UNIT 4 SAMPLING

Structure

4.1 INTRODUCTION

Sampling has been an age old practice in everyday life. Whenever we want to buy a huge quantity of a commodity, we decide about the total lot by simply examining a small fraction of it. It has been established that the sample survey if planned properly, can give very precise information. Since in surveys a part of the population is only surveyed and inference is drawn about the whole population, the results likely to be different from the population values. But the advantage with the sample survey is that this type of error can be measured and controlled and it can be eliminated to great extent by employing properly trained persons in surveys. The other advantage of sample surveys are that it is less time consuming and involves less cost. Usually, the population is too large for the researcher to attempt to survey all of its members. A small, but carefully chosen sample can be used to represent the population. The sample reflects the characteristics of the population from which it is drawn.

After studying this unit, you should be able to

- discuss the meaning and importance of sampling
- describe the steps and criteria involved in selecting a sampling procedure
- distinguish different types of sampling
- explain the process of determination of sampling size

4.2 SAMPLING: MEANING AND CONCEPTS

4.2.1 Meaning of Sampling

According to Levin and Rubin, statisticians use the word, population, to refer not only to people, but, to all items that have been chosen for study. They use the word, sample, to describe a portion chosen from the population.

According to Croach and Housden, a sample is a limited number taken from a large group for testing and analysis, on the assumption that the sample can be taken as representative for the whole group.

According to Boyce, sampling makes an estimate about some of the characteristics of a population. To sample is to make a judgment or a decision about something after experiencing just part of it.

4.2.2 Concepts in Sampling

For clarity and brevity, some concepts and preliminaries of sampling theory, which are used in the study material, are discussed below.

- *Sampling Units and Population:* a unit may be taken as a well defined and identifiable element or a group of elements on which observations can be made. The aggregate of these units is termed as population and the population is said to be finite, if the units are countable. The population is sub-divided into suitable small units known as sampling units for the purpose of sampling. Sampling units may consist of one or more elementary units and each elementary unit belongs to one and one sampling unit.
- *Sampling Frame:* a sampling frame is a list of sampling units with identification particulars indicating the location of the sampling units. A sampling frame represents the population under investigation, and it is the base of drawing a sample. As far as possible, it should be up-to–date, i.e., free from omissions and duplications.
- *Sample:* a fraction of the population is said to constitute a sample. The number of units included in the sample is known as the size of the sample.
- *Sampling Fraction:* the ratio of the sample size, n, to the population size, N, is known as sampling fraction and it is denoted by (n / N).
- *Sampling Procedure/Method:* this is the method of selecting a sample from a population.
- *Census:* this denotes all the elements or units of a population which are used to explain the features of population. It usually refers to complete enumeration of all persons in the population.
- *Population Parameter and Sample Estimator:* any function of the values of units in the population, such as population mean or population variance, is termed a population parameter. There can only be one set of values for a population and the population values are treated as constant. However, the function of the values of the units in the sample, such as sample mean and sample variance, is known as a statistic. The value of the mean and variance differ from sample to sample and, therefore, it is a random variable.

4.2.3 Advantages of Sampling

Some of the key advantages of sampling are:

i) it costs less

ii) takes less time

iii) data are acquired quickly

iv) fewer mistakes are likely

v) a more detailed study can be done.

Now that you have read about the meaning and concept of sampling, answer the following questions in *Check Your Progress 1.*

Check Your Progress 1

Note: a) Write your answer in about 50 words.

b) Check your answer with possible answers given at the end of the unit

1) What do you mean by sampling? What are the advantages of sampling?

.....

.....

.....

.....

.....

2) What is the difference between a parameter and an estimator?

.....

.....

.....

.....

.....

4.3 TYPES OF SAMPLING

There are broadly two types of sampling:

i) Probability sampling

ii) Non-probability sampling

4.3.1 Probability Sampling

A probability sample is one in which each element of the population has a known, non zero chance of being included in the sample. Probability methods include simple random sampling, systematic sampling, and stratified sampling.

1) Simple Random Sample

The random sample entails that each and every individual in a population has an equal chance of being included in the sample and that the selection of one individual is in no way dependent upon the selection of another person. The two popularly used methods in random sampling are

i) draw of lottery

ii) using a random number table.

i) In lottery draw, for example, if we have to select a sample of 25 students from a total of 600 students in a college, then we make separate slips of paper for 600 students and put them in a box and thoroughly mix them.

