CHAPTER FOUR

4. SAMPLING DESIGN

4.1.1 CENSUS AND SAMPLE SURVEY

All item in any field of inquiry constitute a ‘Universe or population’. A complete enumeration of all item in the population is known as a census inquiry. It can be presumed that in such an inquiry, when all item are covered no element of chance is left and highest accuracy is obtained. But in practice this may not be true.

Even the slightest element of bias in such an inquiry will get larger and larger as the number of observation increase. Moreover there is no way of checking element of bias or its extent except through a resurvey or use of sample checks. Beside this type of inquiry involve a great deal of time, money and energy. Therefore when the field of inquiry is large this method become difficult to adopt because of the resource involved. At time this method is practically beyond the reach of ordinary researchers. Perhaps, government is the only institution which can get the complete enumeration carried out.

Even government adopts this in very rare cases such as population census once in a decade. Further many time it is not possible to examine every item in the population and sometimes it is possible to obtain sufficiently result by study only part of total population. In such case there is no utility of census surveys.

However it needs to be emphasized that when the universe is small one, it is no use resorting to a sample survey. When field studies are undertaken in practical life, consideration of time and cost almost invariably lead to a selection of respondents i.e., selection of only few items. The respondent selected should be as representative of the total population as possible in order to produce a miniature cross-section. The selected respondent constitute what is technically called a sample and the selection process is called sampling technique. The survey so conducted is known as sample survey. Algebraically let the population size be N and if a part of size n(which is <N) of this population is selected according to some rule for studying some characteristics of the population, the group consisting of these n units is known as ‘sample’. Research must prepare a sample design for his study i.e., he must plan how a sample should be selected and of what size such a sample would be.

4.1.2 IMPLICATION OF A SAMPLE DESIGN

A sample design is a definite plan for obtaining a sample a given population. It refers to the technique or the procedure the researcher would adopt in selecting item for the sample. Sample design may as well lay down the number of items to be included in the sample i.e., the size of the sample. Sample design is determine before data are collected. There are many sample design from which a researcher can choose. Some design are relatively more precise and easier to apply than others. Researcher must select /prepare a sample design which should be reliable and appropriate for his research study.

4.1.3 STEP IN SAMPLING DESIGN

While developing sampling design the researcher must pay attention to the following points:

(i). Type of Universe: the first step in developing any sample design is to clarify defines the set of objects, technically called the universe, to be studied. The universe can be finite or infinite. In finite universe the number of item is certain, but in case of infinite universe the number of item is infinite i.e., we cannot have any idea about the number of items. The population of the city, the number of workers in a factory and the like are example of finite Universe, whereas number of star in the sky, listeners of specific radio program, throwing of a dice etc. are example of infinite Universe.

(ii). Sampling unit: A decision has to be taken concerning sampling unit before selecting sample. Sampling may be geographical one such as state, district, village, etc., or a construction unit such as house, flat, etc., or it may be social unit such as family, club, school, etc., or it may be individual. The researcher will have to decide one or more of such units that he has to select for his study.

(iii). Source list: it is also known as sampling frame from which sample is to be drawn. It contains the names of all items of a universe (in case of finite universe only). If source list is not available researcher has to prepare it. Such list should be comprehensive, correct, reliable and appropriate. It is extremely important for the source list to be as representative of the population as possible.

(iv). Size of sample: this refers to the number of item to be selected from the universe to constitute a sample. This major problem before a researcher. The size of sample should neither be excessively large, nor too small. It should be optimum. An optimum sample is one which fulfills the requirements of efficiency, representativeness, reliability and flexibility. While deciding the size of sample, researcher must determine the desired precision as also an acceptable confidence level for the estimate. The size of population variance needs to be considered as in case of larger variance usually a bigger sample is needed. The size of population must be kept in view for this also limits the sample size. The parameter of interest in research study must be kept in view, while deciding the size of the sample. Costs to dictate the size of sample that we can draw. As such budgetary constraint must invariably be taken into consideration when we decide the sample size.

