



BHARATHIDASAN UNIVERSITY

Tiruchirappalli- 620024,

Tamil Nadu, India

Programme: M.A.,HUMAN RESOURCE MANAGEMENT

Course Title : Research Methodology

Course Code : 22HRM3CC10

Unit-V

Data Analysis and Report Writing

Dr. T. KUMUTHAVALLI

Associate Professor

Department of Lifelong Learning

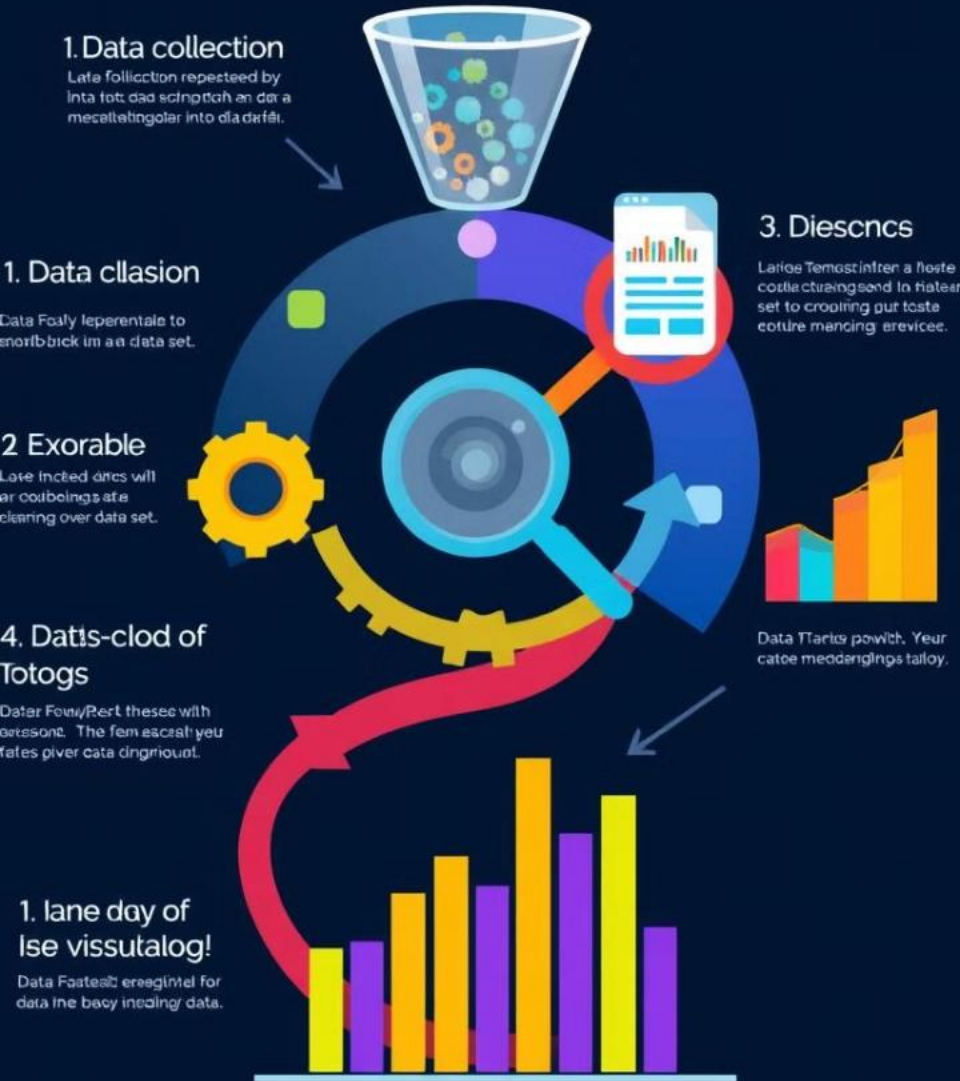
Introduction to Statistics

- Definition of Statistics
- Importance in Data Analysis
- Overview of Presentation Topics

Measure of Central Tendency

- Definition: Measures that represent the center of a data set.
- Importance in summarizing data
- **Mean**
- Definition of Mean
- Formula for Mean: $\mu = \frac{\sum X}{N}$
- Properties and Examples

