UNIT 6 METHODS OF SAMPLING

Structure

6.1 INTRODUCTION

In the last unit we learnt about the issues to be considered in the design and conduct of research studies. Subjects included in these study almost invariably are a sample. Suppose as a researcher, we want to study diarrhoea among children in our country, the target population could be all existing cases of diarrhoea in the country. However, cost and logistic considerations will seldom allow us to study all the subjects. Sampling therfore becomes a natural choice.

What is sampling? What does the process involve? The process of selection demands thorough knowledge of various sampling techniques and data gathering tools. In this unit, we shall familiarize you with the concepts of sample, population and other concepts linked with sampling. We shall also discuss the characteristics of a good sample and the various methods of sampling.

Objectives

After studying this unit, you will be able to:

- define the terms, population, sample and other concepts linked with sampling i.e. sample frame, sample size, sample error etc.,
- discuss the steps in the sampling process and the various methods of sampling,
- define and describe the various types of probability sample,
- define and describe the various types of non-probability sample, and
- enumerate the characteristics of good sample.

6.2 CONCEPT OF SAMPLING

In this section we will first define and describe some terms specially used in the context of sampling. These terms include sample, sampling unit, sampling frame, sample size, parameters and statistics. We begin with an understanding about the concept of sampling.

A "sample" is a miniature representation of and selected from a larger group or aggregate. In other words, the sample provides a specimen picture of a larger whole. This larger whole is termed as the "population" or "universe". In research, this term is used in a broader sense; it is a well defined group that may consist of individuals,

objects, characteristics of human beings, or even the behaviour of inanimate objects, such as, the throw of a dice or the tossing of a coin.

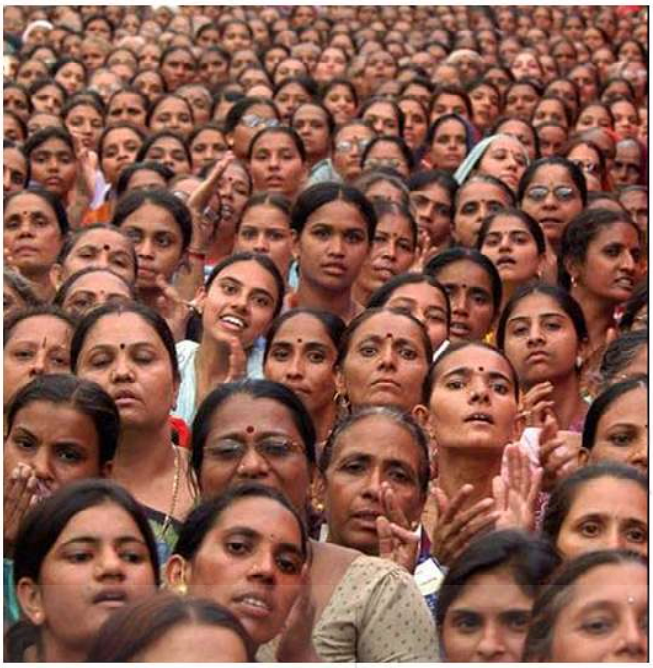

Figure 6.1: Population

It is not possible to include all units of a population in a study in order to arrive at a valid conclusion. Moreover, the sizes of populations are often so large that the study of all the units would not only be expensive but also cumbersome and time consuming. For example, there are more than fifty thousand undergraduate students in IGNOU. For our research, it is impossible to collect information about the study habits of all these students. So, a researcher will have to select a representative few, i.e., a sample from the population of the survey. This process is known as sampling.

Representativeness and Adequacy

Basically there are two requirements of a sample: it has to be 'representative' and 'adequate'. If the nature of the population has to be interpreted from a sample, it is necessary for the sample to be truly representative of the population. Moreover, it calls for drawing a representative 'proportion' of the population. The population may contain a finite number of members or units. Sometimes, the population may be 'infinite' as in the case of air pressure at various points in the atmosphere. Therefore, a population has to be defined clearly so that there is no ambiguity as to whether a given unit belongs to the population or not. Otherwise, a researcher will not know what units to consider for selecting a sample. For example, we want to understand the infant feeding habits of mothers. Here, the population is not well defined: we are not told about the background of the mothers that have to be included in this survey. After all, may be mothers from rural, urban, slum background belonging to different socio-economic groups adopting different practices. Hence, to define it accurately, we have to specify the group as, say, urban slum mothers etc. This would be the *sampling unit*.

The *unit of inquiry* is the subject on which the information is obtained. In other words, the sampling unit is one which is used for selection. For example, in a community survey on stunting among children, the sampling unit could be a family but the unit of enquiry could be a child less than 5 years of age. Note, one sampling unit can have one or two units of enquiry.

