14

SAMPLING

CHAPTER PREVIEW

- Meaning and Types of Sampling
  - A. Probability Sampling Methods
  - B. Nonprobability Sampling Methods
- Need for Sampling
- Fundamental of Sampling
- Factors Influencing Decision to Sample
- Methods of Drawing Random Samples
- Simple Random Sample
- Stratified Random Sample
  - A. Proportionate Stratified Random Sampling
  - B. Disproportionate Stratified Random Sampling
- Area or Cluster Sampling
- Quota Sampling
- Purposive or Judgemental Sampling
- Accidental Sampling
- Systematic Sampling
- Snowball Sampling
- Saturation Sampling and Dense Sampling
- Double Sampling
- Requisites of a Good Sampling Method
- Common Advantages of Sampling Methods
- Sampling Distribution
- Sampling Error

MEANING AND TYPES OF SAMPLING

Nearly all researches—experimental and nonexperimental—in the behavioural sciences, particularly in the fields of psychology, sociology and education—draw some inferences regarding a well-specified and identifiable group on the basis of some selected measures. The well-specified and identifiable group is known as a *population* or *universe* and the selected number of persons or objects is known as a *sample*. The generalized conclusions are technically known as the *statistical inferences*. A population, therefore, may be defined as any identifiable and well-specified group of individuals. All primary school teachers, all college teachers, all university students, all housewives, etc., are examples of populations. A population may be *finite* or *infinite*—a finite population is one where all the members can be easily counted; an infinite population is one whose size is unlimited and therefore, its members cannot be counted. The population of university teachers is an example of a finite population and the population of fishes

in a river is an example of an infinite population because the former can be counted whereas the latter cannot be counted.

Likewise, a population may be *real* or *imaginative*—a real population is one which actually exists and an imaginative population is one which exists only in the imagination. In psychological and educational research, on many occasions, the population is imaginative. A measure based upon the entire population is called a *parameter.*

A sample is any number of persons selected to represent the population according to some rule or plan. Thus, a sample is a smaller representation of the population. A measure based upon a sample is known as a *statistic.*

Before we define the different types of samples (or the methods of sampling), it is essential to define the term `probability', which is the base of sampling theory. The general meaning of probability is *less than certain* and for which there exists some evidence. In sampling theory the term `probability' is used as equivalent to the relative frequency. Thus when we say that the probability of a tail on a single toss of a coin is 1/2, it is meant that when we make several tosses, the relative frequency of a tail will be about 1/2 or 0.5. If one says that the probability of having of a male child is 0.8, it is meant that on previous occasions the relative frequency of the birth of a male child has been 0.8. Probability may be expressed in terms of a fraction or in decimal numbers.

Following Blalock (1960), most sampling methods can be categorized into two—

(A) Probability Sampling Methods

(B) Nonprobability Sampling Methods

A discussion of these two is given below.

A. Probability Sampling Methods

Probability sampling methods are those that clearly specify the probability or likelihood of inclusion of each element or individual in the sample. Technically, the probability sampling methods must satisfy the conditions given below.

(i) The size of the parent population or universe from which the sample is to be taken, must be known to the investigator.

(ii) Each element or individual in the population must have an equal chance of being included in a subsequent sample.

(iii) The desired sample size must be clearly specified.

If, for example, a researcher knows that the population which he is going to study contains 500 elements, or individuals, and if he knows that all the elements (or individuals) are accessible and may be included in a subsequent sample, it can be said that each element (or individual) in the population has an equal chance, that is, 1/500 of a chance of being selected. This constitutes the probability sampling method. In practice, however, sometimes researchers are not able to know for certainty that conditions (i) and (ii) will be satisfied. Sometimes the population studied is so large as to be considered infinite and unknowable for all important and practical purposes.

The positive point of the probability sampling method is that the obtained samples are considered representative, and hence, the conclusions reached from such samples are worth generalization and are comparable to similar populations to which they belong.

The negative point of the probability sampling method is that a certain amount of sampling error exists because the researcher has only a limited element of the entire population. Sampling error refers to the degree to which the sample characteristics approximate the characteristics of the parent population. The *smaller* the sample, the *greater* the sampling error.

The major probability sampling methods are the following:

1. Simple random sampling
2. Stratified random sampling
   (a) Proportionate stratified random sampling
   (b) Disproportionate stratified random sampling
3. Area or cluster sampling

B. Nonprobability Sampling Methods

Nonprobability sampling is one method in which there is no way of assessing the probability of the element or group of elements of the population being included in the sample. In other words, non- probability sampling methods are those that provide no basis for estimating how closely the characteristics of a sample approximate the parameters of the population from which the sample had been obtained. This is because nonprobability samples don't use the techniques of random sampling. Important techniques of nonprobability sampling methods are:

1. Quota sampling
2. Accidental sampling
3. Judgemental or purposive sampling
4. Systematic sampling
5. Snowball sampling
6. Saturation sampling
7. Dense sampling

A discussion of each of these sampling methods will be taken up in a separate section.

NEED FOR SAMPLING

Sampling is needed for a variety of reasons. Some of the important reasons are mentioned below:

(i) Sampling saves time as well as money. A research study based on sampling is completed within a lesser time and incurs less expenditure than a study based upon other criteria.

(ii) A research study based upon sampling is generally conducted by trained and experienced investigators. As such, it provides accuracy in measurement and testing.

(iii) Sampling is also needed because it enables the researcher to estimate the sampling errors and that way it helps in getting information regarding some characteristics of the population.

(iv) Sampling is needed because it remains the only way when a population/universe contains infinitely many members.

(v) Sampling helps in making correct and scientific judgement about the population for which generalization is to be made after the completion of the study.

FUNDAMENTALS OF SAMPLING

There are some fundamental concepts related to sampling that need to be explained. Such important fundamental concepts are *universe / population, sampling frame, sampling design, sampling error, sampling distribution, statistic (c)* and *parameter (s), confidence level* and *significance level.* The meaning of universe/population, sampling error and sampling distribution has been explained elsewhere in this chapter. Consequently, we shall discuss the meaning of only the remaining fundamentals.

(1) Statistic (c) and Parameter (s)

A statistic is a numerical value based upon sample, whereas parameter is the numerical value based upon population. For example, when we calculate mean from a sample, this is called

statistic because it describes the characteristics of a sample. When the same mean is calculated from population, it is called parameter because it describes the characteristics of population. Any sampling analysis aims to obtain the estimate of a parameter from a statistic.

(2) Sampling frame

The elementary units form the basis of the sampling process and such units are called sampling units. A list containing all such sampling units is called a *sampling frame.* In other words, it can be said that the sampling frame consists of a list of items from which sampling is to be finally drawn. In some cases it becomes impossible to draw a sample directly from a population. As such this frame is constructed by the researchers for the purpose of the study or it may consist of some existing list of population. For example, the researcher can use a telephone directory as a frame for conducting an opinion survey in a particular town.