SEARCHIT OF DATA ANALYSIS



Data Analysis and Report Writing

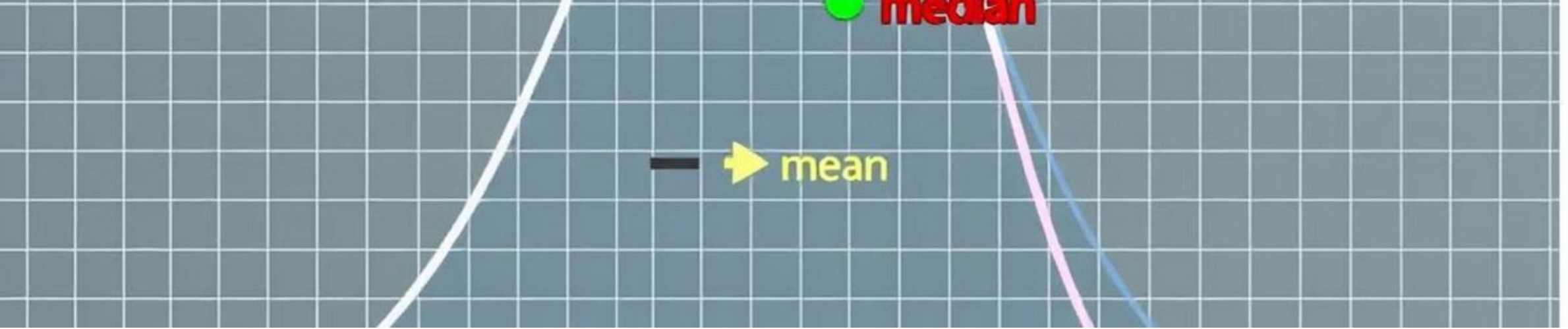
In today's data-driven world, understanding and interpreting data is an essential skill. From businesses to research institutions, the ability to analyze data effectively is crucial for making informed decisions and uncovering hidden insights. This guide will explore key concepts in data analysis and report writing, providing you with the tools and knowledge to confidently navigate the process of turning raw data into actionable intelligence.

Median

- Definition of Median
- How to Find the Median (Odd & Even Data Sets)
- Properties and Examples
- **Mode**
- Definition of Mode
- Example of Mode for Discrete & Continuous Data
- Properties and Limitations

Comparison of Central Tendency Measures

- When to Use Mean, Median, or Mode
- Advantages and Disadvantages
- **Measure of Dispersion**
- Definition of Dispersion: Variability in data
- Why Dispersion is Important



Measure of central tendency

1 Mean

The mean, or average, is calculated by summing all values in a dataset and dividing by the number of values. It's useful for understanding the typical value in a dataset.

2 Median

The median represents the middle value in a dataset when it's ordered from smallest to largest. It's less affected by outliers compared to the mean.

3 Mode

The mode is the value that appears most frequently in a dataset. It's useful for identifying the most common or popular values.

Range

- Definition and Calculation
- Example of Range
- **Variance**
- Definition of Variance
- Formula for Variance: $\sigma^2 = \frac{\sum (X - \mu)^2}{N}$
 $\sigma^2 = \frac{\sum (X - \mu)^2}{N}$
- How Variance Helps in Understanding Data

Standard Deviation

- Definition of Standard Deviation
- Formula for Standard Deviation: $\sigma = \sqrt{\frac{\sum (X - \mu)^2}{N}}$
- Interpretation of Standard Deviation

Orientation to Multi Variate Analysis

What is it?

Multi-variate analysis is a powerful statistical technique used to understand the relationships between multiple variables. It helps us identify patterns, trends, and dependencies between different factors.

Why use it?

This technique is crucial for businesses, researchers, and other professionals seeking to gain comprehensive insights from complex datasets. It can uncover hidden relationships, identify key drivers, and make better predictions.

Common techniques

Popular multivariate analysis techniques include Principal Component Analysis (PCA), Factor Analysis, Cluster Analysis, and Discriminant Analysis. Each technique has its own purpose and application depending on the data and research question.

