2. INTRODUCTION TO SAMPLING METHODS

2.1 Introduction:

Sampling is very often used in our daily life. For example while purchasing food grains from a shop we usually examine a handful from the bag to assess the quality of the commodity. A doctor examines a few drops of blood as sample and draws conclusion about the blood constitution of the whole body. Thus most of our investigations are based on samples. In this chapter, let us see the importance of sampling and the various methods of sample selections from the population.

2.2 Population:

In a statistical enquiry, all the items, which fall within the purview of enquiry, are known as **Population** or **Universe**. In other words, the population is a complete set of all possible observations of the type which is to be investigated. Total number of students studying in a school or college, total number of books in a library, total number of houses in a village or town are some examples of population.

Sometimes it is possible and practical to examine every person or item in the population we wish to describe. We call this a **Complete enumeration,** or **census**. We use **sampling** when it is not possible to measure every item in the population. Statisticians use the word population to refer not only to people but to all items that have been chosen for study.

2.2.1 Finite population and infinite population:

A population is said to be finite if it consists of finite number of units. Number of workers in a factory, production of articles in a particular day for a company are examples of finite population. The total number of units in a population is called population size. A population is said to be infinite if it has infinite number of units. For example the number of stars in the sky, the number of people seeing the Television programs etc.,

2.2.2 Census Method:

Information on population can be collected in two ways – census method and sample method. In census method every element of the population is included in the investigation. For example, if we study the average annual income of the families of a particular village or area, and if there are 1000 families in that area, we must study the income of all 1000 families. In this method no family is left out, as each family is a unit.

Population census of India:

The population census of our country is taken at 10 yearly intervals. The latest census was taken in 2001. The first census was taken in 1871 – 72.

[Latest population census of India is included at the end of the chapter.]

2.2.3 Merits and limitations of Census method:

Mertis:

1. The data are collected from each and every item of the population
2. The results are more accurate and reliable, because every item of the universe is required.
3. Intensive study is possible.
4. The data collected may be used for various surveys, analyses etc.

Limitations:
1. It requires a large number of enumerators and it is a costly method
2. It requires more money, labor, time energy etc.
3. It is not possible in some circumstances where the universe is infinite.

2.3 Sampling:
The theory of sampling has been developed recently but this is not new. In our everyday life we have been using sampling theory as we have discussed in introduction. In all those cases we believe that the samples give a correct idea about the population.
Most of our decisions are based on the examination of a few items that is sample studies.

2.3.1 Sample:
Statisticians use the word **sample** to describe a portion chosen from the population. A finite subset of statistical individuals defined in a population is called a sample. The number of units in a sample is called the **sample size.**

Sampling unit:
The constituents of a population which are individuals to be sampled from the population and cannot be further subdivided for the purpose of the sampling at a time are called sampling units. For example, to know the average income per family, the head of the family is a sampling unit. To know the average yield of rice, each farm owner's yield of rice is a sampling unit.

Sampling frame:
For adopting any sampling procedure it is essential to have a list identifying each sampling unit by a number. Such a list or map is called sampling frame. A list of voters, a list of house holders, a list of villages in a district, a list of farmers etc. are a few examples of sampling frame.

2.3.2 Reasons for selecting a sample:
Sampling is inevitable in the following situations:
1. Complete enumerations are practically impossible when the population is infinite.
2. When the results are required in a short time.
3. When the area of survey is wide.
4. When resources for survey are limited particularly in respect of money and trained persons.
5. When the item or unit is destroyed under investigation.

2.3.3 Parameters and statistics:
We can describe samples and populations by using measures such as the mean, median, mode and standard deviation.
When these terms describe the characteristics of a population, they are called **parameters**. When they describe the characteristics of a sample, they are called **statistics**. A parameter is a characteristic of a population and a statistic is a characteristic of a sample. Since samples are subsets of population statistics provide estimates of the parameters. That is, when the parameters are unknown, they are estimated from the values of the statistics.

In general, we use Greek or capital letters for population parameters and lower case Roman letters to denote sample statistics. [N, μ, σ, are the standard symbols for the size, mean, S.D, of population. n, **x**, s, are the standard symbol for the size, mean, s.d of sample respectively].

2.3.4 Principles of Sampling:

Samples have to provide good estimates. The following principle tell us that the sample methods provide such good estimates:

1. Principle of statistical regularity:

A moderately large number of units chosen at random from a large group are almost sure on the average to possess the characteristics of the large group.

