CHAPTER 3
DATA COLLECTION AND ANALYSIS

Competency:
The learner supports evidence-based decision making through collecting, analyzing interpreting and communicating data using a range of media to solve societal problems

Learning out comes:
The learner should be able to:

- Analyze appropriate methods for data collection in relation to there are of research
- Develop sample tools for data collection in their research
- Establish how best they can present data for their topics
- Understand mathematical and statistical concepts used in data presentation

INTRODUCTION:

What is data?

[6]Data refers to facts, figures, and information collected, stored, and analyzed to gain insights, make informed decisions, or support research. Data can be defined as information in a raw or unprocessed form. It is further defined as a collection of facts and figures used to understand a particular phenomenon.

It is the digital information stored, processed, and transmitted by computers. It means observations and measurements collected through research, experimentation, or observation.

Data can be defined as inputs and outputs of a system, process, or function.

Types of Data

1. Qualitative data: Non-numerical data, such as text, images, or observations.
2. Quantitative data: Numerical data, such as statistics, measurements, or counts.

Uses of Data

1. Informed decision-making: Data informs decisions in various fields, such as business, healthcare, or education.
2. Research: Data is used to support research studies and findings.
3. Analysis: Data is analyzed to identify patterns, trends, and insights.
4. Problem-solving: Data helps identify problems and develop solutions.
5. Predictive analytics: Data is used to predict future trends, behaviors, or outcomes.
6. Personalization: Data is used to personalize experiences, such as product recommendations or targeted advertising.
7. Risk assessment: Data is used to assess risks, such as credit risk or insurance risk.
8. Quality improvement: Data is used to identify areas for improvement and optimize processes.
9. Scientific research: Data is used to support scientific research and discovery.
10. Business intelligence: Data is used to inform business decisions, track performance, and identify opportunities.
11. Public policy: Data is used to inform public policy decisions and evaluate program effectiveness.
12. Healthcare: Data is used to improve patient outcomes, track disease trends, and optimize treatment plans.

[6] W M Harper, *Statistics, sixth edition by financial time pitman publishing*

13.9. Marketing: Data is used to understand customer behavior, track market trends, and measure campaign effectiveness.
10. Artificial intelligence: Data is used to train and improve artificial intelligence models.

Characteristics of Data

1. Accuracy: Data should be accurate and free from errors.
2. Relevance: Data should be relevant to the purpose or question being addressed.
3. Completeness: Data should be comprehensive and include all necessary information.
4. Consistency: Data should be consistent in format and quality.

Data in the Digital Age

1. Big data: Large amounts of data are generated and analyzed to gain insights.
2. Data visualization: Data is visualized to communicate insights and trends.
3. Data-driven decision-making: Data is used to inform decisions in various industries.

QUANTITATIVE DATA

This refers to numerical data that can be measured, counted, or quantified. It is used to describe, analyze, and interpret phenomena using statistical methods.

Characteristics

1. Numerical: Quantitative data is expressed in numbers.
2. Measurable: Quantitative data can be measured using standardized units.
3. Comparable: Quantitative data can be compared using statistical methods.

Examples

1. Age: Measured in years.
2. Height: Measured in meters or inches.
3. Income: Measured in currency units.
4. Test scores: Measured as a numerical score.

Types of quantitative data:

1. **Discrete data:** Countable data, such as number of students in a class.
2. **Continuous data**: Measurable data, such as height or weight.

Advantages of using quantitative data analysis

1. Objectivity: Quantitative data analysis provides an objective view of the data, reducing personal biases.
2. Precision: Quantitative data analysis allows for precise measurements and calculations.
3. Quantitative data analysis can be generalized to larger populations, making it useful for informing policy decisions.
4. Reliability: Quantitative data analysis can be replicated, ensuring consistent results.
5. Comparability: Quantitative data analysis enables comparison of data across different groups, time periods, or contexts.
6. Statistical analysis: Quantitative data analysis allows for the use of statistical methods to identify patterns, trends, and correlations.
7. Efficient data processing: Quantitative data analysis can process large datasets efficiently, saving time and resources.

QUALITATIVE DATA

Qualitative data is non-numerical data that describes qualities, characteristics, or attributes of a phenomenon. It provides rich, contextual insights into people's experiences, perceptions, and behaviors.

Characteristics

1. Non-numerical: Data is not numerical or quantifiable.
2. Descriptive: Data provides detailed descriptions of phenomena.
3. Contextual: Data is often collected in a naturalistic or real-world setting.

Types of Qualitative Data

1. Textual data: Data collected through interviews, focus groups, or open-ended surveys.
2. Observational data: Data collected through observations, such as field notes or ethnographic research.
3. Visual data: Data collected through images, videos, or other visual materials.

