24MA116 **Statistical Data Analysis Using R**

Category: **Skill Development** **2L 0T 2P 3C**

***Pre-requisite*: Elementary Statistics and Computer Knowledge**

Course Description:

The power of data analysis with this course on statistical data analysis using R – language, designed for beginners or those with limited programming experience, this course will equip you to tackle real-world data and learn the fundamentals of R, leading statistical software, and master techniques for data manipulation, exploration, and visualization.

Course Aims and Objectives:

1. gain proficiency in R with the fundamental skills and knowledge to navigate and utilizeR for statistical data analysis.
2. master data manipulation will be adept at importing, cleaning, transforming, and organizing data sets within the R environment.
3. gain the ability to explore and understand your data through techniques like calculating summary statistics, visualizing distributions, and identifying patterns and trends.
4. equip you with the knowledge to perform various statistical tests using R, allowing youto draw meaningful conclusions from your data.
5. learn to create informative and effective charts and graphs using R's visualization capabilities, enabling you to communicate your findings clearly.

Course Outcomes:

At the end of the course, the student will be able to…

CO1: Memorize the basics in R in terms of construct, control statements and string functions (K1)

CO2: Use various operations and apply common function to manipulate and analyze data using basic R syntax.(K3)

CO3: Analyze data to design clear and informative visualizations that effectivelycommunicate patterns and trends to their audience. (K4)

CO4: Calculate and interpret key statistical measures like central tendency, dispersion, and skewness, gaining insights into your data's characteristics. (K4)

CO5: Use the regression model to make predictions about the value of one variable based on the other and understanding the limitations of such predictions. (K4)

Course Structure:

Unit – 1 Introduction to R (Theory 6Hrs + Lab 6 Hrs)

History and overview of R, getting started with R, R nuts and bolts, R editor, workspace, importing and exporting data, data types in R (numeric, logical, character, complex etc.), R objects: vector, matrix, array, list, data frame and factor. R operators: arithmetic, logical and relational.

Unit – 2 Control statements and Functions in R (Theory 6Hrs + Lab 6 Hrs)

Control Structures - if-else, for loops, nested for loops, while loops, repeat loops, **Functions** - functions in R, argument matching, argument, scoping rules of R and loop functions. **Data Frames:** managing data frames with the dplyr package

Unit – 3 Data Visualization in R (Theory 6Hrs + Lab 6 Hrs)

Charts: simple Bar chart, multiple Bar chart, subdivided Bar chart, Pie chart, Line chart, histogram, scatter plot, plotting density function and distribution function. Grammar of graphics (ggplot2 packages).

Unit – 4 Univariate Analysis of Statistical Measures in R (Theory 6Hrs + Lab 6 Hrs)

Measures of Central Tendency: arithmetic mean, median, mode. Skewness based on central values. M**easures of Dispersion**: range, quartile deviation(QD), mean deviation(MD), variance and standard deviation(SD). **Nature of the distribution:** Skewness and Kurtosis

Unit – 5: Linear Model building in R (Theory 6Hrs + Lab 6 Hrs)

Correlation: Introduction Bi-variate data, types of correlation, measure of correlation - Pearson correlation and Spearman's Rank correlation. **Linear Regression**- Introduction, lines of regression, properties of regression coefficients, model accuracy, MAE, MAPE and RSME.

Text Books:

1. Randall Schumacker (2017), Learning Statistics using R, Sage Publication, 1st edition
2. Jared P.Lander (2017), R for Everyone, Pearson Education, 2ND Edition

References:

1. Roger D. Peng, (2015). "R Programming for DataScience" Lean Publishing.
2. R for data science : Import, Tidy, Transform, Visualize, And Model Data (2017) Hadley Wickham, O'Reilly, 1st edition
3. The Book of R: A First Course in Programming and Statistics (2016), Tilman M. Davies, No Starch Press; 1st edition
4. R For Dummies (2015), Andrie de Vries , John Wiley & Sons; 2nd edition.

Web Resources for data sets.

1. https://www.kaggle.com/datasets
2. https://github.com/
3. https://www.r-project.org/
4. https://www.r-project.org/other-docs.html
5. https://www.rstudio.com/

***