(3) Confidence interval and Significance level

The confidence level is the expected percentage of times stipulating that the actual value will fall within the stated precision limits. For example, if we take a confidence interval of 95%, then obviously, it means that there are 95 chances in 100 that the sample results present the true characteristics of the population within a specified precision range against 5 chances in 100 that it does not. Thus confidence level indicates the likehood that the answer will fall within the precision range. Significance level, on the other hand, is the level that indicates the likehood that the answer will fall outside that range. If the confidence level is 95%, then the significance level is 100 – 95 = 5% or .05 and if the confidence level is 99%, the significance level is 100 – 99, that is, 1%. It would not be out of place to mention here that the area of the normal curve within precision limits for the specified confidence level constitutes the acceptance region, and the area the of curve outside these limits in either direction constitutes the rejection regions.

(4) Sampling design

By a sampling design we mean a plan for obtaining a sample from the sampling frame. In other words, it refers to the procedure the researcher would adopt in selecting some sampling units from which inferences about the population is drawn. A good sampling design must possess the following characteristics:

(i) The sample design must provide a sample which would be the true representative of the population

(ii) The sample design should be such that it may result in small sampling error.

(iii) The sample design must allow for controlling systematic bias in the sampling.

(iv) The sample design must be usable in light of the funds available for the research study.

(v) The sample design must be such that the results of the sample study can be applied in general, for the population with a reasonable confidence level.

There are two broad types of sample designs: *probability sampling* and *nonprobebility sampling.* These will be discussed in detail later on in this chapter.

FACTORS INFLUENCING DECISION TO SAMPLE

Any behavioural researcher takes some decision regarding the sampling plan at the time of formulating the research design. His decision to sample is influenced by at least three factors:

1. Size of the population
2. Cost involved in obtaining the elements
3. Convenience and accessibility of the elements

These three may be discussed as follows:

1. Size of the Population

Decision to sample is directly influenced by the size of the population. If the population is small, say, for example, it consists of 300 individuals, the investigator may decide to include all in his study and, therefore, sampling may not be done. On the other hand, if the size of the population is large, say, it consists of 10,000 individuals, he may decide to select a limited number of individuals from the population of 10,000 individuals. Obviously, then it can be said that as the population of elements (or individuals) becomes larger, sampling becomes increasingly very important. It may also be mentioned here that the size of the population is a relative matter. What one investigator regards as a large population, the other may regard it as a small one. In fact, no clear-cut guidance exists for making distinction between large population and small population.

2. Cost Involved in Obtaining the Elements

The investigator is also influenced by the cost likely to be incurred in obtaining the elements from the population. If sampling involves a bigger cost which the investigator can't meet, the decision to sample may be postponed. On the other hand, if sampling involves a cost which the investigator can readily meet, the sampling work is greatly facilitated.

3. Convenience And Accessibility of the Elements

The decision to sample is also influenced by the convenience and accessibility of the elements. Sometimes the investigator may have to deal with a problem with respect to which the elements may not be conveniently available. For example, if the problem deals with investigating the causes of premarital sex relationship among college girls, the investigator may not find it convenient to trace out such girls who could readily tell about their affairs openly. In such cases, sampling may be faulty. On the other hand, some investigators may have access to facilities and staff where a large amount of data could be easily handled. These investigators dare sample even in those cases where the resulting data are complicated and complex.

Obviously, decision to sample effectively is influenced by the size of the population, the anticipated cost of the study and the convenience and accessibility associated with the elements.

METHODS OF DRAWING RANDOM SAMPLES

Randomness is one of the vital points in sampling, particularly, probability sampling (Peatman, 1947). Randomness means that all elements in a given population have an equal and independent chance of being included in the sample. There are several methods through which the investigator maintains randomness in his sampling plans. Some of these methods are given below:

1. Fishbowl Draw Method
2. Using Table of Random Numbers
3. Method of Computer-determined Randomness

A discussion on these methods follows below:

1. Fishbowl Draw Method

This is a very simple method through which the elements from the population can be selected randomly. In this method the entire elements (or individuals) of the population are numbered on slips of paper of equal size, colour, etc. All these slips are folded in one and the same way and are put in a container or bowl. After mixing the slips thoroughly, the investigator, in a blindfolded position, selects one number at a time until the desired sample size is obtained.

Although this method is very simple for random selection of samples, it has some limitations:

(i) This method can't be applied where the size of the population is large. It would be a cumbersome and tedious task to number each element of a population of, say, fifteen thousand cases. Moreover, mixing them in the container or bowl will also pose a big problem.

(ii) This method of random selection is too simple a method to be considered as a scientific one. Here the investigator may select some numbers of slips from the bowl while purposefully excluding the others.

2. *Using the Table of Random Numbers:*

For random sampling the use of the table of random numbers is considered appropriate, easy and scientific. The table of random numbers consists of a continuous row-column sequence of numbers which don't appear in any particular sequence, nor does any number appear more frequently than the other one. For using the table of random numbers the researcher specifies the number of elements in the population and then, numbers them from 1 to *N* where *N* is the total number of elements in the population. Suppose the population size is of 500 and he intends to select 50 cases randomly from this population. He would enter the table at any point. He may move systematically to the right, left, up, down or diagonally, skipping the numbers that are too large and also those that have already been drawn. (This latter event does take place occasionally). He would thus keep on moving systematically through the table until he has selected the random sample of 50 elements.

In Table 14.1 a random selection of 30 cases from a population serially numbered from 1 to 80 has been illustrated. Twelve numbers have been omitted. Numbers 85, 84, 97 and 95 have been omitted because they exceed 80 and numbers 03, 74 and 12 have been omitted because they have duplicated the previous selections. Number 00 has been omitted because our population number starts from 01.

Table 14.1 Table of random numbers and its illustrative use

64755	83885	84122	25920		17696
10302	52289	77436	34430		38112
71017	98495	51308	50374		66591
60012	55605	88410	34879		79655
Sample (*N* = 30 has been taken from a serially numbered population of 80)					
64	59	28	12	85	55
75	20	97	71	03	60
58	17	74	01	74	
38	69	36	79	66	
85	61	34	84	59	
84	03	43	95	16	
12	02	03	51	00	
22	52	81	30	12	

The advantage of using the table of random numbers is that it is easily accessible to the researchers and requires no formal training in using it. However, the disadvantage is that it can't be easily and constructively used when the size of the population exceeds 5 digits.

3. *Method of Computer-determined Randomness:*

This method is used when the size of the population is large. The data is fed in the computer to obtain a random number of elements corresponding to the elements in the population. The investigators note that this method is easier to adopt when computer facilities are available.

