CHAPTER

# 2 Sampling and Inferences

## 2.1 INTRODUCTION

The collection of selected number of individuals, objects or results from the parent universe according to the given rule is known as sample. We know that it is not possible to discuss all the members of the universe separately, because it is very costly and time consuming. So, it becomes necessary to find a rule or process under which by examining only a selected part of universe, we get all the information about it. Thus, these selected members are known as sample.

**For example**: If we buy the sweets from a shop, we examine only a single piece of it. Then this single piece will be treated as a sample.

**Size of Sample :** Number of objects, results or members in the sample is the size of the sample.

## 2.2 SAMPLING

A manner in which we can form a sample from the parent universe (population) is known as sampling.

**For example :**

1. Population of male children born in a particular year.
2. Number of smokers in a particular locality.

## 2.3 TYPES OF SAMPLING

The sample of population can be chosen in four manners or it can be stated as the sampling is of four types :

1. **Random Sampling :** Random sampling is a well known method of sampling. In this method, while choosing the sample, if each member have the same probability of chosen out then this type of sampling is known as random sampling. In short, we can say that when the selection is taken at random, then the sampling is known as random sampling.

   **For example :** If we throw a dice then each number have the same probability of coming out.

2. **Simple Sampling:** Simple sampling is a special case of random sampling. In this, each event has the equal probability of chosen out, in which the probability of choosing an event is free from the previous probability of successes or failure of an event.

   **For example :**

   1. In tossing of a coin, the probability of coming head does not depend upon the previous trial in any manner.
   2. If we want to select a boy from a group of 8 children containing 4 boys and 4 girls, then the probability of choosing a boy is 4/8 and if the child selected is a girl and we does not replace it in the group, then the probability of choosing a boy will be 4/7. Then

this sampling is known as simple sampling without replacement. And if we replace the girl to the group and then select the boy again, then this sampling is known as simple sampling with replacement.

3. **Purposive Sampling :** Purposive sampling is a sampling in which samples are taken under a particular consideration. In this sampling, the investigator select that part of universe by which it conclude its desired result. In this sampling personal individuals have a great chance of chosen out.

   **For example :** If we want to choose the student of 55-60 Kg from a group of students, then the random selection will give the students of all weights. Then, if we choose the students of particular weight, then this is known as purposive sampling.

4. **Stratified Sampling:** The mixture of random sampling and purposive sampling is known as stratified sampling. Sometimes it is impossible to discuss a population by taking random sampling or purposive sampling as a representative sampling. In such cases, we divide the whole universe into distinct parts and then take the sample randomly according to the size of the part. Then this type of sampling is known as stratified sampling.

5. **Cluster Sampling:** A cluster is a randomly selected group. This method is used when units of populations are natural groups such as school of fishes, hospital wards, slums of a town etc. This techniques allows small number of the target population to be sampled while the data provided statistically valid at 95% confidence limit.

**Tippett's Number :** A table constructed by L.H.C. Tippett's which solve the problem of making random sampling such that there is no relation among the numbers used is known as Tipptet's number table.

This table consists of 41600 digit and gives 10400 four figure numbers which are useful in construction of random sampling. The numbers are chosen randomly from this table.

A small part of this table is shown below :

| | | | | | |
|---|---|---|---|---|---|
| 2952 | 6641 | 3992 | 9792 | 7979 | 5911 |
| 3170 | 5624 | 4167 | 9525 | 1545 | 1396 |
| 7203 | 5356 | 1300 | 2693 | 2370 | 7483 |
| 3408 | 2762 | 3563 | 6107 | 6913 | 7691 |
| 0560 | 5246 | 1112 | 9025 | 6008 | 8126 |

**For example** : If we want to select a sample of 8 people from a group of 5000 numbered from 1 to 5000, then we choose the first 8 numbers which not exceed 5000 as

2952, 3992, 3170, 4167, 1545, 1396, 1300, 2693

This is the random sampling according to Tippett.

## FACTS : TO THE POINT

- In a simple random sampling, every member of the population has an equal chance of being selected. No one in the population is favoured over other in the selection process.
- Non-random sample is the sample drawn with a purpose. It does not provide equal opportunity to all members of the population being selected.
- The sample or groups of observations that are being compared are known as treatments.
- Attributes are the non-measurable characteristic which can not be numerically expressed. These are qualitative object.
- The method of stratified sampling is for giving representations to all strata of population such as selecting sample from defined areas, classes, ages, sexes etc. This technique gives more representative sample than simple random sampling in a given large population.
- In stratified sampling, proportionate representative sample from each group is secured and it gives greater accuracy.
- Cluster sampling is most often used to evaluate vaccination.
- Non-random sampling methods are those which do not provide every item in the universe with a known chance of being selected in the sample. The selection process is partially subjective.

## 2.4 CHARACTERISTICS OF A GOOD SAMPLE DESIGN

We can list down the characteristics of a good sample design as under :

1. Sample design must result in a truly representative sample.
2. Sample design must be such which results in a small sampling error.
3. Sample design must be viable in context of funds available for the research study.
4. Sample design must be such so that systematic bias can be controlled in a better way.
5. Sample should be such that the results of the sample study can be applied, in general, for the universe with a reasonable level of confidence.

