Sampling

- **A sample is a finite part of a statistical population whose properties are studied to gain information about the whole (Webster, 1985).**

- **When dealing with people, it can be defined as a set of respondents (people) selected from a larger population for the purpose of a survey.**

- **Researchers typically rely on *sampling* to acquire a section of the population to perform an experiment or observational study.**

- **It is important that the group selected be representative of the population, and not biased in a systematic manner.**

- **The sample should be a representation of the entire population.**

A **sampling design** specifies for every possible sample its probability of being drawn.

Sampling is a means of selecting a subset of units from a target population for the purpose of collecting information. This information is used to draw inferences about the population as a whole.

A procedure or plan drawn up before any data are collected to obtain a sample from a given population. Also known as sampling plan; survey design.

- **When taking a sample from a larger population, it is important to consider how the sample is chosen.**

- **To get a representative sample, the sample must be drawn randomly and encompass the whole population.**

The sampling process comprises several stages:

I. **Defining the population of concern**
II. **Specifying a sampling frame, a set of items or events possible to measure**
III. **Specifying a sampling method for selecting items or events from the frame**
IV. **Determining the sample size**
V. **Implementing the sampling plan**
VI. **Sampling and data collecting**
VII. **Data which can be selected**

Purpose of Sampling

- **To accurately describe the parameters of a population based on the description (statistics) of a set of elements drawn from the population.**

- **To draw conclusions about populations from samples, we must use inferential statistics which enables us to determine a population`s characteristics by directly observing only a portion (or sample) of the population.**

- **We obtain a sample rather than a complete enumeration (a census) of the population for many reasons.**

- **There would be no need for statistical theory if a census rather than a sample was always used to obtain information about populations.**

- **But a census may not be practical and is almost never economical.**

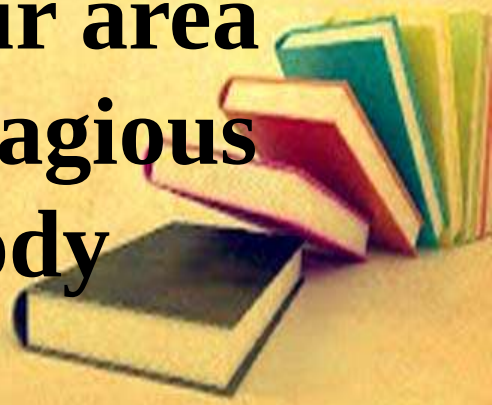
- **There are six main reasons for sampling instead of doing a census.**

- **These are; -**

a) **Economy**
b) **Timeliness**
c) **The large size of many populations**
d) **Inaccessibility of some of the population**
e) **Destructiveness of the observation**
f) **accuracy**

•**Economy** - The economic advantage of using a sample in research Obviously, taking a sample requires fewer resources than a census.

The time factor - A sample may provide you with needed information quickly. For eg, you are a Doctor and a disease has broken out in a village within your area of jurisdiction, the disease is contagious and it is killing within hours nobody



knows what it is. You are required to conduct quick tests to help save the situation. If you try a census of those affected, they may be long dead when you arrive with your results. In such a case just a few of those already infected could be used to provide the required information.

The very large populations - Many populations about which inferences must be made are quite large.

The partly accessible populations - There are Some populations that are so difficult to get access to that only a sample can be used. Like people in prison.

The destructive nature of the observation – Sometimes the very act of observing the desired characteristic of a unit of the population destroys it for the intended use. Good eg of this occur in quality control. For eg, to test the quality of a fuse, to determine whether it is defective, it must be destroyed. To obtain a census of the quality of a lorry load of fuses, you have to destroy all of them. This is contrary to the purpose served by quality-control testing.

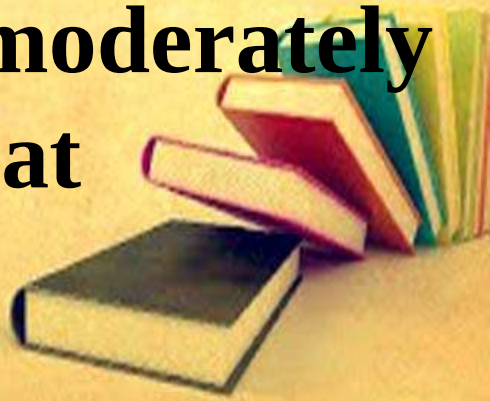
In this case, only a sample should be used to assess the quality of the fuses.

Accuracy and sampling - A sample may be more accurate than a census. A sloppily conducted census can provide less reliable information than a carefully obtained sample.

Basic Principles of Sampling

Theory of sampling is based on the following laws-

- ***Law of Statistical Regularity*** **– This law comes from the mathematical theory of probability. According to King," Law of Statistical Regularity says that a moderately large number of the items chosen at**



random from the large group are almost sure on the average to possess the features of the large group."
According to this law the units of the sample must be selected at random.

