

Unit - 7 SAMPLING DESIGN

Handout- 5

Once the researcher has clearly specified the problem and developed the appropriate design and data-collection instrument, the next step is to select the elements from which information is collected. To effectively undertake research project, the researcher must have the data, without data the analysis may not be convincing. Information (data) can be generated from either population or sample.

The basic idea of sampling is that by selecting some of the elements in the population, we may draw conclusions about the entire population.

Population: It is the totality of object or phenomena under consideration for a specific study. A population is the total collection of elements about which we wish to make some inferences.

Census: A census is a count of all the elements in a population. It is a survey that includes the totality of objects or subjects or phenomenon. But it is not always possible to undertake a census or a complete enumeration of all items in the population particularly when the population is too large. So one has to resort to sample survey to generate the data required for the investigation.

Sample: It is a proper subset or part of population. It is used to represent the population. A population element is the subject on which the measurement is being taken. It is the unit of study.

Sampling: is the procedure of selecting a sample from a population. Sampling aims at obtaining consistent and unbiased estimates of the population. The aim in a sample survey is not just the characteristics of the sample but also that of population from which the sample has been drawn.

Why sample?

There are several reasons for sampling;

- Lower cost
- Greater accuracy of results
- Greater speed of data collection
- Availability of population elements.

SAMPLING CONCEPTS AND TERMINOLOGIES:

Sampling element: The unit of analysis or case in population, it is from which information is collected which provides basis for analysis,

it is the subject on which measurement is being taken, and it can be people, a group or an organization.

Sampling unit: That element or set of elements considered for selection in some stage of sampling. Ex: Population in rural people in Gamo Gofa zone, Woredas, Kebeles etc is sampling units.

Sampling frame: It is the actual list of sampling units from which the sample is selected. It is closely related to the population. *It is the list of elements from which the sample is actually drawn.*

Sampling ratio: Size of the sample / size of population.

Steps in sampling design:

1. Define the target population

It is first necessary to define the target population of the collection of elements about which the researcher wishes to make an inference. Research elements are the objects on which the measurements are taken. The target population may be kids, children, individuals, households, business firms etc. The simpler the definition of target population, the higher the incidence and the easier and less costly it is to find the sample. **Incidence** refers to the percentage of the general population that satisfies the criteria defining the target population. When incidence is high, the cost and time to collect data are minimized.

2. Identify the sampling frame

It is the listing of the elements from which the actual sample will be drawn. Example; telephone directory.

3. Select a sampling procedure

The choice of sampling method depends largely on what the researcher can develop for a sampling frame. For example; a simple random sample requires that a complete, accurate list of population elements by name or other identification code be available.

4. Determine the sample size

It depends on homogeneity of the population (dispersion variance), degree of confidence (how much precision they need), number of sub groups to be studied, cost and time factors.

5. Select the sample elements

The researcher needs to choose the elements that will be included in the study. This depends upon the type of sample being used and the sampling method.

6. Collect the data from the designated elements

Appropriate methods of data collection methods are adopted to gather information from the sample elements to support the study, which leads to further analysis and conclusion to the study.

Characteristics of good sampling:

The logic of the theory of sampling is the logic of induction. i.e., one proceeds from particular (i.e., sample) to the general (i.e., population) and all results are expressed in terms of probability. If the population is identical or homogeneous, no need for careful sampling procedure, any sample would be sufficient. But, when faced with variation or heterogeneity, in the population, more controlled sampling procedures are required. A good sample must be, as representative of the entire population as possible, and ideally it must provide the whole of the information about the population from which the sample has been drawn. So, the ultimate test of a sample design is how well it represents the characteristics of the population it purports to represent. In measurement terms the sample must be valid. *Validity* of the sample depends on two factors: **accuracy and precision**.

- a. **Accuracy:** It is the degree to which bias is absent from the sample. When the sample is drawn properly, some sample elements underestimate the population values being studied and others overestimate them. An accurate unbiased sample is one which the underestimations and overestimations are a balanced among the members of the sample. There is no systematic variance with an accurate sample. **Systematic variance** is defined as the variation in measures due to some known or unknown influences that cause the scores to lean in one direction more than another.
- b. **Precision:** The degree to which the standard errors are minimized. No sample will fully represent its population in all aspects. A sample statistics may be expected to differ from its parameters as a result of random fluctuations inherent in the sampling process. This is referred to as the error of variances or **sampling error**. Precision is measured by the standard errors of estimates, a type of standard deviation measurement; the smaller the standard error of estimates, the higher is the

precision of the sample. The ideal sample design produces a small standard error of the estimate.

