UNIT 12 SAMPLING

Structure

12.1 INTRODUCTION

In Unit 11, we have discussed the meaning and type of hypothesis. We also know by this time that testing the hypothesis is central to the research. In order to accomplish this it is imperative to collect requisite, recent and most relevant data. This helps us to decide from whom and how to collect these data.

The primary purpose of research is to discover principles that have universal application. But to study a whole population in order to arrive at generalizations would be impracticable if not impossible. Some populations are so that their characteristics cannot be measured, because before the measurement could be completed, the populations would have changed. Imagine the difficulty of conducting an experiment with all fifth-grade Indian children [as subjects] on numerical ability. The study of this population would require the services of thousands of researchers, the expenditure of millions of rupees and thousands of class hours. In view of this it becomes imperative to collect the data from a smaller group of the population instead of collecting from the whole population.

There are various ways to achieve this. We will discuss these in this unit.

12.2 OBJECTIVES

After completing the study of this unit, you will be to :

- define the terms 'population and sample',
- justify the need of selecting a sample,
- explain the meaning of probability sampling,
- describe various probability sampling methods,
- explain the meaning of non-probability sampling,
- describe various non-probability sampling methods,
- state the characteristics of a good sample.

12.3 MEANING OF POPULATION AND SAMPLE

A "population" is any group of individuals / units that have one or more characteristics in common which are of interest to the researcher, for a particular research. A population may include all the individuals of a particular type or a more restricted part of that group, e.g. a group of all the university teachers, or a group of male / female university teachers, or distance learners enrolled with IGNOU. For assessing the study habits of adolescent girls in city, all the adolescent girls of that city who are studying in schools and colleges, make up the population for this study.

A "sample" is a small proportion of a population selected for the study. By observing the characteristics of the sample, one can make certain inferences about the characteristics of the population from which it is drawn.

The term sampling refers to the strategies which enable us to pick a subgroup from a larger group and then use the subgroup as a basis for making judgement about the larger group. In order to use such a subgroup to make decisions about the larger group, the subgroup has to resemble the larger group as closely as possible.

12.4 METHODS/DESIGNS OF SAMPLING

In the earlier section we studied the concept of sampling. Now we shall study different sampling methods. The sampling method was used in social sciences research as early as in 1754 by A.L. Bowley. Since then the method is increasingly used. The sampling methods are broadly classified into two types: (i) Probability sampling and (ii) Non probability sampling

12.4.1 Probability Sampling

Probability Sampling is based on some statistical concepts such as the 'Law of Large Numbers', 'Central Limit Theorem and the Normal Distribution' etc. In this type of sampling, the units of the population are not selected at the discretion of the researcher, but by means of certain procedures which ensure that every unit of population has one fixed probability fo being included in the sample. It is also called random sampling.

- The Law of Large Number states that as the sample size becomes large, probability that the estimate differs from the parameter to a greater extent, becomes small. Or in other words a larger number provides a more precise measure of the parameter under consideration. However, one precaution must be taken. While increasing the size of the sample care should be taken to maintain the representiveness of the sample, because a large sample does not automatically guarantee representiveness.

- As per the second concept, sampling distribution approaches normal distribution provided – more the irregular distribution in the population, larger is the sample and sample is selected to avoid biases.

12.4.2 Non-probability Methods

The non-probability methods are based on the judgements of the investigator as the most important element of control. The guiding principles in non-probability methods are – availability of the subjects, the personal judgement of the investigator, and convenience in carrying out the research.

12.4.3 Criteria for Selecting Sampling

i) Because there are various sampling methods it becomes crucial to select appropriate sampling method. Young has suggested three criteria to be considered while selecting a sampling method –

- A measurable or known probability sampling technique should be used to control the risk of errors in the sample estimate.
- Simple, straightforward and workable methods, adapted to available facilities and personnel, should be used.
- Achieving optimum balance between expenditure incurred and maximum of reliable information should be the guiding principle.

ii) The decision whether a probability sampling or a non -probability sampling is to be applied rests on the constraints which are not very different from those stated earlier. These are – objectives of the study, type of study and availability of the resources for the study.

If the objective of the research is to apply the results of the study to a small local group then sampling may not be given as much consideration as in a study the results of which are to be applied to a larger group. In experimental research internal validity is of more concern than the external validity.

