REVA UNIVERSITY
Bengaluru, India

Research Methodology – Unit 3

Course Code: M23DE0205 – Academic year 2024-2025, II Semester MCA (Even Semester)

School of Computer Science and Applications
Dr M Vinayaka Murthy
Professor

COURSE CONTENT

UNIT 1: Research and Types of research: Meaning of Research- Objectives of Research- Motivation in Research. Research methods vs Methodology. Types of research – Descriptive vs. Analytical, Applied vs. Fundamental, Quantitative vs. Qualitative, Conceptual vs. Empirical. Research Process. Criteria of good Research.

UNIT 2: Research Formulation : Defining and formulating the research problem. Selecting the problem - Necessity of defining the problem – Importance of literature review in defining a problem – Literature review – Primary and secondary sources – reviews, treatise, monographs- patents – web as a source – searching the web - Critical literature review – Identifying gap areas from literature review - Development of working hypothesis..

UNIT 3 – SYLLABUS CONTENT

UNIT 3: Data Collection and Analysis: Execution of the research – Observation and Collection of data - Methods of data collection – Modeling, Mathematical Models for research, Sampling Methods- Data Processing and Analysis strategies. Data Analysis with Statistical Packages – Hypothesis-testing, Generalization-and-Interpretation.

LECTURE -1, EXECUTION OF THE RESEARCH: OBSERVATION AND COLLECTION OF DATA

Agenda

- Unit 3 – Syllabus Content
- Execution of the Research
- Observation of Data
- Collection of Data

INTRODUCTION - EXECUTION OF RESEARCH

The execution phase of research involves actively gathering and analyzing data according to a predetermined plan.

This includes collecting data using methods like surveys, interviews, or observations, and then analyzing it to identify patterns and draw conclusions.

EXECUTION OF RESEARCH CONTD.

The execution phase is where the actual collection of data takes place, following the research plan and design.

Careful planning is essential for successful execution, ensuring that the process is smooth, systematic, and leads to valid conclusions.

1. Systematic Planning

Proper planning is crucial before starting research execution. Each step of the research process must be planned to avoid confusion and ensure smooth execution.

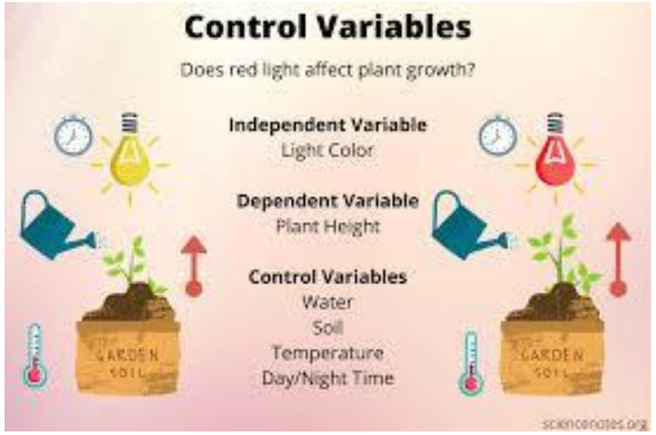
2. Observing Protocols

The researcher must strictly follow the designed research protocols to ensure the validity and reliability of the research. This includes adhering to sampling methods, data collection procedures, and timelines.

EXECUTION OF RESEARCH CONTD.

3. Control Over Variables

In empirical studies, it is important to maintain control over variables to minimize bias and ensure that the results are accurate. This involves ensuring that extraneous variables do not affect the outcome of the research.



4. Monitoring the Research Process

Continuous monitoring and evaluation of the research process are required. This ensures that any deviations from the plan are identified and corrected in a timely manner.

EXECUTION OF RESEARCH CONTD.

5. Execution as per the Plan

The researcher should execute the research exactly as planned without any deviation unless there is a justified reason. Changes should be documented and justified with solid reasoning.

6. Coordination and Cooperation

Execution often requires coordination among multiple team members or collaboration with external parties. Good communication and cooperation help ensure smooth execution.

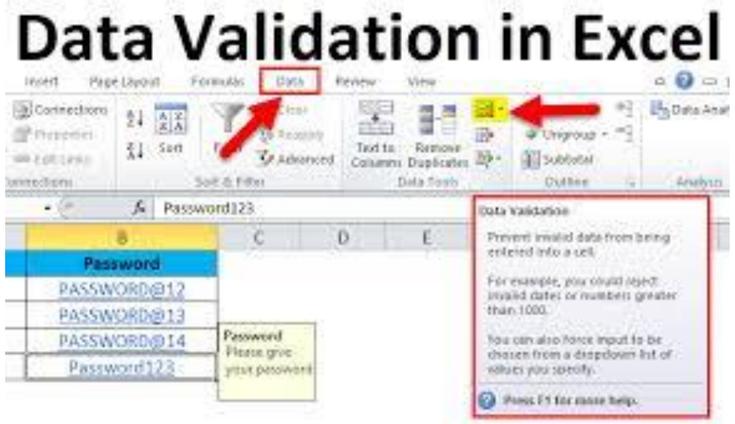
EXECUTION OF RESEARCH CONTD.

7. Ethical Considerations

Ethical principles, including informed consent, confidentiality, and the avoidance of harm to participants, must be adhered to during the execution of the research.

8. Data Validation

validating the collected data to ensure that it is reliable and accurate before moving to the analysis stage.



OBSERVATION OF DATA

Observation is a primary method of data collection in research :

Definition:

Observation involves systematically watching, listening, and recording behavior and events in a particular setting or environment.

Types of Observation:

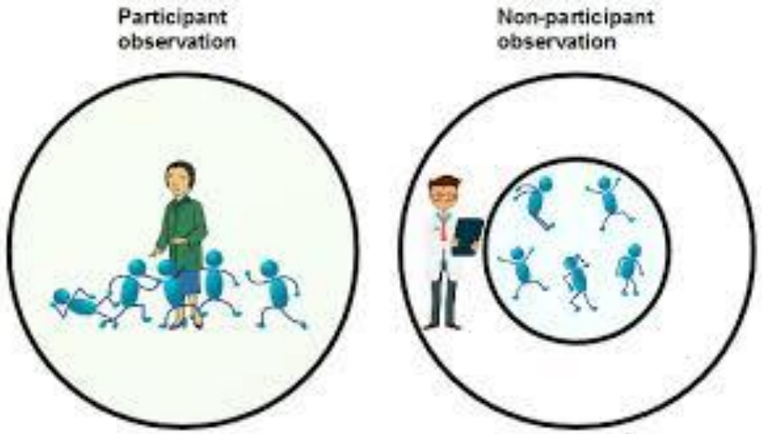
1. Structured Observation: Follows a pre-determined plan, where the researcher knows what is to be observed and recorded.

Unstructured Observation: More flexible and open-ended, with no strict guide on what to observe.

OBSERVATION OF DATA CONTD.

2. Participant Observation: The researcher becomes part of the group or situation being studied.

Non-Participant Observation: The researcher remains detached and simply observes the subject from a distance.

