

18MC212 STATISTICAL PROGRAMMING IN R

Course Description and Objectives:

This course makes a student familiar with the R interactive environment which includes the installation, packages, data types, data structures, user-defined functions and different data format conversions. In addition, it teaches to perform statistical manipulations and programming in a more efficient way when compared with traditional statistical analysis.

Course Outcomes:

The students will be able to:

- Hands on practice to install R studio and R studio interface.
- Use objects, classes, vectors, attributes etc... for a given task.
- Write functions including generic functions using various methods and loops.
- Select and perform manipulations on values as required for a specific statistical analysis.

Skills:

- Create a large corpus using lists and data frames.
- Perform Linear algebra operations on data for finding correlation and covariance.
- Develop functions for data manipulation algorithms.

Activities:

- Taking the large amounts of data and finding vector length, min, mode, max and average of that data.
- Applying list operations on big data for data interpretation.
- Converting list data into data frames and performing various operations on the data frames.
- Reading the data from various formats and converting them into required formats.

Syllabus

UNIT – 1

12 Hours

ESSENTIALS OF R: R program execution, Important R data structures - Vectors, Scalars, Character strings, Matrices, Lists, Data frames, Classes; Vectors – Adding and deleting vector elements, Length, Matrices, Arrays as vectors, Common vector operations, Vectorized operations, Filtering and NA&NULL values.

UNIT – 2

12 Hours

MATRIX AND LIST: Applying functions to matrix rows and columns, Adding and deleting matrix rows, More on vector/matrix distinction, Naming matrix rows and columns, Higher dimension arrays; Lists - Creating Lists, Generating list operations, Accessing list components and values, Applying functions to lists, Recursive list.

UNIT – 3

12 Hours

DATA FRAMES: Creating data frames, Other matrix operations, Merging data frames, Applying functions to data frames, Applying functions to data frames; Factors and labels -Common functions with factors, Working with tables and other functions.

UNIT – 4

12 Hours

R PROGRAM STRUCTURES: Control statements, Arithmetic Boolean operations, Default values for arguments, Return values, Functions are objects, Environment and scope issues, No pointers in R, Recursion.

UNIT – 5

12 Hours

DATA MANIPULATION: Doing math simulation, Functions for statistics, Sorting, Linear algebra operations on matrices and sets, Input output operations, Reading and writing files, String manipulation functions.

List of Experiments:

1. Installation of R software and R studio IDE.
2. R Notations, Basic Operations.
3. Programs which demonstrates R Objects: Data and Object types, Sub setting rules.
4. Atomic Vectors.
5. Experiment to Factor the data.
6. Perform basic matrix operations such as addition, subtraction, dot product, multiplication matrices.
7. Row binding and column binding operations using Arrays.
8. Converting matrices in to Data Frames.
9. Control Structures: Logical Operators, If Statements, If-else Statements.
10. Programs which demonstrates Loops: For Loop, While Loop.
11. Writing user-defined functions for basic operations such as factorial, prime number, and armstrong number.

Text Book:

Norman Matloff, “The Art of R Programming - A Tour of Statistical Software Design”, 1st Edition, no starch press, 2011.

Reference Books:

1. Paul Teetor, “R Cookbook: Proven Recipes for Data Analysis, Statistics, and Graphics”, O'Reilly publishers, 1st Edition, 2011.
2. Andrie de vries, “R for Dummies”, 2nd Edition, John Wiley & Sons, Inc., 2015.