- Action research generally does not require sampling from a larger group. Most of the times sampling is not very essential in historical research also. Whereas survey studies generally have a more rigorous sampling.
- The availability of time, funds, manpower and equipment required is another important consideration in deciding about the size and technique of sampling.

iii) If one is interested in obtaining an estimate of the sampling error, one may resort to probability sampling rather than to a non-probability one.

Check Your Progress

Notes: a) Space is given below for writing your answer.

b) Compare your answer with that given at the end of the unit.

1. State the meaning of the following terms:

Population, Sample, Probability Sampling, Non-probability Sampling.

.....

.....

.....

.....

.....

.....

12.5 PROBABILITY SAMPLING

We know the meaning of and requirement for probability sampling. Now we will take a brief account of different methods of probability sampling.

12.5.1 Simple Random Sampling

Theoretically, this is a method of selecting 'n' units from N units in such a way that everyone in the population of N units, has an equal chance of being selected. This can be done through the following steps –

- defining population by specifying its various limits.
- preparing the sampling frame.
- incorporating the names or serial numbers of individual units in the sampling frame (every unit is to be listed, order does not make any difference).

It is important to note that random sample is not necessarily as identical representation of the population. After this, to get the required 'n' units different techniques are available. Let us get acquainted with these techniques one by one.

i) Lottery Method

After naming or numbering every unit in the population, they are well mixed. The required numbers of units are then drawn from all these well-mixed chits. The individuals/objects with these identification names/numbers are then picked up for inclusion in the sample.

However this technique has some objections. When the population is very large and includes such individuals/objects, which are of such nature that could not be mixed and further if 'well mixing' is not attained despite all efforts, the principle of randomness in the population may be violated.

ii) Random Table Method

In view of the above mentioned objections it is advised to use the random number tables instead of lottery method. The use of random numbers or manual lot drawing

Row → / Column ↓	1	2	3	4	5	...	N
1	2315	7548	5901	8372	5993	...	6744
2	0554	5550	4310	5374	3508	...	1343
3	1487	1603	5032	4043	6223	...	0834
4	3897	6749	5094	0517	5853	...	1695
5	9731	2617	1899	7553	0870	...	0510
6	1174	2693	8144	3393	0862	...	6850
7	4336	1288	5911	0164	5623	...	4036
8	9380	6204	7833	2680	4491	...	2571
9	4954	0131	8108	4298	4187	...	9527
10	3676	8726	3337	9482	1569	...	3880
11	...	...	...	...	...	...	...
12	...	...	...	...	...	...	...
13	...	...	...	...	...	...	...
14	...	...	...	...	...	...	...
15	...	...	...	...	...	...	...
N	3914	5218	3587	4855	4881	...	5042

will be too cumbersome to recommend in case of large population. In such situation, computer generated random selection should be resorted to, in order to save time and labour.

Tables of random numbers have been generated by computers producing a random sequence of digits e.g. random digit table by Rand Corporation and prepared by Kendall & Smith, by Fisher & Yates and by Tippett are frequently used.

The required number of units are selected from such a table in any convenient and systematic way. Now suppose we have select to 20 distance learners for interview from 80-distance learners registered at a study centre. We may start with any column and any row. Because we want 20 numbers i.e. two digit numbers, we have to select only the first two digits from each number. If we select the first column and start from first row then we will get following twenty two digit numbers – 23, 05, 14, 38, 97, 11, 43, ... 61. You will notice that numbers greater than 80 will have to be deleted from this list and for the remaining numbers selecting any other column and the row the procedure will have to be repeated, till we get required number i.e. 20. If any number is repeated in this list, it is to be substituted by selecting the next number. Until a sample of desired size is obtained, the selection procedure is to be continued.

Advantages

1. This method calls for no special expertise and training or even insight. It can be used mechanically by anybody.

However, best results are achieved by adopting simple random sampling method. Still, it is not free from criticism.

Limitations

Practically listing of all the units in the population may not be possible.

1. In case of population with infinite numbers, listing is out of the question.
2. It is difficult though not impossible, but it involves high cost.
3. In case of heterogeneous population the selected random sample may not truly represent the characteristics of the population.