- ***Law of Inertia of Large Numbers –***

According to this law, the other things being equal – the larger the size of the sample; the more accurate the results are likely to be.

Characteristics of the Sampling Technique

1. Much cheaper.
2. Saves time.
3. Much reliable.
4. Very suitable for carrying out different surveys.
5. Scientific in nature.

Advantages of sampling

1. Very accurate.
2. Economical in nature.
3. Very reliable.
4. High suitability ratio towards the different surveys.
5. Takes less time.
6. In cases, when the universe is very large, then the sampling method is the only practical method for collecting the data.

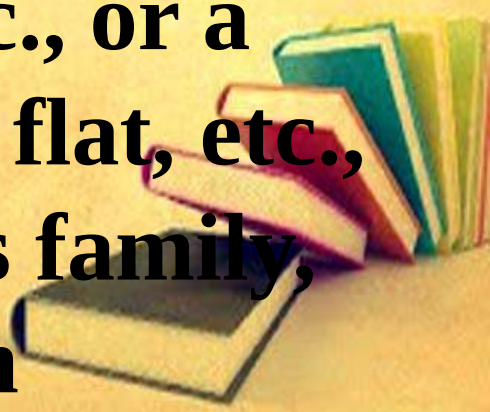
Disadvantages of sampling

1. Inadequacy of the samples.
2. Chances for bias.
3. Problems of accuracy.
4. Difficulty of getting the representative sample.
5. Untrained manpower.
6. Absence of the informants.
7. Chances of committing the errors in sampling.

Sampling Unit

- **A basic concept in sampling theory is the *sampling unit*, which is the minimum unit of observation for information on the operative variables.**

- **Sampling unit may be a geographical one such as state, district, village, etc., or a construction unit such as house, flat, etc., or it may be a social unit such as family, club, school, etc., or it may be an**



individual.

The researcher will have to decide one or more of such units that he has to select for his study.

•The sampling unit must be clearly defined for constructing the sampling frame.

- **By convention in statistics, a capital "N" is used to refer to the number of sampling units making up the universe, and a lowercase "n" for the number of sampling units in the sample itself.**

- **The universe must be defined in the light of the objectives of the survey. It can be expressed in geographical terms (locality, municipality, district) or in sectoral terms (urban population, pottery manufacturers, fuelwood producers).**

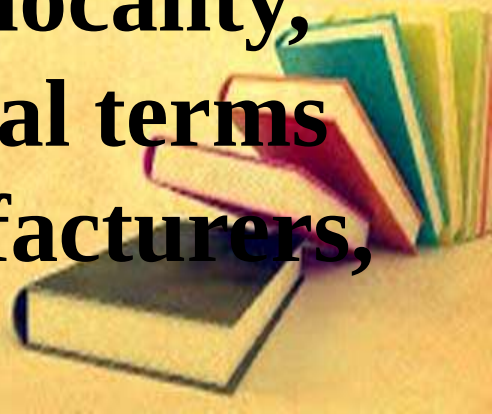
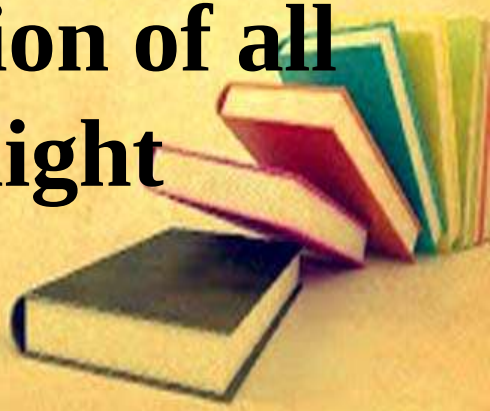


Table 3.1: Sampling unit for thematic group and sector or branch under examination

Group	Sector/branch	Sampling unit
Demand	Residential - urban - rural	Home
	Industrial	Establishment
	Commercial	
	Institutional	
Supply	Direct	Plot
	Indirect	Establishment
Provision	Producers	Individual producers,
	Transport operators	companies
	Commercial suppliers	

Defining the population

- **Target population: a clearly defined population from which the sample will be drawn**
- **Inferential population: a clearly defined population to which our results will be applied**
- **Sampled population: the collection of all possible observation units that might have been chosen in a sample**



Sampling Types

1. Non-Probability Sampling

- **In this type of population sampling, members of the population do not have equal chance of being selected.**

- **Due to this, it is not safe to assume that the sample fully represents the target population..**

- **It is also possible that the researcher deliberately chose the individuals that will participate in the study.**

- **Non-probability population sampling method is useful for pilot studies, case studies,qualitative research, and for hypothesis development.**

- **This sampling method is usually employed in studies that are not**

interested in the parameters of the entire population. Some researchers prefer this sampling technique because it is cheap, quick and easy.