The second issue related to the representation of a sample is to decide about the '*sampling frame*', i.e., listing of all sampling units in the target population. In the above study on stunting, there can be different sampling frames, such as educated/uneducated

mothers, employed/unemployed mothers in the family, etc. The sampling matters should be complete, accurate and up-to-date exhaustive list, and must be drawn before selecting the sample. Preparation of frame requires precise definition of the unit, as well as, the population.

Thirdly, a sample should be unbiased and objective. Ideally, it should provide all information about the population from which it has been drawn. Such a sample, based on the logic of induction, i.e., proceeding from the particular to the general, falls within the range of random sampling errors. This leads us to the results expressed in terms of "probability". Let us understand the concept in greater details.

We know that research studies are always conducted on a fraction of subjects and all subjects are never observed. The knowledge that chills, fever and splenomegaly are common in malaria for example is based on what has been observed in cases over a period. But the cases actually observed or studied never comprise all that occur in the world. Only a fraction of these cases, called sample, are studied. The totality of all subjects or units from which a sample is selected is called *population* in statistical terminology. This also is the target group to which the findings would ultimately apply. We may have population of school going children, population of mentally retarded children, population of blood samples, or simply population of persons residing in an area. One feature of samples is that they tend to provide different picture in repeated sampling. This is called *sampling fluctuation* or *sampling error*. The magnitude of this "error" depends primarily on three factors: (i) the variability between the subjects in the population. The larger the variability, the more is the sampling error; (ii) the size of sample. When the samples include large number of subjects, the picture obtained from one sample is not likely to be very different from another sample of the same size because both tend to be fair representatives. This can not be said for small samples; and (iii) the method of sampling. The subjects should be selected in a manner that a wide spectrum could get adequate representation. Note that sampling "error" is not a mistake but signifies only variation from sample to sample.

We hope the requirement of a sample to be representative must be clear to you by now. A sample should not only provide representativeness, but should also be adequate enough to render stability to its characteristics. This brings us to the concept of sample size. The number of units or subjects sampled for inclusion in the study is called a *sample size*.

What, then, is the ideal size of a sample? We have already studied about the process of sample size determination in Unit 2 earlier. An adequate sample, we learnt, is the one that contains enough cases to ensure reliable results. If the population under study is homogeneous, a small sample is sufficient. However, a much larger sample is necessary, if there is greater variability in the units of population. Thus the procedure of determining the sample size varies with the objects of the study and the nature of the characteristics under study and their distribution in the population. Moreover, the adequacy of a sample will depend on our knowledge of the population, as well as, on the method used in drawing the sample. However, it should be understood that the adequate size of the sample does not automatically ensure accuracy of results.

Finally let us orient ourselves to the concept of parameters and statistics another concept used in the context of sampling.

Parameter and Statistic: Most nutritional studies draw conclusions on the basis of the estimates of either means or proportions. Such quantities of interest are called *parameters* when calculated for the entire population and *statistics* (as plural) when calculated for the sample. The objective of the sampling is to provide statistics which are adequate estimates of the parameters.

Let us take a break here and answer the question included in check your progress exercise 1.

Check Your Progress Exercise 1

1) Define the following terms

a) Sampling: .....

.....

b) Sampling frame: .....

.....

c) Sample size: .....

.....

Next, we shall focus on methods of sampling.

6.3 METHODS OF SAMPLING

In the last section, we suggested that the method used for drawing a sample is significant to arrive at dependable results or conclusions. With this fact in view, here in this section, we shall now talk about the various sampling methods. Sampling methods can be broadly classified into two categories.

i) Probability or a Random Sampling
ii) Non-probability Sampling

A brief review of these sampling methods follows.

6.3.1 Probability Sampling

Probability sampling is based on random selection of units from a population. In other words, the sampling process is not based on the discretion of the researcher but is carried out in such a way that the probability of every unit in the population of being included is the same by starting with a complete sampling frame of all eligible individuals. For example, in the case of a lottery, every individual has equal chance of being selected. Some of the characteristics of a probability sample are:

i) each unit in the sample has some probability of entering the sample,
ii) weights appropriate to the probabilities are used in the analysis of the sample, and
iii) the process of sampling is automatic in one or more steps of the selection of units in the sample.

Probability sampling can be done through different methods, each method having its own strengths and limitations. A brief account of these is given below:

a) ***Simple or Unrestricted Random Sampling***

Simple random sampling is a method of selecting a sample from a finite population in such a way that every unit of the population is given as equal chance of being selected [see item (i) above]. In practice, you can draw a simple random sample unit through the following steps:

i) define the population
ii) make a list of all the units in the population and number them from 1 to n.
iii) decide the size of the sample, or the number of units to be included in the sample.

iv) use either the 'lottery method' or 'random number tables' to pick the units to be included in the sample.

