Statistics and Analytics Lab(20SC02P)
Unit-1
STATISTICAL DATA COLLECTION AND TYPES (04 Hours)

Data: Data refers to information, facts, or statistics.

Data Definition	Examples
Data refers to information, facts, or statistics.	1. Temperature Readings (Hourly temperature in cities)
	2. Student Grades (Grades in different subjects)
	3. Sales Figures (Monthly sales of a company)
	4. Social Media Posts (Content, comments, likes on platforms)
	5. Health Records (Patient's medical history and test results)

Data Collection:
Data collection involves gathering information or data.

Data Collection Definition	Examples
Data collection involves gathering information or data.	1. Survey Questionnaires (Collecting customer feedback)
	2. Sensor Readings (Recording temperature at different locations)
	3. Social Media Monitoring (Analyzing public posts and comments)

Classification of 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

Qualitative Data Definition: Qualitative data refers to non-numerical information that describes qualities, characteristics, and subjective attributes. It provides a deeper understanding of individuals' experiences, opinions, beliefs, and behaviors.

Here's an example table illustrating qualitative data:

Qualitative Data	Description	Example
Interview Quotes	Direct quotes from interviews capturing participants' perspectives and experiences	"I felt excited and empowered after the workshop."
Observations	Descriptive notes on behaviors, interactions, and contexts observed in a natural setting	"The children were engaged in collaborative play activities."
Open-Ended Responses	Free-form responses in surveys or questionnaires that allow participants to provide detailed explanations	"Please describe your most memorable travel experience."
Focus Group Transcripts	Verbatim records of group discussions, capturing participants' opinions, interactions, and insights	"I think the design should be more user-friendly and intuitive."
Narrative Data	Storytelling or narratives that convey personal experiences, perspectives, or anecdotes	"Let me share a story of how I overcame a challenge."

Data collection tools of Qualitative Data

Data Collection Method	Description	Examples
Surveys	Gathering data through structured questionnaires or interviews	Customer satisfaction surveys, employee feedback surveys
Observations	Collecting data by observing and recording behaviors or events	Ethnographic observations, wildlife behavior observations

Interviews	Conducting one-on-one or group interviews to gather information	In-depth interviews with experts, focus group discussions
Experiments	Controlling variables to test hypotheses and collect data	Lab experiments in chemistry, psychology, or biology
Document Analysis	Analyzing existing documents or records for relevant data	Examining financial reports, historical archives
Sensor Technology	Collecting data using various sensors and measuring devices	Weather sensors, fitness trackers, traffic monitoring
Social Media Monitoring	Gathering data from social media platforms and online sources	Analyzing tweets, Facebook posts, YouTube comments
Web Scraping	Extracting data from websites using automated tools	Collecting product prices, customer reviews from e-commerce sites
Focus Groups	Conducting group discussions to elicit insights and opinions	Market research focus groups, product testing sessions

Questionnaires

Questionnaires are structured sets of questions designed to collect data from individuals or groups. They are a common method of data collection in research, surveys, and feedback gathering. Questionnaires can be administered in various formats, including paper-based, online, email, telephone, or in-person interviews. Respondents provide their responses to the questions, allowing researchers to gather quantitative or qualitative data for analysis.

Questionnaire Type	Description	Example
Customer Satisfaction	Measures customers' satisfaction with products or services	Hotel guest satisfaction survey
Employee Feedback	Gathers employees' opinions and feedback on workplace issues	Employee engagement survey
Market Research	Collects data on market trends, consumer preferences, and purchasing behavior	New product research survey
Health Assessment	Assesses individuals' health conditions, habits, and lifestyle	Health risk assessment questionnaire
Social Media Usage	Investigates social media usage patterns, behaviors, and preferences	Social media usage survey

A good survey should have the following features:
- o   Short and simple
- o   Should follow a logical sequence
- o   Provide adequate space for answers
- o   Avoid technical terms

○ Should have good physical appearance such as colour, quality of the paper to attract the attention of the respondent

Survey:

Surveys are a method of data collection used to gather information from a sample of individuals or groups.

They involve systematically asking a set of questions to elicit responses, which are then analysed to draw insights and make informed decisions. Surveys can be conducted through various mediums, including paper-based forms, online platforms, telephone interviews, or in-person interviews.

Here's an example table showcasing different types of surveys and their applications:

Survey Type	Description	Example
Customer Satisfaction	Measures customers' satisfaction with products or services	Restaurant satisfaction survey
Employee Engagement	Assesses employees' satisfaction, motivation, and commitment	Employee engagement survey
Market Research	Collects data on consumer preferences, buying habits, and market trends	Product preference survey
Political Opinion	Gathers public opinions on political candidates, issues, or policies	Election exit poll
Academic Research	Conducts surveys for academic studies and research projects	Student feedback survey

Interviews

Interviews are a method of data collection that involves direct communication between an interviewer and an interviewee.