2. Probability Sampling

- **In probability sampling, every individual in the population have equal chance of being selected as a subject for the research.**

- **This method guarantees that the selection process is completely randomized and without bias.**

- **The advantage of using probability sampling is the accuracy of the statistical methods after the experiment.**

- **It can also be used to estimate the population parameters since it is**

representative of the entire population.

- **It is also a reliable method to eliminate sampling bias.**

- **Probability sampling is also known as 'random sampling' or 'chance sampling'.**

- **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.**

Complex Random Sampling Designs

- **Probability sampling - under restricted sampling techniques.**

- **Also called ‘mixed sampling designs’ since many of such designs may represent a combination of probability and non-probability sampling procedures in selecting a sample.**

Some of the popular complex random sampling designs are as follows:

(i) Systematic sampling
(ii) Stratified sampling
(iii) Cluster sampling
(iv) Area sampling
(v) Multi-stage sampling
(vi) Sampling with probability proportional to size
(vii) Sequential sampling

1. Systematic sampling

- **In some instances, the most practical way of sampling is to select every n^{th} item on a list.**

- **Sampling of this type is known as systematic sampling. An element of randomness is introduced in this using random numbers to pick up a unit to start with.**

- **For eg, in a sample the first item would be selected randomly from the first twenty-five and thereafter every 25^{th} item is selected.**

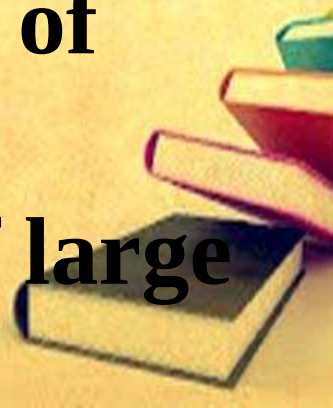
- **Thus, in systematic sampling only the first unit is selected randomly and the remaining units of the sample are selected at fixed intervals.**

•**Although not a random sample in the strict sense, but it is often considered as systematic sample as if it were a random sample.**

Advantages.

- **Improvement over a simple random sample.**
- **easier and less costlier method of sampling**

•**conveniently used even in case of large populations.**



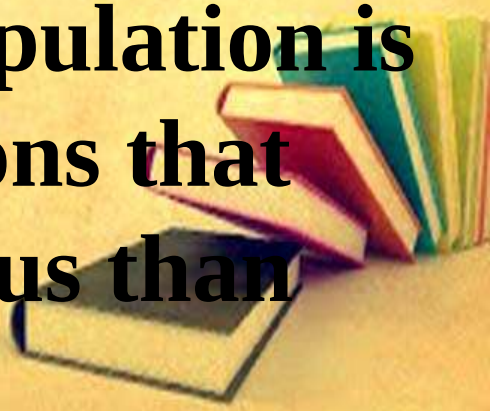
•**Disadvantage**

• **If there is a hidden periodicity in the population, it will prove to be an inefficient method of sampling. For instance, every 25th item produced by a certain production process is defective**

2. 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 into several sub-populations that are individually more homogeneous than**



the total population (the different sub-populations are called ‘strata’) and then we select items 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.**

- **Stratified sampling results in more reliable and detailed information.**

3. 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 clusters), with the ultimate sample consisting of all (or samples of) units in these small areas or clusters.**

Suppose we want to estimate the proportion of machine parts in an inventory which are defective. Also assume that there are 20000 machine parts in the inventory at a 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 surveys in selected clusters. But certainly it is less precise than random sampling.**

- **Cluster sampling is used only because of the economic advantage it possesses; estimates based on cluster samples are usually more reliable per unit cost.**

4. Area sampling

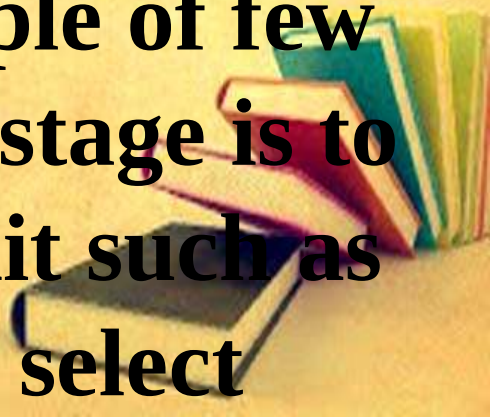
- If clusters happen to be some geographic subdivisions, in that case cluster sampling is better known as area sampling.

- In other words, cluster designs, 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.

5. 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 nationalised 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 a census of all banks within the selected districts, we select certain towns and interview all banks in the chosen towns. This would represent a 3 - stage 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 designs 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 multistage sampling because of sequential clustering, whereas this is not possible in most of the simple designs.