After that, a person is asked to pick up one slip. Here, the probability of each of the student being selected in the sample is 1/600. This procedure is continued till the sample size is acquired.

ii) Another method of simple random sampling is to use a random number table for drawing 25 students from a total of 600 students. The procedure for using a random number table follows.

1) Number each element in the sample frame from 001 to 600.
2) Decide a random starting point in the table. Any point will do. Say second row in the second column (Appendix 1).
3) Look at the first digits at that point, because there are three digit in 600.
4) Then, if the number is less than 600, include it in the sample; if not then look for a number where the first three digits are less than 600.
5) From that point you can move in any direction. Select only three digit numbers that are less than 600, until you have 25 such numbers.

Note: You can move in any direction in the random number table because every digit has been placed in the table at random.

For example, here if we start from the second row in the third column, then, the random numbers are: 31684; 09865; 14491; 34691, continuing till 25 samples are selected.

2) Systematic Random Sample

Designing a Systematic Random Sample is sometimes quite difficult and time consuming and therefore, Systematic Random Sample, like Simple Random Sample, also uses a list of all members of the population in its sampling frame. However, instead of using random numbers to select the sample elements, the researcher applies a skip interval to the list to produce a sample of the required size.

$$\text{Skip interval} = \frac{\text{number of elements in the population}}{\text{the required sample size}}$$

$$K = \frac{N}{n}$$

K = skip interval

N = Universe size

n = Sample size

For example if we have to select a sample of 100 persons from a universe of 1000 population, then the skip is 10. In this case one number between 1 and 10 has to be selected. Suppose 5 is selected, then the first sample would be 5^{th} and the next one 15^{th}, 25^{th}, 35^{th}, 45^{th}, and so on. One of the advantages of this method is that it is more convenient than other methods and simple to design. Again, it is used with very large populations.

3) Stratified Random Sample

In Stratified Random Sampling, the target population of N units is first divided into k subpopulations of N_1, N_2, ..., N_k units. These populations are non-overlapping and together they comprise the whole population. So that $N_1 + N_2 + ... + N_k = N$

The sub-populations are called strata. The number in each stratum should be known. A sample is drawn from each stratum independently. The sample sizes within 'k' strata are denoted by n_1, n_2, ..., n_k respectively. If the total sample size 'n' is to be drawn from the target population than $n_1 + n_2 + ... + n_k = n$

If a simple random sample is drawn in each stratum, the whole procedure is described as stratified random sampling.

Stratified random sampling requires more than making a list of elements (and estimating the number of elements on the list). It also involves ordering that list by sub groups (or strata) and then, to do sampling randomly or systematically within those sub groups. This method of sampling is used for the following reasons.

- It can reduce the errors in the statistical estimates calculated from the sample.
- It allows you to create a sample that is exactly representative of the various sub groups in the population that you find to be of special interest.

For example, the selected village may have households of SC, ST, OBCs, Others, Minority. The village population first may be divided in to smaller sub groups of different sections of population (stratum) and, thus, the village sample may consist of households from each stratum so that sample may contain all the important characteristics of the village population. In the case of SRS, the sample of all strata/ sub groups sometimes may not be included or covered adequately.

- This method helps in conducting and managing a large scale survey to be conducted in a country like India. The agency conducting the survey may have field offices in different locations; each one can supervise the survey for a part of the population.
- The basic idea is that it sub-divides the heterogeneous population into homogeneous sub-populations. If each stratum is homogenous in itself, a precise estimate of any stratum mean can be obtained from a small sample, thus, saving a lot of time and cost.

There are two types of stratified samples.

A **proportionate stratified sample** selects the number of elements from each stratum so that the stratum sample size (n_1, n_2,……., n_k) is proportional to their respective stratum population size (N_1, N_2, ……., N_k).

Consider the following examples:

- A selected village may have households of SC(10%), ST (5%), OBCs (45%), Others (30%), Minority (10%). A village sample of 100 may constitute the households of various casts in the above proportion/ percentage so that the sample may contain all important characteristics of village population.

