Chapter Three: Introduction to Sampling and Sampling Strategy

Chapter Objectives: at the end of this chapter, the student will be able to:

- Describe the Nature and Meaning of Sampling
- Determine sample design for different research approaches
- The Nature and Meaning of Sampling

3.1 What is sampling?

? *Dear students, what do you understand by sampling?*

One of the biggest concern of a researcher in his/her preparation to gather information is the selection of who should be contacted to get information. As a matter of fact not all members of the population in your study area can be communicated. For instance if you like to study the educational background of civil servants in a specific Woreda office where only 20 people are employed, you can easily gather information from each and conclude about them. However, what if you like to study the income status of civil servants in the Woreda where there are more than twenty thousand of them, can you gather data from all of them? Here is when you need to select some of them and conclude about the whole population. In selecting sample from the whole population you need to have a sample design.

? Dear students what is a sample design?

A sample design is a definite plan for obtaining a sample from population. It refers to the technique or the procedure the researcher would adopt in selecting items for the sample. The sample design may as well lay dawn the number of items to be included in the sample i.e. the size of the sample.

Generally sampling refers to the process of selecting a few (a sample) from a bigger group (the sampling population) to become the basis for estimating or predicting the prevalence of … a situation or old come regarding the bigger group. A sample is the sub group of the bigger group that the researcher is interested in.

? Dear students, why do we need to take sample other than dealing with the whole population?

Taking some proportion from the whole population is justifiable from different dimensions. Let us discuss some of them;

i) **Economical in expenditure:** dealing with whole population incurs you large amount of resource so it is feasible to deal with smaller population.
ii) **Greater speed**:- dealing with smaller proportion of the whole population saves time for the researcher.
iii) **Greater scope:-** reducing the larger population in to sample also enables to widen the scope of the information to be collected by incorporating larger categories.
iv) **Practicability:-** taking sample increases the possibility of conducting the research instead of being constrained by lack of resources and time.
v) **Greater accuracy** … sampling increases the accuracy of the data collected since small group is contacted with greater focus.

Even if selecting sample has these advantages it also has disadvantages. After taking sample the data that you gathered may not represent the whole characteristics of the general population from which sample is selected. This gap is filled through estimation or prediction and an error may exist in estimations or predictions.

In a situation where information is gathered from the whole population the possibility of committing error in estimation or prediction is minimum or none at all. Therefore, one important consideration of selecting sample from the total population is tolerance of the possibility of error and use appropriate sampling technique to reduce the margin of error.

3.2 The sampling process

Dear students, there are sequential steps that should be considered in sampling design. These steps involve;

i) **Defining the population:-** The study population or the universe is the population to which the researcher directs his study. This is the whole group to which the researcher

would want to generalize after selecting some amount of sample. In the previous example all the civil servants that reside in the Woreda are your universe or study population. The whole population towards which the researcher directs his/her attention is called as study population usually represented by the letter (N).

ii) **Specify the sampling frame:-** The sampling frame is the physical material from which sample is chosen. This is a document that lists the study population. It may include telephone directory, list of business establishment in a town, list of households and residents in a kebele etc. .. In the previous case for instance to study the income status of civil servants in a particular Woreda you may get the name list of the whole civil servants in the Woreda personnel administration office, which you can use as a sampling frame.

iii) **Specify the sampling unit (element):-** sampling unit represents the basic unit containing the element of the population to be sampled. It may contain one or several population elements. In your case there is only one unity of population & that is civil servants in a Woreda.

iv) **Specify the sampling method:-** refers to the method that you need to follow in selecting the sample from the whole population. For instance, what sampling methods did you use to select only 200 civil servants from a sum total of 20,000. Dear students we will discuss varieties of sampling methods in the coming section.

v) **Determine the sample size:-** Here the amount to be selected from the total population (sample size) is determined. The sample size is the specific population from which you obtain information.

vi) **Specify sampling plan:-** In this step you need to specify operational procedure by which you communicate each sample unit.

vii) **Select the sample:-** here all office and field activities necessary to select the sample is completed and the sample is selected for the study.

The most vital consideration in the choice of the sample is to ascertain its representativeness. The sample size should be representative of the total population (universe). The only mechanism by which the researcher ensures the representativeness of the sample size is to follow accurate procedures and use appropriate sampling

technique. Dear students what if a researcher selects a sample size of 10 people from the total population of one thousand. Does the sample size accurately represent the total population? That is why you need to use different techniques of sampling. Types of sampling techniques will be discussed in the next section.

