UNIT-II

SAMPLING & DATA COLLECTION

SAMPLING: Meaning – objectives – census and sample method- Law of statistical regularity and law of Inertia of large numbers – Methods of sampling – Merits and demerits- **COLLECTION OF DATA:** - Primary and Secondary data – Methods of collecting primary data – qualities of good questionnaire- sources of secondary data- precautions in the use of Secondary data.

SAMPLE-MEANING:

A sample refers to a smaller, manageable version of a larger group. It is a subset containing the characteristics of a larger population. Samples are used in statistical testing when population sizes are too large for the test to include all possible members or observations. A sample should represent the population as a whole and not reflect any bias toward a specific attribute.

OBJECTIVES OF A SAMPLING:

The primary objectives of collecting and analysing a sample are to reveal characteristics of a population as follows:

- To estimate the parameters of the population like means, median, mode, etc.
- Testing validity statements about the population
- Investigating the changes in population over time

In other words, the sampling process involves three main elements – selecting the sample, collecting the information, and also making inferences about the population.

CENSUS METHOD:

A census is the procedure of systematically acquiring and recording information about the members of a given population. The object of a census or complete enumeration is to collect information for each and every unit of the population.

The term census is used mostly in connection with national population and housing censuses; other common censuses including agriculture, business, and traffic censuses. A survey that measures the entire target population is called a census. For example, if the average wage of workers working in a sugar factory is to be calculated, then wage figures would be obtained by dividing the total wages which all the workers received by no of workers working in the sugar industry.

MERITS OF CENSUS METHOD:

- The data are collected from each and every item of the population.
- The results are more accurate and reliable, because every item of the universe is enquired.
- Intensive study is possible.
- The data collected may be used for various surveys, analyses, etc.

DEMERITS OF CENSUS METHOD:

- It requires a large number of enumerators and it is a costly method. Therefore, the government alone can use this method for conducting population census, production census, etc.
- It requires more money, labour, time, energy, etc.
- It is not possible in some circumstances where the universe is infinite.

SAMPLE METHOD:

A Sample is the process of learning about the population based on the basis of a sample drawn from it. In this technique, instead of obtaining information from each and every unit of the universe, only a small representative part of the universe is studied and the conclusion is drawn for the entire universe based on the sample. For example: a doctor examines a few drops of blood and draws conclusion about the blood constitution of the whole body. The values obtained from the study of a sample such as the average, dispersion etc., are known as statistics. On the other hand, such values for the population are called parameters.

ESSENTIALS OF SAMPLING

The following are the essentials of sampling:

- Representativeness
- Adequacy
- Independence
- Homogeneity

Representativeness:

We must select the sample in a manner which represents the universe in its truest sense. Further, if we fail to do so, then we might get misleading results.

Adequacy:

We should also select the size of the sample adequately which represents the parametric characteristics of the population.

Independence:

When we select a sample, we must ensure that we select the items independently and also randomly.

Homogeneity:

This is another important element of a sample investigation. Homogeneity means that there is no basic difference in the nature of the units in the sample and the universe.

MERITS OF SAMPLE METHOD:

Here are some important merits of sampling:

Cost-efficient:

In a sample investigation, the costs associated with the collection of data are less. This is because we collect data only from a fraction of the entire population. Therefore, it is cost-efficient.

Time-efficient:

In sampling, we require less time to collect, analyze, and interpret the data since we are working only on a fraction of the population. Hence, it is time-efficient too.

Reliable:

Usually, the data collected under a sample investigation is reliable because of the use of well-trained and experienced investigators or experts.

Flexible:

When we collect data through sampling, we have a greater scope of flexibility.

Detailed Information:

Since sampling is cost-efficient and also time-efficient, we can collect detailed information about the sample in our survey.

DEMERITS OF SAMPLE METHOD:

While sampling has many merits, there are some demerits associated with it too. Here is a quick look:

- It is impossible to attain a 100 percent accuracy using this process. This is because the investigator draws conclusions about the characteristics of the population using the results that he obtains from the selected sample.
- The results are prone to a sampling error or a random error.
- Experts are required to ensure that the results of a sample investigation are satisfactory.
- Sometimes, the sample does not represent the population correctly. This is because it depends on the attitude and mindset of the investigators.
- If the population has a heterogeneous character, then we cannot use this method.