12.5.2 Systematic Sampling

A variation of the random process of sampling is the systematic sampling. It is the selection of the required number of elements of the population to include in the sample. It involves the following steps :

- Listing the population elements in some order, say alphabetically, meritwise etc.
- Determining the desired number to be selected from the population e.g. 10 % of 1000 means 100 out of 1000.
- Starting with any number from among the numbers 1 to 10 (i.e. 1 to k, both inclusive), to select every 10^{th} (or k^{th}) element from the list. If the number chosen from 1 to 10 is 4 , then the selected numbers will be the 4^{th}, 14^{th}, 24^{th}, ..., 994^{th} elements making the sample with 100 elements.

As the elements are chosen from regular intervals, this technique is also known as sampling by regular intervals, this technique is also known as sampling by regular intervals, sampling by fixed intervals or sampling by every k^{th} unit.

Advantages

1. It is more practical in that it involves less labour.
2. Because it is simpler to perform, it may reduce errors.
3. The procedure is speedy in comparison with simple random sampling.

4. The systematic sample is spread more evenly over the population which makes this method more precise than stratified random sampling.

Limitations

1. Selection of every element other than the first selected randomly is linked with the first element. This makes the process different from the simple random method where selection of every element is independent of other one.
2. When the list of elements has a **periodic arrangement**, there is a risk that the sample interval may coincide with the periodic interval in the list. Suppose, A, B, C, D and E are the 5 schools selected and from each school 100 students are selected. The students from school A are placed starting from 1, from school B starting from 2, from school C starting from 3, from school D starting from 4 and from school E starting from 5 with an interval of 5. Thus the school A students will hold the numbers 1,6,11,16,21,.....496. The school B students will hold the numbers 2,7,12,17,22,.....497. Now in systematic sampling procedure suppose we decide to select 5 % of the total and randomly choose any number from 1 to 5 say '3' then starting from 3 we will have to select every 5^{th} number. These numbers will be 3,8,13,18.....498. Have you noticed that all these numbers belong to school C? Why has it happened so?

The answer is

Because every school is repeated in the list with an interval of '5' and elements are selected with an interval of '5'.

3. Another limitation of the systematic sampling method is the **trend** of the listed population. This is explained below -

Suppose 100 students are listed in the decreasing order of academic merit. We want to draw a sample of 20 students from this using systematic sampling method. 20 out of 100 means the size of interval is '5'. We can draw many samples from this listed population. If we randomly pick up a number from amongst 1 to 5, say 3 then the sample will comprise the elements 3^{rd}, 8^{th}, 13^{th}, 18^{th}.....98^{th} instead , if we randomly pick '5' then the sample will comprise the 5^{th}, 10^{th}, 15^{th},.....,100^{th} elements. Is it not obvious that the two samples will not be comparable in terms of merit ? The mean average of these two samples would be significantly different with respect to merit and other associated variables. Calculations made from such samples cannot pinpoint the sources of variability.

12.5.3 Stratified Sampling

One question may arise in your mind is that :

How to increase the precision of the sample?

By increasing the size of the sample, its precision can be increased.

But this is not the only way. Let us see which is the other one. It is the 'stratified sampling'. The term 'stratified' is very much self-explanatory. It involves dividing the population into such sub-populations(strata) that each one of them is homogeneous within itself.

The steps to be followed in this method are as under –

- deciding upon one or more characteristics on the basis of which strata will be formed e.g. location of schools – rural, urban, suburban, urban-slums, metropolitan etc.
- dividing the population under consideration into strata on the basis of stratification characteristics/criteria.
- listing the units in each stratum separately.

- selecting requisite number of elements from each stratum using appropriate random selection technique.

Thus all the elements selected from all the strata compose the required sample.

Important points to be noted –

i) The criteria for dividing the population into strata should be correlated with the variable being studied.

ii) The criteria should be practical. It should not yield a unwieldy number of strata.

iii) A good measure of the stratification criteria should be available; e.g. if reliable and valid tool of determining socio-economic status is not available , stratification on this basis would lead to confounding of the results.

iv) Selection of the elements at random from each stratum in the same proportion as that of the actual size of the stratum in the population improves the representativeness of the sample and helps in achieving higher efficiency at a reduced cost.

v) In some studies (like census) stratification is not possible before the data have been collected. After collecting the data stratification as per sex, age, educational level is effected. Or a simple random sample of the required size is selected and the classification into strata is observed.