Techniques of sampling

✪ **Overview**

Dear students, the previous section dealt about what sampling is, its advantages & disadvantage and the sampling process in which sequential steps were discussed. Under this section you will be familiarized with different sampling techniques. These techniques are broadly classified into probability sampling techniques & non probability sampling techniques. Each category has further sub categories of specific sampling techniques. For instance probability sampling technique involves simple random sampling, stratified random sampling, cluster sampling, where as the non probability category has Quota sampling, convenience sampling & purposive sampling. Dear students in dealing with all techniques try to see the elements in each & differentiate them from each other.

4.1 Probability sampling techniques.

Probability sampling technique is distinct and important in that it allows each element in the study population equal and in dependent chance of being selected as a sample. Every member of the study population is provided equal chance to be represented in the sample with no personal interest and preference of the researcher. Independence is also important in that each unit would be selected without being influenced by the selection or rejection of another unit. In the example that we discussed previously about civil servants, the researcher's preference should not be reflected in selecting a particular civil servant in to the sample or not that reflects equity. In the case of independence if you select the wife or friends of a particular respondent (sample) to include him in the sample your sample lacks the criteria of independence. Therefore, the selection or rejection of one person should not affect the selection or rejection of another person. Probability sampling techniques are discussed as follows;

4.1.1 Simple Random sampling

? *Dear students how is simple random sampling technique used?*

Simple random sampling is the most commonly used random sampling technique where by each and every element in the total population has equal and independent chance of being selected. In implementing this technique each and every elements in the total population will be allotted with specific number. Later on samples will be drawn in the form of lot until you get the decided amount of sample size. For instance, if you want to study the educational status of students in a class room of 120, you first determine the sample size (assume it is 30), then you give number to each students (from 1-120) & finally draw the numbers until you reach 30 students. Finally, the numbers that have got the chance to be selected will be your sample from whom you collect information. Selection of sample in this case is done with out replacement i.e. once a sample is picked out it will not get another chance of being selected.

4.1.2 Stratified random sampling

? *Dear students how does stratified random sampling differs from simple random technique?*

The quality of data collected in conducting a research study can be evaluated by the extent to which different groups are represented within the study population. The researcher may stratify the study population in to different groups based on the homogenous character that they have. For instance if you like to study the study habit of students in a class room you may categorize the students based on their academic status as outstanding, medium, fair and poor. After this stratification sample will be selected from each non-overlapping groups using the simple random sampling technique. Sample could be selected either proportionally or in the form of disproportional stratified sampling. In the first case sample will be selected from each group based on the number of population in each group. That is, the larger the number in a group the larger sample will be taken from it and vice versa. In the case of disproportionate stratified sampling, sample will be selected from each group irrespective of the number of people in the group.

4.1.3 Cluster sampling

? *Dear students, what is cluster sampling and how can you differentiate it from the previous two techniques?*

Stratified sampling is employed for small number of study population. But when you are required to select sample from large number of population scattered across large geographical area cluster sampling is a better option. In this technique first you select randomly natural groups/units called clusters from the whole study population (universe). Then after all or parts of the units within the clusters are chosen to makeup the sample. Simply stated, the researcher first divides the study population in to groups called clusters and elements could be further selected from each cluster based on simple random sampling. Let us assume the Ethiopian government wanted to study the attitude of Ethiopian population towards the system of decentralization. First the population could be classified in to geographical clusters. For this there are already available 9 regional governments. You might select sample woredas from each regional governments (or from some randomly selected regions) and sample kebeles from each woredas selected by using simple random technique. Finally you might still select people from each kebele through random sampling from whom data will be collected. Dear students as you can see you used different groups and sub groups to select sample. In this example the population is already classified by administrative units but there are cases when the researchers ability is critical to categorize the population in to homogeneous groups at each stage.

4.2 Non – probability sampling techniques

As opposed to probability sampling in non- probability sampling there is no way to estimate or ascertain that each elements in the study population has an equal & independent chance of being included as a sample. This is because sample is selected based on the researchers own preference & based on the criteria of convenience. Therefore, selection of sample in mainly based on other considerations than random selection to represent the whole study population. These methods are also employed when the number of elements in the study population is unknown or cannot be identified by the researcher for different reasons. For instance can you get a real data available on the number of visually impaired people in your Woreda? Here we are going to see three types of non-probability sampling techniques;

4.2.1 Quota sampling.

? ***Dear students what do you know about quota sampling?***

In this type of sampling the researcher specifies to communicate a certain group of people with similar attributes and goes to the place where he can obtain such people. Therefore the samples are usually selected from a place where it is convenient for the researcher to easily get the target population. In this type of sampling people are selected non randomly according to some fixed quota. The two types of quota sampling are proportional and non proportional. In the case of ***proportional quota sampling*** different categories of your study population are represented proportionally in selecting sample. For instance if the total number of your study population is 400 out of which 40% is female and 60% male and your sample size is 100 you will take 60 male and 40 females according to their proportion. In this case we used gender as a basis for category but it could also be educational level, age, religion, race etc… In the case of ***non proportional quota sampling*** the researcher simply specifies the minimum number of sampled units from each category with little concern for numbers that much proportions in the population. This depends on the researchers' judgment of how much unit is enough from each category to assure sufficient representation.