LAWS OF STATISTICS

There are two laws of statistics which form the basis of the theory of sampling. These are:

- The Law of Statistical Regularity
- Law of Inertia of Large Numbers

THE LAW OF STATISTICAL REGULARITY:

This law forms the basis of the theory of probability in statistics. According to this theory, a large random of sample should be taken from a population, and then it is fairly representative of the population.

According to Prof. W.I. King, "The law of statistical regularity formulated in a mathematical theory of probability lays down that a moderately large number of items chosen at random from a very large group are almost sure on the average to have the characteristics of the large group." This law holds well if it meets two conditions:

- The sample is random
- Each and every unit of the population has an equal chance of being selected in the sample.

For Eg: If 50 balls are selected at random from a box containing 400 white and 100 red, then the random from is likely to contain about 40 white and 10 red balls.

THE LAW OF INERTIA OF LARGE NUMBERS:

This law is the corollary of the law of statistical regularity. According to this law, a larger sized sample produces more accurate results. This is because large numbers have more stability, consistency, and steadiness as compared to small numbers.

For Eg: If the probability of falling head or tail is equal i.e. ½ and if we draw a coin 10 times, it is likely that we may not get exactly 5 heads and 5 tails. But at the

same time, if the same experiment is carried out 1000 times, then the chances of 500 heads and 500 tails would be very high.

METHODS / TYPES OF SAMPLING:
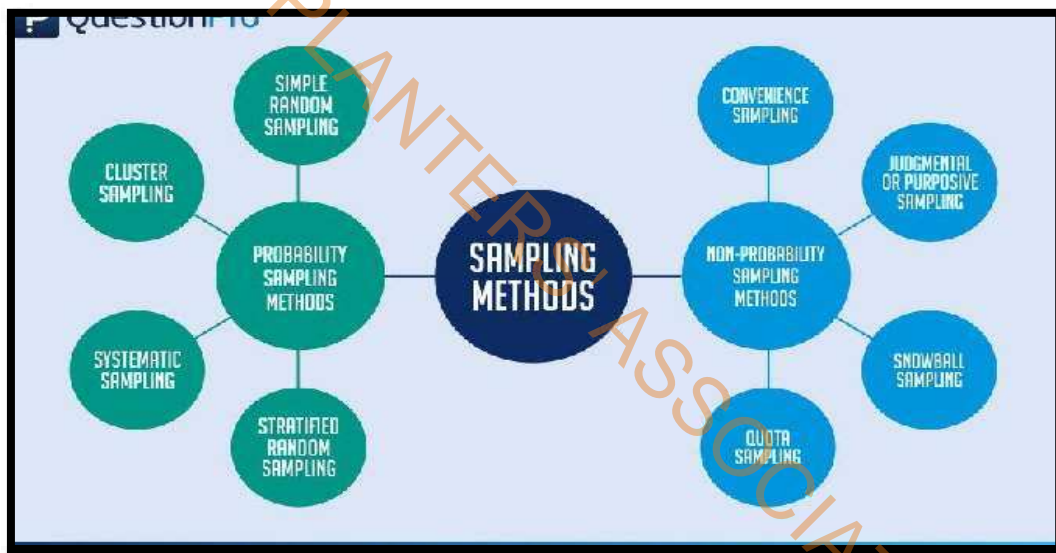
Sampling in market research is of two types:

- **Probability sampling and**
- **Non-probability sampling**

Let's take a closer look at these two methods of sampling.

PROBABILITY SAMPLING:

Probability sampling is a sampling technique where a researcher sets a selection of a few criteria and chooses members of a population randomly. All the members have an equal opportunity to be a part of the sample with this selection parameter.

NON-PROBABILITY SAMPLING:

In non-probability sampling, the researcher chooses members for research at random. This sampling method is not a fixed or predefined selection process. This makes it difficult for all elements of a population to have equal opportunities to be included in a sample.

TYPES OF PROBABILITY SAMPLING WITH EXAMPLES:

Probability sampling is a sampling technique in which researchers choose samples from a larger population using a method based on the theory of probability. This sampling method considers every member of the population and forms samples based on a fixed process.