- Hospital patients are stratified according to age, dividing the population into those who are aged 50years or above, and, those who are under 50. If there are twice as many people aged 50 or above admitted to the hospital as those under 50, a proportionate stratified sample will include twice as many people aged 50 or above.

A **disproportionate stratified sample** selects the number of elements from each stratum so that the stratum sample size is not proportional to the stratum population size. The most common reason for selecting this type of sample is when you want to study a relatively rare but important subpopulation, such as younger patients suffering from heart disease. Proportionate stratification may result in too few elements being selected so that little, if any, statistical analysis can be done. Consequently, even if these patients represent only 1% of the population, you might decide to make them 10% of the final sample. However, once we combine values of all stratums, the size of the higher selected proportion needs to be readjusted which is called weighted estimate.

4) Probability Proportion to Size (PPS) Sample

It has been observed that the elementary units of the population vary in size. Such ancillary information about the size of the unit can be utilized in selecting the sample so as to get better and efficient estimates of the population parameter. For example villages with larger geographical area are likely to have larger area under food crops; therefore, in estimating the production, it would be desirable to adopt a sampling scheme in which villages are selected with probability proportional to geographical area. When units vary in their size and the variable under study is directly related with the size of the unit, the probabilities may be assigned proportional to the size of the unit.

Probability Proportion to Size (PPS) Sampling assures higher probability of selection to sampling unit which are larger in size. This technique was initially used in estimation of crop production, fruits production etc because productivity is directly related with the size of field. In social science surveys also characteristics of village population is influenced by the size of population. The **procedure of selecting the sample** is described below.

Suppose you have to select 5 villages from the list of 10 using PPS sampling. First arrange all villages in ascending or descending order of population size as may be seen in column 2 of the table 1. Then, in the third column, find the cumulative sum of population size and in the fourth column, assign them range of serial numbers as shown below in the table.

Table 4.1: Village population Size

Sl.No.	Village Population Size	Cumulative Sum of Population Size	Cumulative Population Size Interval
1	**2**	**3**	**4**
1	200	200	0001 - 0200
2	250	450	0201 - 0450
3	300	750	0451 - 0750
4	350	1100	0751 - 1100
5	400	1500	1101- 1500
6	450	1950	1501 - 1950
7	500	2450	1951 - 2450
8	550	3000	2451 - 3000
9	600	3600	3001 - 3600
10	650	4250	3601 - 4250
Total	**4250**		

Please notice that the total population of all villages in the target population is a four digit number (4250). Therefore, initially, a random number in four digits, which is less than or equal to the total population of all villages (4250), is selected from the random number table. For example, it is 0331 which will correspond to serial number 2. Next random number is 4320; therefore, it may be discarded. The next number selected is 1296; therefore, it will correspond to serial number 5. The next random numbers may be 1553, 2402 and 3640 which will correspond to serial numbers 6, 8, and 10 respectively. In this way, selected villages will be serial numbers 2, 5, 6, 8, 10.

5) Cluster Sample

Cluster sampling is a sampling technique used when natural groupings are evident in a statistical population. It is often used in marketing research. In this technique, the total population is divided into these known groups (or clusters) and a sample of the groups is selected. Then the required information is collected from the elements within each selected group. This may be done for every element in these groups, or a sub sample of elements may be selected within each of these groups. The technique works best when most of the variation in the population is within the groups, not between them.

Briefly, the procedure for selecting a cluster sample is given below.

- The population is divided into N groups, called clusters.
- The researcher randomly selects n clusters to include in the sample.
- The number of observations within each cluster is known:

  M = M1 + M2 + M3 + ...…… + MN
- Each element of the population can be assigned to one, and only one, cluster.

Cluster sampling should be used only when it is economically justified - when reduced costs can be used to overcome losses in precision. This is most likely to occur in the following situations.

- Constructing a complete list of population elements is difficult, costly, or impossible. For example, it may not be possible to list all elementary units of the populations, for example all households in village, block, etc. However, it would be possible to randomly select a subset of villages, blocks (stage 1 of cluster sampling) and, then, interview the head of family in a house of the selected cluster (stage 2).
- The population is concentrated in natural clusters (city blocks, schools, hospitals, etc.). For example, to conduct personal interviews of operating room nurses, it might make sense to randomly select a sample of hospitals (stage 1 of cluster sampling) and then interview all of the operating room nurses at that hospital. Using cluster sampling, the interviewer could conduct many interviews in a single day at a single hospital. Simple random sampling, in contrast, might require the interviewer to spend all day travelling to conduct a single interview at a single hospital.