Of these three methods, the use of the table of random numbers is more popular for ensuring randomness in the method of selecting a sample.

SIMPLE RANDOM SAMPLE

A simple random sample (also known as an *unrestricted random sample*) may be defined as one in which each and every individual of the population has an equal chance of being included in the sample, and also the selection of one individual is in no way dependent upon the selection of another individual. For example, if we are to select a sample of 10 students from the seventh grade consisting of 40 students, we can write the name (or roll number) of each of the 40 students on separate slips of paper—all equal in size and colour—and fold them in a similar way. Subsequently, they may be placed in a box and reshuffled thoroughly. A blindfolded person, then, may be asked to pick up one slip. Here, the probability of each slip being selected is 1/40. Suppose that after selecting the slip and noting the name written on the slip, he again returns it to the box. In this case, the probability of the second slip being selected is again 1/40. But if he does not return the first slip to the box, the probability of the second slip becomes 1/39. When an element of the population is returned to the population after being selected, it is called *sampling with replacement* and when it is not returned, it is called *sampling without replacement.* Sampling with replacement is wholly feasible except in certain situations where it is seldom used (Cochran, 1963: 19). If sampling with replacement is used, the chance of the same case being selected more than once is increased. In such a situation, the repeated cases may be ignored as is done if a table of random numbers is used in making a selection of cases. Thus random sampling may be defined as one in which all possible combinations of samples of fixed size have an equal probability of being selected.

The major difference between *sampling with replacement* and *sampling without replacement* is mainly concerned with the number of possible samples of size n that could be theoretically drawn. In the case of sampling with replacement the number of possible samples of size n would be greater than the number of possible samples of the same size (from same population) in case of sampling without replacement. This difference can be illustrated through an example. Suppose the size of the population consists of 4 persons, who are named as A, B, C and D. Further, suppose that the investigator wants to select samples of size 2 through the procedure of sampling without replacement. In such a situation, the investigator can maximally draw six samples of size 2 from the population of 4. This could be accomplished with the help of the following equation:

$$\binom{N}{n} = \frac{N!}{(N-n)!\,n!} \tag{14.1}$$

where,

N = the size of parent population

n = the size of the sample

! = factorial

In the above example where $N = 4$ and $n = 2$, the maximum number of sample size of 2 would be 6 as under:

$$\frac{4}{2} = \frac{4!}{(4-2)!\,2!} = \frac{4\times3\times2\times1}{2\times1\times2\times1} = \frac{24}{4} = 6$$

Similarly, where $N = 5$, we can have 10 samples of size 2 as shown:

$$\frac{5}{2} = \frac{5!}{(5-2)!\,2!} = \frac{5 \times 4 \times 3 \times 2 \times 1}{3 \times 2 \times 1 \times 2 \times 1} = \frac{120}{12} = 10$$

But from the same population, we can have 5 samples of size 4 as under:

$$\frac{5}{4} = \frac{5!}{(5-4)!\,4!} = \frac{5 \times 4 \times 3 \times 2 \times 1}{1 \times 4 \times 3 \times 2 \times 1} = \frac{120}{24} = 5$$

But if the investigator has decided to proceed with the technique of sampling with replacement, he can derive the likely number of samples from the given population with the help of the following equation:

$$N^n \tag{14.2}$$

(where meaning of N and n is identical to the meaning) in the case of Equation 14.1.

Suppose the size of the population is 4 and the size of the sample is 2. In such a situation the investigator, following the technique of sampling with replacement, can maximally draw 16 samples, that is, $4^2 = 4 \times 4 = 16$. If the four members of population are named as A, B, C, and D, the sixteen samples of size 2 would be

AA	AB	AC	AD,
BA	BB	BC	BD,
CA	CB	CC	CD,
DA	DB	DC	DD

The case of AA, BB, CC and DD combinations reflects the fact that in sampling with replacement an element or individual once drawn can be drawn again. In actual practice such cases are ignored.

There are some advantages and disadvantages of simple random sampling. The major *advantages* of simple random sampling are given ahead.

1. A sample prepared on the basis of simple random sampling plan is regarded as the representative of the population from which it was drawn. This is because in such a sampling plan all the elements in the population have an equal and independent chance of being included in the sample. A sample drawn in such a way that does not ensure equal chance for all elements to be included in the population rather increases or decreases the likelihood of an element being included, the resultant sample is called a *biased sample.*

2. In simple random sampling the investigator need not know the true composition of the population beforehand. Such a sample theoretically reflects all important characteristics and segments of the population.

3. With a view to understanding and application, simple random sampling is the easiest and simplest technique of all probability sampling plans.

4. Simple random sampling serves as a foundation upon which all other types of random sampling are based because this method of sampling can be readily applied in conjunction with all other probability sampling plans.

5. In simple random sampling the sampling error associated with any given sample drawn can easily be assessed.

6. In simple random sampling the investigator does not commit classification errors because he need not know thoroughly the population characteristics prior to selection of the sample. Classification error meants the error which results from the improper classification of population characteristics or segments.

However, simple random sampling has also some *disadvantages* given below:

1. One of the major disadvantages is that simple random sampling does not ensure that those elements which exist in small numbers in the population will be included in the given sample. This is a very serious limitation. Suppose, in a population of 1,000 people, only 12 persons possess the X trait and the investigator is to draw only 50 cases from 1,000 people. If the investigator wants to include some persons possessing the trait X in his sample, chances are very slim that such persons would be drawn.

2. Another disadvantage of simple random sampling is that it does not fully exploit the knowledge the investigator has concerning the segments of the population. Suppose, for example, the investigator knows the proportion of males and females in the population, proportion of graduates and undergraduates, proportion of persons belonging to high SES and low SES persons in population, etc., he can't utilize these items of information while drawing a sample. This mars the quality of representativeness of the sample being drawn.

3. In the case of simple random sampling the sampling error of a sample of size *n* is greater as compared with the sampling error incurred in the case of a stratified random sample of the same size. This is because the heterogeneity in the case of a random sample is greater than the same in the case of a stratified random sample. In stratified random sampling the sample drawn becomes somewhat typical of the population because it is more or less proportionate to some known characteristics of the population. This fact is ignored in simple random sampling. Hence, the sampling error increases.

Despite these limitations simple random sampling has been preferentially used for assignment of elements randomly to different experimental conditions in psychological experiments as well as in raising a random sample for generalizing the obtained findings.