Advantages

1. In-depth insights: Provides rich, detailed understanding of phenomena.
2. Contextual understanding: Allows researchers to understand the context in which phenomena occur.
3. Flexibility: Qualitative data can be collected through various methods.

Disadvantages

1. Subjective interpretation: Data analysis can be subjective and influenced by researcher bias.
2. Time-consuming: Data collection and analysis can be time-consuming.
3. Limited generalizability: Findings may not be generalizable to larger populations.

Applications

1. Social sciences: Study social phenomena, cultural practices, and human experiences.
2. Healthcare: Understand patient experiences, healthcare practices, and health outcomes.
3. Education: Study teaching methods, learning processes, and educational outcomes.

Analysis Methods

1. Thematic analysis: Identify patterns and themes in the data.
2. Content analysis: Analyze textual or visual data to identify meanings and patterns.
3. Narrative analysis: Examine stories and narratives to understand experiences and meanings.

NOTE: Qualitative data provides valuable insights into complex phenomena, making it a powerful tool for researchers.

DATA COLLECTION

Before any statistical work can be done at all, figures must be collected. The collection of figures is a very important aspect of statistics since any mistakes, errors or biases which arise in collection will inevitably be reflected in conclusions subsequently based on such figures.

NOTE: the conclusion can never be better than the original figures on which it is based. Therefore, unless the original figures are collected properly, any subsequent analysis

will be a waste of time and possibly even disastrous since it may mislead, with serious consequences.

Populations and samples

Before one starts collecting any data at all it is very important to know exactly what one is collecting data about. We call the group of people (or items) about which we want to obtain information the population. The population must be defined very carefully. Thus, if we wish to investigate the performance of girl students in the school X, then our population will be all the girls in school X.

Defining the population may however prove very tricky. For instance, in an inquiry relating to the number of house wives who go out to work, how should house wife be defined? If it is taken to mean a wife with no occupation, other than housekeeping, the survey would show that *no* house wives go out to work! If it means a wife who keeps a house, then the question arises whether a newlywed in a one room flat with her husband away on business nearly all the time is really a house wife. And should we include a wife who has a maid to do most of the work? remember different interpretation will lead to different results in the analysis.

Another problem sometimes associated with population is that its full extent may not be known e.g. The number of people in Uganda with unsuspected HIV/AIDS. Obviously, it is not easy to collect figures for such population.

Samples

A sample is a group of items taken from a population for examination. This is in practice to avoid the inconvenience that comes with having to deal with a whole population.

Sample frame:

A list or source of potential participants, units, or observations from which a sample is drawn.

Purpose

1. Identify population: Define the population of interest.
2. Select sample: Choose a representative sample from the population.
3. Ensure representativeness: Increase the likelihood that the sample accurately represents the population.

Sample Scenario

You are a researcher conducting a study on the impact of social media on mental health among university students. You want to survey a representative sample of students from a university with a total enrollment of 10,000 students.

Task

Define the population and create a sample frame for your study.

Population Definition

The population of interest is all currently enrolled students at the university, including undergraduate and graduate students.

Sample Frame

A suitable sample frame for this study would be the university's enrollment database, which includes a list of all currently enrolled students. This database would provide a comprehensive and up-to-date list of students, allowing for a representative sample to be drawn.

Potential Biases or Limitations

1. Non-response bias: Some students may not respond to the survey, potentially leading to biased results.
2. Database accuracy: The accuracy of the enrollment database may impact the representativeness of the sample.
3. Exclusion of certain groups: Students who are not included in the database (e.g., visiting students) may be excluded from the sample.

Steps to Minimize Bias

1. Use a random sampling method: Ensure that the sample is randomly selected from the sample frame.
2. Increase response rate: Use strategies to encourage students to respond to the survey.
3. Verify database accuracy: Check the accuracy of the enrollment database to ensure it is up-to-date and comprehensive.

Note: By using the university's enrollment database as the sample frame and taking steps to minimize bias, a representative sample of students can be obtained for the study.

Methods of date collection

1. Observations; Participant observation: Researcher participates in the setting. Non-participant observation: Researcher observes without participating.
2. Interviews; Structured interviews: Pre-determined questions. Unstructured interviews: Open-ended conversations. Semi-structured interviews: Combination of structured and unstructured questions.
3. Surveys; Questionnaires: Self-administered or interviewer-administered questionnaires. Online surveys: Web-based surveys.
4. Experiments; Controlled experiments: Manipulate variables in a controlled environment. Field experiments: Conducted in natural settings.
5. Focus Groups; Group discussions: Moderated discussions with a small group.
6. Document Analysis; Review of existing records: Analyze documents, reports, and records.
7. Case Studies; In-depth analysis: Detailed examination of a single case or a small number of cases.
8. Online Data Collection; Social media analysis: Analyze social media data. Web scraping: Collect data from websites.
9. Physiological Measurements; Biometric data: Collect physiological data (e.g., heart rate, blood pressure).