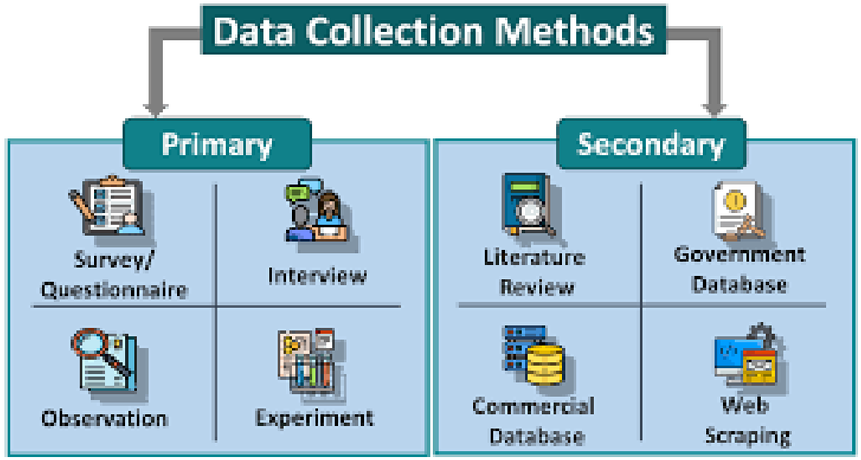


3. Controlled Observation: Takes place in a structured environment, such as a laboratory.

Uncontrolled Observation: Happens in a natural setting without the intervention of the researcher.

COLLECTION OF DATA

Data collection is a critical stage in any research.



Types of Data:

Primary Data: Data collected firsthand by the researcher through methods like surveys, interviews, observation, and experiments.

Secondary Data: (Subset of Primary Data) Data that has already been collected by others, such as in books, reports, websites, and government records.

COLLECTION OF DATA CONTD.

The 5 most common methods for data gathering are,

(a) Document reviews
(b) Interviews
(c) Focus groups
(d) Surveys
(e) Observation or testing.

While each has many possible variations, we will discuss their typical use here. Here are some basic principles to keep in mind when selecting methods.

SUMMARY

- Unit 3 – Syllabus Content
- Execution of the Research:
- Observation of Data
- Collection of Data

PROBLEM FOR PRACTICE

UNIT 3: Data Collection and Analysis: Execution of the research – Observation and Collection of data - Methods of data collection – Modeling, Mathematical Models for research, Sampling Methods- Data Processing and Analysis strategies. Data Analysis with Statistical Packages – Hypothesis-testing, Generalization-and-Interpretation.

LECTURE -2, METHODS OF DATA COLLECTION

- Unit 3 – Syllabus Content
- Execution of the Research
- Observation of Data
- Collection of Data

OBJECTIVE

- Methods of Data Collection

METHOD OF DATA COLLECTION

Observation Method:
Used to directly observe subjects without intervention.
As noted above, it can be participant or non-participant, controlled or uncontrolled, structured or unstructured.

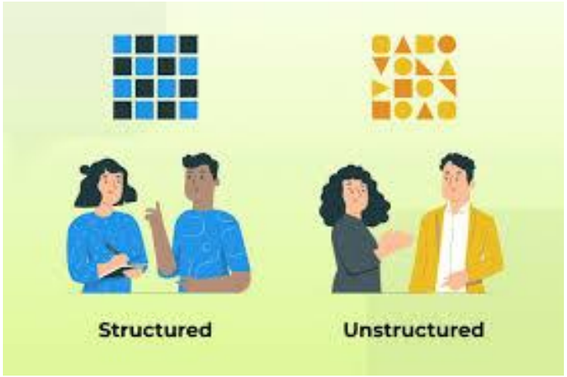
Interview Method:
Personal Interviews: Direct one-on-one interaction with respondents.

Telephone Interviews: Used for quick, short, and geographically dispersed data collection.

METHOD OF DATA COLLECTION

Structured Interviews: Pre-planned set of questions asked in a specific order.

Unstructured Interviews: Open-ended questions that allow for a flexible conversation.



METHOD OF DATA COLLECTION CONTD.

Questionnaire Method:
Involves distributing a list of questions to respondents, either physically or digitally.
Can be structured with close-ended questions or unstructured with open-ended questions.

Schedules:
Similar to questionnaires, but in this method, the researcher or an enumerator fills in the responses during face-to-face interaction.
(A plan for carrying out a process or procedure, giving lists of intended events and times)

METHOD OF DATA COLLECTION CONTD.

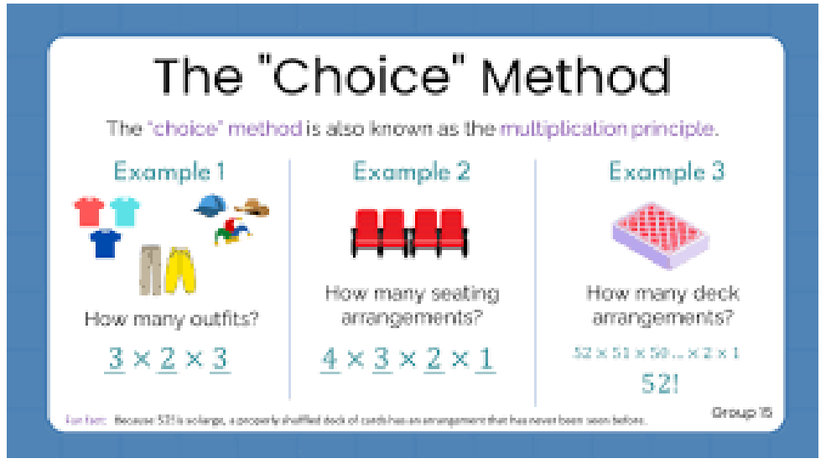
Survey Method : Commonly used in descriptive research, it involves collecting data from a large group of people.

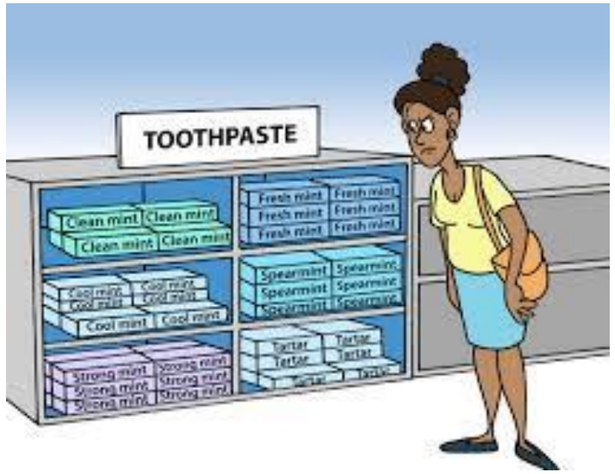
Experiments: Used in scientific research, experiments help establish cause-effect relationships through controlled manipulation of variables.

METHOD OF DATA COLLECTION CONTD.

Choice of Method:

The choice of method for data collection depends on factors like the nature of the research problem, availability of resources, time constraints, and the degree of accuracy required.





METHOD OF DATA COLLECTION CONTD.

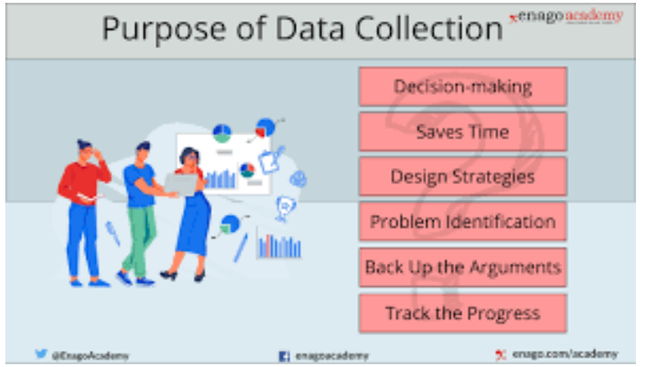
Precautions in Data Collection:

Ensure clarity in the data collection process.