As discussed above, in the cluster sampling method, the primary selecting unit is not a household, rather a natural cluster of households, viz., hamlets in villages, or, created clusters, viz., schools, malls, etc., may be decided. The first list of clusters may be selected using the SRS or the PPS sampling techniques. Then, from each selected cluster, all units, or, some of the units, may be selected as per the required sample size using Stratified Random Sampling or the Systematic Random Sampling techniques.

This sampling technique is quite popular in evaluation surveys in health – it is also called the 30 Cluster Sampling Technique. This is also a rapid method of data collection as the researcher can collect more data in less time due to the decrease in transportation time as compared with other sampling techniques.

4.3.2 Non-Probability Sampling

A **non-probability sample** is one in which a case in a sample is chosen in such a manner that it gives you information for the sample itself and makes it possible to generalize the findings for the population with certain degree of precision. Such a sample is also called a purposive sample. This kind of sampling is primarily used to collect information on market surveys to know the attitude, opinion, behaviour, reactions of individuals. There are many types of non-probability samples, including snowball sampling, convenience, purposive/ judgment, quota sampling, etc.

1) Convenience Sample

The convenience sample is so called because it is relatively easy to obtain and contact. In this method the investigators are usually asked to select the people for the interview in accordance to the instructions from the researcher. The benefit of a convenience sample is that the interviewer can usually get interviews done quickly and cheaply. Convenience sampling is appropriate for exploratory research.

2) Judgments Sample

A judgment sample is similar to that of convenience sample. In a judgment sample, the researcher selects samples that are believed to represent the population. The selection of samples is based on the knowledge of the population and the characteristics which the sample is to represent. It is less costly and very useful for forecasting.

3) Quota Sample

Quota sampling is like stratified sampling. In quota sampling, the population is categorized into several strata which consist of an expected size, and the samples are considered to be important for the population they represent. The advantages of quota sample are that it involves a short time duration, is less costly, and gives moderate representation to a heterogeneous population.

4) Snowball Sample

This is one of the important types of non-probability sampling. In snowball sampling, the investigator encourages the respondents to give the names of other acquaintances and it continues growing in size and chains until the research purpose is achieved. It is also, therefore, known as networking, chain, or referred sampling method. It is very useful in the study of networking and is less costly.

A comprehensive overview of the various types of sampling can be seen in figure 4.1

