODUCTION

In the previous unit you studied about the various research tools and their importance in carrying out a research work.In this unit, you will be introduced to one of the vital components of research, i.e., sampling. Here, the concepts of census, sample, sample size, sampling methods would be explained along with the types of samples and sampling errors. And in the next unit, you would be learning about research data.

theory of sampling deals with how a safe conclusion can be made by minimising the potential errors involved in the whole process of sampling. It, thus, includes two basic issues a) how many sample units should be sufficient for making a safe conclusion i.e. the issue of sample size, and b) once size is known, how those units are to be selected from the population which would represent the population in the most efficient way i.e. the issue of representation. We will take up these two issues systematically. But before that, let us try to understand couple of elementary concepts useful for proceeding further.

7.3 CENSUS AND SAMPLE

As has been pointed out earlier, population is the larger set of study units from which samples are drawn. More specifically, in statistics population is the aggregate of all objects or units under the study. Note that a population can be of two types - finite and infinite. When the population has some known counts of units or objects it is called a finite population. For example students in a college, population of a country, number of accidents occurred in a day, etc. On the other hand, if population consists of units or objects with not known counts, it is termed as infinite population. For instance, number of stars in the sky, population of natural numbers etc. Also note that a population can well be hypothetical, such as number throws of coin say 100 times. In any case, if all units of the population are brought under the study then it is called census or complete enumeration. In census method there is no scope for error since conclusions are made on the basis of all the units under the study. Thus it is the most reliable method of making conclusions. However, the census method possesses some obvious limitations. In the first place census method is highly time consuming and labour intensive. Secondly, it is also not cost effective. Thirdly, the method cannot be applied in case the population happens to be infinite. Still, there are certain specific situations when census method is preferred. For example, when population is of finite type and when information from all units is necessary there is no alternative to census method. Such method becomes absolutely necessary for quality control assessments of industrial production so that no unit is passed onto market with manufacturing defects.

design. The sampling, therefore, is both an art and a science.

7.5 PARAMETERS AND STATISTIC

At this stage, we need to introduce two more concepts - parameters and statistic (not statistics). When we study population characteristics, attributes, features we think of certain statistical measures or constants of these characteristics, attributes and features such as averages, dispersions. The statistical constants of the study attributes in population are known as parameters. Since sample is drawn from the population, there will be similar statistical constants viz. mean, variance etc. for the characteristics, attributes of the sample. The statistical constants obtained in sample are known as statistic (not statistics). Technically, therefore, a population parameter is estimated on the basis of sample statistic, and this process is called sampling. Clearly, value of parameter is dependent on all population units, that of statistic vary according to changes in the sample values.

Now we will proceed to discuss the two fundamental issues of sampling theory - sample size and sample selection.

7.6 SAMPLE SIZE

It has been already pointed out that sampling theory is all about making safe inferences regarding the unknown population. Note that when we try to generalise we do not claim that the population result will be exactly same as the sample results. We actually mean that population results will be somewhat like the sample results. So as long we do not take into account all the units in the population, there is always going to be some difference between the sample and population results. It is important to understand that statistics can not eliminate this error, but can only honestly make it explicit. When we try to make generalisation, we would like to have this error as minimum as possible. To understand the underlying mechanics of sampling theory we need to play with the possible errors.

Let us try to play with our error. Suppose that we have found a person whose height is 5.8 feet. Based on this single information, can we generalise about the possible height of all persons in the population? Well, we can; but the error in this case will be too large. This is because many heights will appear to be probable in the population! Suppose we take a few more

higher variability within themselves we might have a problem in safe generalisation. Second, the sample size and SE are indirectly related that is increase in one would tend to decrease in the other. Therefore, if we want to reduce the error we should have a greater sample size. Third, this notwithstanding, note the square root over the sample size. This means that although increase in sample size would produce lesser error, because of the square root above it very large increase will have only marginal effect on the improvement in error. Getting the standard deviation before hand is not possible though. It is suggested, therefore, that one would conduct a pilot test i.e. pre-test to obtain a possible standard deviation. Once we have the value of we can take a predefined chance of error in generalisation and put these values to the above formula to get the optimal sample size. In practice, we can have plenty of sample size formula, but this hold the key.