For example, in a population of 1000 members, every member will have a 1/1000 chance of being selected to be a part of a sample. Probability sampling eliminates bias in the population and gives all members a fair chance to be included in the sample.

There are four types of probability sampling techniques:

- Simple random sampling
- Cluster sampling
- Systematic sampling
- Stratified random sampling

Simple random sampling:

One of the best probability sampling techniques that helps in saving time and resources, is the Simple Random Sampling method. It is a reliable method of obtaining information where every single member of a population is chosen randomly, merely by chance. Each individual has the same probability of being chosen to be a part of a sample.

For example, in an organization of 500 employees, if the HR team decides on conducting team building activities, it is highly likely that they would prefer picking chits out of a bowl. In this case, each of the 500 employees has an equal opportunity of being selected.

Merits of simple random sampling:

- There is no personal bias. Since the selection of items entirely depends on chance.
- When the size of the sample increases, it will be more representative.
- Sampling errors can be measured.
- It saves time, money and energy.
- We can easily assess the accuracy of these estimates.

Demerits of simple random sampling:

- It is difficult to get the up-to-date list of all the items of the population. So, this method is rarely used.
- The size of the sample to be selected should be more to get the statistical reliability.

Cluster sampling:

Cluster sampling is a method where the researchers divide the entire population into sections or clusters that represent a population. Clusters are identified and included in a sample based on demographic parameters like age, sex, location, etc. This makes it very simple for a survey creator to derive effective inference from the feedback.

For example, if the United States government wishes to evaluate the number of immigrants living in the Mainland US, they can divide it into clusters based on states such as California, Texas, Florida, Massachusetts, Colorado, Hawaii, etc. This way of conducting a survey will be more effective as the results will be organized into states and provide insightful immigration data.

Merits of cluster sampling:

- This method is flexible.
- It enables to cover large area.

- It provides a way of selecting random sample where there is no frame work is available.

Demerits of cluster sampling:

- There is a chance of less accuracy.

- More problems arise in deciding the number of stages.

Systematic sampling:

Researchers use the systematic sampling method to choose the sample members of a population at regular intervals. It requires the selection of a starting point for the sample and sample size that can be repeated at regular intervals. This type of sampling method has a predefined range, and hence this sampling technique is the least time-consuming.

For example, a researcher intends to collect a systematic sample of 500 people in a population of 5000. He/she numbers each element of the population from 1-5000 and will choose every 10th individual to be a part of the sample (Total population/ Sample Size = 5000/500 = 10).

Merits of systematic sampling:

- It is a simple and convenient method of sampling.

- It requires less time and energy.

- The results are more reliable.

- This method can be used for infinite population.

Demerits of systematic sampling:

- There is a chance of personal bias

- This method can be applied only when population can be arranged in same order.

↓ This method is less representative.

Stratified random sampling:

Stratified random sampling is a method in which the researcher divides the population into smaller groups that don't overlap but represent the entire population. While sampling, these groups can be organized and then draw a sample from each group separately.

For example, a researcher looking to analyse the characteristics of people belonging to different annual income divisions will create strata (groups) according to the annual family income. Eg – less than $20,000, $21,000 – $30,000, $31,000 to $40,000, $41,000 to $50,000, etc. By doing this, the researcher concludes the characteristics of people belonging to different income groups. Marketers can analyse which income groups to target and which ones to eliminate to create a roadmap that would bear fruitful results.

Merits of stratified random sampling:

↓ It is more representative.
↓ It ensures greater accuracy.
↓ It studies special characteristics of different parts of the population.
↓ Greater geographical concentration reduces time and expenses.
↓ This method is more suitable when the population is heterogeneous.

Demerits of stratified random sampling:

↓ The results may not be reliable if proper stratification is not done.
↓ Stratification requires more time, energy and money.
↓ It is difficult to select items at random from each stratum.

Types of non-probability sampling with examples

The non-probability method is a sampling method that involves a collection of feedback based on a researcher or statistician's sample selection capabilities and not on a fixed selection process. In most situations, the output of a survey conducted with a non-probable sample leads to skewed results, which may not represent the desired target population. But, there are situations such as the preliminary stages of research or cost constraints for conducting research, where non-probability sampling will be much more useful than the other type.