STRATIFIED RANDOM SAMPLE

In stratified random sampling the population is, first, divided into two or more strata, which may be based upon a single criterion such as sex, yielding two strata—male and female, or upon a combination of two or more criteria such as sex and graduation, yielding four strata, namely, male undergraduates, male graduates, female undergraduates and female graduates. These divided populations are called *subpopulations*, which are non-overlapping and together constitute the whole population. Having divided the population into two or more strata, which are considered to be homogeneous internally, a simple random sample for the desired number is taken from each population stratum. Thus in stratified random sampling the stratification of population is the first requirement. There can be many reasons for stratification in a population. Two of them are mentioned below.

1. Stratification tends to increase the precision in estimating the attributes of the whole population. If the whole population is divided into several internally homogeneous units, the chances of variations in the measurements from one unit to another are almost nil. In such a situation a precise estimate can be made for each unit and by combining all these estimates, we can make a still more precise estimate regarding the population.

2. Stratification gives some convenience in sampling. When the population is divided into several units, a person or group of persons may be deputed to supervise the sampling survey in each unit. Or, the possibility is that the institution conducting the sampling survey may have field branches to supervise the survey in each part or unit of the population.

Stratified random sampling is of two types:

(A) Proportionate stratified random sampling

(B) Disproportionate stratified random sampling

A discussion of these two types of sampling plans is as follows.

A. Proportionate stratified random sampling

As its name implies, in this sampling plan the researcher stratifies the population according to the known characteristics of the population and, subsequently, randomly draws the individuals in a similar proportion from each stratum of the population. Suppose, the investigator knows the classwise distribution of population of the students in a university as given ahead.

Table 14.2 Classwise distribution of a population of 10,000 students

Class	Composition of population	Proportion of each class
IA	$N_1 = 3{,}000$	0.30
BA	$N_2 = 4{,}000$	0.40
MA	$N_3 = 2{,}000$	0.20
M Phil.	$N_4 = 1{,}000$	0.10
	$N_t = 10{,}000$	1.00

Now, if the investigator has decided to draw a sample of, suppose, 1,000, he will include the number of students from each stratum in similar proportion. Thus, for drawing individuals from each stratum of the population in a proportionate way, he will go through a procedure as under:

Number of students from IA Class = 1000 × .30 = 300

Number of students from BA Class = 1000 × .40 = 400

Number of students from MA Class = 1000 × .20 = 200

Number of students from M Phil. Class = 1000 × .10 = 100

$\Sigma = 1000$

The proportionate breakdown of a sample of 1,000 students has been shown in Table 14.3.

Table 14.3 A proportionate breakdown of a sample of 1,000 students

Class	Composition of a population	Proportion of each class
IA	$N_1 = 300$	.30
BA	$N_2 = 400$	.40
MA	$N_3 = 200$	.20
M Phil.	$N_4 = 100$	.10
	$N_t = 1{,}000$	1.00

Having decided the proportion of each stratum in the sample in proportion to the population stratum and the number of students to be included in each stratum, the investigator goes on randomly drawing the students in each stratum until the desired sample size is reached. This is easily done with the help of either a table of random numbers or any suitable technique of randomness.

Proportionate stratified random sampling has many research applications. It is readily and easily applied in military studies where ranks are likely to affect other variables in the research project. Likewise, such sampling is considered important in the study of school systems and

administrative services where the population remains divided into many strata. To make the sample representative of the population it is essential to have a proportionate representation.

Proportionate stratified random sampling has some advantages and disadvantages. The important advantages are given below:

1. Proportionate stratified random sampling increases the representativeness of the sample drawn. The representativeness is strengthened because of the fact that this method of sampling ensures that those elements that exist in a few numbers are also included proportionately in the sample.

2. In proportionate stratified random sampling, the sampling error is minimized because the sample drawn possesses all the necessary characteristics of the parent population.

3. Proportionate stratified random sampling eliminates the necessity of weighing the elements according to their original distribution in the population. Since the sample is proportionately stratified according to several salient features, the sample is weighted automatically. The weight of any particular value is the frequency of it in the population.

Despite these advantages, the proportionate stratified random sampling has some limitations or disadvantages as indicated below:

1. Proportionate stratified random sampling is a difficult method. This method assumes that the researcher knows the composition and distribution of population well before the actual sampling starts. This is often an unrealistic assumption.

2. This method is a time-consuming method because the samples are drawn according to the proportion of each stratum in population.

3. This method has the probability of classification error. Since the investigator is required to identify several strata, he is likely to classify an element or individual of the population in such a way as to put him into the wrong strata. This will affect the validity and reliability of the obtained results.

B. Disproportionate stratified random sampling

Disproportionate stratified random sampling bears similarity with the proportionate stratified random sampling. The only difference is that the substrata of the drawn sample are not necessarily distributed according to their proportionate weight in the population from which they were randomly selected. In fact, some of the strata of the population may be overrepresented or some underrepresented. Sampling disproportionately means that

(a) either the investigator will give equal weight to each of the substrata, or

(b) he will give greater representation to some substrata and not enough weight to other substrata of the population in the sample to be drawn.

Suppose the investigator divides a given population of 10,000 individuals into 6,000 males and 4,000 females. If he has decided to draw a sample of 1,000 individuals from the set of 10,000 and if he draws randomly both the males and the females in equal number, say, 500 each, it will constitute the example of a disproportionate stratified random sampling. But if he randomly draws 600 males and 400 females in his sample, it will constitute the example of proportionate stratified random sampling. From this example it becomes obvious that in the disproportionate stratified random sample the investigator tries to give equal weight to each stratum, that is, he tries to draw equal number of individuals from each stratum. In doing so he overrepresents one stratum while underrepresents the other strata. In this example, when he randomly draws 500 males and 500 females, he is overrepresenting a female stratum and underrepresenting a male stratum.

Disproportionate stratified random sampling has both advantages and disadvantages. Its major advantages are indicated as follows:

1. Disproportionate stratified random sampling is comparatively less time-consuming than proportionate stratified random sampling because here the investigator is not worried about making proportionate representation of each stratum of the population.

2. In disproportionate stratified random sampling the investigator is able to give weight to the particular groups of elements that are not represented as frequently in the population as compared with other elements.

The major disadvantages of disproportionate stratified random sampling are given below:

1. In this method of sampling certain stratum of the population is overrepresented and some other strata are underrepresented in the samples drawn. The assignment of greater weight to one set of elements or overrepresenting one stratum of the population may introduce some bias in the sample. Such type of sample may not be a true representative.

2. Disproportionate stratified random sampling assumes that the investigator knows the composition of the original population. This automatically meants that this method of sampling cannot be used in a situation where the investigator has no idea about the composition of the original population.

3. In this method of sampling the investigator is required to classify the population into different substrata and from each substratum, he takes more or less, equal number of cases. In drawing a sample from such a classified set, he may actually misclassify certain elements, that is, he may put element A into stratum X whereas A might belong to stratum Z.