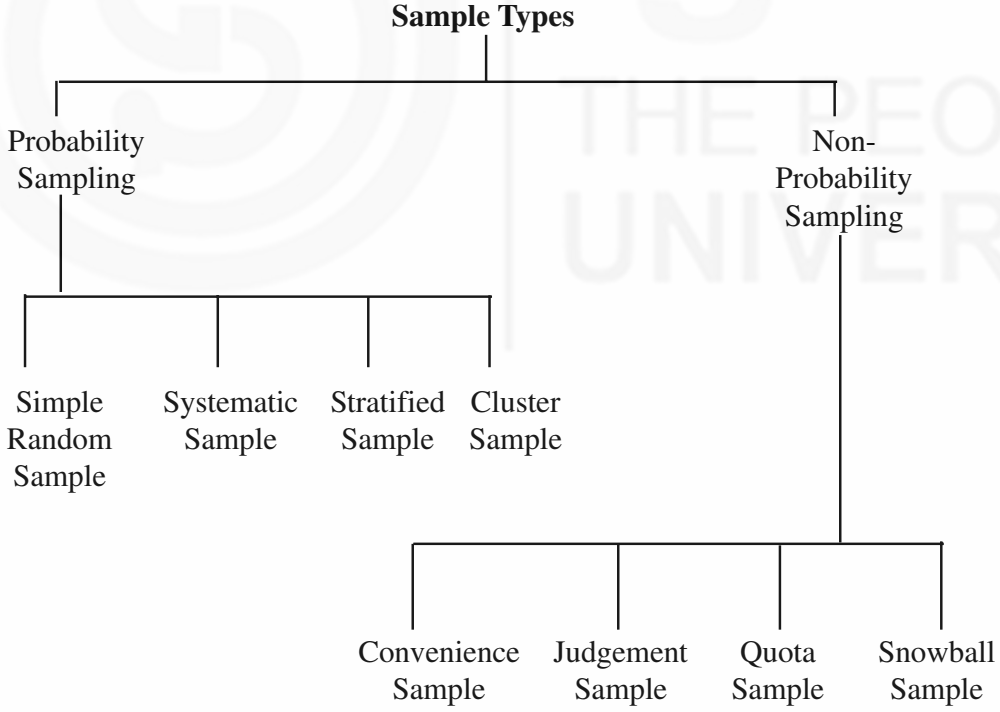
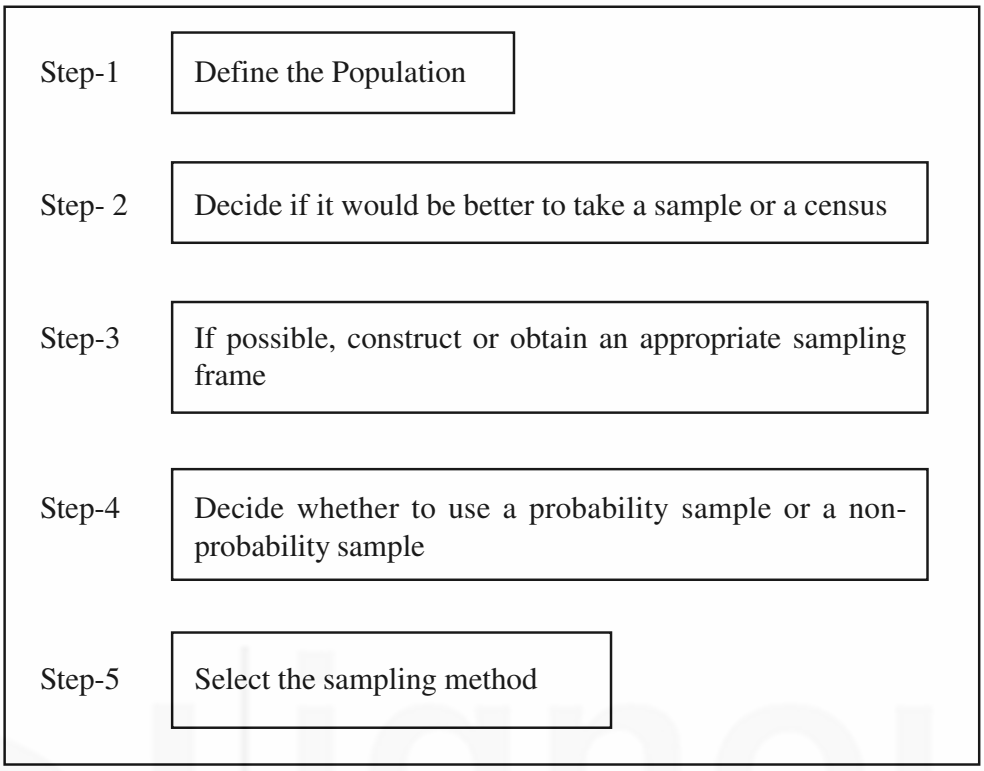


Fig. 4.1: Types of Sample

4.4 SAMPLE DESIGN PROCESS

The sample design process follows five steps as given in Box-1

Box 4.1: Sample Design Process



Source: John Boyce (www.mhhe.com/av/boyceze)

Step-1: Define the Population

We use the word, population, frequently in our day-to-day conversations, for example, 'The population of India', or, 'The population of Punjab', or, 'The population of Kerela'. However, the meaning of the word, population, in research is different from what we use in day-to-day conversation. A research population may be defined as 'a clearly defined group of entities that have some characteristics in common'. This means the kind of people on whom we wish to base our research project. Sometimes, in research, we use the word, universe, instead of population.

In a research project, our intention is to learn or infer something about the population. Whether we would use a sample or a population has to be clearly defined. For example, if we want to conduct a study on road safety, then the task of defining the population for a survey would be whether we should

- interview only the people who drive two wheelers
- interview only the people who drive four wheelers
- interview the pedestrians
- interview only who are hand rickshaw pullers or ride bi-cycle.