Avoid leading questions that may bias respondents.

Establish rapport with participants to gain more honest and accurate responses.

Ensure data collection tools are reliable and valid.



SUMMARY

- Methods of Data Collection

PROBLEM FOR PRACTICE

UNIT 3: Data Collection and Analysis: Execution of the research – Observation and Collection of data - Methods of data collection – Modeling, Mathematical Models for research, Sampling Methods- Data Processing and Analysis strategies. Data Analysis with Statistical Packages – Hypothesis-testing, Generalization-and-Interpretation.

LECTURE -3, MODELING, MATHEMATICAL MODELS FOR RESEARCH

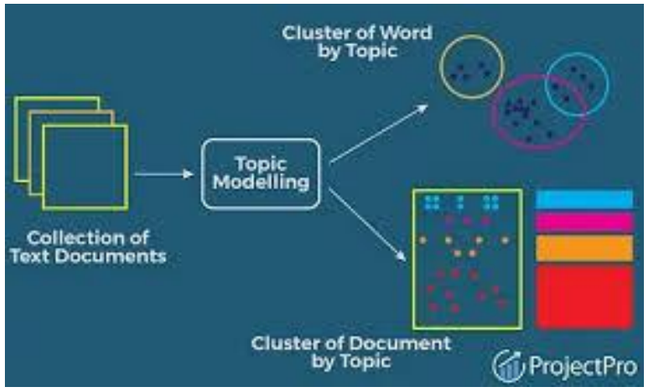
- Methods of Data Collection

OBJECTIVE

- Modeling for Research
- Mathematical Models for Research
- Applications of Mathematical Models
- Importance of Mathematical Modeling

MODELING FOR RESEARCH

Topic modeling is a natural language processing (NLP) technique used to discover the hidden thematic structure within a collection of documents, such as research articles.



It involves automatically identifying and categorizing these themes, which can be represented as distributions over a fixed vocabulary.

MODELING FOR RESEARCH CONTD.

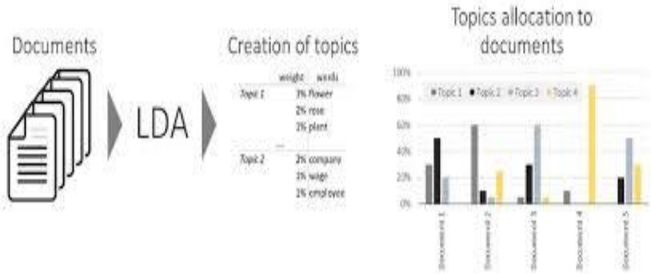
How Topic Modeling Works:

1. Data Preparation:

The first step involves gathering and preparing the research articles. This may include cleaning the text data, removing irrelevant words (stop words), and potentially stemming or lemmatizing the words.

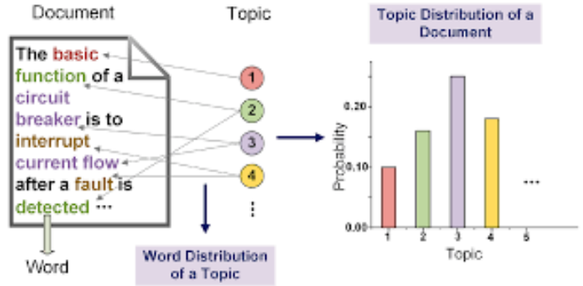
2. Model Selection:

Various topic modeling algorithms exist, with Latent Dirichlet Allocation (LDA) being a popular choice. Other options include Non-negative Matrix Factorization (NMF), BERTopic, and Top2Vec.



3. Topic Extraction:

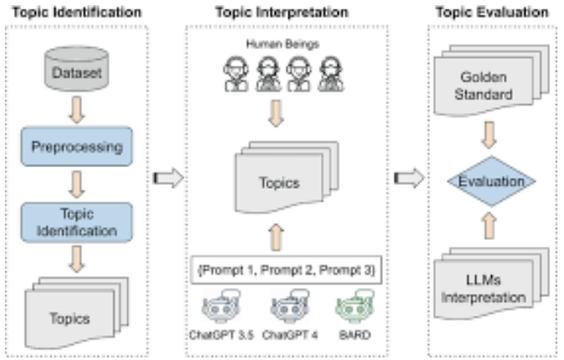
The chosen algorithm analyzes the text data, identifying and grouping words that frequently co-occur, thus revealing underlying topics.



MODELING FOR RESEARCH CONTD.

4. Topic Interpretation:

Researchers then analyze the identified topics, examining the keywords associated with each topic to understand its meaning and relevance.



5. Analysis and Visualization:

Topic modeling results can be analyzed further using various techniques, such as visualizing the relationships between topics using inter topic distance maps or exploring the evolution of topics over time.

Summarization:

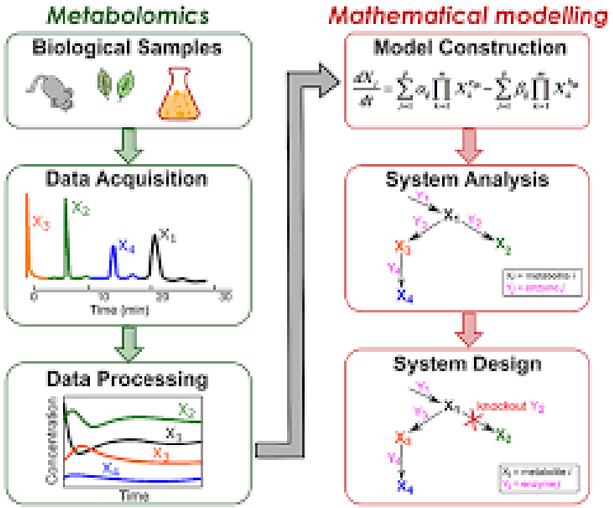
Topic modeling helps summarize large volumes of text data, making it easier to grasp the main themes of a research area

MATHEMATICAL MODELS FOR RESEARCH

Mathematical models are essential tools in research across various fields, providing a framework to represent real-world phenomena using mathematical language and concepts.

They allow researchers to analyze, predict, and understand complex systems, often expressed through equations, graphs, or other mathematical structures.

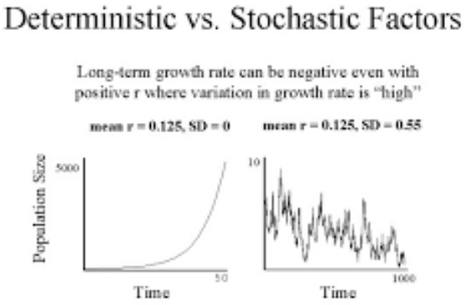
These models are crucial for gaining insights, making predictions, and optimizing processes.



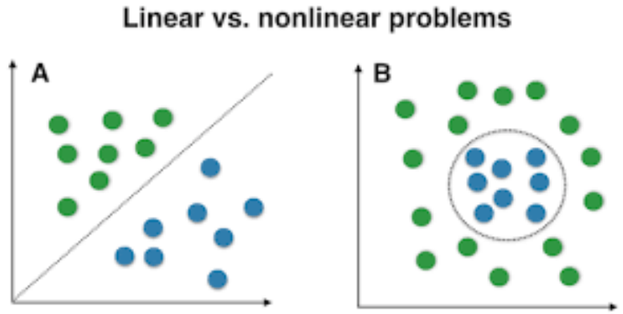
MATHEMATICAL MODELS FOR RESEARCH CONTD.

