ECO-SE-3014: Data Collection and Presentation

UNIT 2: Use of Data

Data and Sources

What is Primary Data?

Primary data is the kind of data that is collected directly from the data source without going through any existing sources.

Primary data is often reliable, authentic, and objective in as much as it was collected with the purpose of addressing a particular research problem. It is noteworthy that primary data is not commonly collected because of the high cost of implementation.

A common example of primary data is the data collected by organizations during market research, product research, and competitive analysis. This data is collected directly from their original sources which in most cases are the existing and potential customers.

Sources of Primary Data:

The sources of primary data are primary units such as basic experimental units, individuals, households. Following methods are used to collect data from primary units usually and these methods depend on the nature of the primary unit.

- Personal Investigation

The researcher conducts the experiment or survey himself/herself and collected data from it. The collected data is generally accurate and reliable. This method of collecting primary data is feasible only in case of small scale laboratory, field experiments or pilot surveys and is not practicable for large scale experiments and surveys because it take too much time.

- Through Investigators

The trained (experienced) investigators are employed to collect the required data. In case of surveys, they contact the individuals and fill in the questionnaires after asking the required information, where a questionnaire is an inquiry form having a number of questions designed to obtain information from the respondents. This method of collecting data is usually

employed by most of the organizations and its gives reasonably accurate information but it is very costly and may be time taking too.

- Through Questionnaire

The required information (data) is obtained by sending a questionnaire (printed or soft form) to the selected individuals (respondents) (by mail) who fill in the questionnaire and return it to the investigator. This method is relatively cheap as compared to "through investigator" method but non-response rate is very high as most of the respondents don't bother to fill in the questionnaire and send it back to investigator.

- Through Local Sources

The local representatives or agents are asked to send requisite information that provides the information based upon their own experience. This method is quick but it gives rough estimates only.

- Through Telephone

The information may be obtained by contacting the individuals on telephone. It's a quick and provide accurate required information.

- Through Internet

With the introduction of information technology, the people may be contacted through internet and the individuals may be asked to provide the pertinent information. Google survey is widely used as online method for data collection now a day.

What is Secondary Data?

Secondary data is the data that has been collected in the past by someone else but made available for others to use. They are usually once primary data but become secondary when used by a third party.

Secondary data are usually easily accessible to researchers and individuals because they are mostly shared publicly. This, however, means that the data are usually general and not tailored specifically to meet the researcher's needs as primary data does.

Some common sources of secondary data include trade publications, government statistics, journals, etc. In most cases, these sources cannot be trusted as authentic.

Sources of Secondary Data:

The secondary data may be available from the following sources-

- Government Organizations
- Semi-Government Organization
- Municipal committees, District Councils, Commercial and Financial Institutions like banks etc.
- Teaching and Research Organizations
- Research Journals and Newspapers
- Internet

The differences between the primary and secondary data:

- Definition

  Primary data are those that are collected for the first time. Secondary data refer to those data that have already been collected by some other person.

- Originality

  Primary data are original because these are collected by the investigator for the first time. Secondary data are not original because someone else has collected these for his own purpose.

- Nature of Data

  Primary data are in the form of raw materials. Secondary data are in the finished form.

- Reliability and Suitability

  Primary data are more reliable and suitable for the enquiry because these are collected for a particular purpose. Secondary data are less reliable and less suitable as someone else has collected the data which may not perfectly match our purpose.

- Time and Money

  Collecting primary data is quite expensive both in the terms of time and money. Secondary data requires less time and money; hence it is economical.

Sampling and its Types

A sample is a group of units selected from the population used to draw valid conclusions about the larger group. Sampling is the process of selecting a subset of observations from an entire population of interest so that characteristics from the subset (sample) can be used to draw conclusions or making inferences about the entire population.

Probability sampling:

Probability sampling involves selection of samples based on probability and involves random selection. In this case every unit in the population has a chance or probability of being selected in the sample. In fact, the probability of being selected can be determined.

The different probability sampling or random sampling methods are:

(i) Simple Random Sampling

(ii) Systematic Random Sampling

(iii) Stratified Random Sampling

(iv) Cluster Sampling and

(v) Multistage Sampling

- Simple random sample: Every member and set of members has an equal chance of being included in the sample. Technology, random number generators, or some other sort of chance process is needed to get a simple random sample.
- Stratified random sample: The population is first split into groups. The overall sample consists of some members from every group. The members from each group are chosen randomly.
- Cluster random sample: The population is first split into groups. The overall sample consists of every member from some of the groups. The groups are selected at random.
- Systematic random sample: Members of the population are put in some order. A starting point is selected at random and every n^{th} member is selected to be in the sample.

- Multistage sampling: It can be a complex form of cluster sampling because it is a type of sampling which involves dividing the population into groups (or clusters). Then, one or more clusters are chosen at random and everyone within the chosen cluster is sampled.

Non-probability sampling:

In non-probability sampling the sample selection is not random; hence, the samples are not derived on the basis of probability.

Non probability sampling methods can be divided into convenience or accidental and purposive sampling approaches. Convenience sampling is the most common form of sampling used in social science research and simply involves the selection of units that are available for study. In purposive sampling the sample is selected with an unambiguous purpose and thus the choice of sampling units depends entirely on the discretion of the investigator.

Differences between Census and Sampling Method:

Census	Sampling
A systematic method that collects and records the data about the members of the population is called Census.	Sampling refers to a portion of the population selected to represent the entire group, in all its characteristics.
Enumeration is complete in census.	Enumeration is partial in sampling.
Each and every unit of the population is studied.	A part of the population is studied.
It is a time consuming process.	It is a time saving process.
Expensive method	Economic method
Reliable and accurate	Less reliable and less accurate
It is appropriate for population of heterogeneous nature.	It is appropriate for population of homogeneous nature.