Four types of non-probability sampling explain the purpose of this sampling method in a better manner:

- Convenience sampling
- Judgmental or purposive sampling
- Snowball sampling
- Quota sampling

Convenience or chunk sampling:

This method is dependent on the ease of access to subjects such as surveying customers at a mall or passers-by on a busy street. It is usually termed as convenience sampling, because of the researcher's ease of carrying it out and getting in touch with the subjects. Researchers have nearly no authority to select the sample elements, and it's purely done based on proximity and not representativeness. This non-probability sampling method is used when there is time and cost limitations in collecting feedback. In situations where there are resource limitations such as the initial stages of research, convenience sampling is used.

For example, start-ups and NGOs usually conduct convenience sampling at a mall to distribute leaflets of upcoming events or promotion of a cause – they do that by standing at the mall entrance and giving out pamphlets randomly.

Judgmental or purposive sampling:

Judgemental or purposive samples are formed by the discretion of the researcher. Researchers purely consider the purpose of the study, along with the understanding of the target audience. For instance, when researchers want to understand the thought process of people interested in studying for their master's degree. The selection criteria will be: "Are you interested in doing your masters in …?" and those who respond with a "No" are excluded from the sample.

Snowball sampling:

Snowball sampling is a sampling method that researchers apply when the subjects are difficult to trace. For example, it will be extremely challenging to survey shelter less people or illegal immigrants. In such cases, using the snowball theory, researchers can track a few categories to interview and derive results. Researchers also implement this sampling method in situations where the topic is highly sensitive and not openly discussed—for example, surveys to gather information about HIV Aids. Not many victims will readily respond to the questions. Still, researchers can contact people they might know or volunteers associated with the cause to get in touch with the victims and collect information.

Quota sampling:

In Quota sampling, the selection of members in this sampling technique happens based on a pre-set standard. In this case, as a sample is formed based on specific attributes, the created sample will have the same qualities found in the total population. It is a rapid method of collecting samples. This method is similar to stratified sampling. The selection of the sample will depends on personal judgement.

SOURCE OF DATA COLLECTION:

Collection of data refers to systematic recording of results either by counting or by enumeration. The entire structure of statistical analysis for any enquiry is based on systematic collection of data. Broadly speaking, there are two types of data;

- **Primary data**
- **Secondary data**

PRIMARY DATA:

It is the data, which is collected for the first time by investigators or enumerators working under his supervision to serve a particular purpose. Such a data is of original in nature. In India, the organizations such as CSO (Central Statistical Organization), Census of India, NSS (National Sample Survey), and RBI (Reserve Bank of India) collect and publish the primary data and so they are primary sources of data.

METHODS OF COLLECTING PRIMARY DATA:

Primary data may be obtained by applying any of the following methods:

- Direct personal interviews,
- Indirect oral interviews,
- Information from correspondents,
- Mailed questionnaire method
- Schedule sent through enumerators and
- Telephonic survey.

DIRECT PERSONAL INTERVIEW:

In this method, there is a face-to-face contact between interviewer and the informants. The interviewer asks them questions related to the survey and collect the required information. The information thus obtained is first-hand or original in character.

MERITS:

➢ Response is more encouraging as most people are willing to supply information when approached personally.

➢ The information obtained by this method is likely to be more accurate.

➢ Questions about which the informant is likely to be sensitive can be carefully sandwiched between other questions by the interviewer. He can twist the questions keeping in mind the informant's reactions.

➢ The language of communication can be adjusted to the status and educational level of the person interviewed, thus avoiding inconvenience and misinterpretation on the part of the informant.

DE-MERITS:

➢ It may be very costly where the number of persons to be interviewed is large and they are spread over a wide area.

➢ There are greater chances of personal prejudice and bias when comparing with other methods.

➢ Untrained or poorly trained interviewers may spoil the entire work.

➢ The time required to collect information by this method is more than other methods.

INDIRECT ORAL INTERVIEWS:

Under this method of collecting data, the investigator contacts the third parties called witness capable of supplying the necessary information. Enquiry committees and commissions appointed by the governments generally adopt this method to get people's views. For eg: clues about thefts or murders are obtained by the police by interrogating third parties who are supposed to have knowledge about the case under investigation. This method is very popular in practice. This system is more suitable, where the area to be studied is large. It is used when direct information cannot be obtained.