2. Principle of Inertia of large numbers:

Other things being equal, as the sample size increases, the results tend to be more accurate and reliable.

3. Principle of Validity:

This states that the sampling methods provide valid estimates about the population units (parameters).

4. Principle of Optimization:

This principle takes into account the desirability of obtaining a sampling design which gives optimum results. This minimizes the risk or loss of the sampling design.

The foremost purpose of sampling is to gather maximum information about the population under consideration at minimum cost, time and human power. This is best achieved when the sample contains all the properties of the population.

Sampling errors and non-sampling errors:

The two types of errors in a sample survey are sampling errors and non - sampling errors.

1. Sampling errors:

Although a sample is a part of population, it cannot be expected generally to supply full information about population. So there may be in most cases difference between statistics and parameters. The discrepancy between a parameter and its estimate due to sampling process is known as **sampling error.**

2. Non-sampling errors:

In all surveys some errors may occur during collection of actual information. These errors are called Non-sampling errors**.**

2.3.5 Advantages and Limitation of Sampling:

There are many advantages of sampling methods over census method. They are as follows:

1. Sampling saves time and labor.
2. It results in reduction of cost in terms of money and man hour.
3. Sampling ends up with greater accuracy of results.
4. It has greater scope.
5. It has greater adaptability.
6. If the population is too large, or hypothetical or destroyable sampling is the only method to be used.

<u>The limitations of sampling are given below:</u>

1. Sampling is to be done by qualified and experienced persons. Otherwise, the information will be unbelievable.

2. Sample method may give the extreme values sometimes instead of the mixed values.
3. There is the possibility of sampling errors. Census survey is free from sampling error.

2.4 Types of Sampling:
The technique of selecting a sample is of fundamental importance in sampling theory and it depends upon the nature of investigation. The sampling procedures which are commonly used may be classified as:

1. Probability sampling.
2. Non-probability sampling.
3. Mixed sampling.

2.4.1 Probability sampling (Random sampling):
A probability sample is one where the selection of units from the population is made according to known probabilities. (eg.)

Simple random sample, probability proportional to sample size etc.

2.4.2 Non-Probability sampling:
It is the one where discretion is used to select 'representative' units from the population (or) to infer that a sample is 'representative' of the population. This method is called **judgment or purposive** sampling. This method is mainly used for opinion surveys; A common type of judgment sample used in surveys is quota sample. This method is not used in general because of prejudice and bias of the enumerator. However if the enumerator is experienced and expert, this method may yield valuable results. For example, in the market research survey of the performance of their new car, the sample was all new car purchasers.

2.4.3 Mixed Sampling:
Here samples are selected partly according to some probability and partly according to a fixed sampling rule; they are termed as mixed samples and the technique of selecting such samples is known as **mixed sampling**.

2.5 Methods of selection of samples:

Here we shall consider the following three methods:

1. Simple random sampling.
2. Stratified random sampling.
3. Systematic random sampling.

1. Simple random sampling:
A simple random sample from finite population is a sample selected such that each possible sample combination has equal probability of being chosen. It is also called unrestricted random sampling.

2. Simple random sampling without replacement:
In this method the population elements can enter the sample only once (ie) the units once selected is not returned to the population before the next draw.

3. Simple random sampling with replacement:
In this method the population units may enter the sample more than once. Simple random sampling may be with or without replacement.

2.5.1 Methods of selection of a simple random sampling:

The following are some methods of selection of a simple random sampling.

a) Lottery Method:
This is the most popular and simplest method. In this method all the items of the population are numbered on separate slips of paper of same size, shape and color. They are folded and mixed up in a container. The required numbers of slips are selected at random for the desire sample size. For example, if we want to select 5 students, out of 50 students, then we must write their names or their roll numbers of all the 50 students on slips and mix them. Then we make a random selection of 5 students.

This method is mostly used in lottery draws. If the universe is infinite this method is inapplicable.

b) Table of Random numbers:
As the lottery method cannot be used, when the population is infinite, the alternative method is that of using the table of random numbers. There are several standard tables of random numbers.

1. Tippett's table
2. Fisher and Yates' table
3. Kendall and Smith's table are the three tables among them.

A random number table is so constructed that all digits 0 to 9 appear independent of each other with equal frequency. If we have to select a sample from population of size N= 100, then the numbers can be combined three by three to give the numbers from 001 to 100.