For example, you may use the lottery method to draw a random sample by using a set of 'n' tickets, with numbers '1 to n' if there are 'n' units in the population. After shuffling the tickets thoroughly, the sample of a required size, say x, is selected by picking the required x number of tickets. The units which have the serial numbers occurring on these tickets will be considered selected. The assumption underlying this method is that the tickets are shuffled so that the population can be regarded as arranged randomly. Similarly, while selecting 500 students from the total population of 50,000 undergraduate students of IGNOU, you will write the roll numbers of all the students on small pieces of paper. Jumble the chits well and then choose five hundred roll numbers.

The best method of drawing a simple random sample, however, is to use a *table of random numbers*. These random number tables have been prepared by (Fisher and Yates, 1967). After assigning consecutive numbers to the units of population, the researcher starts at any point on the table of random numbers and reads the consecutive numbers in any direction horizontally, vertically or diagonally. If the read out number corresponds with the one written on a unit card, then that unit is chosen for the sample.

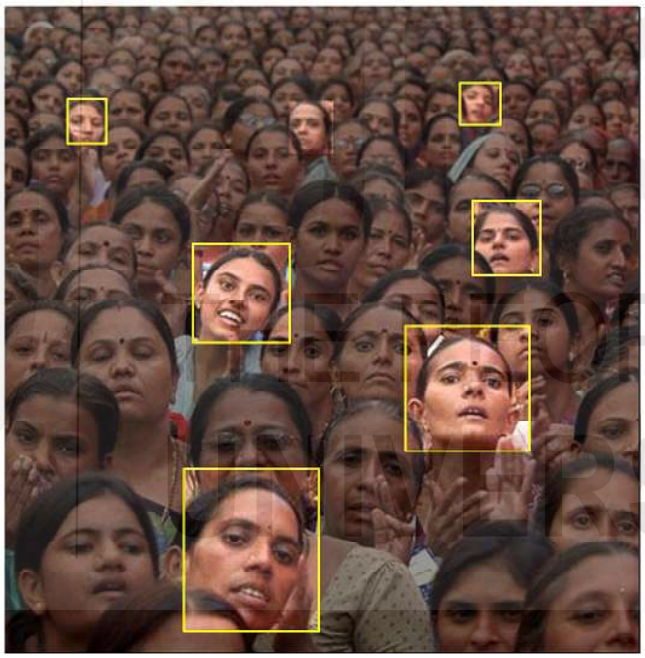

Figure 6.2: Probability sample

Let us, suppose that a sample of 5 health centers is to be selected at random from a serially numbered population of 60 health centers. Using a part of a table of random numbers reproduced here in Table 6.1, five two digit numbers (as the total population of study centers, 60, is a two digit figure) are selected.

If you start with the first row and the first column, 23 is the first two-digit number, 05 is the next number and so on. Any point can be selected to start with the random numbers for drawing the desired sample size. Suppose the researcher selects column 1 from row 4, the number to start with 83. In this way he/she can select first 5 numbers from the column starting with 83.

The sample, then is as follows:

~~83~~	~~75~~
53	33
40	01
05	26

Let us consider how we selected these numers. In selecting the sample of 5 study centers, two numbers, 83 and 75, need to be deleted as they are bigger than 60, the size of the population. The processes of selection and deletion are stopped after the required number of five units get selected.

Table 6.1: An abbreviated table of random numbers

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

The selected numbers are 53,40,05,33 and 01. If any number is repeated in the table, it may be substituted by the next number from the same column. The researcher will go on to the next column until a sample of the desired size is obtained.

From the example above, it must be evident that when the scheme is such that each unit of the population has the same chance of being included in the sample, it is called *simple random sampling*. Simple random sampling, ensures the best results. However, from a practical point of view, a list of all the units of a population is not possible to obtain. Even if it is possible, it may involve a very high cost which a researcher or an organization may not be able to afford. Therefore, simple random sampling is difficult to realize. Also, in case of a heterogeneous population, a simple random sample may not necessarily represent the characteristics of the total population, even though all selected units participate in the investigation. Thus a disadvantage with simple random sampling is that it does not guarantee adequate representation of different segments of the target population. When adequate representation of various segments is required, the method of selection should be stratified sampling as described next.

b) ***Stratified Sampling***

Stratified random sampling takes into account the stratification of the main population into a number of sub-populations, each of which is homogenous with respect to one or more characteristic(s). Having ensured this stratification, it provides for selecting randomly the required number of units from each sub-population. The selection of a sample from each sub-population may involve random or any other mode of selection. The steps involved in the stratified sampling are given as follows:

i) Deciding upon the relevant stratification criteria such as sex, geographical region, age, courses of study, etc.

ii) Dividing the total population into sub-populations based on the stratification criteria.

iii) Listing the units separately in each sub-population.

iv) Selecting the requisite number of units from each sub-population by using an appropriate random selection technique.

v) Consolidating the sub-samples for making the main sample.