6. Sampling with probability proportional to size

- **In case the cluster sampling units do not have the same number or approximately the same number of elements, it is considered appropriate to use a random selection process where the probability of each cluster being included in the sample is proportional to the size of the cluster.**

For this purpose, we have to list the number of elements in each cluster irrespective of the method of ordering the cluster. Then we must sample systematically the appropriate number of elements from the cumulative totals.

The actual numbers selected in this way do not refer to individual elements, but indicate which clusters and how many from the cluster are to be selected by

simple random sampling or by systematic sampling.

- **The results of this type of sampling are equivalent to those of a simple random sample and the method is less cumbersome and is also relatively less expensive.**

7. Sequential sampling

- **This sampling design is some what complex sample design.**

- **The ultimate size of the sample under this technique is not fixed in advance, but is determined according to mathematical decision rules on the basis of information yielded as survey progresses.**

•**This is usually adopted in case of acceptance sampling plan in context of statistical quality control.**

• **When a particular lot is to be accepted or rejected on the basis of a single sample, it is known as single sampling; when the decision is to be taken on the basis of two samples, it is known as double sampling and in case the decision rests on the basis of more than two samples but the number of samples is certain and**

decided in advance, the sampling is known as multiple sampling.

- **But when the number of samples is more than two but it is neither certain nor decided in advance, this type of system is often referred to as sequential sampling.**

Characteristics of Good Samples

1. Representives:

- A sample is a subset of the population or universe

- The sample must be representatives of the universe. Therefore, the researcher must select the sample members who have the characteristics of the universe.

2. Focus on Objectives:

- **The sample size must be selected depending upon the research objectives.**

- **For instance, if a research is undertaken 'to find out the impact of inflation on the poor then the sample size would be larger, as there are more poor' households in India. (In 2008, there were about 147 million households of poor people in India yearly income of less than Rs. 90,000).**

3. Flexibility

- **The sample size should not be rigidly followed. The sample size can be modified depending upon the circumstances. For instance, the sample size may be reduced, if sufficient information is already available or if there is a limitation of time and funds. However, sample size may be increased, if proper information is not available from the current sample.**

4. Method of Sampling

- **The researcher must select proper method of sampling depending on the purpose.**

5. Proper Selection of Sample Unit:

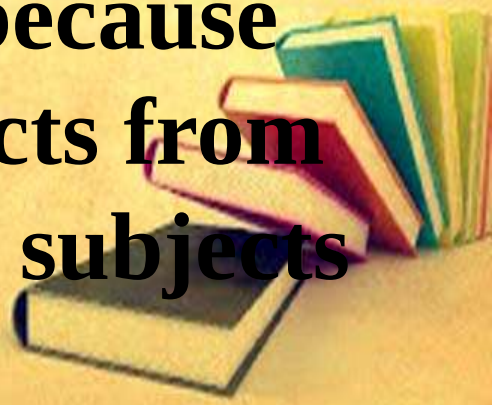
- **The sample unit must be appropriate. The universe comprises of the elements, and each element can be further divided into units.**

For instance, if a study is conducted to study job satisfaction among bank employees, then bank employees comprise the universe. The element of universemay comprise of bank employees / manager in rural banks, and in urban banks. The sample unit may include male and female employees, junior or senior employees.

Sampling Error & ways to reduce them

- **Sampling error is the deviation of the selected sample from the true characteristics, traits, behaviors, qualities or figures of the entire population.**

- **Sampling process error occurs because researchers draw different subjects from the same population but still, the subjects have individual differences.**



- **The most frequent cause of the said error is a biased sampling procedure.**

- **Every researcher must seek to establish a sample that is free from bias and is representative of the entire population.**

- **In this case, the researcher is able to minimize or eliminate sampling error.**

- **Another possible cause of this error is chance.**

- **The process of randomization and probability sampling is done to minimize sampling process error but it is still possible that all the randomized subjects are not representative of the population.**

- **The most common result of sampling error is systematic error wherein the results from the sample differ significantly from the results from the entire population. It follows logic that if the sample is not representative of the entire population, the results from it will most likely differ from the results taken from the entire population.**

Sampling error can also occur in following conditions

1. **Where the researcher does not understand who she should survey/ study – this type of error is population specification error.**

2. **Where the sample design is not properly chosen /occurs when the wrong sub-population is used to select a sample. This type of error is sample frame error.**

3. Selection error — this occurs when respondents self select their participation in the study / only those that are interested respond.

•Selection error can be controlled by going extra lengths to get participation.

Sampling errors can be controlled by

(1) careful sample designs,
(2) large samples,
(3) multiple contacts to assure representative response.
(4) Minimize the potential for bias in the selection of the sample through random sampling.
(5) Ensure the sample is representative of the population by implementing a stratification protocol.

(6) Replicate your study by taking the same measurement repeatedly, using more than one subject or multiple groups, or by undertaking multiple studies.