## 2.5 TECHNIQUES FOR RANDOM SAMPLING

The techniques for random sampling can be of three types :

**1.** Random sampling by lottery system.

**2.** By arranging the whole numbers according to a rule and then selecting the individuals in a sequence

**3.** By random number method

**1. Random sampling by lottery system :** The lottery system is used by three ways. In first case, we make the pieces of a paper of same size such that we cannot differentiate them. Now, we numbered them according to the individuals and then we mix them. A chit is drawn out and the process will go on until we get the required number of chits equal to the sample size.

The individuals corresponding to the chits form a sample. This method of random sampling is known as chit method.

Now, the second method of lottery system is card method. In this, cards are used in place of chits. All the cards are numbered according to the sample and then shuffled. One card is drawn out and then the cards are reshuffled and the process will go on until we get the cards equal to the sample size. Now the individuals corresponding to card form a sample.

The third method is lottery system. This method is almost same as chit method. In this method, chits are placed in similar containers and these containers are rotated in a rotating drum. Then these containers are picked up one by one until the sample size is obtained. These individuals corresponding to container form a sample. This method is known as lottery system for random sampling.

**2. By arranging the whole numbers according to a rule and then selecting the individuals in a sequence :** Here, at first we arrange the whole universe in a particular manner and then select the individuals in a sequence.

**For example :** If we want to select 10 boys from a group of 250, then we arrange the whole group according to a rule (according to height, weight or names). Now, if we select every $25^{th}$ boy, we get a random sampling of 10 boys.

**3. By random number method :** This method completely depends upon the random number table. Fisher, Kendall, Mahalnobis and Tippett have published such type of tables.

According to Tippett, if we want to find out a sample of 8 people from a group of 5000, then we number them from 1 to 5000 and then select first 8 numbers which are less than 5000. The people corresponding to these numbers form a random sample.

### 2.5.1 MERITS AND DEMERITS OF RANDOM SAMPLING

#### MERITS

**M(1)** This method of obtaining data is more scientific because there are less chances for personal bias in sampling from the population and chances of selection of every item is equal.

**M(2)** Sampling error can be measured easily.

**M(3)** Method of random sampling is more economical.

**M(4)** Due to the randomness, the theory of probability is applicable.

**Demerits**

**DM(1)** This method requires a complete list of all the items of the population. Sometimes, such up-to-date lists are not available in many situations.

**DM(2)** In case of small sample, it will not be true representative of the population.

**DM(3)** If the units of population are spread over a large area, this method can not be used.

**DM(4)** Random sampling will not work where only certain data are accessible.

## 2.6 STRATIFIED RANDOM SAMPLING

Let us suppose the given population is heterogeneous (non-homogeneous) in nature. Then, entire heterogeneous population is divided into a number of homogeneous group, which is called strata or subpopulation.

Let population of $N$ units be divided into $l$ sub populations of $N_1, N_2, ..., N_l$ units respectively such that each subpopulations (strata) are non-overlapping and together they form the whole population. Thus, we can write $\sum_{i=1}^{l} N_i = N$

When the subpopulation (strata) have been determined a sample is drawn from each stratum. The drawing are made independently, in different strata, the sample sizes within the strata are denoted by $n_1, n_2, ..., n_k$ respectively, *i.e.*, $\sum_{i=1}^{l} n_i = n$

**Remarks**

- If simple random sampling is taken in each stratum, the whole procedure is known as "Stratified Random Sampling".
- The sample which is the set of all the sampling units drawn from each stratum is known as stratified sampling.

### 2.6.1 PRINCIPLES OF STRATIFICATION

In stratifying a population, following are the principles :

1. The stratification of population should be done such that strata are homogeneous.
2. The strata should be non-overlapping and together they must form the whole population.
3. Sometimes, when it is difficult to stratify with respect to characteristic under study, administrative convenience may be considered as the basis for stratification.
4. It will be better to treat each subpopulation as a stratum.

**FACTS : TO THE POINT**

- Systematic random sampling is done in case of large, scattered and heterogeneous populations, when complete list of population is available.
- Like simple random sampling, in systematic random sampling, at first all the sampling units are arranged with numbers. A random starting point at the begining of an ordered population is chosen and then rest of the sample is selected by taking every $n^{th}$ sampling point.
- Stratified sampling is used in heterogeneous population and when we desire to have information about the distribution of a particular variable.
- In stratified random sampling, at first, the entire heterogeneous population is divided into small homogeneous groups (called strata), then from each group, required number of study subjects is selected by simple or stratified random sampling in population to its original size.
- In stratified random sampling, groups or strata, should be mutually exclusive (*i.e.* all members should belong to more than one group) and exhaustive (*i.e.* all members should belong to same group).
- Cluster sampling involves dividing the specific population of interest into geographically distinct groups of clusters, such as neighbourhoods or families.
- Cluster sampling method is used when units of population are natural groups or clusters like blocks, villages, slums, schools etc.
- If the clusters are geographic regions, then sampling is called area sampling.