Types of Mathematical Models:

Deterministic vs. Stochastic:
Deterministic models assume a fixed outcome for a given set of inputs, while stochastic models incorporate randomness and probability.



Linear vs. Nonlinear:
Linear models have a straightforward, proportional relationship between variables, while nonlinear models exhibit more complex interactions.

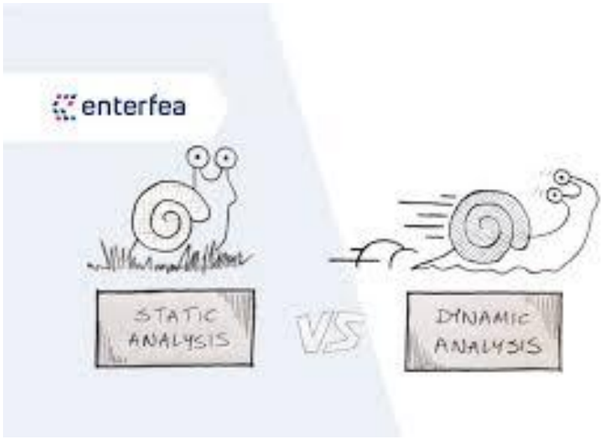


MATHEMATICAL MODELS FOR RESEARCH CONTD.

Discrete vs. Continuous:
Discrete models deal with countable, distinct values, while continuous models handle values that can change smoothly over a range.

Static vs. Dynamic:
Static models represent a system at a single point in time, whereas dynamic models describe how a system evolves over time.

Examples:
Common types include linear programming, dynamic programming, discrete optimization, and stochastic programming models.



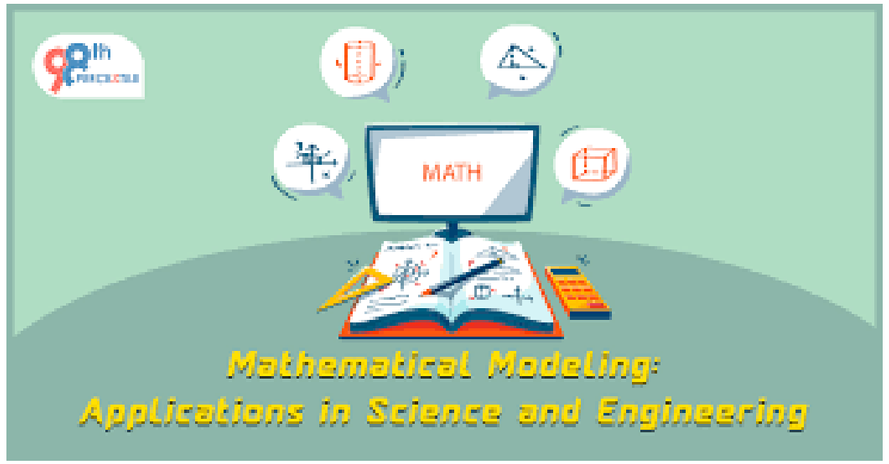
APPLICATIONS OF MATHEMATICAL MODELS

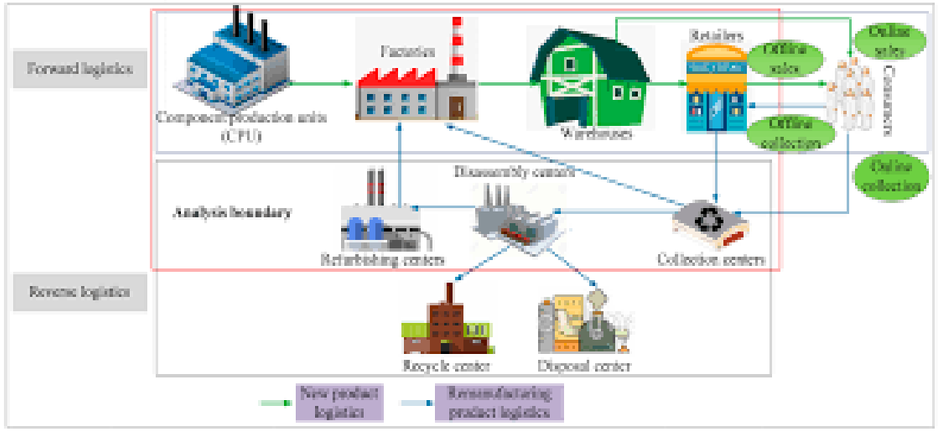
Scientific Research

Engineering

Economics and Finance

Social Sciences





IMPORTANCE OF MATHEMATICAL MODELING

Importance of Mathematical Modeling:

Prediction and Forecasting:

Models can be used to predict future outcomes based on current data and assumptions.

Understanding Complex Systems:

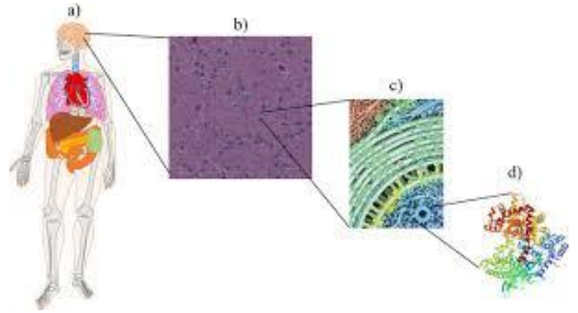
They help researchers break down complex systems into manageable components and understand their interactions.

Optimization and Decision Making:

Models can be used to identify optimal solutions for various problems, aiding in decision-making processes.

Hypothesis Testing:

Models can be used to test different hypotheses and refine our understanding of a system.



SUMMARY

- Modeling for Research
- Mathematical Models for Research
- Applications of Mathematical Models
- Importance of Mathematical Modeling

PROBLEMS FOR PRACTICE

UNIT 3: Data Collection and Analysis: Execution of the research – Observation and Collection of data - Methods of data collection – Modeling, Mathematical Models for research, Sampling Methods- Data Processing and Analysis strategies. Data Analysis with Statistical Packages – Hypothesis-testing, Generalization-and-Interpretation.

LECTURE -4, SAMPLING METHODS

- Modeling for Research
- Mathematical Models for Research
- Applications of Mathematical Models
- Importance of Mathematical Modeling