(v). Parameters of interest: In determining the sample design, one must consider the question of the specific population parameters which are of interest. For instance we may be interested in estimating the proportion of person with some characteristics in the population, or we may be interested in knowing some average or the other measure concerning the population. There may also be important sub-groups in the population about whom we would like to estimate. All this has a strong impact upon the sample design we would accept.

(vi). Budgetary constraint: Cost consideration, from practice point of view, have a major impact upon decision relating to not only the size of the sample, but also to the type of sample. This fact can even lead to the use of a non- probability sample.

(vii). Sampling Procedure: Finally, the researcher must decide the type of sample he will use i.e., he must decide about the technique or procedure stand for the sample design itself. There are several sample designs out of which the researcher must choose one for his study. Obviously he must select that design which for a given sample size and for a given cost, has a smaller sampling error.

4.1.4 Criteria of Selecting a Sampling Procedure

In this context one must remember that two costs are involved in sampling analysis viz.., the cost of collecting data and the cost of an incorrect inference resulting from the data. Researcher must keep in view the two cause of incorrect inference viz.., systematic bias and sampling error. A systematic bias result from errors in the sampling procedures, and it cannot be reduced or eliminated by increasing the sample size. At best the causes responsible for these errors can be detected and corrected. Usually a systematic bias is the result of one or more of the following factors:

1. **Inappropriate sampling frame:** If the sampling frame is inappropriate i.e., a biased representation of the universe it will result in a systematic bias
2. **Defective measuring device:** If the measuring device is constantly in error, it will result in systematic bias. In survey work, systematic bias can result if the questionnaire or the interviewer is biased. Similarly, if the physical measuring device is defective thee will be systematic bias in the data collected through such a measuring device.
3. **Non-respondents**: If we are unable to sample all the individual initially included in the sample there may arise a systematic bias. The reason is that in such a situation the likelihood of establishing contact or receiving a response from an individual is often correlated with the measure of what is to be estimated.
4. **Indeterminacy principle:** Sometimes we find that individual act differently when kept under observation than what they do when kept in non- observed situations. For instance, if workers are aware that somebody is observing them in course of a work study on the basis of which the average length of time to complete a task will be determined and accordingly the quota will be set for piece work, they generally tend to work slowly in comparison to the speed with which they work if kept unobserved. Thus the indeterminacy principle may also be a cause of a systematic bias.
5. **Natural bias in the reporting of data:** natural bias of respondent in the reporting of data is often the cause of a systematic bias in many inquiries. There is usually a downward

bias in the income data collected by government taxation department, whereas we find an upward bias in income data collected by some social organization. People in general understate their income if asked about it for tax purpose, but they overstate the same if asked them for social status or their affluence. Generally in psychological survey people tend to give what they think is the correct answer rather than revealing their true feeling.

Sampling error are the random variation in the sample estimation around the true population parameters. Since they occur randomly and are equally likely to be in either direction their nature happens to be of compensatory type and the expected value of such errors happens to be equal to zero. Sampling error decrease with the increase in the size of the sample, and it happen to be of a smaller magnitude in case of homogeneous population.

Sampling error can be measured for a given sample design and size. The measurement of sampling error is usually called the ‘precision of the sampling plan’. If we increase the sample size the precision can be improved.

4.2 Type of Sampling Design

There are different type of sample design based on two factors viz.., the representation basis and the element selection technique. On the representation basis the sample may be probability sampling or it may be non- probability sampling. Probability sampling is based on the concept of random selection, whereas non- probability sampling is is ‘non-random’ sampling.