## 2.7 SYSTEMATIC RANDOM SAMPLING

In this sampling method, we used partly the arbitrariness and partly randomness. Let us suppose $N$ units of populations are numbered $1, 2, ..., N$ in some order. Let $N = nk$, where $n$ is the sample size and $k$ is an integer (called the sampling interval). Then, a random number less than or equal to $k$ be selected and every $k^{th}$ unit thereafter. The resulting sample is called $k^{th}$ systematic sample and such a procedure is known as linear systematic sampling.

Now, if $N \neq nk$ and every $k^{th}$ unit be included in a circular manner till the whole list is exhausted, it is known as circular systematic sampling.

The systematic random sampling is used if a complete and up-to-date sampling is available. Also, under many situations, systematic sampling provides estimates which are more efficient than those obtained with simple random sampling without replacement.

## 2.8 LIMITATIONS OF SAMPLING

Some limitations of sampling theory are as follows :

1. Proper care should be taken in the planning, otherwise the results obtained might be inaccurate and misleading.
2. If the information is required about each and every unit of the universe, there is no way but to resort to complete enumeration. Also, if time and money are not important, a complete census may be better than any sampling method.
3. In the absence of the services of trained and qualified personnel and sophisticated equipments for its planning, execution and analysis, the result of sample survey are not trustworthy.

## EXERCISE 2.1

1. Write short notes on the following :
   (i) Census method
   (ii) Sampling method
   (iii) Systematic random sample
2. Write the advantages and disadvantages of the following :
   (i) Census method
   (ii) Sampling method
3. Describe the followings :
   (i) Simple random sampling
   (ii) Stratified random sampling
4. Discuss various methods of selecting a random sample.
5. What are the essentials of sampling.

## OBJECTIVE EVALUATION

### MULTIPLE CHOICE QUESTIONS (CHOOSE THE MOST APPROPRIATE ONE)

1. In a city, every $6^{th}$ house is selected for study, this is which type of sampling:
   (a) systematic random sampling
   (b) simple random sampling
   (c) stratified random sampling
   (d) none of these **(AIIMS–1994)**
2. Immunization status in an area is checked by:
   (a) simple random sampling
   (b) cluster sampling
   (c) systematic random sampling
   (d) none of these **(JIPMER–1993)**
3. Which one of the following is true about simple sampling?
   (a) every person has an equal chance of selection
   (b) groups are not equally distributed
   (c) also known as systematic random sampling
   (d) none of these **(AIIMS–2001, 09)**
4. Which of the following are random sampling methods?

(a) simple random sampling
(b) cluster sampling
(c) stratified random sampling
(d) all are true (AIIMS–1996, 2000)

5. Which of the following is true about simple random sampling?
(a) each person has a known and equal chance of being selected
(b) error most frequent
(c) adjacent samples should be drawn
(d) none of these (AIIMS–1996, 98, 2000)

6. At first one family unit is chosen at random and then every 5th family is chosen, this is:
(a) systematic random sampling
(b) simple random sampling
(c) stratified random sampling
(d) none of these

7. About the cluster sampling, which of the following is not true?
(a) It is independent of the result
(b) It is done for evaluating the immunization status
(c) Sample size is same as simple random sampling
(d) None of these (AIIMS–2007, 08)

8. The number of patients required in a clinical trial to treat a specific disease increases as:
(a) The dropout rate increases
(b) The significance level increases
(c) The incidence of the disease decreases
(d) None of these (AIIMS–2002)

## Answers

1. (a) 2. (b) 3. (a) 4. (d) 5. (a) 6. (a) 7. (c) 8. (b)

## Archive

1. What is sampling? Define different types of sampling with their advantages and disadvantages. (WB–2013)
2. Write the short not on the following:
a. Sampling (IPGME–2009, BURDWAN–2008, BANKURA SAMMILIANI–2008, RGUHS–2009, 11, CSSM–2003, 04, GUHAWATI–2015)
b. Sampling techniques (MIDNAPUR–2008)
c. Random sampling (CALCUTTA–2011)
d. Cluster sampling (RGUHS–2007, 09)
3. Define simple random sampling. (RGUHS–2005, CSSM–2013)
4. Define stratified random sampling. (RGUHS–2008, CSSM–2009, 11)

## More to Know

- Entire group of the study population is known as universe or whole population.
- Each member of the whole population is called 'sampling unit'.
- A list where all individuals from the whole population are drawn up is called 'sampling figure'.
- Sample is a small representative part of the whole population.
- Statistic is the summary value of the sample.
- Quality that describe a population characteristic is known as parameter.
- Simple random sampling involves use of random numbers.
- In systematic random sampling, stratification of population is done.
- Probability sampling is superior to non-probability sampling.
- Cluster sampling may be of different types.
- Sample is a part of the universe.
- Stratified random sampling is applicable in heterogeneous population.
- Sample size in cluster sampling is applicable in heterogeneous population.