

UNIT-1 STATISTICAL DATA COLLECTION AND TYPES (04 HOURS)**Data:**

Definition: Data is a collection of information gathered by observations, measurements, research or analysis. They may consist of facts, numbers, names, figures or even description of things. Data is organized in the form of graphs, charts or tables.

Data Collection: The data collection is a process of gathering information from all the relevant sources to find a solution to the research problem. It helps to evaluate the outcome of the problem. The data collection methods allow a person to conclude an answer to the relevant question. Most of the organizations use data collection methods to make assumptions about future probabilities and trends. Once the data is collected, it is necessary to undergo the data organization process.

Classification of Data

The main sources of the data collections methods are “Data”. A data can be classified into two types, namely primary data and the secondary data. Depends on the type of data, the data collection method is divided into two categories namely,

- Primary Data or primary data collection methods
- Secondary Data or Secondary data collection methods

Primary Data

Primary data or raw data is a type of information that is obtained directly from the first-hand source through experiments, surveys, or observations. The primary data is further classified into two types. They are

- Qualitative Data
- Quantitative Data

Qualitative Data

If the data is classified on the basis of the qualitative characteristics or attributes is called qualitative data.

It does not involve any mathematical calculations. This method is closely associated with elements that are not quantitative. This qualitative data collection method includes interviews, questionnaires, observations, case studies survey focus group discussion etc. There are several methods to collect this type of data. They are also called data collection tools.

Data collection tools of Qualitative Data

1. Questionnaires
2. Survey
3. Interviews
4. Focus group discussion.

Questionnaires

In this method, the set of questions are mailed to the respondent. They should read, reply and subsequently return the questionnaire. The questions are printed in the definite order on the form. A good survey should have the following features:

- Short and simple
- Should follow a logical sequence
- Provide adequate space for answers
- Avoid technical terms
- Should have good physical appearance such as colour, quality of the paper to attract the attention of the respondent

Survey:

A survey is a process of data gathering involving a variety of data collection methods, including a questionnaire. Questionnaires are a list of questions either an open ended (Questions in which the respondent answers in his own words or close - ended (Question in which respondent selects one or more options from pre-determined set of responses for which the respondent give answers). Enumerator goes to the respondents, asks them the questions from the Performa in the order listed, and records the responses in the space provided Questionnaire is the most commonly used method in survey.

Interviews

The Interview Method of collecting data involves presentation of oral-verbal stimuli and reply in terms of oral – verbal responses. It requires the interviewer asking questions in a face-to-face contact with the person. The method of collecting data in terms of oral or verbal responses. It is achieved in two ways, such as

- **Personal Interview** – In this method, a person known as an interviewer is required to ask questions face to face to the other person. The personal interview can be structured or unstructured, direct investigation, focused conversation etc.
- **Telephonic Interview** – In this method, an interviewer obtains information by contacting people on the telephone to ask the questions or views orally.

Focus group discussion

The opposite of quantitative research which involves numerical based data, this data collection method focuses more on qualitative research. It falls under the primary category for data based on the feelings and opinions of the respondents. This research involves asking open-ended questions to a group of individuals usually ranging from 6-10 people, to provide feedback.

Quantitative Data Collection Methods

The classification of the units on the basis of quantitative characteristics or variable (such as weight, wages, age in years, numbers of children, phone number etc) is called quantitative data collection method. It is based on mathematical calculations using various formats like close-ended questions, correlation and regression methods, mean, median or mode measures. This method is cheaper than qualitative data collection methods, and it can be applied in a short duration of time.

Quantitative data is also known as numerical data which represents the numerical value (i.e., how much, how often, how many). Numerical data gives information about the quantities of a specific thing. Some examples of numerical data are height, length, size, weight, and so on. The quantitative data can be classified into two different types

based on the data sets. The two different classifications of numerical data are, discrete data and continuous data.

Discrete Data

Discrete data can take only discrete values. Discrete information contains only a finite number of possible values. Those values cannot be subdivided meaningfully. Here, things can be counted in the whole numbers.

Example: Number of students in the class

Continuous Data

Continuous data is data that can be calculated. It has an infinite number of probable values that can be selected within a given specific range.

Example: Temperature range

Secondary Data Collection Methods

Secondary data is the data which has been already collected and analysed by someone other than the actual user. It means that the information is already available, and someone analyses it. The secondary data includes magazines, newspapers, books, journals etc. It may be either published data or unpublished data.

Published data are available in various resources including

- Government publications
- public records
- Historical and statistical documents
- Business documents
- Technical and trade journals

Unpublished data includes

- Diaries
- Letters
- Unpublished biographies etc.

Data cleaning.

Data cleansing or data cleaning is the process of detecting and correcting corrupt or inaccurate records from a record set, table, or database and refers to identifying incomplete, incorrect, inaccurate or irrelevant parts of the data and then replacing, modifying, or deleting the dirty or coarse data

There are several methods for cleaning data depending on how it is stored along with the answers being sought.