Thus, stratification improves the representativeness of a sample by introducing a secondary element of control. However, the efficiency of the stratified random sample depends on the allocation of sample size to the strata. Rendering proportional weight-age to each criterion improves it further by allowing the use of a smaller sample and by helping in achieving higher efficiency at a reduced cost.

Stratified random sample is very useful when lists of unit or individuals in the population are not available. It is also useful in providing more accurate results than simple random sampling. For example, in a study on malnutrition, as assessed by weight for age, as predictor for death in children with specified infections, it is necessary that children of different nutritional status (such as normal, moderate and severe malnutrition) are included in the sample. The procedure for sampling therefore should be to first divide the subjects in the frame by nutritional status (i.e. normal, moderate, malnutrition) and then draw independent random samples of size, say 20 from each division. Such division of the frame is called *stratification* and each division is a *stratum*. The investigator can decide how many units are to be selected from different strata.

Sometimes stratification is not possible before collecting the data. The stratum to which a unit belongs may not be known until the researcher has actually conducted the survey. Personal characteristics such as sex, social class, educational level, age etc., are examples of such stratification criteria. The procedure in such situations involves taking of a random sample of the required size and then classifying the units into various strata. The method is quite efficient provided the sample is reasonably large. i.e., more than 20 in every stratum.

Next, let us get to know about systematic sampling, yet anohter sampling technique.

c) ***Systematic Sampling***

Systematic sampling provides a more even spread of the sample over the population list and leads to greater precision. Systematic random sampling is therefore a sampling method in which the first unit is randomly selected and the others automatically included on the basis of the sampling interval. The process involves the following steps:

i) Make a list of the population units based on some order – alphabetical, seniority, street number, house number or any such factor.

ii) Determine the desired sampling fraction, say 50 out of 1000; and also the number of the K^{th} unit. [K = N/n = 1000/50 = 20].

iii) Starting with a randomly chosen number between 1 and K, both inclusive, select every K^{th} unit from the list. Here our k^{th} unit is 20. Therefore, select every 20th member from the list starting from a randomly chosen number. If in the above example the randomly chosen number is 4, the sample shall include the 4^{th}, 24^{th}, 44^{th}, 64^{th} units in each of the series going upto the 984^{th} unit.

This method provides a sample as good as a simple random sample and is comparatively easier to draw. For example, patients coming to a diabetic clinic, every K^{th} after the random first can be easily selected for inclusion in the study. However, this method suffers from the following drawbacks because of departure from randomness in the arrangement of the population units.

i) *Periodic effects*: Populations with more or less definite periodic trend are quite common. Patients attendance at out patient department (OPD) over seven days in a week or mothers' attendance at a health clinic over seven days in a week, sales of a store over twelve months in a year and flow of road traffic past a particular traffic point on a road over 24 hours area a few examples to show periodic trend or cycle fluctuation in a given population. In such cases systematic sample may not represent the population adequately or remain effective all the time.

ii) *Trend*: Another handicap of systematic sampling emerges from the fact that very often 'n' is not an integral multiple of 'k'. This leads to a varying number of units in the sample from the same finite population.

Suppose a population of 100 counselors is listed according to seniority and a researcher wants to select a sample of 20. First he/she divides 100 by 20 to get 5 as the size of the interval. Suppose he/she picks 4 at random from 1 to 5 as a starting number. Then, he/she selects each 5^{th} name at 9,14,19,… until he/she draws the desired 20 names. If he/she picks 2 as the starting point, another sample would consist 2,7,12,… In the latter sample each counselors seniority is lower than his/her counterpart in the former sample. The mean average of these two samples would be significantly divergent as regards seniority and other associated variables. Many such samples can be drawn by taking different starting points but there will be greater variation among them.

Thus, the 'periodic effects' and 'trend' of the listed population unduly increase the variability of the samples, and calculations made from such samples cannot show the sources of variability. The main advantages of systematic sampling are:

a) It involves simple calculations.

b) It is less expensive than random sampling.

Like simple random sampling, systematic random sampling also fails to give adequate representation to different segments of the population.

Next, let us get to learn about the cluster sampling.

d) ***Cluster Sampling***

Cluster sampling is used when the population under study is infinite, where a list of units of population does not exist, when the geographic distribution of units is scattered, or when sampling of individual units is not convenient for several administrative reasons. It involves division of the population of elementary units into groups or clusters that serve as primary sampling units. A selection of the clusters is then made to form the sample. Thus, in cluster sampling, the sampling unit contains groups of elements or clusters instead of individual members or items in the population.