Choosing a Method

In choosing a data collection method, consider:

1. Research question: Align method with research objective.
2. Population: Choose method suitable for target population.
3. Resources: Consider time, budget, and equipment.

Note: Each method has strengths and limitations. Selecting the right method ensures valid and reliable data collection.

a) Direct observation method:

Direct observation involves systematically watching and recording behavior, events, or phenomena in their natural setting. This is the best method of collecting data as it reduces the chance of incorrect data being recorded.

Characteristics

1. Firsthand data collection: Researcher directly observes and records data.
2. Natural setting: Observations occur in real-world environments.
3. Systematic recording: Data is recorded using a structured protocol.
4. Rich, detailed data: Provides nuanced insights into behavior and context.
5. Contextual understanding: Allows researchers to understand the setting and its impact.
6. Reduced bias: Minimizes reliance on self-reported data.

Disadvantages

1. Observer bias: Researcher's presence or biases may influence observations.
2. Time-consuming: Requires significant time and resources.
3. Ethical considerations: Ensure participant consent and privacy.

Applications

1. Social sciences: Study human behavior, social interactions, and cultural practices.
2. Healthcare: Observe patient behavior, treatment outcomes, or healthcare practices.
3. Education: Study teaching methods, student behavior, or learning environments.

Best Practices

1. Clear observation protocol: Establish a structured plan for data collection.
2. Trained observers: Ensure observers are trained to minimize bias.
3. Accurate recording: Use systematic recording methods.

NB: Direct observation provides valuable insights into real-world phenomena, making it a powerful data collection method.

Sample scenario activity

You are conducting a study to observe and record the behavior of children in a playground setting. Your goal is to understand how children interact with each other and with the playground equipment.

Task

Design a direct observation study to collect data on children's behavior in the playground. Consider the following:

1. *What specific behaviors would you observe and record?*
2. *How would you ensure the accuracy and reliability of the data?*
3. *What steps would you take to minimize observer bias?*

Possible Observations

1. Social interactions: Play behavior, sharing, cooperation, conflict.
2. Physical activity: Use of playground equipment, physical play.
3. Emotional expressions: Displays of joy, frustration, or other emotions.

Your Response Should:

1. Describe the observation protocol.
2. Explain data recording methods.

3. Discuss strategies to minimize bias.

Note: this question requires applying direct observation methods to collect data in a specific context.

b) Interviewing

Interviews involve direct, face-to-face conversations between the researcher and participant to gather in-depth information.

Types of Interviews

1. Structured interviews: Pre-determined questions asked in a specific order.
2. Unstructured interviews: Open-ended conversations with flexible questioning.
3. Semi-structured interviews: Combination of structured and unstructured questions.

Advantages

1. In-depth insights: Rich, detailed data from participants' perspectives.
2. Flexibility: Allows for follow-up questions and probing.
3. Contextual understanding: Provides nuanced understanding of participants' experiences.

Disadvantages

1. The first disadvantage of interviewing is that inaccurate or false data may be given to the interviewer. The reason may be:
   - Forgetfulness
   - Misunderstanding the question
   - Or deliberate intent to mislead
2. Time-consuming; It Requires significant time and resources.
3. Interviewer bias: Researcher's presence or biases may influence responses.
4. Participant bias: Participants may not respond truthfully.

Applications

1. Qualitative research: Explore experiences, perceptions, and meanings.
2. Case studies: Gather detailed information about a specific case.
3. Needs assessment: Understand needs, preferences, and concerns.

Best Practices

1. Clear questioning: Use concise, relevant questions.
2. Active listening: Pay attention to participants' responses.
3. Note-taking: Accurately record responses.

Note: Interviews provide valuable insights into participants' thoughts, feelings, and experiences, making them a powerful data collection method

Sample scenario activity:

You are conducting a study to understand the experiences of entrepreneurs who have successfully launched their own businesses in a specific industry. You want to gather in-depth information about their challenges, strategies, and lessons learned.

Task

Design an interview-based study to collect data from these entrepreneurs. Consider the following:

1. *What type of interview would be most suitable (structured, unstructured, or semi-structured)?*
2. *What questions would you ask to gather relevant data?*
3. *How would you ensure the validity and reliability of the data collected?*

Possible Interview Questions

1. Background: Can you describe your background and experience in the industry?
2. Challenges: What were some of the biggest challenges you faced when launching your business?
3. Strategies: What strategies did you use to overcome these challenges?
4. Lessons learned: What are some key lessons you've learned throughout your entrepreneurial journey?