4.2.1 Non- probability Sampling:

Non-probability is the sampling procedure which does not afford any basis for estimating the probability that each item in the population has of being included in the sample. Non-probability sampling is also known by different name such as deliberate sampling, purposive sampling and judgmental sampling. In this type of sampling item for the sample are selected deliberately by the researcher; his choice concerning the item remains supreme. In other words, under non- probability sampling the organization of the inquiry purposively choose the particular units of the universe for constituting a sample on the basis that the small mass that they so select out of a huge one will be typical or representative of the whole. For instance if economic condition of people living in a state are to be studied a few towns and village may be purposively selected for intensive study on the principle that they can be representative of the entire state. Thus the judgment of the organization of the study plays an important part in this sampling design.

In such design personal element has a great chance of entering into the selection of the sample. The investigator may select a sample which shall yield results favorable to his point of view and if that happens, the entire inquiry may get vitiated. Thus there is always the danger of bias entering into this type of sampling technique. But in the investigators are impartial, work without bias and have the necessary experience so as to take sound judgment, the result obtained from an analysis of deliberately selected sample may be tolerably reliable. However in such sampling there is no assurance that every element has some specifiable chance of being included. Sampling error in this type of sampling cannot be estimated and the element of bias, great or small, is always there. As such this sampling design in rarely adopted in large inquires of importance. However in small inquires and researches by individual, this design may be adopted because of the relative advantage of time and money inherent in this method of sampling.

Quota sampling also an example of non-probability sampling. Under quota sampling the interviewers are simply given quotas to be filled from the different strata, with some restriction on how they are to be filled. In other words the actual selection of the item for the sample is left the interviewers discretion. This type of sampling is very convenient and is relatively inexpensive. But the sample so selected certainly do not possess the characteristic of random sample. Quota samples are essentially judgment samples and inferences drawn on their basis are not amenable to statistical treatment in a formal way.

4.2.2 Probability Sampling

Probability sampling is also known as ‘random sampling’ or ‘chance sampling’. Under this sampling design every item of the universe has an equal chance of inclusion in the sample. It is, so to say a lottery method in which individual units are picked up from the whole group not deliberately but by some mechanical process. Here it is blind chance alone that determine whether one item or the other is selected.

The result obtained from probability or random sampling can be assured in terms of probability i.e., we can measure the error of estimating or the significance of result obtained from a random sample and this fact bring out the superiority of random sampling design over the deliberate sampling design. Random sampling ensures the law of statistical regularity which states that if on an average the sample chosen is a random one, the sample will have the same composition and characteristics as the universe. This is the reason why random sampling is considered as the best technique of selecting a representative sample.

Random sampling from a finite population refers to that method of sample selection which gives each possible sample combination an equal probability of being picked up and each item in the entire population to have an equal chance of being included in the sample. This applies to sampling without replacement i.e., once an item is selected for the sample it cannot appear in the sample again.

Sampling with replacement is used less frequently in which procedure the element selected for the sample is returned to the population before the next element is selected. In such a situation the same element could appear twice in the same sample before the second element is chosen. In brief the implications of random sampling (or simple random sampling) are :

(a) It give each element in the population an equal probability of getting in to the sample and all choices are independent of one another.
(b) It give each possible sample combination an equal probability of being chosen. Keeping this in view we can define a simple random sample or simply random

sampling) from a finite population as a sample which is chosen in such a way that each of the NCn possible samples has the same probability, 1/NCn of being selected. To make it more clear we take a certain finite population consisting of six elements (say a, b, c, d, e, f) i.e., N=6. Suppose that we want to take a sample of size n=3 from it. Then there are 6C3=20 possible distinct samples of the required size, and they consist of the elements abc, abd,abe, abf, acd, ace, acf, ade, adf, aef, bcd, bce, bcf, bde, bdf, bef, cde, cdf, cef and def. if we choose one of these samples in such a way that each has the probability 1/2o being chosen we will then call this a random sample.