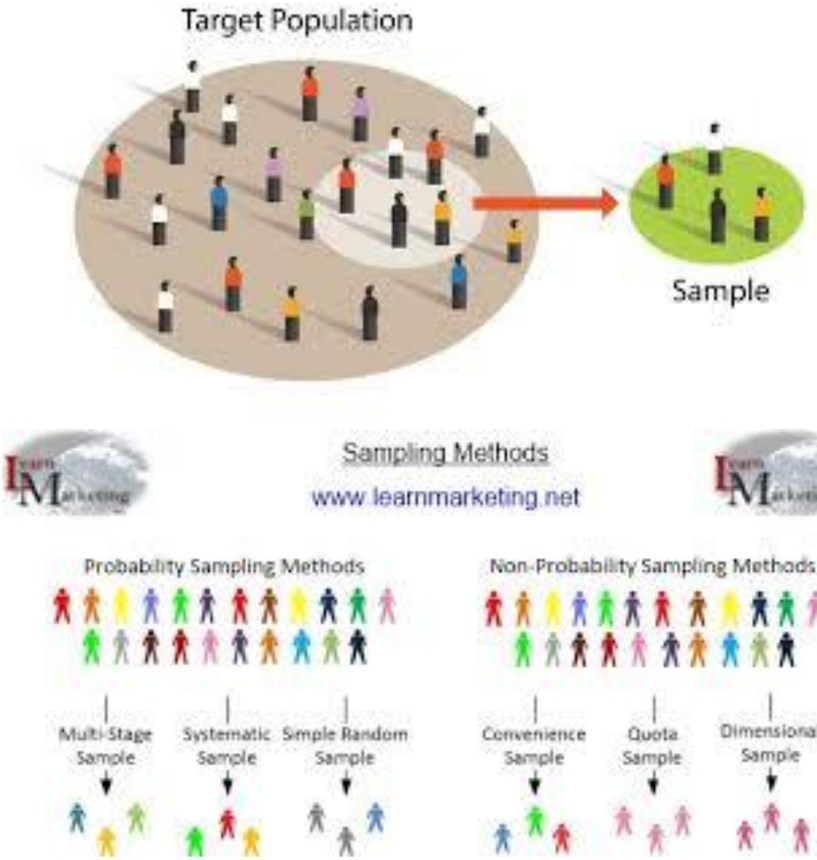
OBJECTIVE

- Sampling Methods
- Probability Sampling Methods
- Non-Probability Sampling Methods
- Choosing a Sampling Methods

SAMPLING METHODS

Sampling methods in research refer to the techniques used to select a subset of individuals or items from a larger population to study and draw conclusions about that population.

These methods are broadly categorized into probability sampling
(where each member has a known chance of selection) and
non-probability sampling
(where selection is based on other criteria).



PROBABILITY SAMPLING METHODS

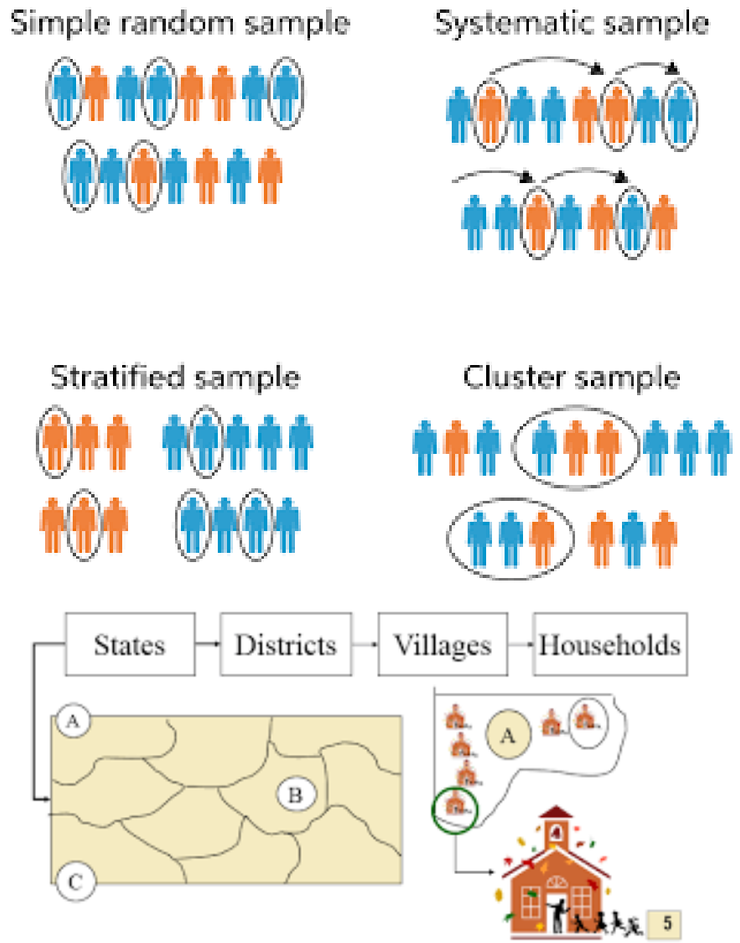
Simple Random Sampling: Each member of the population has an equal and random chance of being selected, similar to drawing names from a hat.

Systematic Sampling: Items are selected at a fixed interval after a random starting point.

Stratified Sampling: The population is divided into subgroups (strata), and then samples are randomly selected from each stratum.

Cluster Sampling: The population is divided into clusters, and then entire clusters are randomly selected.

Multistage Sampling: Combines different probability sampling techniques.



NON-PROBABILITY SAMPLING METHODS

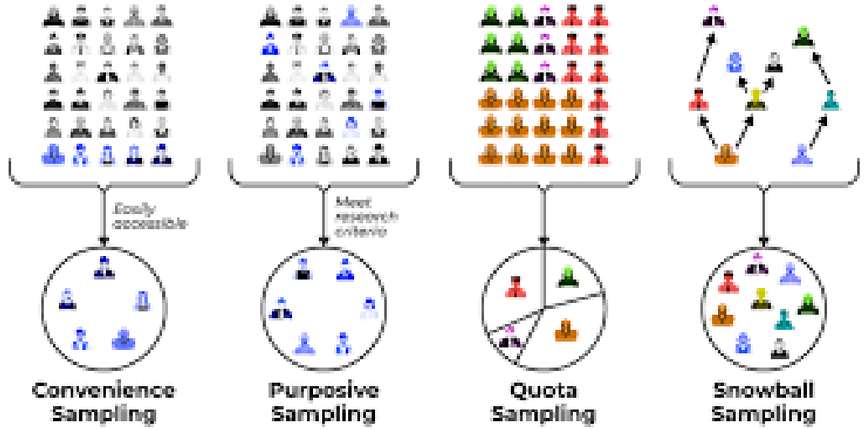
Convenience Sampling: Participants are selected based on ease of access and availability.

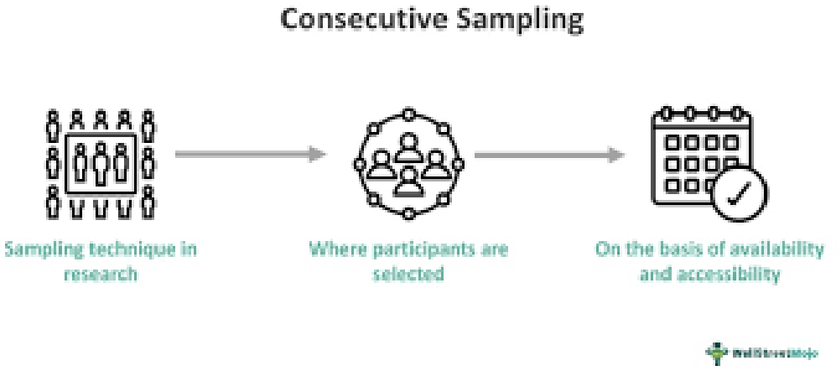
Purposive Sampling: Participants are chosen based on specific characteristics relevant to the research question.

Snowball Sampling: Participants are identified through referrals from existing participants.

Quota Sampling: Participants are selected to meet predetermined quotas based on specific characteristics.