In fact, all sample size formula is based on this fundamental relation amongst the sample size, standard error with some more assumptions. Consider the following example for explanations.

Example:

From a pilot survey it is found that sample standard deviation of the study attribute is 15. What should be the sample size for making conclusion with a standard error of 1.5? How much more units have to be sampled if one attempts to make the SE half of 1.5? What does this suggest? Further, how can one draw the inference on the basis of this sample?

Solution:

Fundamental relation amongst the SE, SD and sample size is given by

$$SE = \frac{\sigma}{\sqrt{n}}$$

Now for SE=1.5 and =15, the above formula will give n = 100.

From the sample formula for SE=0.75 (i.e. half of 1.5) and =15, the sample size n needs to be = 400.

This means that for reducing error by 0.75 one would need 300 more units in the sample. This means that in order to improve small accuracy, a large no of units are needed to be sampled. [Also note that for improving accuracy level by 0.1, one has to sample at least 115 units. Therefore, whether one takes 100 and 112 units in the sample, does not

The population is first divided into strata and then from each strata sample units are selected at random. In this type of sampling sample units can be selected proportionally or dis-proportionally. For instance, if male and female constitute 60 and 40 percent in the population same proportion should be reflected in the sample as well. When it is done it is called proportional, else the allocation is called disproportional. The sample can also be selected at various stages. For example in the first stage we sample the states, then districts and then villages. This type of sampling is called multi-stage sampling. When area based divisions are created and samples are taken for instance left and right side of the road, north and south side of rive etc we have cluster sampling. In cluster sampling ideally each cluster should contain same number of units. It may be noted that difference between the stratified and cluster sample lies in the fact that stratification is done on the basis of attributes so that within strata units are homogenous with respect to the stratification attribute, whereas cluster is merely a geographical division and within the divisions there is no homogeneity.

There is another type of sampling methods where a fixed rule is applied rather than any probability scheme. One such method is called systematic sampling. In systematic sampling first unit is selected at random and then every unit is selected at fixed interval determined by the ratio of number of population unit and expected sample size. For instance if we want to select a sample of size 10 from 500 units fixed interval will be 50; the first unit will be selected at random from 1 to 50 and then every 50th unit will be selected until 10 units are selected. One can, however, be imaginative and combine different methods to make the sample more representative of the population. Such samples are called mixed sample.

7.8 SAMPLING AND NON-SAMPLING ERROR

It is evident from above that a sample typically entails only a part of the population and as such, sample result is different from the true population value. The difference is termed as error. Unless and until all units of population is considered this error is going to remain in sample. Since the error is consequent on sample itself this error is called sampling error. Or in other words, the sampling error emerges due to the fact that sample does not consider all units of the population. Reasons behind sampling error can be

of sampling. We have discussed the fundamental relations as to how an appropriate sample size can be determined. We have also discussed various techniques of sampling methodology for drawing representative samples from population. We have highlighted benefits of sampling. We have also understood the difference between sampling and non-sampling errors.



7.10 FURTHER READING

- "Fundamental of Statistics"; S C Gupta, S Chand and Company, New Delhi
- "Fundamental of Statistics"; A M Gun, M K Gupta, B Dasgupta, Vol. I, World Press
- "Basic Statistics"; A L Nagar, Oxford University Press
- "Basic Statistics"; B L Agarwal, New Age International
- "Statistics without Tears", D Rowntree, Penguin Books



7.11 ANSWERS TO CHECK YOUR PROGRESS

Ans to Q No 1: In any case, if all units of the population are brought under the study then it is called census or complete enumeration. On the other hand, sample study relies on a finite subset of population and then tries to generalise the results for the entire population.

Ans to Q No 2: subjective sample

Ans to Q No 3: mixed

Ans to Q No 4: The statistical constants of the study attributes in population are known as parameters.

Ans to Q No 5: Standard error

Ans to Q No 6: A pilot test is a pre-test to obtain a possible standard deviation.

Ans to Q No 7: If the population can be divided into segments based on some characteristics and we expect that these characteristics have some systematic relations in the result under the study we usually