Despite these limitations, the use of disproportionate stratified random sampling is common in social science researches. In fact, because of ease and convenience it is more popular than proportionate stratified random sampling.

AREA (OR CLUSTER) SAMPLING

Area (or cluster) sampling is another important method of probability sampling. Such sampling method has its origin in the field of agriculture. Farming experiments that were conducted to determine the effect of various kinds of fertilizers, soil treatments and a variety of planting methods on crop yield, mostly used this method of sampling. In social sciences application of area sampling has been extensive in survey research and field research.

In area sampling, generally, geographical divisions of territory, community, neighbourhood, cities, states, etc., are made on a map and a certain number of them is drawn at random and called sample. The investigator or interviewer proceeds to interview all elements of the randomly drawn areas or clusters. That is the reason why this method of sampling is also known as *Cluster sampling.*

Suppose the investigator wants to assess the attitude of the people of Tamil Nadu towards family planning. For this, it will be convenient for the investigator to have the map of Tamil Nadu before him and then divide it into various sections according to a number of vertical and horizontal grid lines drawn across the total area. He will then number each section from 1 to *N*, *N* being equal to the total number of sections. With the help of the table of random numbers, he will draw a specified number of sections to constitute the sample that he will finally study. The investigator will, then, interview all persons or members of families living in those sections. If any drawn section contains extremely different types of families, again, a random selection from among those families can be done and finally interviewed. In this way according to necessity further subdivision and selection of samples can be done to different stages. This is called *multi-stage sampling.*

As given hereunder area sampling has many research utilities or applications.

(i) Large scale surveys of political, religious and social behaviour are easily conducted by area sampling.

(ii) Where somehow lists of specific individuals are unobtainable or are inaccessible, area sampling becomes the best method of sampling.

(iii) Public opinion polls are easily and smoothly conducted using an area sampling plan.

Area sampling has some advantages and disadvantages. The important *advantages* are given below:

1. When larger geographical areas are to be covered, it is easier to use area sampling than any other method of probability sampling. It is easier in the sense that the investigator need not have the list of the individuals inhabiting a given area. He simply draws some geographical sections randomly; and, subsequently, he interviews all the families or persons living in the randomly drawn sections. In this sense, area sampling is easier than the other two preceding methods of probability sampling.

2. In area sampling respondents can readily be substituted for other respondents within the same random section. This is permitted because clusters of elements are sampled and not the individuals.

3. Area sampling saves both time and money. The investigator can concentrate his efforts in one specific region and, thus, can save time. The cost of such sampling is much less as compared to the other methods of probability sampling. The investigator need not travel great distances to interview specific individuals residing at random points in a certain geographical region.

4. Area (or cluster) sampling possesses the trait of flexibility. In a multi-stage area sampling design the investigator can successfully employ different forms of sampling in several successive stages. For example, after drawing several sections randomly in the first stage, the investigator may decide to stratify each of the first-stage units and select sections from some strata to constitute the second stage units. The investigator may use a third type of sampling plan in the third or fourth stage.

5. Still another advantage of area sampling is that the respondents can readily be substituted for other respondents within the same random section. This further increases the degree of flexibility in the area sampling.

Area sampling has some disadvantages too. The important disadvantages are given below:

1. In area sampling the degree of sampling error is usually high. This makes the sampling much less fruitful or dependable in comparison to other methods of probability sampling.

2. In area sampling there is no correct way to ensure that each sampling unit included in an area sample will be of equal size. In fact, here the researcher has little control over the size of each cluster. This introduces bias into the samples.

3. In area sampling it is also difficult to ensure that the individuals included in one cluster are independent of other randomly drawn clusters. It may be that an individual is interviewed in one cluster and next morning he travels to another area that falls within another randomly selected section.

Despite these disadvantages, area sampling is a common and popular method of sampling in behavioural researches.

QUOTA SAMPLING

Quota sampling is one of the important types of non-probability sampling methods which is apparently similar to stratified random sampling. In quota sampling the investigator recognizes

the different strata of population and from each stratum he selects the number of individuals arbitrarily. This constitutes the quota sample.

Suppose the investigator knows that the population of the individuals he is going to study has three strata in terms of SES (socio-economic status)—high, middle, and low. Further, suppose he knows that there are 1,000 people in high SES, 7,000 people in middle SES and 2,000 people in lower SES. Thus, the population consists of 10,000 individuals. If he wants to select 1,000 individuals and, finally, selects 100 individuals from high SES, 700 from middle SES and 200 from lower SES, according to his convenience (and not randomly), this constitutes quota sample.

It is obvious that there is similarity between quota sampling and proportionate stratified random sampling with the only difference that in the former the final selection of individuals is not random, whereas in the latter the final selection of individuals is random. Keeping in view this similarity, quota sampling is sometimes referred to as "*the poor man's proportionate stratified sample*", because such sampling seeks to ensure that specific elements are included and represented in subsequent collection of them.

Quota sampling has both advantages and disadvantages. The major *advantages* of quota sample are stated below:

1. Quota samples are the most satisfactory means when quick and crude results are desired.

2. This method of sampling is convenient and less costly than many other methods of sampling, whether probability or non-probability.

3. Quota sampling, to a greater extent, can guarantee the inclusion of individuals from different strata of population.

However, quota sampling has some *disadvantages* also as given below:

1. In quota sampling there is no means of establishing randomness. As such, the selected samples remain no longer representative of the population. The conclusion, therefore, lacks external validity or generalizability.

2. In quota sampling the investigators or interviewers get ample opportunity to select the most accessible individuals influencing their friends and relatives. Such readily accessible individuals may not be typical of the population they are going to study.

3. Quota sampling is amenable to classification error. Here the interviewer or the investigator bases his classification of respondents on the way they apparently look to him. In fact, he possesses no knowledge concerning the way respondents should be classified. The investigator remains ignorant of many important variables that might otherwise be used in classifying them. All these tend to make quota sampling less dependable and reliable.

4. In quota sampling, the researcher, to a greater extent, controls one variable such as sex or caste, etc., but he can't control other variables that may have both theoretical or practical significance. This mars the dependability of quota sampling.

Despite these limitations, quota sampling is a popular method among non-probability methods of sampling, because it enables the researcher to introduce a few controls into his research plan.

PURPOSIVE OR JUDGEMENTAL SAMPLING

Purposive sample, a kind of non-probability sample is one which is based on the typicality of the cases to be included in the sample. The investigator has some belief that the sample being handpicked is typical of the population or is a very good representative of the population. A purposive sample is also known as a *judgemental sample* because the investigator on the basis of his impression makes a judgement regarding the concerned cases, which are thought to be typical of the population. For studying attitudes towards any national issue, a sample of

journalists, teachers and legislators may be taken as an example of purposive samples because they can more reasonably be expected to represent the correct attitude than other classes of persons residing in the country. Before the start of general elections, purposive samples are often taken in an attempt to forecast the national elections. The investigator selects the persons from those states whose election results on previous polls have approximated the actual results and thus, have been typical of the whole population.