[See Appendix for the random number table]

Procedure to select a sample using random number table:

Units of the population from which a sample is required are assigned with equal number of digits. When the size of the population is less than thousand, three digit number 000,001,002, ….. 999 are assigned. We may start at any place and may go on in any direction such as column wise or row- wise in a random number table. But consecutive numbers are to be used.

On the basis of the size of the population and the random number table available with us, we proceed according to our convenience. If any random number is greater than the population size N, then N can be subtracted from the random number drawn.

This can be repeatedly until the number is less than N or equal to N.

Example 1:

In an area there are 500 families. Using the following extract from a table of random numbers select a sample of 15 families to find out the standard of living of those families in that area.

4652 3819 8431 2150 2352 2472 0043 3488
9031 7617 1220 4129 7148 1943 4890 1749
2030 2327 7353 6007 9410 9179 2722 8445
0641 1489 0828 0385 8488 0422 7209 4950

Solution:

In the above random number table we can start from any row or column and read three digit numbers continuously row-wise or column wise.

Now we start from the third row, the numbers are:

203 023 277 353 600 794 109 179
272 284 450 641 148 908 280

Since some numbers are greater than 500, we subtract 500 from those numbers and we rewrite the selected numbers as follows:

203 023 277 353 100 294 109 179
272 284 450 141 148 408 280

c) Random number selections using calculators or computers:

Random number can be generated through scientific calculator or computers. For each press of the key get a new random numbers. The ways of selection of sample is similar to that of using random number table.

Merits of using random numbers:

Merits:
1. Personal bias is eliminated as a selection depends solely on chance.
2. A random sample is in general a representative sample for a homogenous population.
3. There is no need for the thorough knowledge of the units of the population.
4. The accuracy of a sample can be tested by examining another sample from the same universe when the universe is unknown.
5. This method is also used in other methods of sampling.

Limitations:

1. Preparing lots or using random number tables is tedious when the population is large.
2. When there is large difference between the units of population, the simple random sampling may not be a representative sample.
3. The size of the sample required under this method is more than that required by stratified random sampling.
4. It is generally seen that the units of a simple random sample lie apart geographically. The cost and time of collection of data are more.

2.5.2 Stratified Random Sampling:

Of all the methods of sampling the procedure commonly used in surveys is stratified sampling. This technique is mainly used to reduce the population heterogeneity and to increase the efficiency of the estimates. Stratification means division into groups. In this method the population is divided into a number of subgroups or strata. The strata should be so formed that each stratum is homogeneous as far as possible. Then from each stratum a simple random sample may be selected and these are combined together to form the required sample from the population.

Types of Stratified Sampling:
There are two types of stratified sampling. They are **proportional** and **non-proportional**. In the proportional sampling equal and proportionate representation is given to subgroups or strata. If the number of items is large, the sample will have a higher size and vice versa.

The population size is denoted by N and the sample size is denoted by ‘ n’ the sample size is allocated to each stratum in such a way that the sample fractions is a constant for each stratum.

That is given by n/N = c. So in this method each stratum is represented according to its size.

In non-proportionate sample, equal representation is given to all the sub-strata regardless of their existence in the population.

Example 2:

A sample of 50 students is to be drawn from a population consisting of 500 students belonging to two institutions A and B.

The number of students in the institution A is 200 and the institution B is 300. How will you draw the sample using proportional allocation?

Solution:
There are two strata in this case with sizes $N_1 = 200$ and $N_2 = 300$ and the total population

$N = N_1 + N_2 = 500$

The sample size is 50.

If n_1 and n_2 are the sample sizes,

$$n_1 = \frac{n}{N} \times N_1 = \frac{50}{500} \times 200 = 20$$

$$n_2 = \frac{n}{N} \times N_2 = \frac{50}{500} \times 300 = 30$$

The sample sizes are 20 from A and 30 from B. Then the units from each institution are to be selected by simple random sampling.

Merits and limitations of stratified sampling:

Merits:
1. It is more representative.
2. It ensures greater accuracy.
3. It is easy to administer as the universe is sub - divided.
4. Greater geographical concentration reduces time and expenses.
5. When the original population is badly skewed, this method is appropriate.
6. For non – homogeneous population, it may field good results.

Limitations:
1. To divide the population into homogeneous strata, it requires more money, time and statistical experience which is a difficult one.

2. Improper stratification leads to bias, if the different strata overlap such a sample will not be a representative one.

2.5.3 Systematic Sampling:
This method is widely employed because of its ease and convenience. A frequently used method of sampling when a complete list of the population is available is **systematic sampling**.

It is also called **Quasi-random sampling**.

