MY LEARNING JOURNEY IN STATISTICAL METHODS (SEMESTER 1)

24pbd022 | Mauli Patel | SXCA | BDA(2024-2026)

What I Learned – Statistical Methods

Semester 1 | M.Sc. Big Data Analytics

24pbd022 Mauli Patel

St. Xavier's College, Ahmedabad

OVERVIEW OF THE SUBJECT

- Data preprocessing, EDA, and visualization

- **Descriptive statistics:** Central tendency, dispersion

- Categorical data analysis using chi-square tests

- Correlation and regression

- Basics of time series analysis

- Hands-on practice with **R Programming**

24pbd022 | Mauli Patel | SXCA | BDA(2024-2026)

DATA COLLECTION & VISUALIZATION

- **Scales of measurement:** Nominal, Ordinal, Interval, Ratio

- Designing effective data collection formats

- Cleaning and treating missing values

- Visualizing data with:

  Histograms

  Box plots

  Frequency tables

DESCRIPTIVE STATISTICS

- **Measures of Central Tendency:**

  Mean, Median, Mode

- **Measures of Dispersion:**

  Range, Variance, Standard Deviation, IQR

- **Shape of Distribution:**

  Skewness and Kurtosis

- **Tools used:** summary(), sd(), boxplot() in R

CORRELATION & REGRESSION

- **Correlation:**

  Pearson & Spearman correlation

- **Regression:**

  Simple Linear Regression model

  Equation: $Y = a + bX + \varepsilon$

  Interpretation of slope, intercept, R^2

  Visualized using scatterplots in R

CONTINGENCY TABLES & CHI-SQUARE TEST

- Two-way tables for categorical data

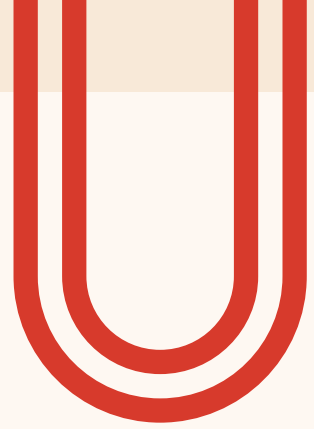- Chi-square test for independence

- Hypothesis testing:

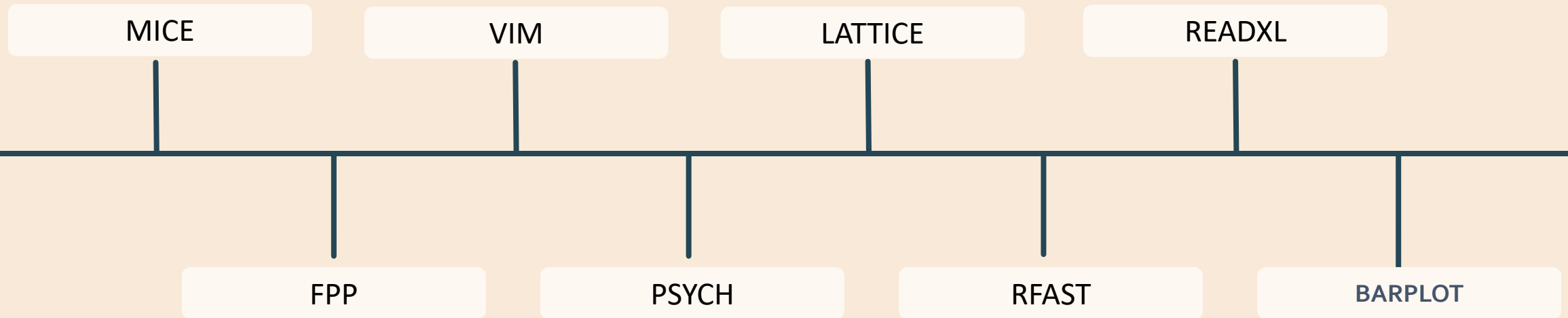 H_0: No association

  H_1: Association exists

- R command: chisq.test(table)

INTRODUCTION TO TIME SERIES

- Components: Trend, Seasonality, Cyclic, Random

- Decomposition of time series data

- Basic smoothing techniques

- Stationarity & autocorrelation concepts

- Worked with ts(), decompose(), acf() in R

HANDS-ON MINI-PROJECTS

- COVID-19 Time Series Analysis
- Retail Sales EDA
- Chi-square Analysis on Survey Data

KEY TAKEAWAYS

- Strong understanding of statistical foundations

- Developed skills in data exploration & interpretation

- Applied theoretical concepts using real-world data in R

- Built visualizations and performed hypothesis testing

Thank You

- Questions?

- Let's connect: LinkedIn | GitHub | Email