4.2.2 Accidental Haphazard (convenience) sampling.

Accidental sampling is usually based on the convenience in accessing the sampling population. In this case a researcher may go out on the street, pick any body and collect data about a certain issue, until the sample size reaches to a certain predetermined level. The sampling is accidental in that, the person communicated may not have any information regarding the issue raised by the researcher. For instance a journalist may go out on the street to gather data from people about the new environmental policy of the government. He may communicate with students, taxi drivers, street vendors, civil servants etc… walking on the street, many of these respondents may not have information on the issue. Therefore it differs from quota sampling in that it does not specify a specific target group from which information can be collected.

4.2.3 Purposive or Judgment sampling

Dear students how do you differentiate purposive sampling from the other two non-random sampling techniques?

The main consideration of the researcher in employing this technique is the determination of who can provide him with detailed information about the issue under the study. Therefore, the researcher uses his own opinion to go to those people who have rich information and are willing to provide him. Still the sampling frame and the representative ness of the information collected is unknown. It is mainly applicable when it is used by an expert who knows the population very well, to identify a typical case, to construct a historical reality, to describe about an unfamiliar phenomenon etc…

a. Sample size determination

Dear students, you should also be concerned about how to select your sample size. You have to ask yourself 'how big a sample should I select?', 'what should be my sample size?' and 'how many cases do I need?' Sample size is among one of the most crucial aspects that determines the accuracy of your findings. Representativness of the sample units to the total study population is one of the issues that have to be dealt carefully. In qualitative research the concern is mainly to explore and describe issues or situations and the question of sample size is less important. Data is collected until it reaches to saturation point. Ones the researcher convinced that he/she has got sufficient new information data collection will be stopped. But sample size determination for quantitative study takes some very important issues in to consideration like the level of confidence in testing the results, the degree of accuracy with which the population is needed to be tested and level of variation or standard deviation. Dear students, discussing these points is beyond the scope of this module since there is another course to deal with quantitative methods of data analysis.

Determining sample design

All the items under consideration in any field of inquiry constitute a 'universe' or 'population'. A complete enumeration of all the items in the 'population' is known as a census inquiry. It can be presumed that in such an inquiry when all the items 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 observations increases. Moreover, there is no way of checking the element of bias or its extent except through a resurvey or use of sample checks. Besides, this type of inquiry involves a great deal of time, money and energy. Not only this, census inquiry is not possible in practice under

many circumstances. For instance, blood testing is done only on sample basis. Hence, quite often we select only a few items from the universe for our study purposes. The items so selected constitute what is technically called a sample.

The researcher must decide the way of selecting a sample or what is popularly known as the sample design. In other words, a sample design is a definite plan determined before any data are actually collected for obtaining a sample from a given population. Thus, the plan to select 12 of a city's 200 drugstores in a certain way constitutes a sample design. Samples can be either probability samples or non-probability samples. With probability samples each element has a known probability of being included in the sample but the non-probability samples do not allow the researcher to determine this probability. Probability samples are those based on simple random sampling, systematic sampling, stratified sampling, cluster/area sampling whereas non-probability samples are those based on convenience sampling, judgment sampling and quota sampling techniques. A brief mention of the important sample designs is as follows:

(i) *Deliberate sampling*: Deliberate sampling is also known as purposive or non-probability sampling. This sampling method involves purposive or deliberate selection of particular units of the universe for constituting a sample which represents the universe. When population elements are selected for inclusion in the sample based on the ease of access, it can be called *convenience sampling*. If a researcher wishes to secure data from, say, gasoline buyers, he may select a fixed number of petrol stations and may conduct interviews at these stations. This would be an example of convenience sample of gasoline buyers. At times such a procedure may give very biased results particularly when the population is not homogeneous. On the other hand, in *judgement sampling* the researcher's judgement is used for selecting items which he considers as representative of the population. For example, a judgement sample of college students might be taken to secure reactions to a new method of teaching. Judgement sampling is used quite frequently in qualitative research where the desire happens to be to develop hypotheses rather than to generalise to larger populations.