Qualities of a Good Measure of Dispersion

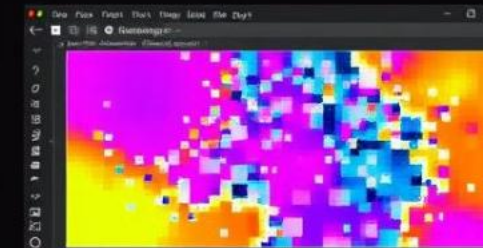
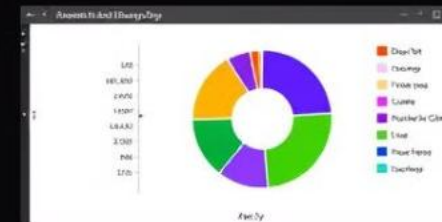
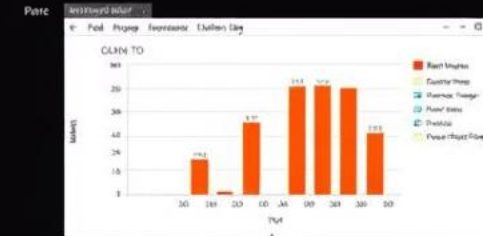
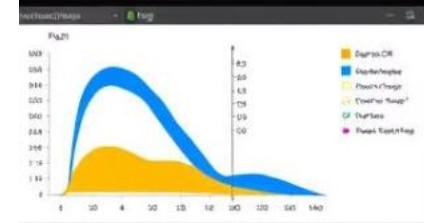
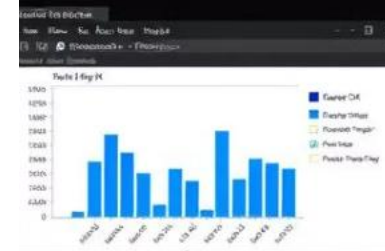
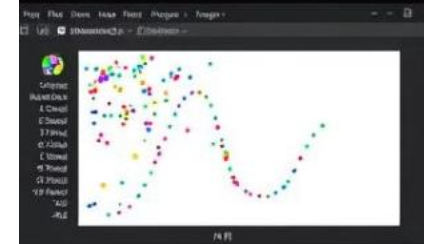
- Consistency
- Sensitivity to Extreme Values
- Representing Variability Effectively
- **Test of Significance**
- Definition: Determining if a result is statistically significant
- Why is it important in hypothesis testing?

Parametric vs Non-Parametric Tests

- **Parametric Tests:** Assumes data follows a specific distribution (e.g., t-test, ANOVA)
- **Non-Parametric Tests:** Does not assume a specific distribution (e.g., Chi-square, Mann-Whitney U test)

Different Software and its application

Software	Application
R	Open-source statistical programming language, language, widely used for data analysis, visualization, and machine learning.
Python	Versatile programming language with powerful powerful libraries for data analysis, including including NumPy, Pandas, and Scikit-learn.
SPSS	Statistical software package used for data analysis, hypothesis testing, and predictive modeling.
Tableau	Data visualization and business intelligence software, used for creating interactive dashboards and reports.



Examples of Parametric Tests

- **T-test:** Testing for differences between two groups
- **ANOVA:** Testing for differences among more than two groups

Correlation

- Definition: Measuring the strength and direction of the relationship between two variables
- **Pearson Correlation:** Formula $r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$
 $r = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum (X_i - \bar{X})^2 \sum (Y_i - \bar{Y})^2}}$
- Interpretation of r value

Types of Correlation

- Positive Correlation
- Negative Correlation
- Zero Correlation

Rank Correlation

- Definition of Rank Correlation
- **Spearman's Rank Correlation Coefficient**
- Formula and Example Calculation

Factor Analysis

- Definition: A technique to identify underlying factors or dimensions in data
- When and why is factor analysis used?

Purpose of Factor Analysis

- Data Reduction
- Identifying Patterns
- Exploring Complex Data Sets
- **Orientation to Multivariate Analysis**
- Overview of Multivariate Analysis
- Types: MANOVA, Principal Component Analysis (PCA), Cluster Analysis
- Applications in Various Fields

Summary

- Recap of Key Points: Measures of Central Tendency, Dispersion, Significance Testing, Correlation, Factor Analysis, Multivariate Analysis
- Overview of Software Applications in Statistics
- Importance of Report Writing