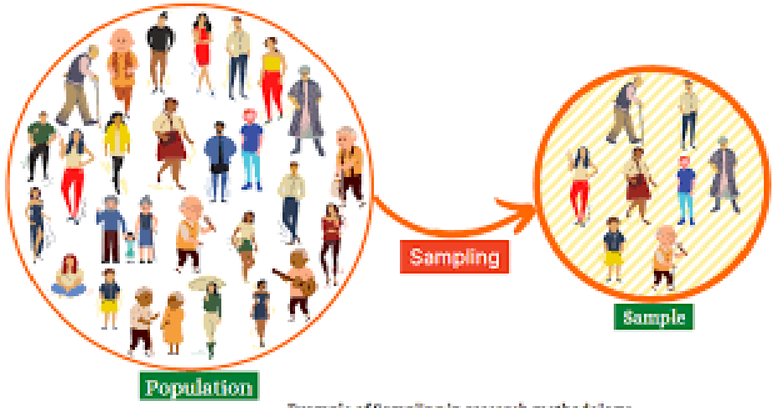
Consecutive Sampling: Participants are selected as they become available until the desired sample size is reached.





CHOOSING A SAMPLING METHODS

The selection of a sampling method depends on the research question, the nature of the population, and practical considerations like time, resources, and accessibility.



Example of Sampling in research methodology.

Probability sampling is generally preferred for making statistical inferences about the entire population, while non-probability sampling is often used when random sampling is not feasible or practical.

SUMMARY

- Sampling Methods
- Probability Sampling Methods
- Non-Probability Sampling Methods
- Choosing a Sampling Methods

PROBLEMS FOR PRACTICE

UNIT 3: Data Collection and Analysis: Execution of the research – Observation and Collection of data - Methods of data collection – Modeling, Mathematical Models for research, Sampling Methods- Data Processing and Analysis strategies. Data Analysis with Statistical Packages – Hypothesis-testing, Generalization-and-Interpretation.

LECTURE -5, DATA PROCESSING AND ANALYSIS STRATEGIES

- Sampling Methods
- Probability Sampling Methods
- Non-Probability Sampling Methods
- Choosing a Sampling Methods

OBJECTIVE

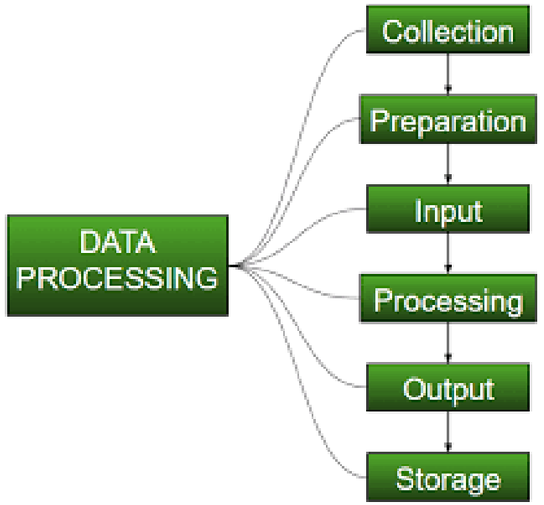
- Data Processing
- Key Stages of Data Processing
- Analysis strategies
- Key Data Analysis Strategies
- Specific Data Analysis Techniques

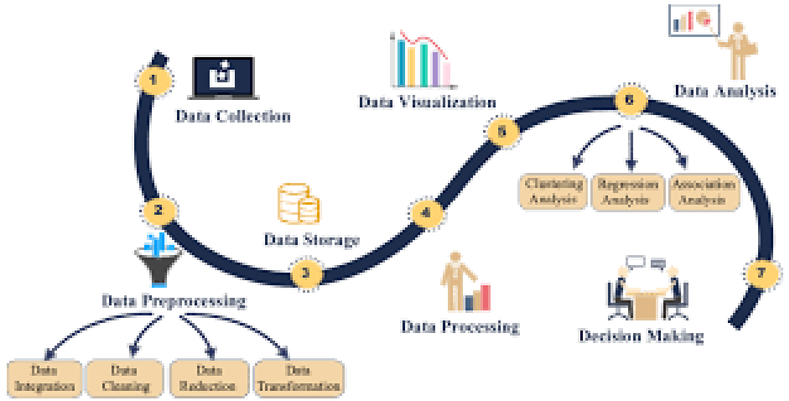
DATA PROCESSING

Data processing is the systematic manipulation of data to convert it from its raw, unstructured form into a more useful and informative state.

It involves collecting, cleaning, and analyzing data to extract insights and make it usable for various applications.

This process is crucial for informed decision-making, efficient operations, and driving innovation across diverse fields.





KEY STAGES OF DATA PROCESSING

Collection: Gathering raw data from various sources.

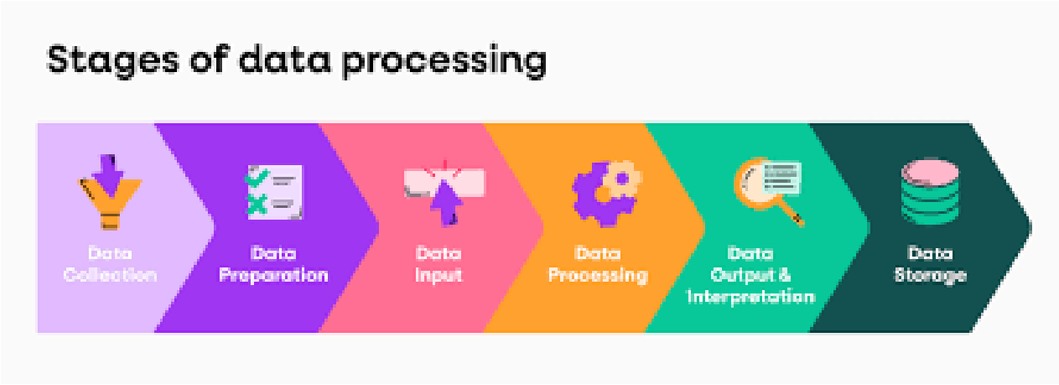
Preparation: Cleaning, filtering, and transforming the data to remove errors and inconsistencies.

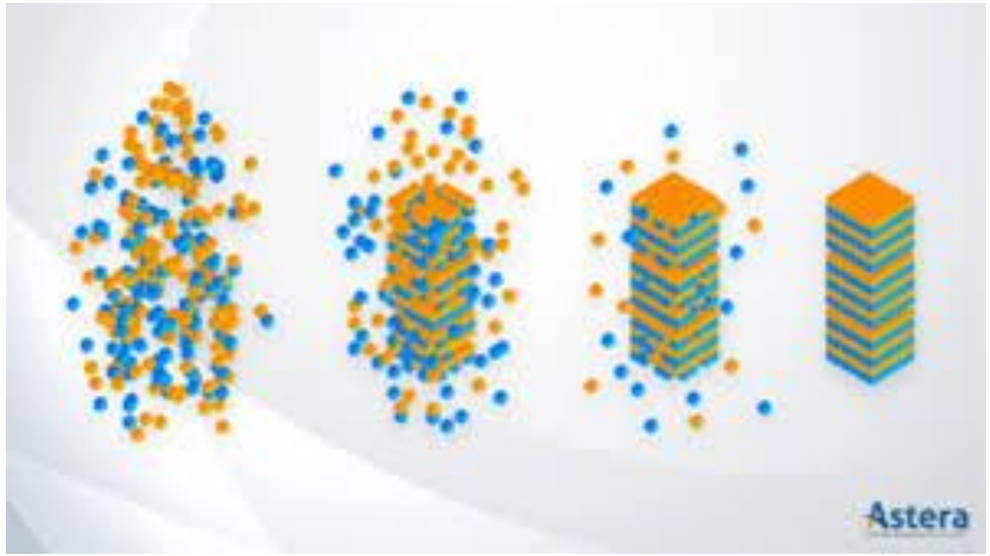
Input: Entering the prepared data into a system for processing.