Therefore, judging a population is the starting of the sampling process.

Step-2: Decide whether to take a sample or a census

After judging the population, the next step in the sampling process is to decide whether to take a sample or a population in your research project. In a census, usually every member of the population is interviewed.

While in a sample method only selected members of the population are included. From the census we obtain data that are called population parameter, and from the sample we obtain statistics in a parameter. A parameter is a measurement of a characteristic of a population, while a sample statistic is used as an estimate of a population parameter.

Usually, in individual research, we use census when the population is small, and sample when the population is large.

Step-4: Decide whether to use a probability sample or a non-probability sample

The fourth step in the sampling process is whether to use probability sampling or non-probability sampling. In a probability sample, the sample elements are chosen by random selection, while in non-probability sampling, each sample element is chosen according to whether the researcher decides that it should be included or not.

Step-5: Select the sampling method

Last, but not least in the sampling process, is the selection of the sampling method. In the probability sampling method, the following four principal kinds of probability sampling are used: the simple random sample, the systematic sample, the stratified sample, and the cluster sample. The main non-probability sampling methods are the convenience sample, the judgment sample, the quota sample, and the snowball sample.

Sampling Frame

A sampling frame is a listing of all the elements from which you will draw the sample. In the ideal situation, the sampling frame will include all elementary units in the target population. A list of employees in an organization can create a sampling frame that exactly matches the population of interest. You should try to ensure that the sampling frame has the following characteristics.

- It is actually created from the target population.
- It is as complete a list as possible of the elements in the population.

In this section, you have read about the various types of sampling and the sample design process. Now, answer the following questions in *Check Your Progress 2.*

Check Your Progress 2

Note: a) Write your answer in about 50 words.

b) Check your answer with possible answers given at the end of the unit

1) What is a stratified sample?

.....

.....

.....

.....

.....

2) What is a cluster sample and when is it best used?

.....

.....

.....

.....

.....

.....

3) What do you understand by a quota sample?

.....

.....

.....

.....

.....

.....

4.5 ERRORS IN SAMPLING

Many mistakes and errors in social science research happen because of misleading and biased sampling. A sample which does not represent the population is called a biased sample. According to Yule and Kendal, "Bias may be due to imperfect instruments, the personal qualities of the observer, defective techniques and other cases. Like experimental error, it is difficult to eliminate entirely, but usually may be reduced to relatively small dimensions by taking proper care." There are two types of errors such as sampling errors and non-sampling errors. These are discussed below:

i) Sampling Error

By definition, when you have collected a sample from a population, you have less than complete information about the population. This, in turn, means that there is a chance that the sample statistics you calculate, (for example, the mean of a variable, a frequency distribution, etc.) may not be an unbiased estimate of the population parameter.

The error in the sample estimate is not an intrinsic impediment to analysis. For probability samples, sampling theory allows you to calculate the expected amount of error given a particular sample size, sampling method, and the specific statistic of interest. In general terms, the sampling error for a statistic can be defined as:

$$\text{Standard error} = \sqrt{\frac{Variance}{n}} = \frac{\text{sd}}{\sqrt{\text{n}}}$$

Where n refers to the number of respondents (sample size).

As the sample size increases, the standard error of a statistic decreases; as the variance, or dispersion, of a statistic increases, so does its sampling error.

Sampling error decreases rapidly as the sample size increases from a few hundred to about 1000 respondents. However, there is rarely any reason to select larger samples while comparing the increased cost of survey with reduction in sampling error (see 'Calculating the Sample Size', in next section).

The formula for the standard error of a proportion is simple and easy to apply:

Standard error = $\sqrt{\frac{p \times (1-p)}{n}}$

Here, p represents the proportion of successes (favourable response, those who received the benefits), {q = (1-p)} represents the proportion of failures (those who did not receive the benefits), and n is the total number of respondents. The standard error of a statistic is greatest when p and (1-p) are equal, which occurs when each is 0.50, or 50%, of the sample.

ii) Non-Sampling Error

Before discussing how to determine sample size, we will briefly review other sources of error in surveys. When you read a a news article that reports the results of a national poll, the error in the estimates is always listed, derived, generally speaking, from Equation 6.2. However, experienced survey researchers know that errors due to other sources are typically greater than the error due to sampling alone. Following are some other types of errors.