Purposive sampling has some advantages and disadvantages. The important *advantages* are given below:

1. Since purposive sampling does not involve any random selection process, it is somewhat less costly and more readily accessible to the investigator.
2. Purposive sampling is a very convenient method of sampling as compared to other methods of non-probability sampling.
3. Purposive sampling guarantees that those individuals will be included in the sample that are relevant to the research design. The investigator does not get such guarantee in any other methods of non-probability sampling.

Purposive sampling has some *disadvantages* also as given below:

1. In purposive sampling there is no way to ensure that the sample is truly random or representative of the population despite the belief in typicality of the sample by the investigator. This inhibits his ability to generalize the findings.
2. In purposive sampling too much emphasis is placed on the ability of the investigator to assess which elements or individuals are typical of population and which are not. This leaves ample scope for introducing subjectivity in the sampling. Once he has selected the sample he assumes that errors arising from his selection method will be minimal, but actually there is no legitimate stand for arguing his position.
3. In the case of purposive sampling the inferential statistics can't be used legitimately, because, under all inferential statistical techniques, there is an assumption of randomness. This criticism also applies to other forms of non-probability sampling method.

ACCIDENTAL SAMPLING

Accidental sampling, also known as *incidental sampling*, is another popular method of non-probability sampling plan. It refers to a sampling procedure in which the investigator selects the persons according to his convenience. Here he does not care about including the people with some specific or designated trait, rather he is mainly guided by convenience and economy. Heiman (1995) has named it as convenience sampling. This is a crude method of sampling, and the investigator knows that little can be generalized from the sample thus drawn.

An investigator may take students of class X into research plan because the class teacher of that class happens to be his friend. This illustrates accidental sampling.

Accidental sampling has some advantages and disadvantages. The important *advantages* are given below:

1. Accidental sampling is the most convenient method of sampling.
2. This method of sampling possesses the trait of economy. This method saves time, money and labour of the investigator.

However, accidental sampling has the undernoted *disadvantages* as well:

1. From accidental samples nothing can be generalized with confidence because the samples remain no longer representative of the population.
2. In accidental sampling the investigator gets ample opportunity to show his bias and prejudice in selecting the individuals. As such, this method of sampling is not much dependable.

3. In accidental sampling the probability of sampling error is high. Therefore, the validity and reliability of this method are badly affected.

Despite these disadvantages, it will not be an exaggeration to say that in many psychological and sociological researches this method of sampling is frequently used.

SYSTEMATIC SAMPLING

Systematic sampling is another method of non-probability sampling plan, though the label 'systematic' is somewhat misleading in the sense that all probability sampling methods are also systematic sampling methods. Due to this, it often sounds that systematic sampling should be included under one category of probability sampling, but in reality this is not the case.

Systematic sampling may be defined as drawing or selecting every *n*th person from a predetermined list of elements or individuals. Selecting every 5th roll number in a class of 60 students will constitute systematic sampling. Likewise, drawing every 8th name from a telephone directory is an example of systematic sampling. If we pay attention to the systematic sampling plan, it becomes obvious that such a plan possesses certain characteristics of randomness (first element selected is a random one) and at the same time, possesses some non-probability traits such as excluding all persons between every *n*th element chosen.

Systematic sampling has some advantages and disadvantages. The important *advantages* are mentioned below:

1. Systematic sampling is relatively a quick method of obtaining a sample of elements. If the investigator has a short time schedule, this method of sampling eliminates several steps otherwise taken in different methods of sampling.

2. Systematic sampling makes it very easy to check whether every *n*th number or name has been selected. In case, there occurs any error in counting, that is, if the investigator selects 6th number instead of 5th number, his sample will not be seriously affected.

3. Systematic sampling is very easy to use. In fact, it is much simpler than having to employ a table of random numbers for drawing the sample or fixed quota from each stratum of the population in order to have proportional representation.

Despite these advantages, there are some *limitations* of systematic sampling as indicated overleaf:

1. Systematic sampling ignores all persons between every *n*th element chosen. Obviously, then, it is not a probability sampling plan.

2. In systematic sampling, the sampling error increases if the list is arranged in a particular order, say, the list increases or decreases with respect to some trait such as age, education, income, caste, etc. In such a situation a bias will be introduced in the sample, because more individuals are likely to be drawn from one group, particularly from that group which is represented most frequently on the group. Such type of bias may also result when the list is alphabetical (Blalock, 1972).

Despite these limitations, systematic sampling is frequently used in psychological and sociological researches because this method of sampling possesses the trait of both probability and non-probability sampling.

SNOWBALL SAMPLING

Snowball sampling, which is a non-probability sampling method, is basically indirectly sociometric. It is defined as having all the persons in a group or organisation identifying their friends who in turn identify their friends and associates until the researcher observes that a constellation of friendships converges into some type of a definite social pattern. Some selected behaviour is usually used as the basis of contact and/or association. Obviously, then, snowball

sampling is used for obtaining an impression of informal social relations among individuals. In fact, the reality is that sometimes the researcher seeks to study a 'hidden' population which is not easily identifiable, as when he deals with drug addicts, hardened criminals, prostitutes, and so on. In such cases, snowball sampling is used. Here the researcher identifies one potential subject and from him, he obtains the name and addresses of other subjects and from them, he obtains still other potential subjects so that the growth of the sample tends to build or 'snowball'. Coleman et al. (1957) used the snowball sampling successfully in a study of diffusion of medical information among physicians. The purpose was to determine how physicians come to use certain medical products such as drugs and related supplies. Snowball sampling revealed the pattern of diffusion of information. Besides, such sampling can also locate the physicians' cliques through which such information is revealed. It may answer the questions: Do the physicians read about new drugs in journals or hear about it in some medical seminar? If they hear about it in some medical seminar or convention, whom do they contact among their physician friends regarding it? How does information about a given drug spread among physician fellows? And so on.

Snowball sampling has important research applications in relatively small business and industrial organisations where *N* is expected not to exceed 100. Such sampling is more convenient to the studies of social change and diffusion of information among specific segments of social organizations.

Snowball sampling has some advantages and disadvantages. The important *advantages* are given below:

1. Snowball sampling, which is primarily a sociometric sampling technique, has proved very important and is helpful in studying small informal social group and its impact upon formal organizational structure.

2. Snowball sampling reveals communication pattern in community organization concepts like community power; and decision-making can also be studied with the help of such sampling technique.