For example, for the purpose of selecting a sample in a study to assess the prevalence of goiter in school children of a state, a researcher must enlist all schools in various

public health centre (PHC) areas, urban blocks, census town etc. and select randomly a 10 per cent sample (say) of the schools or clusters of units. She/he either uses all the children as the sample or randomly select a few of them. Thus each school constitutes a primary sampling unit. The children in each primary unit are the final sampling units in this prevalence study. Thus if *n* clusters out of N are randomly selected, all subjects in the selected cluster are investigated. This is called cluster random sampling. Cluster random sampling is sometimes considered a rapid assessment method. The World Health Organization (WHO) recommends this kind of sampling to estimate the percentage of children immunized in a community.

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

Besides the sampling techniques described above, we will focus here on two other sampling techniques: multi-stage sampling and probability proportion to size sampling.

e) ***Multi-stage Sampling***

Multi-stage sampling is used in large scale surveys for a more comprehensive investigation. The researcher may have to use two, three or even four stage sampling. For example, in surveys mailed questionnaires are generally used to gather information from people living in widely scattered areas. Although the method is cost effective, partially completed questionnaires may introduce a bias due to which a representative sample cannot be obtained. To overcome this bias, two-stage sampling has to be used. A second sample from non-respondents is selected at random by contacting them personally. In this way the consistency of the data obtained from the first sample can also be verified. Similarly, if a researcher goes for a national survey of anganwadi workers, he/she can draw a sample of five states representing northern, eastern, southern, western and central regions. From these five states, all the districts can be enlisted out of which a sample of 30 to 40 districts can be drawn randomly. Out of this, all the ICDS projects in different districts can be enumerated. A random sample of about 300 to 400 ICDS projects is then drawn. Further, a random sample of about 1500-2000 anganwadi workers are drawn for the survey. The successive random sampling of states, districts, ICDS projects and finally anganwadi workers also provides a multi-stage sample.

Multi-stage sampling is advantageous as the burden on the respondents is lessened, it is cost effective, time saving and efficient in formulating the sub-sample data. However, this method is recommended only when it seems impractical to draw a simple random sample.

Finally, a word about probability proportion of size sampling.

f) ***Probability Proportion to Size Sampling***

When the units vary in size, it is better to select a sample in such a way that the probability of selection of units is proportional to its size. For example, a particular community has a population of 200 infants and another one has 100. While drawing a sample, the first community will have double the representation as compared to the second community. Such a sample is known as probability proportion to size sample or PPS sample.

With this we end our study of probability or random sampling. Next, we shall move on to non-probability sampling methods, but first we shall try to answer the question included in check your progress exercise and assess our understanding of the topic discussed so far.

Check Your Progress Exercise 2

1) List the various types of probability sampling.

.....

.....

.....

2) How is stratified random sampling different than simple random sampling?

.....

.....

.....

Now, let us study about non-probability sampling.

6.3.2 Non-probability Sampling

Non-probability sampling is based on the judgment of the researcher. The guiding factors in non-probability sampling include the availability of the units, the personal experience of the researcher and his/her convenience in carrying out a survey. Since these samples are not prepared through random sampling techniques, they are known as non-probability samples. Depending on the technique used, non-probability samples are classified into *purposive, incidental* and *quota samples*. A brief description of these samples is given below.

a) ***Purposive Sample***

Purposive sampling is one in which no random component is present and specific subjects are intentionally included. Volunteers are an example.

A purposive sample is also known as a judgment sample. This type of sample is chosen because there are good reasons to believe that it is a representative of the total population. This also reflects certain controls identified as representative areas like a city, state or district; representative characteristics of individuals like age, sex, marital status, etc; or types of groups like school administrators, elementary school teachers, secondary school teachers, college teachers, etc.