- Measurement errors, caused by poorly written questions, poorly designed questionnaires, respondent errors in completing questionnaires, and so on.
- Non-response errors, caused because the respondents are not a representative subset of the population.
- Data coding errors, caused, by errors in coding and entering the data.

Of these error sources, the first two are typically more severe. In mail surveys, non-response error is often the most serious problem.

There are two critical characteristics of these non sampling errors. First, as mentioned above, their sum is often greater than the sampling error. Second, and more insidious, these errors are often impossible to estimate for any one survey, especially measurement and non-response errors. Consequently, using Equation 6.1 and Equation 6.2 to estimate the error in a statistics often provides a false sense of security.

Experienced survey researchers take this fact into account by being more cautions in discussing survey results than the sampling error alone would indicate, and you should do the same. Ideally, the other sources of error would balance themselves out so that errors in one direction negate errors in the other directions, but you cannot assume that this is the case.

4.6 DETERMINATION OF SAMPLE SIZE

The sample size can be determined by:

i) Using a formula

ii) Using a table

4.6.1 Determining Sample Size Using a Formula

(when population is greater than 10,000)

$$nf = \frac{n}{1+(n/N)}$$ (when population is less than 10,000)

n, nf = desired sample size

Z = the standard normal deviate

p = the portion in the target population estimated to have a particular characteristic. If there is no reasonable estimate, then use 50 percent (.50).

q = 1-p

d= degree of accuracy desired, usually set at .05 or occasionally at .02.

n = the estimate of the population size

Z at 99% confidence level i.e. at 1% level of significance = 2.58

Z at 95% confidence level i.e. at 5% level of significance = 1.96

Z at 90% confidence level i.e. at 10% level of significance = 1.65

Exaple: (when population is more than 10,000)

If the proportion of target population with a certain characteristic is .50, the Z statistic 1.96 and we desire accuracy at the 0.05 level, then the sample size is

$$n = \frac{(1.96)^2\ (.50\times.50)}{(0.05)^2}$$

$$= \frac{3.84\times0.25}{(.0025)}$$

$$= \frac{0.96}{0.0025}$$

$$= 384$$

If we use the more convenient 2.0 for the Z statistic, then the sample size will be smaller.

=

= 286

4.6.2 Determining Sample Size by Using a Table

Another way to determine sample size is to rely on published tables which provide the sample size for a given set of criteria. Table 1 presents sample size values that will be appropriate for many common sampling problems. The table includes sample sizes for both continuous and categorical data assuming alpha levels of .10, .05, or .01.

Table 4.1: Table for Determining Minimum Returned Sample Size for a Given Population Size for Continuous and Categorical Data

	Sample size					
	Continuous data (margin of error= .03)			Categorical data (margin of error= .05)		
Population size	alpha=.10 t=1.65	alpha=.05 t=1.96	alpha=.01 t=2.58	p=.50 t=1.65	p=.50 t=1.96	p=.50 t=2.58
100	46	55	68	74	80	87
200	59	75	102	116	132	154
300	65	85	123	143	169	207
400	69	92	137	162	196	250
500	72	96	147	176	218	286
600	73	100	155	187	235	316
700	75	102	161	196	249	341
800	76	104	166	203	260	363
900	76	105	170	209	270	382
1,000	77	106	173	213	278	399
1,500	79	110	183	230	306	461
2,000	83	112	189	239	323	499
4,000	83	119	198	254	351	570
6,000	83	119	209	259	362	598
8,000	83	119	209	262	367	613
10,000	83	119	209	264	370	623

In this session you studied about errors in sampling and determination of sample size. Now, answer the questions given in *Check Your Progress 3*.

Check Your Progress 3

Note: a) Write your answer in about 50 words.

b) Check your answer with possible answers given at the end of the unit

1) What is sampling error?

.....

.....

.....

.....

.....

2) How the sample size is determined using the formula?

.....

.....

.....

.....

.....

4.7 LET US SUM UP

In this unit, we discussed the meaning and various concepts in sampling particularly of sample and population. There is also a detailed discussion on the sample types and sample design process. There are two types of sampling such as probability and non-probability sampling. The types of probability sampling are the Simple Random Sample, the Systematic Sample, the Stratified Sample, and the Cluster Sample, while different types of non-probability sample are the Convenience Sample, the Quota Sample, the Judgment Sample, and the Snowball Sample. The unit also discusses various steps of the sampling design process. This is followed by two of the very important concepts of sampling: the determination of sample size and errors in sampling.