Merits:

- It is simple and convenient.
- It saves time, money and labour.
- It can be used in the investigation of large area.
- The information is unbiased.

De-merits:

- The information cannot be relied because of the absence of direct contact.
- Interview with improper man will spoil the results.
- The careless attitude of the informant will affect the degree of accuracy.
- Witnesses may colour the information according to their interests.

INFORMATION FROM CORRESPONDENTS:

The investigator appoints local agents or correspondents in different places to collect information. The correspondents collect and transmit the information to the central office where the data are processed. For eg: Newspaper agencies generally

adopt this method. The correspondents may be paid or honorary persons but generally they are paid. This method is adopted in various departments of government. Also correspondents in different places supply information relating to such events as accidents, strikes etc, to head office.

Merits:

- Extensive information can be had.
- It is the most cheap and economical method.
- Speedy information is possible.
- It is useful where information is needed regularly.

De-merits:

- The information may be biased.
- Degree of accuracy cannot be maintained.
- Uniformity cannot be maintained.
- Data may not be original.

MAILED QUESTIONNAIRE METHOD:

Under this method, a well-designed questionnaire is mailed to the informants with a request to fill it up and return the same within the specific time schedule. It is pointed out in the covering letter to the respondents that information supplied by them in the questionnaire will be kept strictly confidential. The investigators send questionnaire along with self-addressed envelope to respondents for quick and better response. This type of method is very popular among research workers, private agencies, etc.

MERITS:

- It is most economical method in terms of time, money and manpower.

- It may be used more effectively in cases where scope of enquiry is very wide and extensive.

- Direct information from respondents, leaves less scope for personal bias.

DE-MRITS:

- This method does not work where majority of respondents are literate.

- There is low degree of reliability of the information supplied by informants.

- This method is not flexible.

SCHEDULE OR QUESTIONNAIRE SENT THROUGH ENUMERATORS:

Under this method of enquiry, the trained enumerators personally visit the informants and explain the objectives of enquiry ask questions and record their replies. This method is commonly used by big business houses, research institutions and large public undertakings.

MERITS:

- It enables to extract information from all type of respondents literate and illiterate.

- The information collected by this method is more accurate and reliable.

- This method is ideal for extensive surveys.

- In the case of poor response, the enumerators can personally persuade the informants to supply information.

DE-MERITS:

 - ➢ It is the most expensive method because the enumerators who collect information are paid workers.

 - ➢ It is time consuming

 - ➢ It is hectic and tiring

 - ➢ The success of this method depends on skill of enumerators, object of enquiry, etc.

TELEPHONIC SURVEY:

Under this method investigator instead of presenting himself before the informants contact them on telephone and collect the desired information.

MERITS:

 - ➢ It is convenient and less time consuming.

DE-MERITS:

 - ➢ It is subjective in nature.

SECONDARY DATA:

It refers to the data which is originally collected and published by the authorities other than who require it. Such data is already available from the government publications, research study, journals or newspapers. Main sources of secondary data may be classified in the following two categories:

 - ➢ **Published sources**

 - ➢ **Un- published sources**

PUBLISHED SOURCES:

Published sources of secondary data basically refers to various national / international organizations / agencies which collect data and publish the statistical data relating to business, trade, labour, price , consumption, production, investment, savings, population, unemployment, banks and financial institutions, corporations etc, are some of the important sources of secondary data are:

- Publications of the central and state governments, of foreign governments and international bodies like IBRD, IMF, ILO, WTO, and WHO etc.

- Publications of semi-government organization eg: Reserve Bank of India Bulletin, reports on currency and finance.

- Publications of CSO and other research bodied (ICAR & IARI, Delhi).

- Publications of various chambers of commerce, trade associations and co-operative societies.

UNPUBLISHED SOURCES:

It covers all those sources of secondary data where records are maintained by private agencies or business firms for their own use and are restrictedly available for use of general public. Data collected by research institutions are also included in the category of unpublished sources of secondary data.