Categories of sampling procedures:

There are two categories of sampling procedures: Random (Probabilistic) and Non-random (Non-Probabilistic).

Probability sampling: Is a sampling technique in which every member of the population will have a known, non-zero or equal probability of selection.

Non-Probabilistic sampling: is a sampling technique in which units of the sample are selected on the basis of personal judgment or convenience.

I. Probability sampling techniques:

1. **Simple random sampling:** a simple random sample is a sample selected in such a way that every element in the population has the same chance of being chosen, every sample of size 'n' has the same chance to be chosen. Each population element has a known and equal chance of being selected. It is same as the practice of picking lottery winners. The disadvantage of this method it is too expensive to interview a national face-to face sample based on simple random method, it requires listing of the entire population of interest.
2. **Systematic sampling:** A systematic random sampling is a sample which contains every 'i' th element of the population. The first element is chosen randomly, the rest systematically. It is one of the most widely used probability sampling technique, its major advantages are its simplicity and flexibility, it may result in error when the population is not uniform or homogenous or where there are systematic arrangements, it is less expensive and easy to carry out. Example: from 20 elements if we want to take four, then i th element = $20/4 = 5^{\text{th}}$ thus every 5th element of the group of 20 elements after the 1st sample is taken is included in our sample.
3. **Stratified random samples:** A stratified sample is a probability sample that is distinguished by the following two step procedure;

- (1) The parent population is divided into mutually exclusive and exhaustive sub sets.
- (2) A simple random sample of elements is chosen independently from each group or subset.

The subsets into which the universe elements are divided are called strata or subpopulations. The division is mutually exclusive and exhaustive. This means that every population element must be assigned to one and only one stratum and that no population elements are omitted in the assignment procedure.

Most populations can be segregated into a number of mutually exclusive subpopulations dividing the population to non-overlapping group (strata) is called stratification. After a population is divided into the appropriate strata a simple random sample can be taken from each stratum. The stratified sampling technique is particularly useful when we have heterogeneous populations. With stratification each stratum is homogenous internally and heterogeneous with other strata.

The researcher chooses a stratified random sample due to the following reasons;

- To increase a sample's statistical efficiency
- To provide adequate data for analyzing the various subpopulations and
- To enable different research methods and procedures to be used in different strata.

The researcher must still decide whether to select;

- a) a proportionate stratified sample,
- b) a disproportionate stratified sampling.

With a **proportionate stratified sampling** the number of observations in the total sample is allocated among the strata in proportion to the relative number of elements in each stratum in the population. The sample drawn is proportionate to the stratum's share of the total population.

Disproportionate stratified sampling involves balancing the two criteria of strata size and strata variability. With a fixed sample size, strata exhibiting more variability are sampled more than proportionately to their relative size. Conversely, those strata that are very homogenous are sampled less than proportionately.

4. **Cluster Sampling:** If the total area of interest happens to be 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 areas. The total area can be divided into groups or clusters of elements and some of the groups or clusters are selected randomly. This is

cluster sampling. Cluster sampling reduces cost by concentrating surveys in selected areas.

Cluster sampling differs from stratified sampling in several ways;

Stratified sampling	Cluster sampling
➤ We divide the population into a few sub groups, each with many elements in it. The subgroups are selected according to some criterion that is related to the variables under study. ➤ We try to secure homogeneity within subgroups and heterogeneity between subgroups. ➤ We randomly choose elements from within each subgroup.	➤ We divide the population into many subgroups, each with a few elements in it. The subgroups are selected according to some criterion of ease or availability in data collection. ➤ We try to secure heterogeneity within subgroups and homogeneity between subgroups. ➤ We randomly choose a number of subgroups, which we then study in depth.

5. Area sampling

When research involves population identified from a geographical area, it is advisable to use area sampling. It is an important form of cluster sampling. Suppose the investigator is interested in estimating the consumption per household in the city of Chicago, and how consumption is related to family income. One approach is;

1. Choose a simple random sample of 'n' blocks from the population of N blocks.
2. Determine consumption and income for all households in the selected blocks and generalize the sample relationships to the larger population.

This is also called one stage area sampling.

In the two stage area sampling the selected areas themselves can be sub sampled.