4.2.2.1 Random Sample from an infinite population

So far we have talked about random sampling, keeping in view only the finite populations. But what about random sampling in context of infinite population? It is relatively difficult to explain the concept of random sample from an infinite population. However, a few examples will show the basic characteristics of such a sample. Suppose we consider the 20 throws of a fair dice as a sample from the hypothetically infinite population which consists of the results of all possible throws of the dice. If the probability of getting a particular number, say 1, is the same for each thrown and the 20 thrown are all independent, then we say that the sample is random. Similarly it would be said to be sampling from an infinite population if we sample with replacement from a finite population and our sample would be considered as a random sample if in each draw all elements of the population have the same probability of being selected and successive draws happen to be independent. In brief one can say that the selection of each item in a random sample from an infinite population is controlled by the same probability and that successive selections are independent of one another.

4.2.2.2 Complex random sampling design

Probability sampling under restricted sampling techniques, as stated above, may result in complex random sampling designs. Such design may as well be called 'mixed sampling design' for many of such designs may represent a combination of probability and non- probability sampling procedure in selecting a sample. Some of the popular complex random sampling designs are as follows:

(i). Systematic sampling: In some instance, the most practical way of sampling is to select every i^{th} item on a list. Sampling of this type is known as systematic sampling. An element of randomness is introduced into this kind of sampling by using random number to pick up the unit with which to start. For instance if a 4% sample is desired the first item would be selected randomly from the first twenty- five and thereafter every 25^{th} item would automatically be included in the sample. Thus in systematic sampling only the first unit is selected randomly and the remaining units of the sample are selected at fixed intervals. Although a systematic sample is not a random sample in the strict sense of the term, but it is often considered reasonable to treat systematic sample as if it were a random sample.

Systematic sampling has certain plus points. It can be taken as an improvement over a simple random sample in as much as the systematic sample is spread more evenly over the entire population. It is an easier and less costly method of sampling and can be conveniently used even in case of large populations. But there are certain dangers too in using this type of sampling. If there is hidden periodicity in the population, systematic sampling will prove to be an inefficient method of sampling.

For instance, every 25^{th} item produced by certain production process is defective. If we are to select a 4% sample of the item of this process in systematic manner, we would either get all defective item or all good item in our sample depending upon the random starting position. If all element of the universe are ordered in a manner representative of the total population, i.e., the population list is in random order, systematic sampling is considered equivalent to random sampling. But if this is not so then the result of such sampling may at times not be very reliable. In practice systematic sampling is used when lists of population are available and they are of considerable length.

(ii). Stratified Sampling: if a population from which a sample is to be drawn does not constitute a homogeneous group, stratified sampling technique is generally applied in order to obtain a representative sample. Under stratified sampling the population is divided in to several sub-population that are individually more homogeneous than the total population (the different sub-population are called strata) and then we select item from each stratum to constitute a sample. Since each stratum is more homogeneous than the total population, we are able to get more precise estimates for each stratum and by estimating more accurately each of the component parts, we get better estimate of the whole. In brief stratified sampling result in more reliable and detailed information.

The following three questions are highly relevant in the context of stratified sampling:

(a) How to form strata?
(b) How should item be selected from each stratum?
(c) How many items be selected from each stratum or how to allocate the sample size of each stratum?

Regarding the first question, we can say that the strata be formed on the basis of common characteristic(s) of the item to be put in each stratum. This means that various strata be formed in such a way as to ensure elements being most homogeneous within each stratum and most heterogeneous between the different strata. Thus, strata are purposively formed and are usually based on past experience and personal judgment of the researcher. One should always remember that careful consideration of the relationship between the characteristics of the population and the characteristics to be estimated are normally used to define the strata. At times pilot study may be conducted for determining a more appropriate and efficient stratification plan. We can do so by taking small samples of equal size from each of the proposed strata and then examining the variances within and among the possible stratification, we can decide an appropriate stratification plan for our inquiry.