4.8 KEYWORDS

Sample : A sample is simply a subset of a larger aggregation, i.e., typically a population and it contains all the characteristics of a population,

Sampling : The process of selection of subjects/study elements to create a sample for collecting information about a population.

Standard Error : This is the expected amount of error while estimating the specific statistic of interest, using a particular sample size and sampling method with respect to actual population value.

Sampling Error : While collecting information from a sample, there is a chance that the sampling statistics may not be equal to the same values in the population. The error is that the sample does not contain complete information about the population.

Confidence Interval : This gives the probability of the sample estimate falling within the interval.

Sample Size : The number of elementary units in a sample is called a sample size.

4.9 REFERENCES AND SELECTED READINGS

Daroga S. and F.S. Chaudhary (2002), *Theory & Analysis of Sample Survey Designs,* New Age International Publishers Ltd.

Goldberg H. I. (1991), *Survey Sampling: In Epidemiological Approach to Reproductive Health*, WHO.

Henderson R. H and Sundaresan T (1982), *Cluster Sampling to Assess Immunisation Coverage: A Review of Experience with a Simplified Sampling Method*, WHO Bulletin.

Pandurang V. S. (1984), *Sampling Theory of Surveys with Applications,* Iowa State Press.

Cochran W. G. (1977), *Sampling Technique*, John Willey & Sons.

4.10 CHECK YOUR PROGRESS – POSSIBLE ANSWERS

Check Your Progress 1

1) What do you mean by sampling? What are the advantages of sampling?

   Generally, a sample implies a small representative of a large whole. This sampling method is frequently used in social science research to save time. Some of the key advantages of sampling are: (i) it costs less; (ii) takes less time; (iii) data are sometimes wanted quickly; (iv) fewer mistakes are likely; (v) a more detailed study can be done.

2) What is the difference between a parameter and an estimator?

   Any function of the values of units in the population, such as the population mean or population variance, is termed, a population parameter. There can only be one set of values for a population, there population values are treated as constant. However, the function of the values of the units in the sample, such as the sample mean and sample variance is known as a statistic. The value of the mean and variance differs from sample to sample and, therefore, it is a random variable.

Check Your Progress 2

1) What is stratified sampling?

In *stratified sampling*, the target population of N units is first divided into k subpopulations of units. These populations are non-overlapping and together they comprise the whole population, so that

The sub-populations are called strata. The number in each stratum should be known. A sample is drawn from each stratum independently. The sample sizes within 'k' strata are denoted by respectively If the total sample size n is to be drawn from the target population then

If a simple random sample is drawn in each stratum, the whole procedure is described as *stratified random sampling*.

2) What is cluster sampling and when is it best used?

Cluster sampling is a sampling technique used when natural groupings are evident in a statistical population. It is often used in marketing research. In this technique, the total population is divided into these known groups (or clusters) and a sample of the groups is selected. Then, the required information is collected from the elements within each selected group. This may be done for every element in these groups or a sub sample of elements may be selected within each of these groups. The technique works best when most of the variation in the population is within the groups, not between them.

3) What do you understand by quota sample?

Quota sampling is like stratified sampling. In quota sampling, the population is categorized into several strata which consist of an expected size and they are considered to be important for the population they are supposed to represent. The advantages of the quota sample are: shorter time duration, less costly, and gives moderate representation to a heterogeneous population.

Check Your Progress 3

1) What is sampling error?

By definition, when you have collected a sample from a population, you have less than complete information about the population. This, in turn, means that there is a chance that the sample statistics you calculate, (for example, the mean of a variable, a frequency distribution, etc.) may not be unbiased estimate of the population parameter. This error is called sampling error.

2) How is the sample size determined using the formula?

The calculation of the sample size is concerned with the number of respondents required. To determine the number to select for the sample drawn from the sampling frame, you must estimate the non-response rate. The actual sample size to be drawn is:

So, if any survey organization decides that they need 700 respondents, and the expected response rate from the population is 50%, then 700/0.50, or 1400, customers must be drawn from the sampling frame.