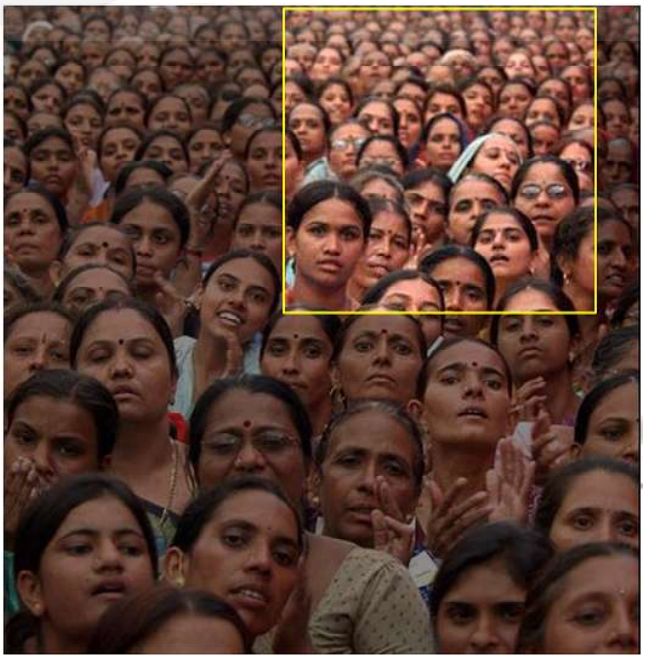

Figure 6.3: Non-probability sample

A purposive sample differs from stratified random sample in that the actual selection of the units to be included in the sample in each group is done purposively rather than by random methods. This approach comes in handy where it is necessary to include a very small number of units in the sample. For example, for study of 'gifted' children, the researcher, on the basis of his/her past experience, selects certain individuals giving extra ordinary performance in school while excluding all others from the sample.

Next, let us learn about incidental sample.

b) ***Incidental Sample/Convenience Sample***

The term incidental sample, also known as accidental sample, is applied to samples that have been drawn because of the easy availability of units. An investigator employed in IGNOU for example may select learners enrolled for the Master's Programme in Dietetics and Food Service Management while conducting a study on higher education, as these learners are readily available and fulfill the conditions of the study. But, neither of the two reasons may be of the investigator's choice. Therefore, such casual groups rarely constitute random samples of any definable population.

The merits of this procedure are mainly the convenience of obtaining units, the ease of testing and completeness of the data collected. However, it is the limitations that have defined population and no randomization has actually been done. Therefore, any attempt to arrive at generalized conclusion in such cases will be erroneous and misleading.

Lastly let us focus on quota sample.

c) ***Quota Sample***

Quota sample is another type of non-probability sample. It involves the selection of sample units within each stratum or quota on the basis of the judgment of the researcher rather than on calculable chance of the individual units being included in the sample. Suppose a national survey has to be done on the basis of quota sampling. The first step in quota sampling would be to stratify the population region wise like rural/urban, administrative districts etc. and then fix a quota of the sample to be selected. In the initial stage quota sampling is similar to stratified sampling. However, it may not necessarily employ random selection procedure in the initial stage in exactly the same way as probability sampling. The essential difference between probability sampling and quota sampling lies in the selection of the final sampling units. The quota is usually determined by the proportion of the groups. Suppose a researcher wants to study the attitude of university teachers towards distance education. First of all, he/she may stratify the university teachers in the category of sex and then as professors, readers and lecturers. Later, he/she may fix quota for all these categories. In this way, the quota sample would involve the use of strata but selection within the strata is not done on a random basis.

The advantages of quota sampling are, its being less expensive, convenient, and more suitable in the case of missing or incomplete sampling frames.

The non-probability samples are generally considered to be convenient when the sample to be selected is small and the researcher wants to get some idea of the population characteristics within a short time. In such cases, the primary objective of the researcher is to gain insight into the problem by selecting only those persons who can provide maximum insight into the problem.

However, the following are some inherent limitations of non-probability sampling methods:

i) No statistical theory has been devised to measure the reliability of results derived through purposive or other non-random samples. Hence, no confidence can be placed in the data obtained from such samples and results cannot be generalized for the entire population.

ii) The selective sampling based on convenience affects the variance within the group, as well as, between the groups. Further, there is no statistical method to determine the margin of sampling errors.

iii) Sometimes such sample are based on an obsolete frame which does not adequately cover the population.

d) ***Snowball Sample***

This type of sampling is used when there are not enough potential participants or they are hard to find. This is also known as chain referral sampling in which each research participant recruit other participant from their familiarity or known circle suited for the study. Thus, this sample group is said to grow like a rolling snowball. For example, if a researcher wants to enroll diabetic patients in the study, he/she will takeup a diabetic patient as a subject and would ask him, if he knows any other subject/ subjects having diabetes. He would now enroll the referred subjects and further asks them for more subjects. Likewise, the chain keeps on increasing till the desired number is met.

With snowball sample we end our study of the non-probability sampling methods used in research. We hope you are now familiar with the different sampling methods and the situations when to use these methods. The next sub section also highlights the considerations basic to choice of sampling methods.

6.3.3 Choice of the Sampling Method

The choice of sampling methods depends on several considerations unique to each individual project. These include issues related to the definition of population, availability of information about the structure of the population, the parameters to be estimated, the objectives of the analysis including the degree of precision required, and the availability of financial and other resources. This calls for appropriate selection of a sample for the conduct of any research study. The characteristics of a good sample are enumerated next.