Advantages

1. Stratified random sampling is very useful when a list of the elements in the population is not available.
2. It is the most applicable method of sampling when the population is heterogeneous.

12.5.4 Cluster Sampling

When listing of population or at least the total strength of the population is possible and available or in other words when the population is **finite** we saw that the sampling methods applicable were – simple random, systematic or stratified random sampling.

But what to do when the population is **infinite ?**

In such a case , the method applicable is called as **'cluster sampling'**.

Cluster sampling is used when the population under study is relatively infinite, where the list of the elements is not available, the elements/units are geographically scattered or when sampling of individual elements is not required or is not convenient.

A cluster is an intact group as available in the field. It is not formed by the researcher for the purpose of data collection. For example a school complex (a group of schools) is a cluster. Some such clusters are selected to make a sample. Here a sampling element/unit is a group/cluster.

In social survey, the cluster sampling is described as **'area sampling'**.

This method involves following steps –

- deciding the nature of the cluster required.
- indentifying/locating such clusters to make the population.
- selecting the clusters in required number at random.

Advantages

This method of sampling is economic, especially when the cost of measuring a unit is relatively small.

Limitations

When the sampling unit is to be an individual element/unit or number in the population, this method is not applicable.

12.5.5 Multi-stage and Multi-phase Sampling

In the multi-stage sampling selection of different types of sampling units such as some Districts in a State, some Taluka places in those Districts and then some schools, is involved at different sampling stages. Whereas in the multi-phase sampling, the researcher is concerned with the same type of sampling unit at each phase but some members are asked for more information than others, e.g. information regarding study habits of distance learners can be collected from 100 distance learners through a questionnaire and 20 out of them can be interviewed for more information.

The main distinction between the multi-stage and the multi-phase sampling is the use of unit of sampling at different levels.

Advantages

1. In both the methods burden on respondents is reduced.
2. Relative cost also gets reduced.
3. Two-phase sampling is useful in studying rare cases.
4. In two-phase sampling resulting gain in precision is more due to possibility of getting more information in details.

Check Your Progress

Notes: a) Space is given below for writing your answers.
b) Compare your answers with those given at the end of the unit.

2. Match the Pairs.

I) *Sampling Method*	*Limitations*
i) Simple Random	a) Sampling unit is not an individual element
ii) Systematic	b) Not applicable to heterogeneous population
iii) Cluster	c) Listing the elements in sub-population necessary
iv) Stratified random	d) Periodic arrangement of elements
	e) Periodic arrangement of elements

II) *Sampling Method*	*Special Feature*
i) Stratified random	a) Every unit in the population has equal chance of being selected
ii) Simple random	b) Spread more evenly over the population
iii) Systematic	c) Useful in case of heterogeneous population
iv) Cluster	d) Applicable in case of infinite population
	e) Same type of sampling unit at each phase

3. Distinguish between multi-stage and multi-phase sampling methods.

.....

.....

.....

12.6 NON-PROBABILITY SAMPLING

In 12.4.2 earlier we have understood the meaning of non-probability sampling. It is reiterated here. Where the selection of the units in the sample is based on the researcher's judgement and not on equal or known probability, it is the non-probability sampling method. A non-probability sample is termed as **'non-random sample'** due to the very fact that it is selected through non-random method. The main feature of such a sample is the lack of control of the sampling error on account of which this method of sampling is referred to as **'uncontrolled sampling'** method. This description of the non-probability sampling should not be taken in negative sense. In spite of all this many a times it is the demand of the situation to go for non-probability sampling method. Let us now study the different non-probability sampling methods one by one.

12.6.1 Incidental Sampling

Incidental Sampling is also known as accidental or convenience sampling. When a readily or easily available group is selected as a sample, it is termed as an **'incidental sample'**. A teacher-educator e.g. may select the students from a school situated in the same campus which serves as a practicing school for the concerned college of education, to experimentally find the effectiveness of concept attainment model to teach a mathematical concept say, a quadrilateral.

Advantages

The administrative convenience of obtaining sample for the study, the ease of testing, saving in time, completeness of the data collected are some of the merits of this method.