In respect of the second question, we can say that the usual method, for selection of item for the sample from each stratum, resorted to is that of simple random sampling. Systematic sampling can be used if it is considered more appropriate in certain situations.

Regarding the third question, we usually follow the method of proportional allocation under which the size of the sample from different strata are kept proportional to the size of the strata. That is if pi represent the proportion of population included in stratum i, and n represent the total sample size, the number of element selected from stratum i is n. pi. To illustrate it, let us suppose that we want a sample of size N1=4000, N2=2400 and N3=1600. Adopting proportional allocation, we shall get the sample sizes as under for the different strata: For strata with N1=4000 we have p1= 4000/8000 and hence n1= n*p1=30*(4000/8000)=15. Similarly, for strata with N2=2400, we have n2 =n*p2 =30(2400/8000)=9, and for strata with N3= 1600, we have n3= n*p3 30*(1600/8000)= 6

Thus, using proportional allocation the sample sizes for different strata are 15, 9 and 6 respectively which is in proportion to the sizes of the strata viz., 4000: 2400: 1600. Proportional allocation is considered most efficient and an optimal design when the cost of selecting an item is equal for each stratum, there is no difference in within stratum variances, and the purpose of sampling happens to be to compare the difference among the strata, then equal sample selection from each stratum would be more efficient even if the strata differ in size,

(iii). Cluster sampling: If the total area of interest happens to be a big one, a convenient way in which a sample can be taken is to divide the area into a number of smaller non-overlapping areas and then to randomly select a number of these smaller areas (usually called cluster), with the ultimate sample consisting of all (or samples of) units in these small areas or clusters.

Thus in cluster sampling the total population is divided into a number of relatively small subdivisions which are themselves clusters of still smaller units and then some of these clusters are randomly selected for inclusion in the overall sample. Suppose we want to estimate the proportion of machine parts in the inventory which are defective. Suppose we want to estimate the proportion of machine parts in an inventory which are defective. Also assume that there are 20,000 machine parts in the inventory at the given point of time, stored in 400 cases of 50 each. Now using a cluster sampling we would consider the 400 cases as clusters and randomly select ‘n’ cases and examine all the machine parts in each randomly selected case.

Cluster sampling, no doubt, reduces cost by concentrating survey in selected clusters. But certainly it is less precise than random sampling. There is also not as much information in ‘n’ observation within a cluster as there happens to be in ‘n’ randomly drawn observations. Cluster sampling is used only because of the economic advantage it possesses; estimate based on cluster samples are usually more reliable per unit cost.

(iv). Area sampling: If cluster happen to be some geographical subdivision, in the case cluster sampling is better known as area sampling. In other words, cluster design, where the primary sampling unit represents a cluster of units based on geographic area, are distinguished as area sampling. The plus and minus points of cluster sampling are also applicable to area sampling.

(v). Multi-stage sampling: Multi-stage sampling is a further development of the principle of cluster sampling. Suppose we want to investigate the working efficiency of nationalized banks in India and we want to take a sample of few banks for this purpose. The first stage is to select large primary sampling unit such as states in a country. Then we may select certain districts and interview all banks in the chosen districts. This would represent a two- stage sampling design with the ultimate sampling units being clusters of districts.

If instead of taking census of all banks within the selected districts, we select certain towns and interview all banks in the chosen towns. This would represent a three-stage sampling design. If instead of taking a census of all banks within the selected towns, we randomly sample banks from each selected town, then it is a case of using a four- stage sampling plan. If we select randomly at all stage, we will have what is known as ‘multi-stage random sampling design’. Ordinarily multi-stage sampling is applied in big inquires extending to a considerable large geographical area, say the entire country.

There are two advantages of this sampling design viz.,

(a).it is easier to administer than most single stage design mainly because of the fact that sampling frame under multi-stage sampling is developed in partial units

(b). A large number of units can be sampled for a given cost under multi stage sampling because of sequential clustering, whereas this is not possible in most of the simple design.