Processing: Applying algorithms, calculations, or other operations to analyze and manipulate the data.

Output: Presenting the processed data in a desired format, such as reports, visualizations, or other usable forms.

Storage: Saving the processed data for future use and analysis.





IMPORTANCE OF DATA PROCESSING

Informed Decision-Making

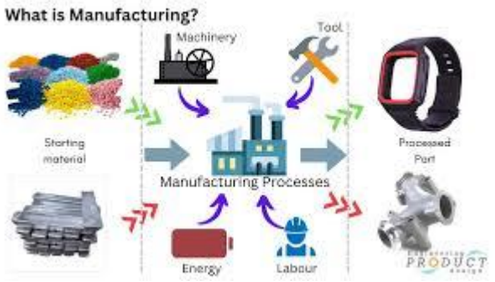
Operational Efficiency

Insights and Discovery

Innovation

Examples of Data Processing

- Business Intelligence
- Scientific Research
- Financial Transactions
- Healthcare
- Social Media
- Manufacturing

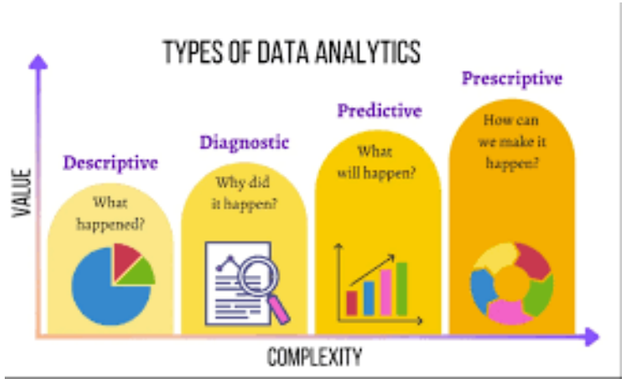


DATA ANALYSIS STRATEGIES

Data analysis strategies involve employing various techniques to examine, interpret, and draw conclusions from data.

These strategies can be broadly categorized into descriptive, diagnostic, predictive, and prescriptive analysis, with specific techniques like regression, clustering, and time series analysis used within these categories.

Effective data analysis requires a clear objective, appropriate data collection and preparation, suitable analytical methods, and clear communication of findings.



KEY DATA ANALYSIS STRATEGIES

Data Collection and Preparation

Gather relevant data from various sources and ensure it's clean, accurate, and in a usable format.

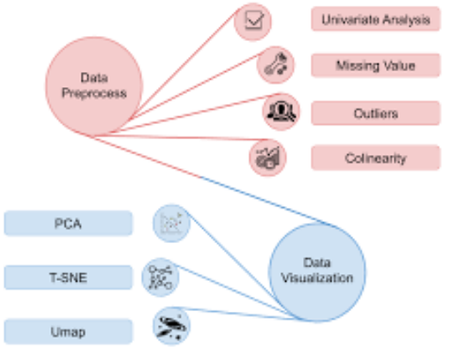
Choosing the Right Techniques

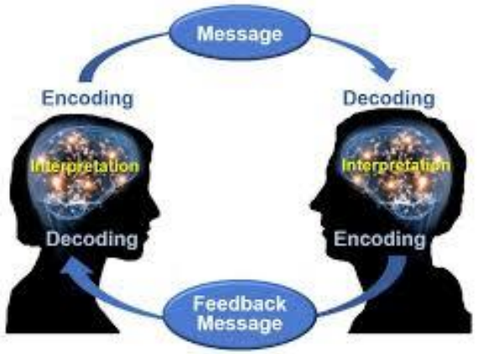
Select the most appropriate methods for the data and the objective, considering descriptive, diagnostic, predictive, or prescriptive analysis.

KEY DATA ANALYSIS STRATEGIES CONTD.

Data Exploration and Visualization

Use techniques like exploratory data analysis and visualization to understand patterns, trends, and relationships within the data.





Interpretation and Communication

Translate the results into meaningful insights and effectively communicate them to stakeholders.

SPECIFIC DATA ANALYSIS TECHNIQUES.

Descriptive Analysis
Diagnostic Analysis
Predictive Analysis
Prescriptive Analysis
Regression Analysis
Clustering
Time Series Analysis
Sentiment Analysis
Cohort Analysis
Factor Analysis

SUMMARY

- Data Processing
- Key Stages of Data Processing
- Analysis strategies
- Key Data Analysis Strategies
- Specific Data Analysis Techniques

PROBLEMS FOR PRACTICE

UNIT 3: Data Collection and Analysis: Execution of the research – Observation and Collection of data - Methods of data collection – Modeling, Mathematical Models for research, Sampling Methods- Data Processing and Analysis strategies. Data Analysis with Statistical Packages – Hypothesis-testing, Generalization-and-Interpretation.

LECTURE -6, DATA ANALYSIS WITH STATISTICAL PACKAGES & HYPOTHESIS-TESTING

- Data Processing
- Key Stages of Data Processing
- Analysis strategies
- Key Data Analysis Strategies
- Specific Data Analysis Techniques

OBJECTIVE

- Data Analysis with Statistical Packages
- Key Functions of Statistical Packages.
- Popular Statistical Packages
- Hypothesis-Testing

DATA ANALYSIS WITH STATISTICAL PACKAGES

Statistical packages are software tools used for analyzing numerical data, enabling researchers to perform various statistical calculations and visualizations.

These packages offer a wide range of functionalities, including data management, descriptive statistics, inferential statistics, regression analysis, and more.

Popular examples include SPSS, SAS, and R.

KEY FUNCTIONS OF STATISTICAL PACKAGES

Data Management:

Statistical packages allow users to import, clean, and transform data from various sources.

Descriptive Statistics:

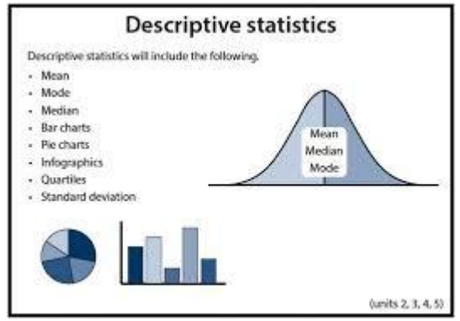
They compute summary statistics like mean, median, standard deviation, and create tables and charts to describe the data

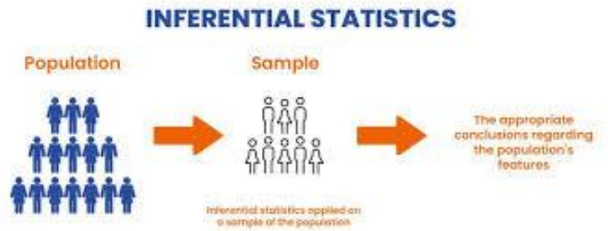
Inferential Statistics:

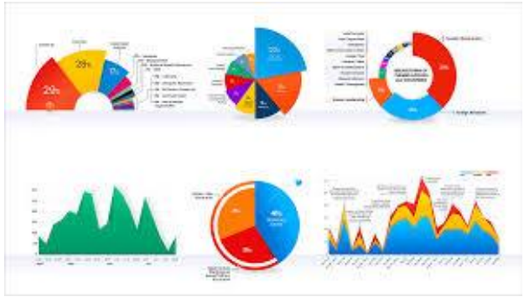
These packages perform hypothesis tests, regression analysis, and other statistical procedures to draw inferences about populations based on sample data.