Example;

Consider a universe of 100 blocks. Suppose that there are 20 households per each block. Assume that a sample of 80 households is required from this total population of 2,000 households. The overall sampling fraction is thus $80/100 = 1/25$. There are a number of ways by which the sample can be selected;

1. Selecting 4 blocks and 20 households per block. (one-stage area sampling)

2. Selecting 10 blocks and 8 households per block (two-stage area sampling)

II. Non Probability sampling: The probability of selecting the population elements is unknown. We cannot estimate any range within which to expect the population parameter. There are a variety of ways to choose persons or cases to include in the sample. Despite accepted superiority of probability sampling methods, non probability sampling maybe used when probability sampling is prohibitively expensive and when precise representation is not necessary.

- It is used because of cost and time requirements
- It is used if there is no desire to generalize a population parameter
- The total population may not be available for the study in certain cases.
- It involves personal judgment somewhere in the selection process.

Different non-probability methods could be identified as follows:

1. **Convenience sampling (accidental samples):** Select anyone who is convenient, it can produce, ineffective, highly unrepresentative sample and is not recommended, it has high bias and systematic errors. It is least reliable but cheap and easy to collect. There is no control to ensure precision. Ex: The person on the street interviewed for a television programme.
2. **Purposive:** When one draws a non-probability sample that confirms to a certain criteria, it is called purposive sampling. It occurs when one picks sample members to conform to some criteria. *Purposive sampling can be of two types; judgment sampling and quota sampling.*
 - a) **Judgment sampling:** It uses the judgment of experts in selecting cases or it selects cases with specific purpose in mind. But the researcher does not know whether the case selected represents the population: Ex: in the study of standard of living, the cost of electricity, refrigerator, video recorder, satellite dish cannot be included for all people in Ethiopia.
 - b) **Quota sampling:** Quotas are assigned to different strata group. The logic behind quota sampling is that certain relevant characteristics describe the dimensions of the population. In quota sampling, a researcher first identifies categories of people then decides how many to get in each category. Thus the number of people in various categories is fixed. It gives no assurance that the sample is a representative on the variable

being studied. As there is no element of randomization, the extent of sampling error cannot be estimated.

3. **Snowball sampling:** Also called network, chain referral or computational sampling is a method for identifying and sampling or selecting the case in a network. It begins with one or a few people or cases and spreads out on the basis of links to the initial case. Snowball sampling is a judgment sample used to sample special populations. The researcher locates an initial set of respondents. These respondents are used as informants to identify others with desired characteristics.

Problems in sampling:

- a) **Non sampling errors:** This refers to :

- (i) **Non-coverage error:** this refers to sample frame defects: Ex: omission of part of the population; soldiers, students, people in hospital are typically excluded from national survey (but not series). It is serious in telephonic surveys since those who do not have a telephone are excluded. It also occurs when the cost used for sampling are incomplete.
- (ii) **The wrong population is sampled:** Be sure that the group being sampled is drawn from the population. Ex: drawing a sample of college students generalize about all college age persons.
- (iii) **Non-response error:** The response rate is low. Some people refuse to be interviewed because they are too busy, or simply do not trust the interviewer or they may not be interested to.
- (iv) **Instrumental errors:** instrument device to collect data (Ex: questionnaire), Ex: when questionnaire is badly worded or asked, leading questions or carelessly worded questions may be misinterpreted.
- (v) **Interviewer errors:** When some characteristics of the interviewer (age, sex, etc) affects the way in which the respondent answers questions. Ex: Questions about racial discriminations might be differently answered depending on the racial group of the interviewer.

- b) **Sampling errors:** It is a random variation in the sample estimates around the true population parameter. It is calculated only for probability sampling. Random sampling allows unbiased estimates of sampling error. The measurement of sampling error is usually called the precision of the sampling plan.

What sample size is required?

The sample size should be a function of the variation in the population parameters under study and the estimating precision needed by the researcher. Some principles that influence sample size includes;

- The greater the dispersion or variance within the population, the larger the sample must be to provide estimation precision.
- The greater the desired precision of the estimate, the larger the sample must be.
- The narrower the interval range, the larger the sample must be.
- The higher the confidence level in the estimate, the larger the sample must be.
- The greater the number of subgroups of interest within a sample, the greater the sample size must be, as each subgroup must meet the minimum sample size requirements.
- If the calculated sample size exceeds 5 percent of the population, sample size may be reduced without sacrificing precision.