Limitations

Since there is no well-defined population and no random sampling method is applied to select the sample, the standard error formulae apply with a high degree of approximation. Hence no valid generalization can be drawn. Any attempt at generalization based on such data and conclusion thereof will be misleading.

12.6.2 Purposive Sampling

Another non-probability sampling method is 'purposive sampling'. In this method samples are expressly chosen because in the light of available information they resemble some larger group with respect to one or more characteristics. The controls/criteria for categorization in such samples are usually identified as representative areas such as a state, a district, a city etc or representative characteristics of individuals such as age, sex, socio-economic status etc. or representative types of groups such as elementary school teachers, secondary school teachers, college teachers, university teachers etc. These controls/ criteria may be further sub- divided e.g. the group of college teachers can be divided into male and female teachers or teachers in science/arts/commerce colleges etc.

Have you noticed that upto this stage the controls are somewhat similar to stratification criteria? After deciding upon the category required for the research, the researcher has to select the sample.

Actual selection of the units for inclusion in the sample is done purposively and not randomly; e.g. in order to tackle the problem of indiscipline only the undisciplined students are selected as the sample excluding others, on the basis of past experience.

Advantages

1. This method of sampling is useful where a small sample is required.
2. It is focused on solving problems of particular groups.

Limitations

This method is applicable only for the selection of samples including typical/special cases such as 'best teacher award winners' from the population of teachers or 'meritorious past students of the school' from the population of the past students.

12.6.3 Quota Sampling

This is another method of non-probability sampling. It involves the selection of the sample units within each stratum, on the basis of the judgement of the researcher. What distinguishes it from probability sampling is that, once the strength of the sample (e.g. how many women teachers from among the college teachers) is decided which forms the **'quota'**, the choice of the actual units to fit into this framework is left to the researcher.

Quota Sampling is thus a method of stratification sampling in which selection of sample units within the stratum is non- random.

These quotas are determined by the proportion of the groups, e.g. in order to study the attitude of school teachers towards environment education, first of all the school teachers will be stratified into men and women teachers, quotas for these strata will be fixed and will be selected (not randomly).

12.7 CHARACTERISTICS OF A GOOD SAMPLE

1. It is free from error due to bias or due to deliberate selection of the units of the sample.
2. There is no substitution of originally selected unit by some other more convenient way.
3. It does not suffer from incomplete coverage of the units selected for study.
4. It includes such units, which are as far as possible independent.
5. It represents the population in the strict sense that it is a miniature or replica in all respects of the population from which it has been drawn. This should at least apply to the characteristics directly under the investigation or those likely to affect these characteristics indirectly.
6. It is adequate or sufficient in size to allow confidence in the stability of its characteristics. An adequate sample is one that contains enough cases to ensure reliable results.

Check Your Progress

Notes: a) Space is given below for writing your answers.
b) Compare your answers with those given at the end of the unit.

4. What do you understand by an incidental sample?

.....

.....

.....

5. Compare and contrast the purposive and quota sampling methods.

.....

.....

.....

6. State any five characteristics of a good sample.

.....

.....

.....

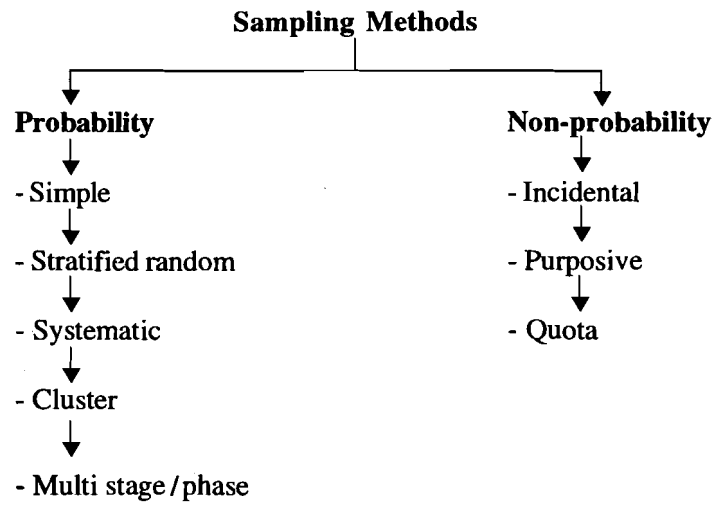
12.8 LET US SUM UP

In this Unit you got acquainted with the concept of population and sample. We elaborated the two types of sampling, namely, the probability and non-probability sampling.

