

Sampling Theory

MODULE I

LECTURE - 1

INTRODUCTION

DR. SHALABH
DEPARTMENT OF MATHEMATICS AND STATISTICS
INDIAN INSTITUTE OF TECHNOLOGY KANPUR

Statistics is the science of data.

Data are numerical values containing some information.

Statistical tools can be used on a data set to draw statistical inferences. These statistical inferences are in turn used for various purposes. For example, government uses such data for policy formulation for the welfare of the people, marketing companies use the data from consumer surveys to improve the company and to provide better services to the customer, etc. Such data is obtained through sample surveys. Sample surveys are conducted throughout the world by governmental as well as non-governmental agencies. For example, “National Sample Survey Organization (NSSO)” conducts surveys in India, “Statistics Canada” conducts surveys in Canada, agencies of United Nations like “World Health Organization (WHO), “Food and Agricultural Organization (FAO)” etc. conduct surveys in different countries.

Sampling theory provides the tools and techniques for data collection keeping in mind the objectives to be fulfilled and nature of population.

There are two ways of obtaining the information

1. Sample surveys

2. Complete enumeration or census

Sample surveys collect information on a fraction of total population whereas in census, the information is collected on the whole population. Some surveys, e.g. economic surveys, agricultural surveys etc. are conducted regularly. Some surveys are need based and are conducted when some need arises, e.g., consumer satisfaction surveys at a newly opened shopping mall to see the satisfaction level with the amenities provided in the mall .

Sampling unit:

An element or a group of elements on which observations can be taken is called a sampling unit. The objective of the survey helps in determining the definition of sampling unit.

For example, if the objective is to determine the total income of all the persons in the household, then the sampling unit is household. If the objective is to determine the income of any particular person in the household, then the sampling unit is the income of the particular person in the household. So the definition of sampling unit depends and varies as per the objective of the survey. Similarly, in another example, if the objective is to study the blood sugar level, then the sampling unit is the value of blood sugar level of a person. On the other hand, if the objective is to study the health conditions, then the sampling unit is the person on whom the readings on the blood sugar level, blood pressure and other factors will be obtained. These values will together classify the person as healthy or unhealthy.

Population:

Collection of all the sampling units in a given region at a particular point of time or a particular period is called population. For example, if the medical facilities in a hospital are to be surveyed through the patients, then the total number of patients registered in the hospital during the time period of survey will be the population. Similarly, if the production of wheat in a district is to be studied, then all the fields cultivating wheat in that district will constitute the population. The total number of sampling units in the population is the population size, denoted generally by N . The population size can be finite or infinite (N is large).

Census:

Complete count of population is called census. The observations on all the sampling units in the population are collected in a census. For example, in India, the census is conducted at every tenth year in which observations on all the persons staying in India is collected.

Sample:

One or more sampling units are selected from the population according to some specified procedure.

A sample consists only of a portion of the population units.

In the context of sample surveys, a collection of units like households, people, cities, countries etc. is called a **finite population**.

A census is a 100% sample and it is a complete count of the population.

Representative sample:

All salient features of population are present in the sample.

It goes without saying that every sample is considered as a representative sample.

For example, if a population has 30% males and 70% females, then we also expect the sample to have nearly 30% males and 70% females.

In another example, if we take out a handful of wheat from a 100 Kg. bag of wheat, we expect the same quality of wheat in hand as inside the bag. Similarly, it is expected that a drop of blood will give the same information as all the blood in the body.

Sampling frame:

List of all the units of the population to be surveyed constitutes the sampling frame. All the sampling units in the sampling frame have identification particulars. For example, all the students in a particular university listed along with their roll numbers constitute the sampling frame. Similarly, the list of households with the name of head of family or house address constitutes the sampling frame. In another example, the residents of a city area may be listed in more than one frame - as per automobile registration as well as the listing in the telephone directory.

Ways to ensure representativeness:

There are two possible ways to ensure that the selected sample is representative.

1. Random sample or probability sample:

The selection of units in the sample from a population is governed by the laws of chance or probability.

The probability of selection of a unit can be equal as well as unequal.

2. Non-random sample or purposive sample:

The selection of units in the sample from population is not governed by the probability laws.

For example, the units are selected on the basis of personal judgment of the surveyor. The persons volunteering to take some medical test or to drink a new type of coffee also constitute the sample on non-random laws.

Another type of sampling is **Quota Sampling**. The survey in this case is continued until a predetermined number of units with the characteristic under study are picked up.

For example, in order to conduct an experiment for rare type of disease, the survey is continued till the required number of patients with disease are collected.

Advantages of sampling over complete enumeration:

Reduced cost and enlarged scope

Sampling involves the collection of data on smaller number of units in comparison to complete enumeration, so the cost involved in the collection of information is reduced. Further, additional information can be obtained at little cost in comparison to conducting another survey. For example, when an interviewer is collecting information on health conditions, then he/she can also ask some questions on health practices. This will provide additional information on health practices and the cost involved will be much less than conducting an entirely new survey on health practices.

Organization of work

It is easier to manage the organization of collection of smaller number of units than all the units in a census. For example, in order to draw a representative sample from a state, it is easier to manage to draw small samples from every city than drawing the sample from the whole state at a time. This ultimately results in more accuracy in the statistical inferences because better organization provides better data and in turn, improved statistical inferences are obtained.

Greater accuracy

The persons involved in the collection of data are trained personals. They can collect the data more accurately if they have to collect smaller number of units than large number of unites in a given time.

Urgent information required

The data from a sample can be quickly summarized.

For example, the forecasting of the crop production can be done quickly on the basis of a sample of data than collecting first all the observations.

Feasibility

Conducting the experiment on smaller number of units, particularly when the units are destroyed, is more feasible.

For example, in determining the life of bulbs, it is more feasible to fuse minimum number of bulbs. Similarly, in any medical experiment, it is more feasible to use less number of animals.