3. The method of snowball sampling is amenable to various scientific sampling procedures at various stages such as use of random numbers or computer determination.

Despite these advantages, snowball sampling has some *limitations* also as described below:

1. Snowball sampling becomes cumbersome and difficult when *N* is large or say it exceeds 100.

2. This method of sampling does not allow the researcher to use probability statistical methods. In fact, the elements included in sample are not randomly drawn and they are dependent on the subjective choices of the originally selected respondents. This introduces some bias in the sampling.

Despite these limitations, snowball sampling is used in the study of diffusion of some specific information and social changes.

SATURATION SAMPLING AND DENSE SAMPLING

Coleman (1959) has emphasized these two types of sampling techniques which are used less frequently as compared to other techniques of sampling. Saturation sampling is defined as drawing all elements or individuals having characteristics of interest to the investigator. Drawing all physicians having at least the age of 45 (from a small community), would be called saturation sampling. Dense sampling is a method of sampling which lies somewhere between simple random sampling and saturation sampling. When the researcher selects 50% or more from the population and takes a majority of individuals having specified traits or characteristics which are of interest to him, it is called dense sampling. For example, if the researcher selects 500 to 600 students from a population of 1,000 students, it will constitute dense sampling.

The advantage of saturation sampling and dense sampling is that these are convenie methods of sampling where *N* of the population does not exceed 1,000. In case it exceeds, the methods of sampling become cumbersome and inconvenient.

DOUBLE SAMPLING

Double sampling, as its name implies is defined as drawing a sample of individuals from anoth sample of them. Suppose the investigator randomly draws a sample of 1,000 from a populatio having $N = 10{,}000$. From this 1,000 individuals, he again randomly draws a sample of 300 fo further study. This is called double sampling.

Suppose the researcher is studying the attitude of newly married couples towards famil planning through mail questionnaire. For this, he mailed 1,000 questionnaires to one thousan couples residing in different localities. He finds that only 40% (that is 400) questionnaires wer returned. From these 400 persons, he draws randomly a sample of 100 and mails anothe questionnaire to get their in-depth knowledge towards the different techniques of famil planning. This is called double sampling.

For double sampling to be an effective part of the research, it is essential that the sampling plan followed from the beginning should maximize the representativeness of the individuals drawn subsequently. This method has the disadvantage of taking much time and labour o the researchers.

REQUISITES OF A GOOD SAMPLING METHOD

Any sampling method to be good and scientifically sound must possess at least the two following properties:

1. It must ensure the representativeness of the sample.
2. It must ensure the adequacy of the sample.

The first requirement of a sampling method is that the selected samples must be representative of the population. What is meant by representativeness of the sample? In the words of Kerlinger (1973: 119), "Ordinarily, representative means to be typical of a population, that is, to exemplify the characteristics of the population... In research a representative sample means that the sample has approximately the characteristics of the population relevant to the research in question." Truly speaking, an investigator cannot be certain about the representativeness of a sample unless the entire population is tested. But this is ordinarily not feasible. Hence, he must satisfy himself otherwise. A sample can be regarded as being representative of the population when it possesses all the relevant characteristics of the population in about the same proportion as is in the population. If sex is the relevant characteristic of a population of students of a particular university and if the population has a male: female ratio of 60:40, a representative sample would be one where the ratio of males and females is also 60:40. For ensuring representativeness of samples it is also essential that the population should be clearly defined and a suitable and appropriate definition of the observations, which are to constitute samples, be framed. An observation must be defined in the light of the definition of population otherwise the result will be a biased sample. Suppose that the population is defined as all those females who are widows. Further suppose that the observation is defined in terms of selection of those who admit that they are widows. In this case, the resulting samples will not be representative ones because observation has been defined differently from the population. Obviously, this definition of observation will automatically exclude those who refuse to admit that they are widows and therefore, the sample will be a biased one.

The second requirement of good sampling is that the samples must be adequate. A sample is said to be adequate when it is of sufficient size. There is no rigid rule regarding the size of a sample. But a larger sample, in general, is preferred because it tends to reduce the error, which is

the difference between the population value and the sample value. The mean of a sample taken from a population will not ordinarily be equal to the mean of that population. The relationship between the size of the sample and the error, as defined above, is negative. The larger the size of a sample, the smaller the error and vice versa. The negative relationship between the size of a sample, and the error is demonstrated in Figure 14.1. When the size of the sample is increased, the form and the statistics of the sample distribution become a close approximation to those of population. Hence, the error is reduced and therefore, a large sample is preferred. But a very large sample may not yield that result because it tends to create several problems, which sometimes become difficult to solve.

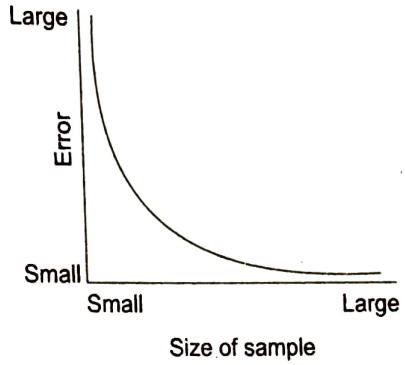

Fig. 14.1 Relationship between the size of sample and the error

COMMON ADVANTAGES OF SAMPLING METHODS

The purpose of any sound sampling procedure is to make sampling more efficient and easy. In any behavioural research the whole population is not examined and an estimate of the characteristics of population is made on the basis of the characteristics of the sample drawn from the population. Investigators rely upon samples and the conclusion is generalized. Apart from serving this function of generalization, sampling has some additional advantages.

1. *Accuracy is increased*: When one is entrusted with examining sample (and not the population), naturally, a great volume of the work is reduced. Consequently, one can easily supervise the task assigned to the sample and the results can also be analyzed with great care. Not only this, the services of trained personnel can be utilized at a lower cost when the number of cases to be processed is low. The overall results of all this is that the accuracy of the investigation in increased.

2. *Cost is reduced*: When the data are to be collected from a limited number of cases rather than from a larger population, this naturally entails a reduced cost to the investigator. Not only this, the time taken in processing of the data is also reduced. When one is dealing with only a limited number of data, the chances of accidental error in statistical calculation are also minimised.

3. *Speed is increased*: When the data are collected from a sample rather than from the entire universe, the investigator can do his work speedily without doing any injustice to the investigation. This is one of the vital considerations specially when the information about the sample is needed urgently.

SAMPLING DISTRIBUTION

In any sampling situation, three types of distributions are of common interest to the research investigator: *the sample distribution, the population distribution*, and *the sampling distribution*. The meaning of and distinction between these three types of distributions may be explained through an illustration.