Selection procedure:
The whole sample selection is based on just a random start.

The first unit is selected with the help of random numbers and the rest get selected automatically according to some pre designed pattern is known as **systematic sampling**. With systematic random sampling every K_{th} element in the frame is selected for the sample, with the starting point among the first K elements determined at random. For example, if we want to select a sample of 50 students from 500 students under this method K_{th} item is picked up from the sampling frame and K is called the **sampling interval**.

Sampling interval, K = $\frac{N}{n} = \frac{Population\ Size}{Sample\ Size}$

$$K = \frac{500}{50} = 10$$

K = 10 is the sampling interval. Systematic sample consists in selecting a random number say i K and every K^{th} unit subsequently. Suppose the random number ‘ i’ is 5, then we select 5, 15, 25, 35, 45,……. The random number ‘i’ is called random start. The technique will generate K systematic samples with equal probability.

Merits:
1. This method is simple and convenient.
2. Time and work is reduced much.
3. If proper care is taken result will be accurate.
4. It can be used in infinite population.

Limitations:
1. Systematic sampling may not represent the whole population.
2. There is a chance of personal bias of the investigators.

Systematic sampling is preferably used when the information is to be collected from trees in a forest, house in blocks, entries in a register which are in a serial order etc.

Exercise – 2

I. Choose the best Answer:

1. Sampling is inevitable in the situations
   (a) Blood test of a person
   (b) When the population is infinite
   (c) Testing of life of dry battery cells
   (d) All the above

2. The difference between sample estimate and population parameter is termed as
   (a) Human error
   (b) Sampling error
   (c) Non-sampling error
   (d) None of the above

3. If each and every unit of population has equal chance of being included in the sample, it is known as:
   (a) Restricted sampling
   (b) Purposive sampling
   (c) Simple random sampling
   (d) None of the above

4. Simple random sample can be drawn with the help of
   (a) Slip method
   (b) Random number table
   (c) Calculator
   (d) All the above

5. A selection procedure of a sample having no involvement of probability is known as
   (a) Purposive sampling
   (b) Judgment sampling
   (c) Subjective sampling
   (d) All the above

6. Five establishments are to be selected from a list of 50 establishments by systematic random sampling. If the first number is 7, the next one is
   (a) 8
   (b) 16
   (c) 17
   (d) 21

II. Fill in the blanks:

7. A population consisting of an unlimited number of units is called an _____ population

8. If all the units of a population are surveyed it is called _____

9. The discrepancy between a parameter and its estimate due to sampling process is known as _____

10. The list of all the items of a population is known as _____

11. Stratified sampling is appropriate when population is _____

12. When the items are perishable under investigation it is not possible to do _____

13. When the population consists of units arranged in a sequence would prefer _____ sampling

14. For a homogeneous population, _____ sampling is better than stratified random sampling.

III. Answer the following questions:

15. Define a population

16. Define finite and infinite populations with examples

17. What is sampling?

18. Define the following terms
    (a) Sample
    (b) Sample size
    (c) Census
    (d) Sampling unit
    (e) Sampling frame

19. Distinguish between census and sampling

20. What are the advantages of sampling over complete enumeration?

21. Why do we resort to sampling?

22. What are the limitations of sampling?

23. State the principles of sampling

24. What are probability and non-probability sampling?

25. Define purposive sampling. Where it is used?

26. What is called mixed sampling?

27. Define a simple random sampling.

28. Explain the selection procedure of simple random Sampling.

29. Explain the two methods of selecting a simple random sampling.

30. What is a random number table? How will you select the random numbers?

31. What are the merits and limitations of simple random sampling?

32. What circumstances stratified random sampling is used?

33. Discuss the procedure of stratified random sampling. Give examples.

34. What is the objective of stratification?

35. What are the merits and limitations of stratified random sampling?

36. Explain systematic sampling

37. Discuss the advantages and disadvantages of systematic random sampling

38. Give illustrations of situations where systematic sampling is used.

39. A population of size 800 is divided into 3 strata of sizes 300, 200, 300 respectively. A stratified sample size of 160 is to be drawn from the population. Determine the sizes of the samples from each stratum under proportional allocation.

40. Using the random number table, make a random number selection of 8 plots out of 80 plots in an area.

41. There are 50 houses in a street. Select a sample of10 houses for a particular study using systematic sampling.

IV. Suggested activities:

42. (a) List any five sampling techniques used in your environment
(b) List any five situations where we adopt census method.(i.e) complete enumeration).