Your Response Should

1. Describe the interview approach.
2. Develop relevant interview questions.
3. Discuss strategies to ensure data validity and reliability.

Note: This question requires applying interview methods to collect in-depth data from participants.

c) Surveys;

Surveys involve collecting data through self-report measures, such as questionnaires or online forms, to gather information from a sample of participants.

Types of Surveys

1. Structured surveys: Pre-determined questions with fixed response options.
2. Unstructured surveys: Open-ended questions allowing participants to respond freely.
3. Mixed-mode surveys: Combination of structured and unstructured questions.

Advantages

1. Cost-effective: Surveys can reach a large sample size at a relatively low cost.
2. Efficient: Surveys can be administered quickly and easily.
3. Anonymity: Participants can respond anonymously, increasing honesty.

Disadvantages

1. Response bias: Participants may not respond accurately or truthfully.
2. Sampling bias: Sample may not be representative of the population.
3. Limited depth: Surveys may not capture complex or nuanced information.

Applications

1. Quantitative research: Gather numerical data on attitudes, behaviors, or demographics.
2. Market research: Understand customer preferences and opinions.
3. Social sciences: Study social phenomena, attitudes, and behaviors.

Best Practices

1. Clear questioning: Use concise, relevant questions.
2. Pilot testing: Test the survey with a small sample to ensure validity.
3. Random sampling: Ensure the sample is representative of the population.

Note: Surveys provide a valuable means of collecting data from a large sample size, making them a popular data collection method.

Designing a questionnaire:

If a questionnaire is to be used as a basis for interviewing, the following points should be observed:

- The questionnaire should be short and simple. It is far better that there are many short questions than there are a few long questions.

- The questions should be unambiguous
- The best questions are those that allow a preprinted answer to be ticked
- The questionnaire should be as short as possible
- The question should be neither irrelevant nor too personal
- Leading questions should not be asked e.g. Do you agree that all sensible people use XYZ soap?
- The questions should follow a logical sequence. Thus enable the respondent to understand its purpose and thus the quality of his/ her answers may be improved.

Sample scenario activity:

You are conducting a study to understand the factors that influence customer satisfaction with online shopping platforms. You want to collect data on customers' experiences, preferences, and pain points.

Task

Design a questionnaire to collect data from online shoppers. Consider the following:

1. *What demographic information would you collect (e.g., age, gender, income)?*
2. *What questions would you ask to understand customers' experiences with online shopping (e.g., frequency, product types, payment methods)?*
3. *How would you assess customer satisfaction (e.g., rating scales, open-ended questions)?*

Sample questionnaire

Sample Questionnaire on Customer Satisfaction with Online Shopping Platforms

(*Tick the box of your choice*)

Section 1: Demographic Information

1. What is your age?
   - Under 18 ☐
   - 18-24 ☐
   - 25-34 ☐
   - 35-44 ☐
   - 45-54 ☐
   - 55 or older ☐
2. What is your gender?
   - Male ☐
   - Female ☐
3. What is your annual income?
   - Less than $25,000 ☐
   - $25,000-$49,999 ☐
   - $50,000-$74,999 ☐
   - $75,000 or more ☐

Section 2: Online Shopping Habits

1. How often do you shop online?
   - Daily ☐
   - Weekly ☐
   - Monthly ☐
   - Rarely ☐
   - Never ☐
2. What types of products do you usually purchase online? (Select all that apply)
   - Electronics ☐
   - Clothing ☐
   - Home goods ☐
   - Books ☐
   - Other (please specify) ☐

Section 3: Customer Satisfaction

1. How satisfied are you with the overall online shopping experience on our platform?
   - Very satisfied ☐
   - Somewhat satisfied ☐
   - Neutral ☐
   - Somewhat dissatisfied ☐
   - Very dissatisfied ☐
2. How would you rate the usability of our website?
   - Very easy ☐
   - Somewhat easy ☐
   - Neutral ☐
   - Somewhat difficult ☐
   - Very difficult ☐

Section 4: Pain Points

1. What challenges have you faced while shopping online on our platform? (Select all that apply)
   - Technical issues ☐
   - Payment security concerns ☐
   - Difficulty finding products ☐
   - Slow delivery ☐
   - Other (please specify) ☐
2. Do you have any suggestions for improving our online shopping platform?

Section 5: Additional Comments

Is there anything else you'd like to share about your online shopping experience on our platform

d) Focused group discussion

A focused group discussion (FGD) is a qualitative research method where a small, diverse group of people discuss a specific topic or issue, guided by a moderator.