Suppose in a small town the total number of graduate parents is 30,000. Suppose further that we are interested in knowing the mean fertility rate among these parents. Owing to time and money considerations, we cannot examine the whole population and therefore, we decide to examine only a random sample of 300 parents. The observed data may be arranged in frequency distribution and the desired statistics may be calculated. This type of distribution is known as the *sample distribution* because it shows the distribution of data obtained from sample drawn randomly from the population. Now, if with the help of any social organization or institution we are able to record the fertility rate of 30,000 parents and arrange them in frequency distribution, this distribution is termed as the *population distribution*. In reality, the forms and the parameters of a population are not ordinarily known. Therefore, an estimate about these two properties of the population is made on the basis of the form and parameters of the sample distribution. If the form of sample distribution is normal, we expect that the form of the population distribution will also be normal. If the mean and the standard deviation of the fertility rate of the 300 graduate parents are 3.89 and 1.02 respectively, we assume that the mean and the standard deviation of the population distribution are also 3.89 and 1.02 respectively. If we take all possible samples of a given size (in this example 100 samples consisting of 300 parents each can be taken) we will get different sample distributions. For illustration, we will get 100 different sample distributions, that is, 100 different means and 100 different standard deviations. If we group the 100 means and the 100 standard deviations into frequency distributions separately, we will get the sampling distribution of mean and the sampling distribution of standard deviation. Thus, the sampling distribution is a *theoretical* or *idealized distribution*, which is obtained if all possible samples of the given size are taken from the population

SAMPLING ERROR

As the reader must have gathered by now, sampling refers to the process of drawing a limited number of cases from a population. Whatever the sampling procedure used, the sample must represent, as far as possible, all characteristics of the population that are related to the problem under study.

A statistical measure calculated from a sample is known as a statistic and the corresponding measure which is usually unknown and is based directly upon the population is known as a parameter. A statistic is frequently referred to as an obtained measure or "derived measure". A mean calculated from a sample of a population is called "statistic mean" and the value of mean based directly upon the population is known as "parameter mean". A sampling error is the difference between the parameter and the statistic. As Lindquist (1968:8) has put it, "A sampling error is the difference between a parameter and an estimate of that parameter which is derived from a sample."

A sampling problem is said to exist whenever the conclusions obtained on the basis of a limited number of individuals (called the sample) are applied to a larger number of individuals (called the population). Such generalized conclusions are known as *statistical inferences*. The higher the sampling error, the poorer the statistical inference. The term "sampling error" does not suggest any mistake in the process of sampling itself, rather it merely suggests the chance variations which are inevitable when a number of randomly selected samples are taken from the same population. Due to this sampling error, making an estimate about the population characteristics from the sample characteristics no longer remains an exact process. It has been observed that a sample statistic does not ordinarily agree with that obtained from the second, third or fourth sample or with the population parameter. The amount of fluctuation (or the sample error) depends upon the following two factors:

(i) Variability in the population

(ii) Size of the sample

If the nature of the population is such that it is not stable and tends to vary, naturally, the successive samples taken from the same population over time will tend to differ. Generally, a random sample provides a good approximation to the population and the accuracy of approximation depends upon the size of the sample. As the size of the sample is increased, the form and the statistics of the sample distribution come nearer to those of the population. For example, when the size of the sample is 100, the distribution as represented by the frequency polygon is usually smooth enough to give a very clear idea about the form of the population distribution.

There are certain statistics which tend to fluctuate less than others in successive samples and hence, the sampling error is less. For example, in sampling from a normal distribution the mean and the standard deviation fluctuate less than other measures of central tendency and measures of variability. Hence, these two are more reliable than other similar measures because they approximate more closely the corresponding population parameters and hence, tend to fluctuate less from sample to sample and thus produce the least sampling error.

In most sampling situations there are three types of important distributions which are of concern to the reader in order to have a clear-cut idea about the application of the sampling error. They are: the sample distribution, the population distribution and the sampling distribution. These three terms may be explained with an example: Suppose, we wish to know the mean intelligence score of workers (in the age range of 20–45 years) in the Tata Iron and Steel factory. Further suppose that the total number of workers in this given age range is 30,000. Owing to time and money considerations, we may decide to take a random sample of 300 and administer a standardized test of intelligence. The obtained scores on the test by 300 workers expressed in frequency distribution will be called the sample distribution. Thus, the sample distribution is an observed distribution which can be described by any descriptive statistics. Hence, sample distribution is the concrete distribution in any sampling situation.

However, if we decide to administer the test of intelligence on the whole population, that is, on 30,000 workers, we shall be getting a set of 30,000 scores. If we convert this set into frequency distribution, this will be known as the population distribution. Ordinarily, the form and the parameters of the population distribution are not known. However, we can infer them on the basis of the sample distribution. If the sample distribution is normal, it can be inferred that the population distribution is also normal, ascribing any irregularities in the sample distribution to the sampling fluctuations.

The inferences drawn regarding population distribution on the basis of the sample distribution may not be correct due to sampling errors. In the above example we do not have the fixed mean and the fixed standard deviation which would be obtained again in the second or third sample of 300 from the population. If we were to take, say, 100 samples of 300 each, we would be getting 100 different means, 100 varying standard deviations. If all the 100 means and 100 standard deviations are grouped in frequency distribution separately, we would have an exact sampling distribution of mean and the standard deviation in the population. As this is not done, sampling distribution is a theoretical or idealized distribution which may be thought of as a distribution which would be obtained if the values of a given statistic were computed in all possible numbers of a sample of given size actually taken from the population.

Statisticians have been able to determine the sampling distribution of the majority of commonly used statistics in the samples taken from the normal population. When the sampling distribution of a statistic is known, it is possible for the investigator to determine the relative

frequency with which different sample values tend to occur in random sampling from population in which the statistics have some of them assumed or hypothesized. After knowing the relative frequency, the investigator is able to know the probability that a particular sample value has resulted from the sampling fluctuations or chance. Ultimately, by knowing the probability figure, he will be able to judge the soundness of a hypothesis regarding the value of the statistic in a population.

☑ Review Questions

1. What do you mean by sampling? Discuss the factors that influence the decision to sample.
2. What is sampling distribution? Discuss the methods of drawing random samples.
3. Citing some examples discuss the relative advantages and disadvantages of any two probability sampling methods.
4. Citing some examples discuss the relative advantages and disadvantages of any two non-probability sampling methods.
5. What is `systematic sampling'? How does it differ from `cluster sampling'? Discuss the relative advantages and disadvantages of systematic sampling.
6. Explain the importance of sampling in research. Discuss any two methods of sampling.
7. What are the requisites of a good sample? State the advantages of randomizing in scientific research.
8. What is the greatest danger in using accidental sample in research?
9. When is cluster sampling likely to be preferred most?
10. Why would a researcher choose stratified random sampling instead of simple random sampling?

✣