Data Visualization:

Statistical packages offer tools for creating various plots and charts to visualize data patterns and relationships







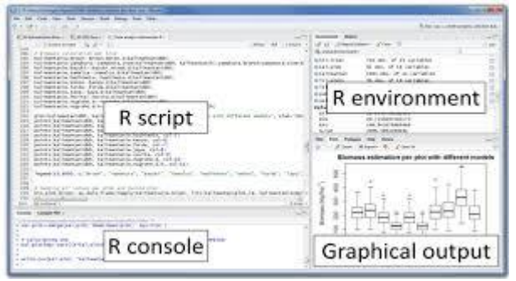
POPULAR STATISTICAL PACKAGES

SPSS (Statistical Package for the Social Sciences)
Widely used in social sciences, healthcare, and marketing research, SPSS is known for its user-friendly interface and comprehensive features.

SAS (Statistical Analysis System):
A powerful tool for handling large and complex datasets, SAS is popular in fields like clinical trials, healthcare, and biostatistics.

R
An open-source language and environment for statistical computing and graphics, R is favored for its flexibility and vast collection of statistical packages.

Excel
While not a dedicated statistical package, Excel includes modules for basic statistical analysis like descriptive statistics and regression



HYPOTHESIS-TESTING

The following is a procedure for testing a statistical hypothesis

Step 1: Set up the null hypothesis H_0.

Step 2: Set up the alternative hypothesis H_1.

Just negation of null hypothesis. i.e., $H_1 = \sim H_0$.

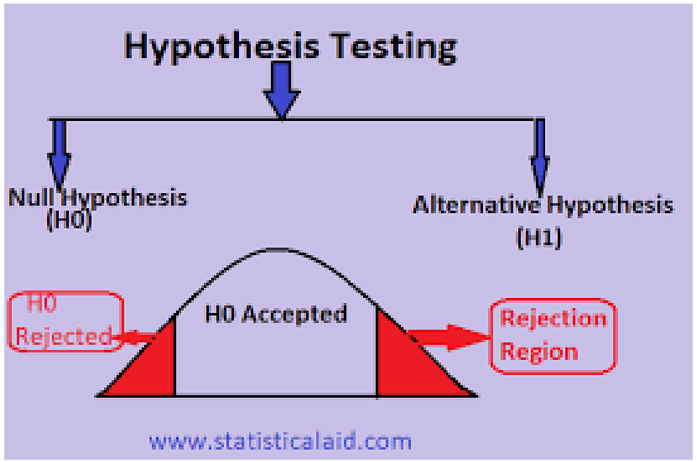
Step 3: Compute the test statistic | Z |

Here a) Single mean

b) Difference of two means

c) Single proportion

d) Equal of two proportions



Step 4: Choose the critical value 'k', appropriate level of significance α

in general α is 5%, k =1.96 or 1%, k = 2.58

Step 5: Compare |Z| and k

If $|Z| \leq k$, then H_0 is accepted

or If $|Z| > k$, then H_1 is accepted

Conclusion: which is accepted should mention (write)

SUMMARY

- Data Analysis with Statistical Packages
- Key Functions of Statistical Packages.
- Popular Statistical Packages
- Hypothesis-Testing

PROBLEMS FOR PRACTICE

UNIT 3: Data Collection and Analysis: Execution of the research – Observation and Collection of data - Methods of data collection – Modeling, Mathematical Models for research, Sampling Methods- Data Processing and Analysis strategies. Data Analysis with Statistical Packages – Hypothesis-testing, Generalization-and-Interpretation.

LECTURE -7, GENERALIZATION AND INTERPRETATION

- Data Analysis with Statistical Packages
- Key Functions of Statistical Packages.
- Popular Statistical Packages
- Hypothesis-Testing

OBJECTIVE

- Generalization
- Interpretation of Research Findings
- Practical consideration for Generalization and Interpretation

GENERALIZATION

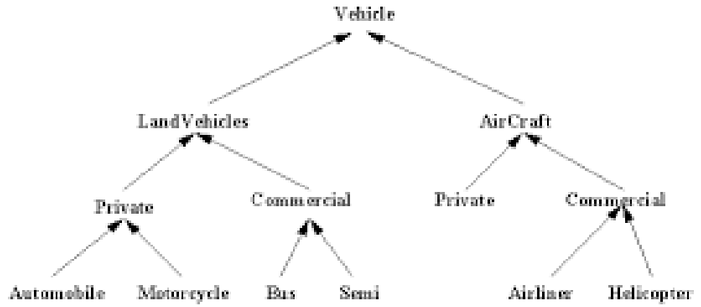
Generalization: The process of extending research findings from a sample to the broader population from which the sample was drawn.

Purpose:

1. **Applicability:** To determine whether results observed in a sample can be applied to a larger population.
2. **Relevance:** Ensures that findings are not unique to the sample but reflect broader trends or patterns.

Conditions for Generalization:

1. **Representativeness:** The sample must accurately represent the population.
2. **Sample Size:** Larger samples generally provide more reliable estimates.
3. **Sampling Method:** Random sampling methods enhance the ability to generalize findings.



INTERPRETATION OF RESEARCH FINDINGS

Interpretation: The process of explaining and understanding the meaning of research results in the context of the research questions and objectives.

Steps in Interpretation:

1. **Contextualization:** Relate findings to the theoretical framework and research questions.
2. **Pattern Recognition:** Identify patterns, trends, and relationships in the data.
3. **Comparative Analysis:** Compare findings with existing literature and theories.

Common Interpretative Errors:

1. **Overgeneralization:** Making broad claims based on limited or non-representative data.
2. **Confirmation Bias:** Interpreting data in a way that confirms pre-existing beliefs or hypotheses.

PRACTICAL CONSIDERATIONS FOR GENERALIZATION AND INTERPRETATION

1. **Validity and Reliability:**
    1. **Validity:** Ensure that the research accurately measures what it intends to measure (internal validity).
    2. **Reliability:** Ensure consistent results across different studies or measurements.
2. **Implications for Practice:**
    1. **Policy and Decision-Making:** Use findings to inform practical decisions and policy-making.
    2. **Future Research:** Identify areas for further investigation to address limitations and explore additional questions.
3. **Reporting and Transparency:**
    1. **Transparency:** Clearly report methods, limitations, and the extent of generalization.
    2. **Ethical Considerations:** Consider the ethical implications of generalizing and interpreting findings.

GENERALIZATION CONTD.

Factor Affecting Generalization

Sample Representativeness:

1. **Importance:** The sample should reflect the diversity of the population.
2. **Techniques:** Stratified sampling, random sampling to ensure representation.

Context and Setting:

1. **Contextual Factors:** The setting of the research (e.g., geographical location, time) can affect generalizability.
2. **External Validity:** The extent to which findings apply to different settings, populations, or times.

GENERALIZATION CONTD.

Homogeneity vs. Heterogeneity:

1. **Homogeneous Samples:** May limit the generalizability of findings to diverse populations.

2. **Heterogeneous Samples:** Improve the ability to generalize findings across various subgroups.

SUMMARY

- Generalization
- Interpretation of Research Findings
- Practical consideration for Generalization and Interpretation

PROBLEMS FOR PRACTICE

UNIT 3: Data Collection and Analysis: Execution of the research – Observation and Collection of data - Methods of data collection – Modeling, Mathematical Models for research, Sampling Methods- Data Processing and Analysis strategies. Data Analysis with Statistical Packages – Hypothesis-testing, Generalization-and-Interpretation.