6.3.4 Characteristics of a Good Sample

A good sample should have the characteristics of (i) Representativeness and (ii) Adequacy, as already described earlier in the unit.

It is essential that the sample should be 'representative' of the population if the information from the sample is to be generalized for that population. The term representative sample means an ideal 'miniature' or 'replica' of the population from which it has been drawn.

A good sample should also be 'adequate' or of sufficient size to allow confidence in the stability of its characteristics. An adequate sample is considered to be one that contains enough cases to ensure reliable results. Hence, planning the size of the sample in advance is very important. It varies with the nature of the characteristics under study and its distribution.

It may be mentioned that representativeness and adequacy do not automatically ensure accuracy of results. The sampling and data collection techniques need to be selected and employed carefully to obtain higher degrees of precision in results and generalizations about the population.

Check Your Progress Exercise 3

1) Describe the various types of non-probability sample.

.....

.....

2) Discuss the characteristics of a good sample.

.....

.....

.....

6.4 KEY POINTS AT A GLANCE

The basic concepts discussed in this unit are summarized herewith:

1) A population is a well defined group of units: individuals, objects, attributes, qualities, characteristics, traits of human beings, etc.

2) A sample is a small representation of a population. It is a miniature picture of the entire group from which it has been selected.

3) To obtain a representative sample, you must select the unit in a specified way. This process is called sampling. It usually involves the following four steps: (i) Defining the population; (ii) Listing the population; (iii) Selecting a representative sample; and (iv) Obtaining an adequate sample.

4) Sampling methods can be classified into two broad categories: (i) Probability sampling and (ii) Non-probability sampling.

5) In probability 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 a population has the same probability of being included in the sample.

6) Sample or unrestricted random sampling, systematic sampling, stratified sampling, cluster sampling, multi-stage sampling and probability proportion to size sampling are the six main types of probability sampling. In all these types each unit in the sample has some known probability of entering the sample.

7) In simple or unrestricted random sampling each unit of the population is given an equal chance of being selected, and the selection of any one unit is in no way tied to the selection of any other. The law of chance is allowed to operate freely in the selection of such samples and carefully controlled conditions are created to ensure that each unit in the population has an equal chance of being included in the sample.

8) The researcher may use the lottery method or table of random numbers for drawing a simple random sample.

9) Simple random sampling ensures best results. However, it is neither feasible nor possible if the lists of the units do not exist or if such lists are incomplete.

10) If there is more heterogeneity among the units of the population, a simple random sample may not necessarily represent the characteristics of the total population even if all selected units participate in the investigation.

11) In systematic sampling, a researcher generally starts with a list in which all the N units of the population are listed in alphabetical or in any other order. To select a sample of size n, the researcher has to select a unit at random from the first $k = (N/n)$ units of the list and then every subsequent k^{th} unit is selected.

12) A systematic sample is as good as a simple random sample and is comparatively more convenient to draw. However, the characteristics of "trend", "cyclical fluctuations" and "periodic effects" of a listed population unduly increase the variability of samples.

13) When the units in a sample are proportional to their presence in the population, the sampling is said to be stratified.

14) When a population is stratified, the units within each stratum are more or less homogenous than the units within the entire population.

15) Stratified random sampling is very useful when lists of units in the population are not available. This method has been found practical even for small finite populations when cent per cent response is difficult to secure within the desired time.

16) Stratified random sampling provides more accurate results than simple random sampling only if stratification results in greater homogeneity within the strata than in the whole population taken as one unit. It is particularly useful in opinion survey studies.

17) Cluster sampling is used when the population under study is infinite, where a list of units of the population does not exist, when the geographical distribution of units is scattered, or when sampling of individual units is not convenient for various practical purposes.

18) Cluster sampling involves division of the population of elementary units into groups of elements or clusters instead of individual members or items in the population.

19) Cluster sampling is economical, especially when the cost of measuring a unit is relatively small and cost of reaching it is relatively large.

20) Multi-stage sampling is used in large scale surveys for a more comprehensive investigation. In this type of sampling, the researcher may have to use two, three or even four stages of sampling.

21) Multi-stage sampling is comparatively convenient, less time consuming and less expensive. However, an element of sample bias gets introduced because of the unequal size of some of the selected sub-samples.

22) When the units vary in size, it is better to select a sample in which the probability of selection of a unit is proportional to its size. This sample is known as probability proportion to size sample or PPS sample.

23) Non-probability sampling is based on the judgment of the researcher. Its guiding principles are: (i) availability of sampling units; (ii) personal experience of the researcher, and (iii) the researcher's convenience in conducing the research. Since this type of sampling does not involve the principle of probability, it is called non-probability sample.

24) Non-probability sampling provides (i) purposive samples, (ii) incidental samples, and (iii) quota samples.

