

- Midterm Exam
- Programming Assignments
- Final Exam

Mapping of COs to Assessment Rubrics :

	Internal Exam	Assignment/ Quiz	Seminar	End Semester Examinations
CO-1	✓		✓	✓
CO-2	✓		✓	✓
CO-3	✓	✓		✓
CO-4	✓	✓		✓

8. R PROGRAMMING FOR DATA SCIENCE

Discipline	COMPUTER SCIENCE				
Course Code	UK2DSCCSC107				
Course Title	R PROGRAMMING FOR DATA SCIENCE				
Type of Course	DSC/VOCATIONAL				
Semester	II				
Academic Level	1				
Course Details	Credit	Lecture per week	Tutorial per week	Practical per week	Total Hours/Week
	4	3 hours	-	2 hours	5 hours

Pre-requisites	Knowledge in basic concepts of Statistics and Probability is desirable.
Course Summary	This course provides fundamental concepts of data analytics, R language & data visualization

Detailed Syllabus:

Module	Unit	Content	Hrs (L + P)
I	Introduction to R Programming		15
	1	Basic Interaction with R, Using R as a Calculator, functions, Control Structures, factors, data frames	
	2	Using R as a Calculator, functions, Control Structures, factors, data frames	
	3	Data pipelines, coding and naming conventions.	
	4	Data Manipulation: Reading Data, Manipulating and tidying Data with deplorer	
II	Visualizing Data		15
	5	Visualizing Data: Basic Graphics, The Grammar of Graphics and the ggplot2 Package	
	6	Figures with multiple plot	
	7	Working with Large Datasets	
	8	Expressions, Basic Data Types, Data Structures, Control Structures, Functions, Recursive Functions	
III	Advanced R Programming		15
	9	Working with Vectors and Vectorizing Functions	
	10	Advanced Functions, Functional Programming	
	11	Function Operations: Functions as Input and Output, Building an R Package	
	12	Creating an R Package, R oxygen	
IV	Data analysis using R		15

	13	Exploratory data analysis using R functions –sqrt, range, sort, minimum, maximum, median, average, standard deviation, skewness, variance	
	14	Correlation and covariance between Power tests- Bivariate Analysis-Paired sample t-test, t-test to compare means-one mean and two means	
	15	One factor ANOVA comparing means across several groups, 2-way ANOVA. Simple linear regression.	
V	Flexi Module (Not included for End semester Examination)		15
	16	Supervised Learning: Machine Learning, Supervised Learning, Regression versus Classification, Inference versus Prediction	
	17	Unsupervised Learning: Clustering, k-Means Clustering, Hierarchical Clustering	
	18	Object Oriented Programming: Immutable Objects and Polymorphic Functions, Data Structures, Classes	

TEXT BOOK

1. Thomas Mailund, Beginning Data Science in R, Data Analysis, Visualization, and Modelling for the Data Scientist, APress
2. Keen, K. J. . Graphics for statistics and data analysis with R. CRC Press, 2010.
3. Tony Fischetti, Data Analysis with R.
4. Joseph Schmuller, Statistical Analysis with R for dummies.

LAB EXERCISES

- 1) Find roots of a quadratic equation using the R program.
- 2) Calculate simple interest by creating function in R program
- 3) Copy spreadsheet data to clipboard and from clipboard transfer to table. Sort the data in ascending order; find average and standard deviation. [Hint dat <- ead.table("clipboard", header=TRUE)].
- 4) Read the student name and mark from a text file and store it in a table. Find maximum, minimum, average, median and standard deviation of marks. Display histogram and barplot.
- 5) Read the salesman name and sales amount from a CSV file. Display the average and standard deviation of sales. Visualize using plot and box plot of the sales amount. Inspect the boxplot and comment on the presence of outliers
- 6) The profit of a company on five products is given. Find the average profit of the company using the R function. Plot the data using plot, hist and pie graphs. Write the screen output to text files [Hint: use the function sink ()]
- 7) Create dataset of age of 50 students using rnorm() with n=50, mean=3.1 and sd=0.04 and conduct one sample t-test at significance level of 0.05, to check the validity of the statement “

the average age of students joining the play school is 3 years”. Display this diagram. Interpret the result.

- 8) A table contains expenditure and profit of a company. Conduct Pearson correlation test using R to find the correlation of expenditure on profit. Display data using line graph using ggplot()
- 9) A shopkeeper has data on the sales per day of one month. He introduced a new scheme in the next month. He wants to check whether there are any significant differences in average sales of the current month and the previous month. Display boxplot for both the data and interpret the result [Hint create suitable dataset using rnorm() and conduct 2 Sample t-test].
- 10) Crop yield and quantity of fertilizer used in an agricultural field is given. Conduct one-way ANOVA test to check whether the quantity of fertilizer used has any impact on the crop yield. Interpret the result.

Course Outcomes

No.	Upon completion of the course the graduate will be able to	Cognitive Level	PSO addressed
CO-1	Demonstrate the basic features used in R Programming	Ap	PSO-1, 3
CO-2	Illustrate the concepts of data visualization and its usage in various scenarios.	Ap	PSO-1, 3
CO-3	Examine different functions used in advanced R Programming	Ap	PSO-1, 3
CO-4	Experiment with different ways of Data Analysis techniques	Ap	PSO-1, 2, 3

R-Remember, U-Understand, Ap-Apply, An-Analyse, E-Evaluate, C-Create

Note: 1 or 2 COs/module

Name of the Course:R PROGRAMMING FOR DATA SCIENCE

Credits: 3:0:1 (Lecture:Tutorial:Practical)

CO No.	CO	PO/PSO	Cognitive Level	Knowledge Category	Lecture (L)/ Tutorial (T)	Practical (P)
CO-1	Demonstrate the basic features used in R Programming	PO-4, 6, 7 PSO-1, 3	Ap	F, C, P	L	P
CO-2	Illustrate the concepts of data visualization and its usage in various scenarios.	PO-4, 6, 7 PSO-1, 3	Ap	F, C, P	L	P

CO-3	Examine different functions used in advanced R Programming	PO-4, 6, 7 PSO-1, 3	Ap	F, C, P	L	P
CO-4	Experiment with different Data Analysis techniques.	PO-1, 2, 4, 6, 7 PSO-1, 2, 3	Ap	F, C, P,	L	P

F-Factual, C- Conceptual, P-Procedural, M-Metacognitive

Mapping of COs with PSOs and POs :

	PO1	PO 2	PO 3	PO 4	PO 5	PO6	PO 7	PO 8	PSO 1	PSO2	PSO 3	PSO 4
CO 1	-	-	-	1	-	1	1	-	2	-	1	-
CO 2	-	-	-	1	-	1	2	-	2	-	1	-
CO 3	-	-	-	1	-	2	2	-	2	-	1	-
CO 4	1	1	-	1	-	2	2	-	2	1	1	-

Correlation Levels:

Level	Correlation
-	Nil
1	Slightly / Low
2	Moderate / Medium
3	Substantial / High