(ii) *Simple random sampling*: This type of sampling is also known as chance sampling or probability sampling where each and every item in the population has an equal chance of inclusion in the sample and each one of the possible samples, in case of finite universe, has the same probability of being selected. For example, if we have to select a sample of 300 items from a universe of 15,000 items, then we can put the names or numbers of all the 15,000 items on slips of paper and conduct a lottery. Using the random number tables is another method of random sampling. To select the sample, each item is assigned a number from 1 to 15,000. Then, 300 five digit random numbers are selected from the table. To do this we select some random starting point and then a systematic pattern is used in proceeding through the table. We might start in the 4th row, second column and proceed down the column to the bottom of the table and then move to the top of the next column to the right.

When a number exceeds the limit of the numbers in the frame, in our case over 15,000, it is simply passed over and the next number selected that does fall within the relevant range. Since the numbers were placed in the table in a completely random fashion, the resulting sample is random. This procedure gives each item an equal probability of being selected. In case of infinite population, the selection of each item in a random sample is controlled by the same probability and that successive selections are independent of one another.

(iii) *Systematic sampling*: In some instances the most practical way of sampling is to select every 15th name on a list, every 10th house on one side of a street and so on. Sampling of this type is known as systematic sampling. An element of randomness is usually introduced into this kind of sampling by using random numbers to pick up the unit with which to start.

This procedure is useful when sampling frame is available in the form of a list. In such a design the selection process starts by picking some random point in the list and then every *n*th element is selected until the desired number is secured.

(iv) *Stratified sampling*: If the population from which a sample is to be drawn does not constitute a homogeneous group, then stratified sampling technique is applied so as to obtain a representative sample. In this technique, the population is stratified into a number of non overlapping subpopulations or strata and sample items are selected from each stratum. If the items selected from each stratum is based on simple random sampling the entire procedure, first stratification and then simple random sampling, is known as *stratified random sampling*.

(v) *Quota sampling*: In stratified sampling the cost of taking random samples from individual strata is often so expensive that interviewers are simply given quota to be filled from different strata, the actual selection of items for sample being left to the interviewer's judgement. This is called quota sampling. The size of the quota for each stratum is generally proportionate to the size of that stratum in the population. Quota sampling is thus an important form of non-probability sampling. Quota samples generally happen to be judgement samples rather than random samples.

(vi) *Cluster sampling and area sampling*: Cluster sampling involves grouping the population and then selecting the groups or the clusters rather than individual elements for inclusion in the sample. Suppose some departmental store wishes to sample its credit card holders. It has issued its cards to 15,000 customers. The sample size is to be kept say 450. For cluster sampling this list of 15,000 card holders could be formed into 100 clusters of 150 card holders each. Three clusters might then be selected for the sample randomly. The sample size must often be larger than the simple random sample to ensure the same level of accuracy because is cluster sampling procedural potential for order bias and other sources of error is usually accentuated. The clustering approach can, however, make the sampling procedure relatively easier and increase the efficiency of field work, specially in the case of personal interviews.

Area sampling is quite close to cluster sampling and is often talked about when the total geographical area of interest happens to be big one. Under area sampling we first divide the total area into a number of smaller non-overlapping areas, generally called geographical clusters, then a number of these smaller areas are randomly selected, and all units in these small areas are included in the sample. Area sampling is specially helpful where we do not have the list of the population concerned. It also makes the field interviewing more efficient since interviewer can do many interviews at each location.

(vii) *Multi-stage sampling*: This is a further development of the idea of cluster sampling. This technique is meant for big inquiries extending to a considerably large geographical area like an entire country. Under multi-stage sampling the first stage may be to select large primary sampling units such as states, then districts, then towns and finally certain families within towns. If the technique of random-sampling is applied at all stages, the sampling procedure is described as multi-stage random sampling.

(viii) *Sequential sampling*: This is somewhat a complex sample design where the ultimate size of the sample is not fixed in advance but is determined according to mathematical decisions on the basis of information yielded as survey progresses. This design is usually adopted under acceptance sampling plan in the context of statistical quality control.

In practice, several of the methods of sampling described above may well be used in the same study in which case it can be called mixed sampling. It may be pointed out here that normally one should resort to random sampling so that bias can be eliminated and sampling error can be estimated.

But purposive sampling is considered desirable when the universe happens to be small and a known characteristic of it is to be studied intensively. Also, there are conditions under which sample designs other than random sampling may be considered better for reasons like convenience and low costs.

The sample design to be used must be decided by the researcher taking into consideration the nature of the inquiry and other related factors.