

UNIT-1

STATISTICAL DATA COLLECTION AND TYPES

DATA

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.

TYPES OF DATA

Statistical data is classified in 3 groups. They are

1. Qualitative data
2. Quantitative data
3. Mixed data

• QUALITATIVE DATA

Data that is represented either in a verbal or narrative format is qualitative data. These types of data are collected through focus groups, interviews, opened ended questionnaire items, and other less structured situations. A simple way to look at qualitative data is to think of qualitative data in the form of words.

This data type isn't necessarily measured using numbers but rather categorized based on properties, attributes, labels, and other identifiers. Numbers like national identification number, phone number, etc. are however regarded as qualitative data because they are categorical and unique to one individual.

Examples of qualitative data include sex (male or female), name, state of origin, citizenship, etc.

Qualitative Data can be divided into two types, namely;

- i. Nominal
- ii. Ordinal Data

▪ Nominal data

In statistics, nominal data is a classification of categorical variables that do not provide any quantitative value. It is sometimes referred to as labeled or named data.

For example, a researcher may need to generate a database of the phone numbers and location of a certain number of people. An online survey may be conducted using a closed open-ended question. Example: Enter your phone number with country code.

▪ Ordinal data

In statistics ordinal data is a type of qualitative data where the variables have natural, ordered categories and the distances between the categories are not known.

For example, ordinal data is said to have been collected when a customer inputs his/her satisfaction on the variable scale — "satisfied, indifferent, dissatisfies"

Another example is the data collected from asking a question with a Likert scale is ordinal.

An organization creates an employee exit questionnaire which primarily highlights this question: "How will you rate your experience working with us?"

- Very great
- Great
- Bad
- Very bad

UNIT-1

STATISTICAL DATA COLLECTION AND TYPES

• **QUANTITATIVE DATA**

These can be measured and not simply observed. They can be numerically represented and calculations can be performed on them.

It is also known as numerical data which represents the numerical value i.e., how much, how often, how many.

Quantitative variables are divided into two types: discrete and continuous.

- 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

- On the other hand, 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

For simplicity, we usually referred to years, kilograms (or pounds) and centimeters (or feet and inches) for age, weight and height respectively. However, a 28-year-old man could actually be 28 years, 7 months, 16 days, 3 hours, 4 minutes, 5 seconds, 31 milliseconds, 9 nanoseconds old.

• **MIXED DATA**

When both qualitative and quantitative data are mixed for analysis then the data type becomes mixed data.

Examples for data types

Qualitative Data

- Family Partnership Agreements
- Social Service logs
- Advisory group minutes
- Policy Council minutes
- Newspaper articles

Quantitative Data

- Child performance tracking
- Health data tracking systems

Mixed Data

- Enrolments info
- Enrolments & transition records
- Surveys (i.e., Parent, Teacher/Staff, Farmer) becomes mixed data.

DATA COLLECTION

Data collection is the systematic approach to gathering and measuring information from a variety of sources to get a completed and accurate picture of an area of interest.

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.

Types of data collection

❖ **Primary Data**

These are the data that are collected for the first time by an investigator for a specific purpose. Primary data are 'pure' in the sense that no statistical operations have been performed on them and they are original. An example of primary data is the Census of India.

UNIT-1

STATISTICAL DATA COLLECTION AND TYPES

❖ **Secondary Data Collection Methods**

Secondary data is data which has been already collected and analyzed 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

Purpose of data collection:

- To obtain information
- To keep on record
- To make decisions about important issues

In order to successfully manage the data collection process, programs need a plan that addresses the following

- What types of data are most appropriate to answer the questions?
- How much data are necessary?
- Who will do the collection?
- When and where will the data be collected?
- How will the data be compiled and later stored

DATA COLLECTION TOOLS:

Data collection tools refer to the devices/instruments used to collect data, such as a paper questionnaire or computer-assisted interviewing system. Case Studies, Checklists, Interviews, Observation sometimes, and Surveys or Questionnaires are all tools used to collect data.

Some of the data collection tools are:

- Questionnaires.
- Survey.
- Interviews.
- Focus group discussion.

I. 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.

Types of questionnaires

- **Structured Questionnaires:** Structured questionnaires collect quantitative data. The questionnaire is planned and designed to gather precise information. It also initiates a formal inquiry, supplements data, checks previously accumulated data, and helps validate any prior hypothesis.

UNIT-1

STATISTICAL DATA COLLECTION AND TYPES

• **Unstructured Questionnaires:** Unstructured questionnaires collect qualitative data. They use a basic structure and some branching questions but nothing that limits the responses of a respondent. The questions are more open-ended to collect specific data from participants.

Types of questions

- **Open-Ended Questions:** Open-ended questions help collect qualitative data in a questionnaire where the respondent can answer in a free form with little to no restrictions.
- **Close-ended question/ Dichotomous:** The dichotomous question is generally a “yes/no” close-ended question. This question is usually used in case of the need for necessary validation. It is the most natural form of a questionnaire.
- **A single response question** is the one where the respondent will have only one option to give the response, ex is either yes or no, true or false, male or female etc
- **A multiple response question** possesses multiple options for the respondents to give the response.
- **A Rating question** provides an option for the respondent to compare different items using a common scale.

For example, please rate each of the following objects on the rating scale 1-10.

- **A Ranking question** on the other hand asks the respondents to compare a list of different objects to one another.
- **Multiple-Choice Questions:** Multiple-choice questions are a close-ended question type in which a respondent has to select one (single-select multiple-choice question) or many (multi-select multiple choice question) responses from a given list of options.
- **Pictorial Questions:** This question type is easy to use and encourages respondents to answer. It works similarly to a multiple-choice question. Respondents are asked a question, and the answer choices are images. This helps respondents choose an answer quickly without over-thinking their answers, giving you more accurate data.

II. Survey

Survey is a method of gathering information from a sample of people, traditionally with the intention of generalizing the results to a larger population.

Surveys provide a critical source of data and insights for nearly everyone engaged in the information economy, from businesses and the media to government and academics.

How to Design a Survey:

- Define your research question and goals
- Identify who you'll be surveying
- Design and pre-test surveys
- Select a sample to survey
- Send out your survey
- Analyze the data

UNIT-1

STATISTICAL DATA COLLECTION AND TYPES

III. Interviews

An interview is a face-to-face conversation between two individuals with the sole purpose of collecting relevant information to satisfy a research purpose. 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.

IV. 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.

DATA CLEANING

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.

Data cleaning is not simply about erasing information to make space for new data, but rather finding a way to maximize a data set's accuracy without necessarily deleting information. For one, data cleaning includes more actions than removing data, such as fixing spelling and syntax errors, standardizing data sets, and correcting mistakes such as empty fields, missing codes, and identifying duplicate data points.

What are the different types of data issues?

- **Duplicate data:** There are 2 or more identical records. This may cause misrepresentation of inventory counts/duplication of marketing collateral or unnecessary billing activities.
- **Conflicting Data:** When there are same records with different attributes, it means data is conflicting. For example, a company with different versions of addresses may cause delivery issues.
- **Incomplete Data:** The data that has missing attributes.
- **Invalid Data:** Data attributes are not conforming to standardization. For example, 9-digit phone number records rather than 10 digits.

What are the benefits of data cleaning?

Almost all modern business processes involve data. Subsequently, when data cleaning is seen as an important organizational effort, it can lead to a wide range of benefits for all. Some of the biggest advantages include:

- Streamlined business practices:
- Better decisions: Better data = better decisions.
- Increased productivity:
- Faster sales cycle