43. Select a sample of students in your school (for a particular competition function) at primary, secondary higher secondary levels using stratified sampling using proportional allocation.

44. Select a sample of 5 students from your class attendance register using method of systematic sampling.

POPULATION OF INDIA 2001

India/State/ Union territories*	POPULATION OF INDIA 2001			Population Variation 1991-2001	Sex ratio (females per thousand males)
	PERSONS	MALES	FEMALES		
INDIA 1,2	1,027,015,247	531,277,078	495,738,169	21.34	933
Andaman & Nicobar Is.*	356,265	192,985	163,280	26.94	846
Andhra Pradesh	75,727,541	38,286,811	37,440,730	13.86	978
Arunachal Pradesh	1,091,117	573,951	517,166	26.21	901
Assam	26,638,407	13,787,799	12,850,608	18.85	932
Bihar	82,878,796	43,153,964	39,724,832	28.43	921
Chandigarh*	900,914	508,224	392,690	40.33	773
Chhatisgarh	20,795,956	10,452,426	10,343,530	18.06	990
Dadra & Nagar Haveli*	220,451	121,731	98,720	59.20	811
Daman & Diu*	158,059	92,478	65,581	55.59	709
Delhi*	13,782,976	7,570,890	6,212,086	46.31	821
Goa	1,343,998	685,617	658,381	14.89	960
Gujarat 5	50,596,992	26,344,053	24,252,939	22.48	921
Haryana	21,082,989	11,327,658	9,755,331	28.06	861
Himachal Pradesh 4	6,077,248	3,085,256	2,991,992	17.53	970
Jammu & Kashmir 2,3	10,069,917	5,300,574	4,769,343	29.04	900
Jharkhand	26,909,428	13,861,277	13,048,151	23.19	941
Karnataka	52,733,958	26,856,343	25,877,615	17.25	964
Kerala	31,838,619	15,468,664	16,369,955	9.42	1,058
Lakshadweep*	60,595	31,118	29,477	17.19	947
Madhya Pradesh	60,385,118	31,456,873	28,928,245	24.34	920
Maharashtra	96,752,247	50,334,270	46,417,977	22.57	922

Manipur	2,388,634	1,207,338	1,181,296	30.02	978
Meghalaya	2,306,069	1,167,840	1,138,229	29.94	975
Mizoram	891,058	459,783	431,275	29.18	938
Nagaland	1,988,636	1,041,686	946,950	64.41	909
Orissa	36,706,920	18,612,340	18,094,580	15.94	972
Pondicherry*	973,829	486,705	487,124	20.56	1,001
Punjab	24,289,296	12,963,362	11,325,934	19.76	874
Rajasthan	56,473,122	29,381,657	27,091,465	28.33	922
Sikkim	540,493	288,217	252,276	32.98	875
Tamil Nadu	62,110,839	31,268,654	30,842,185	11.19	986
Tripura	3,191,168	1,636,138	1,555,030	15.74	950
Uttar Pradesh	166,052,859	87,466,301	78,586,558	25.80	898
Uttaranchal	8,479,562	4,316,401	4,163,161	19.20	964
West Bengal	80,221,171	41,487,694	38,733,477	17.84	934

Notes:

1. The population of India includes the estimated population of entire Kachchh district, Morvi, Maliya-Miyana and Wankaner talukas of Rajkot district, Jodiya taluka of Jamanagar district of Gujarat State and entire Kinnaur district of Himachal Pradesh where population enumeration of Census of India 2001 could not be conducted due to natural calamity.
2. For working out density of India, the entire area and population of those portions of Jammu and Kashmir which are under illegal occupation of Pakistan and China have not been taken into account.
3. Figures shown against Population in the age-group 0-6 and Literates do not include the figures of entire Kachchh district, Morvi, Maliya-Miyana and Wankaner talukas of Rajkot district, Jodiya taluka of Jamanagar district and entire Kinnaur district of Himachal Pradesh where population enumeration of Census of India 2001 could not be conducted due to natural calamity.
4. Figures shown against Himachal Pradesh have been arrived at after including the estimated figures of entire Kinnaur district of Himachal Pradesh where the population enumeration of Census of India 2001 could not be conducted due to natural calamity.
5. Figures shown against Gujarat have been arrived at after including the estimated figures of entire Kachchh district, Morvi, Maliya-Miyana and Wankaner talukas of Rajkot district, Jodiya taluka of Jamnagar district of Gujarat State where the population enumeration of Census of India 2001 could not be conducted due to natural calamity.