We discussed in details various probability sampling methods including simple random sampling, stratified random sampling, systematic sampling, cluster sampling, and multi stage/ phase sampling.

You also got an explanation of the non-probability sampling methods including incidental sampling, purposive sampling and quota sampling.

At the end of the Unit we have tried to state the characteristics of a good sample.



12.9 UNIT-END ACTIVITY

- Get five Theses/Dissertations based on different research methods e.g. case study, survey, research etc. Study the sampling methods adopted in these studies and prepare a chart including the following columns – Research method, Sampling method, Sample size, rationale for adopting the sampling method.

12.10 POINT FOR DISCUSSION

- How do you determine the sample size when different research methods are adopted?

12.11 SUGGESTED READINGS

- Aggarwal, Y.P. (1998): *Better Sampling*. New Delhi: Sterling Publisher Private Ltd.
- Best, John W. and Kahn, James V. (1992): *Research in Education*. New Delhi: Prentice Hall of India Pvt. Ltd.

- Dalen, Van. Deobold, & Meyer, William. J. (1962): *Understanding Educational Research: An Introduction*. New York: McGraw Hill Book Company Inc.
- Kerlinger, Fred N. (1993): *Foundation of Behavioural Research : Educational and Psychological Enquiry*. New York: Holt, Rinehart and Winston.
- Koul, Lokesh. (1997): *Methodology of Educational Research*. New Delhi: Vikas Publishing House Pvt. Ltd.
- Upasani, K.N. (1987): *Conducting Educational Research*. Examination Reforms Unit, S.N.D.T. Women's Univerisity, Pune: Kalpana Mudranalaya,
- Vockell, Edward L. (1983): *Educational Research*. New York: MacMillan Publishing Co. Inc.

12.12 ANSWERS TO CHECK YOUR PROGRESS

1. *Population* – A group of individuals / units having one or more characteristics in common which are of interest to the researcher for a particular research.

   Sample – A small representative proportion of a population selected for a particular research.

   Probability Sampling – Sampling based on some statistical concepts such as the 'Law of Large Numbers', 'Central Limit Theorem', and the 'Normal Distribution' is known as probability sampling.

   Non-probability Sampling – Sampling based on the judgements of the researcher as the most important element of control is known as non-probability sampling.

2. I) i) – (b) II) i) – (c)
   ii) – (e) ii) – (a)
   iii) – (a) iii) – (b)
   iv) – (c) iv) – (d)

3. The main distinction between multi-stage and multi-phase sampling is the use of unit of sampling at different levels. In multi-stage, sampling is done at various levels such as national, state, district level. In multi-phase, sampling units are of the same type at each phase only a few of them are asked for more information than others.

4. When a readily or easily available group is selected as per the convenience of the researcher, it is termed as the 'incidental sample'.

5. i) *Similarity* – Purposive and quota sampling both include stratification.

   ii) *Difference* – In purposive sampling actual selection of the units from a stratum, to be included in the sample is done purposively rather than by random methods.

   In quota sampling quota is usually determined by the proportion of the strata and quota within the strata is selected as per the availability and convenience and not randomly.

6. Characteristics of a good sample –

   i) Free from error due to bias or deliberate selection of some units.

   ii) Originally selected units are not substituted and incomplete coverage of units is not involved.

   iii) As far as possible independent units are included.

   iv) It is a smaller image of the population.

   v) It is adequate in size.

12.13 GLOSSARY

Parameter : It is a population value representing any trait or characteristic of the population as a whole.

Sampling : It is the process of selecting a sample from the population.

Sampling Bias : When the mean of the sampling distribution does not coincide with the parameter, it is said to be biased.

Sampling Distribution : It is the relative frequency distribution of an infinity of determinations of the value of this statistic.

Sampling Error : It is the difference between the value of parameter and that of corresponding statistic or the difference between population value and sample value of a characteristic.

Sampling Frame : It is a complete, accurate, and up-to-date list of all the units in a population.

Statistic : It is the sample value representing any trait or characteristic of the members of the sampling.