DIFFERENCE BETWEEN PRIMARY AND SECONDARY DATA:

Sr. No	Points	Primary Data	Secondary Data
1.	Meaning	Data collected by researcher himself	Data collected by other persons.
2.	Originality	Original or unique information	Not original or unique information.
3.	Adjustment	Doesn't need adjustment, is focused	Needs adjustment to suit actual aim.
4.	Sources	Surveys, observations, experiments	Internal records, Govt. published data, etc.
5.	Type of data	Qualitative data	Quantitative data
6.	Methods	Observation, experiment, interview	Desk research method, searching online, etc.
7.	Reliability	More reliable	Less reliable
8.	Time consumed	More time consuming	Less time consuming
9.	Need of investigators	Needs team of trained investigators	Doesn't need team of investigators
10.	Cost effectiveness	Costly	Economical
11.	Collected when	Secondary data is inadequate	Before primary data is collected
12.	Capability	More capable to solve a problem	Less capable to solve a problem
13.	Suitability	Most suitable to achieve objective	May or may not be suitable
14.	Bias	Possibility of bias exist	Somewhat safe from bias
15.	Collected by	Researcher or his agents	Persons other than who collects primary data
16.	Precaution to use	Not Necessary	Quite necessary

REQUISITES OF A GOOD QUESTIONNAIRE:

➢ A covering letter should be enclosed with every questionnaire.

➢ The questions should be simple and short to understand and the number of questions should be small.

➢ Questions should be arranged logically.

➢ Ambiguous questions should be avoided.

➢ As far as possible, personal questions about income, sales-tax paid, etc., should be avoided.

> The necessary instructions to the informants should be included.

> Objective type questions and 'yes or no' questions should be included.

> A questionnaire should be pre-tested and cross-checked before mailing it.

PRECAUTIONS IN THE USE OF SECONDARY DATA:

Secondary data are collected by someone and used by others. So, the investigators should check about the reliability, suitability, accuracy and adequacy before using secondary data. It is because the secondary data may be inadequate and unsuitable. Thus, the following precautions should be adopted while using secondary data:

RELIABILITY OF SECONDARY DATA:

The success or failure of any research work depends on the reliability of data. If the collected data are accurate and reliable then the results become right and such type of investigation shows reality. If the collected data are wrong or insufficient, the result becomes wrong and the consequences become harmful. Thus, before using secondary data, we should check whether the data are reliable or not. While testing the reliability, the following things should take into consideration.

a) Who collected the data, what were the objectives of collecting data?

b) Was the investigator experienced, capable, honest, and unbiased?

c) Whether the appropriate techniques of collecting data were applied or not?

d) What degree of accuracy was maintained in the data?

e) Was there checking about the accuracy of data?

f) In which time period and in what sort of condition were the data collected?

SUITABILITY OF DATA:

While using secondary data, we have to examine thoroughly, whether the data are suitable for our investigation or not. The suitability of the data can be tested by comparing the nature, objectives, and scope of the present inquiry with those of the original inquiry. If the nature of the data is different, such data will not be suitable for the study and can't be a valid result.

ADEQUACY OF DATA:

Adequacy of the data is to be judged in the light of requirement of the survey and geographical area covered by the available data; for example, if the purpose of enquiry is to study the working conditions of workers in sugar industry in India and if the data is available only to the State of Tamil Nadu, then such data is insufficient and cannot be used.

ACCURACY OF DATA:

The accuracy of conclusion depends mainly on the accuracy of the collected data. Therefore, the investigator must see whether secondary data are accurate and can be used for the study.

Thus, these are the precautions in the use of secondary data.

QUESTIONS:

1. Distinguish between Primary and secondary data.

2. Discuss the various methods of collecting primary data.

3. What are the requisites of a good questionnaire?

4. Define secondary data. State the sources of secondary data.

5. Point out the precautions in the use of secondary data.

6. Discuss the methods of data collection.

7. Write a short note on: Law of inertia of large numbers and the law of statistical regularity.

8. Elucidate / describe the methods of sampling.

9. What is meant by census method and state its merits and demerits.

10. List out the essentials of sampling.

11. What is meant by sample method and state its merits and demerits.

12. What is meant by systematic sampling and write its merits and demerits.

13. State the meaning of direct interview method and state its merits and demerits.