They allow for in-depth exploration of topics, gathering rich qualitative data through conversations. Interviews can be conducted in-person, over the phone, or through video conferencing.

Here's a table showcasing different types of interviews, their characteristics, and an example:

Interview Type	Description	Example
Structured Interview	Follows a predetermined set of questions asked in a fixed order. Questions are standardized to ensure consistency.	Job interview with predefined questions
Semi-Structured Interview	Combines predetermined questions with room for open-ended discussion. The interviewer has flexibility to explore additional topics.	Research interview with a mix of fixed and open-ended questions
Unstructured Interview	Freestyle conversation with no predetermined set of questions. The interviewer allows the conversation to flow naturally, exploring various topics.	Qualitative research interview on personal experiences

| Group Interview | Involves multiple participants being interviewed simultaneously. Participants interact and provide insights based on each other's responses. | Focus group discussion on consumer preferences |
| Telephone/Remote Interview | Conducted over the phone or via video conferencing. Enables remote interviews without the need for physical proximity. | Remote research interview with a subject matter expert |

Focus group discussion

The characteristics of focus group discussions and an example:

Focus Group Discussion	Description	Example
Definition	A qualitative research method involving a small group of participants engaging in a guided discussion on a specific topic.	Focus group discussion on consumer preferences
Group Size	Typically consists of 6-10 participants, allowing for diverse perspectives and interactive discussions.	Focus group discussion on new product development
Moderator	A skilled moderator guides the discussion, ensures all participants have a chance to express their opinions, and probes for deeper insights.	Focus group discussion on marketing strategies
Interactive and Dynamic	Participants engage in a back-and-forth conversation, responding to each other's comments, building on ideas, and generating new insights collectively.	Focus group discussion on public policy issues
Qualitative Data Collection	Focus group discussions provide rich qualitative data, capturing participants' thoughts, feelings, beliefs, and shared experiences.	Focus group discussion on social media usage

Quantitative Data Collection Methods

Definition: Quantitative data refers to numerical information that can be measured and analyzed statistically. It involves objective measurements or counts, allowing for precise calculations and quantitative analysis.

different types of quantitative data along with examples:

		Time taken to complete a task (e.g., 2.3 minutes, 6.8 minutes)

Quantitative Data Type	Description	Example
Continuous Data	Represents measurements on a continuous scale without any gaps or interruptions	Height measurements (e.g., 165 cm, 180 cm)
		Temperature readings (e.g., 25.5°C, 36.7°C)
Discrete Data	Consists of separate, distinct values or categories	Number of children in a family (e.g., 0, 1, 2)
		Number of defects in a product (e.g., 0, 1, 2, 3, ...)
		Number of customer complaints (e.g., 0, 1, 2, 3, ...)
Binary Data	Represents data with only two possible outcomes	Yes/No responses
		True/False answers
		Presence/Absence of a characteristic
Ordinal Data	Represents data with a specific order or ranking	Ratings of satisfaction (e.g., 1 to 5 scale)
		Performance rankings (e.g., Excellent, Good, Fair, Poor)
		Educational levels (e.g., High School, Bachelor's, Master's)

Secondary Data Collection Methods

Existing data collected by someone else for a different purpose.

Defining secondary data, its types, and the methods used for collecting secondary data:

Secondary Data Definition	Types	Collection Methods
Existing data collected by someone else for a different purpose.	Published Sources: Books, journals, reports.	Literature Review: Analyzing existing research and publications.
	Government Sources: Data collected by government agencies.	Online Research: Extracting data from online sources and databases.

	Institutional Sources: Data collected by organizations or institutions.	Data Aggregation: Compiling data from multiple sources.
	Online Databases: Data available through online platforms.	Government Reports: Using data from official reports and surveys.
	Historical Data: Data collected in the past and archived.	Organizational Records: Utilizing data from organizational records.

- Government publications
- public records
- Historical and statistical documents
- Business documents
- Technical and trade journals Unpublished data includes
- Diaries
- Letters
- Unpublished biographies etc.

Data cleaning.

Definition: Data cleaning, also known as data cleansing or data scrubbing, is the process of identifying and correcting or removing errors, inconsistencies, and inaccuracies in datasets.

It involves the detection and resolution of problems to ensure that data is accurate, complete, and reliable for analysis.

Here's an example illustrating data cleaning:

Data Cleaning Task	Description	Example
Removing Duplicates	Identifying and eliminating duplicate entries in the dataset	Removing multiple entries of the same customer in a sales dataset
Handling Missing Values	Dealing with missing data points by imputing or removing them	Filling in missing age values with the median age in a survey dataset
Standardizing Formats	Ensuring consistent formats for variables or fields in the dataset	Converting date formats to a uniform format (e.g., YYYY-MM-DD)
Correcting Typos	Rectifying spelling errors or typographical mistakes in the data	Correcting misspelled names or addresses in a customer database
Handling Outliers	Identifying and addressing extreme or unusual data points that deviate significantly from the norm	Removing or adjusting values that are beyond a certain threshold