25) A purposive sample is arbitrarily selected because there is good evidence that it is a representative of the total population. The evidence is based on researcher's experience.

26) An incidental sample is generally used with those groups which are selected because of the easy or ready availability of sample units.

27) A quota sample involves selection of the sample units within each stratum or quotas on the basis of the judgment of the researcher rather than on calculable chance of being included in it.

28) Non-probability samples are very convenient in situations where the sample to be selected is very small and the researcher wants to get some idea of the characteristics of a population in a shorter time.

29) Non-probability samples have certain limitations. No valid generalizations can be made beyond the sample studied. These samples depend exclusively on uncontrolled factors and the researcher's insight. Hence, the sampling error of such samples is hardly determinable.

30) The choice of an appropriate sampling method by a researcher depends upon many factors. These include (i) defining the population, (ii) availability of information about the structure of population, (iii) the parameters to be estimated, (iv) the objectives of the analysis including degree of precision required, and (v) the availability of financial and other resources.

31) Representativeness and adequacy are the major characteristics of a good sample.

6.5 LET US SUM UP

In this unit we discussed the concept of population and sample, and the two methods of sampling, namely, probability and non-probability sampling.

Under 'probability sampling' we discussed it various types such as simple sampling or unrestricted random sampling, systematic sampling, stratified sampling, cluster sampling and multi-stage sampling.

Under 'non-probability' sampling we discussed purposive sample, incidental sample, quota sample and also touched upon the choice of sample.

We ended this unit with a description of the characteristics of a good sample: representativeness and adequacy.

6.6 GLOSSARY

Population : a population is any group of individuals or units that have one or more characteristics in common and are of interest to the researcher. It may consists of all the units or individuals of a particular type or a more restricted part of that group.

Sample : a sample is a small proportion of a population selected for analysis. By observing the sample, certain inferences may be drawn about the population. Samples are not selected haphazardly, but deliberately, so that the influence of chance or probability can be estimated.

Probability : probability is the ratio of the number of ways in which a favoured way can occur to the total number of ways the event can occur. It may range from zero, when there is no chance whatever, of the favoured event, to 1.0, where there is absolute certainty that nothing else could happen.

Probability Sampling : in probability sampling, the units of a population are not selected at the discretion of the researcher but by means of certain procedures which ensure that every unit of the population has one fixed probability of being included in the sample. It is a procedure of drawing the units of a population in such a way that every unit has an equal and independent chance of being included in the sample.

Non-probability Sampling: in non-probability sampling, the units are selected at the discretion of the researcher. The researcher uses his/her judgment or experience while selecting the sample.

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

Representative Sample : A representative sample is one that matches with its corresponding population with respect to the characteristics important for the research.

Parameter : measures which describe a population are called parameters.

Statistics : the measures estimated from the samples are called statistics.

Sampling Error : the 'statistics' estimated from samples tend to differ more or less from sample to sample drawn from the same population due to sampling fluctuations. On the other hand, the 'parameter' is considered to have a fixed reference value. It is not possible to compute parameter but there is a statistical procedure to forecast the parameter from sample statistics provided certain conditions have been satisfied. The difference between the sample estimate (statistics) and the population value (parameter) is called the 'sampling error'.

Biased Sample : a sample that is not representative is known as a biased sample. Biases may be due to imperfect tools or instruments, personal qualities of the researcher, defective techniques or other causes.

6.7 ANSWERS TO CHECK YOUR PROGRESS EXERCISES

Check Your Progress Exercise 1

1) a) Sampling is the process of selecting a sample which is a small representation of a large whole or group. A sample should represent truly and adequately the larger whole. A sampling frame should be complete, accurate, up-to-date, unbiased and objective.

   b) List of all sampling units in the target population is called a sampling frame.

   c) The number of units or subjects sampled for inclusion in the study in called sample size.

Check Your Progress Exercise 2

1) Sampling or unrestricted random sampling, systematic sampling, stratified sampling. cluster sampling, multi-stage sampling and probability proportion to size sampling (PPS) are the important types of probability sampling.

2) Stratified random sampling takes into account the stratification of main population into a number of sub-population and then draw independent simple random samples from each division.

   On the other hand when the scheme is such that each unit of the population has the same chance of being included in the sample, it is called simple random sampling.

Check Your Progress Exercise 3

1) Non-probability sampling includes (i) purposive sample, (ii) incidental sample and (iii) quota sample.

   - Purposive sample is useful when we have to include a very small number of units in a given sample.
   - Incidental sample is generally applied in the case where units are easily or readily available.

- Quota sample involves the selection of sample units within each stratum or quota on the basis of the judgment of the researcher. Sample units are not included accidentally or at random.

2) A good sample must be
   - representative of the population chosen
   - adequate and
   - accurate