

Name of Faculty	:	Faculty of Computer Science & Applications
Name of Program	:	Master of Computer Application in Data Science
Course Code	:	2MDS01
Course Title	:	R Programming for Data Science
Type of Course	:	Professional core
Year of Introduction	:	2023-24

Prerequisite	:	Knowledge of basic programming concept.
Course Objective	1	Aims to provide exposure in terms of Statistical Analysis, Hypothesis Testing, Regression and Correlation using R programming language.
	2	Aim to implement fundamental programming tasks
	3	How to use R for effective data analysis and visualization.
Course Outcomes	:	The student should be able to
	CO1	To understand the basic programming concepts of R programming language.
	CO2	To understand the data structures in R Statistical computing programming language.
	CO3	To understand the important packages and functions in R statistical computing programming language.
	CO4	To understand the importance of R in statistical analysis and customizing the analysis.
	CO5	To understand the impact of R in the current analytical organization over proprietary statistical software.

Teaching and Examination Scheme

Teaching Scheme (Contact Hours)			Credits	Examination Marks				Total Marks
				Theory Marks		Practical Marks		
L	T	P	C	SEE	CIA	SEE	CIA	
3	0	2	4	70	30	30	20	150

Legends: L-Lecture; T-Tutorial/Teacher Guided Theory Practice; P- Practical, C - Credit, SEE - Semester End Examination, CIA - Continuous Internal Assessment (It consists of Assignments/Seminars/Presentations /MCQ Tests, etc.)

Course Content

Unit No.	Topics	Teaching Hours	Weightage	Mapping with CO
1	Introduction To R Environment: History and development of R Statistical computing programming language, installing R and R studio, getting started with R, creating new working directory, changing existing working directory, understanding the different data types, installing the available packages, calling the installed packages, arithmetic operations,	07	20%	CO1

	variable definition in R, simple functions, vector definition and logical expressions, matrix calculation and manipulation using matrix data types, workspace management.			
2	Data Structures, Looping and Branching: Introduction to different data types, vectors, atomic vectors, types and tests, coercion, lists, list indexing, function applying on the lists, adding and deleting the elements of lists, attributes, name and factors, matrices and arrays, matrix indexing, filtering on matrix, generating a covariance matrix, applying function to row and column of the matrix, data frame - creating, coercion, combining data frames, special types in data frames, applying functions: <code>lapply()</code> and <code>sapply()</code> on data frames, control statements, loops, looping over non vector sets, arithmetic and Boolean operators and values, branching with <code>if</code> , looping with <code>for</code> , <code>if-else</code> control structure, looping with <code>while</code> , vector based programming.	08	20%	CO2
3	R - Object Oriented Programming: Introduction to object oriented concepts in R, basics of S3 classes - S3 Generic functions, OPP in linear model functions, writing S3 classes, using inheritance, introduction to S4 classes, writing S4 Classes, implementing a generic function on an S4 Classes, comparison of S3 and S4 classes, management of objects - listing objects, removing specific objects from the existing function and working directory, saving the collection of objects with <code>save()</code> function.	07	20%	CO3
4	R for Statistics: Descriptive statistics - mean (arithmetic, geometric and harmonic), median, mode for raw and grouped data, measure of dispersion - range, standard deviation, variance, coefficient of variation, testing of hypothesis - small sample test, large sample test - for comparing mean, proportion, variance, correlation and regression - significance of correlation and regression coefficients, chi-square test, non-parametric test, Analysis of Variance for one way variation and two	07	20%	CO3 CO4

	variation – with and without interaction.			
5	R with C, C++ and Python: Introduction to C and C++ programming concepts, writing C/C++ functions to be called from R, preliminaries of R to C and C++ programming languages, some mathematical programming example with R and C/C++, compiling and running the code, debugging R/C code, introduction to Python and its components, installing packages related with python in R, syntax of RPy packages.	07	20%	CO5

Suggested Distribution of Theory Marks Using Bloom's Taxonomy

Level	Remembrance	Understanding	Application	Analyse	Evaluate	Create
Weightage	20	30	30	20	0	0

NOTE: This specification table shall be treated as a general guideline for the students and the teachers. The actual distribution of marks in the question paper may vary slightly from above table.

Suggested List of Experiments/Tutorials

Sr. No.	Name of Experiment/Tutorial	Teaching Hours
1	Install and configure R, set working directory.	01
2	Install Packages and calling installed packages	01
3	R studio environment and functionalities of R studio.	01
4	Implement basic R operations (data input, missing values, importing data into R using different formats : xlsx, CSV, Text files).	02
5	Use R as a calculator.	01
6	Explore various functionalities of dataframes.	01
7	Create data set using data frames, list and tables.	02
8	Calculate the remainder after dividing 31079 into 170166719.	01
9	Calculate the interest earned after 5 years on an investment of \$2000,	01
10	Assuming an interest rate of 3% compounded annually.	01

Major Equipment/ Instruments and Software Required

Sr. No.	Name of Major Equipment/ Instruments and Software
1	R IDE; R Studio
2	I3/ I5 processor; 8GB RAM; 250GB HDD

Suggested Learning Websites

Sr. No.	Name of Website
1	https://www.udemy.com/course/r-programming/
2	https://www.analyticsvidhya.com/blog/2016/02/complete-tutorial-learn-data-science-scratch/

Reference Books

Sr. No.	Name of Reference Books
1	Introduction to Scientific Programming and Simulation using R - Owen Jones, Robert Maillardet and Andrew Robinson, CRC Press
2	Advanced R - Hadley Wickham, CRC Press.
3	The art of R programming - Norman Matloff, no starch Press, San Francisco.