SIR C.R.R.COLLEGE OF ENGINEERING, ELURU
DEPARTMENT OF INFORMATION TECHNOLOGY

COURSE HANDOUT
First Semester 2019 – 2020

Course Code : R1621051
Course Title : STATISTICS WITH R PROGRAMMING
Course Coordinator : K.PURNA PRAKASH
Team of instructors : SINGLE INSTRUCTOR

Course description:

In this course students will learn about the fundamentals of computers and understand the various steps in Program development. It provides the syntax and semantics of R Programming Language. This course makes how to write modular and readable R Programs. It also makes to write programs using structured programming approach in R to solve problems.

Scope and objectives:

After taking the course, students will be able to

- Use R for statistical programming, computation, graphics, and modeling,

- Write functions and use R in an efficient way,

- Fit some basic types of statistical models

- Use R in their own research,

- Be able to expand their knowledge of R on their own.

Prerequisite:

Basic programming knowledge, Mathematics, Probability concepts

Required Textbook:

1) The Art of R Programming, Norman Matloff, Cengage Learning

2) R for Everyone, Lander, Pearson

Reference Books

1) R Cookbook, PaulTeetor, Oreilly.

2) R in Action,Rob Kabacoff, Manning

Syllabus

II Year - I Semester   **L T P C**
                                                                              4 0 0 3

STATISTICS WITH R PROGRAMMING

UNIT-I:

Introduction, How to run R, R Sessions and Functions, Basic Math, Variables, Data Types, Vectors, Conclusion, Advanced Data Structures, Data Frames, Lists, Matrices, Arrays, Classes**.**

UNIT-II:

R Programming Structures, Control Statements, Loops, - Looping Over Non-vector Sets,- If-Else, Arithmetic and Boolean Operators and values, Default Values for Argument, Return Values, Deciding Whether to explicitly call return- Returning Complex Objects, Functions are Objective, No Pointers in R, Recursion, A Quick sort Implementation-Extended Extended Example: A Binary Search Tree.

UNIT-III:

Doing Math and Simulation in R, Math Function, Extended Example Calculating Probability-Cumulative Sums and Products-Minima and Maxima- Calculus, Functions Fir Statistical Distribution, Sorting, Linear Algebra Operation on Vectors and Matrices, Extended Example: Vector cross Product- Extended Example: Finding Stationary Distribution of Markov Chains, Set Operation, Input /output, Accessing the Keyboard and Monitor, Reading and writer Files,

UNIT-IV:

Graphics, Creating Graphs, The Workhorse of R Base Graphics, the plot() Function – Customizing Graphs, Saving Graphs to Files.

UNIT-V:

Probability Distributions, Normal Distribution- Binomial Distribution - Poisson Distributions Other Distribution, Basic Statistics, Correlation and Covariance, T-Tests,-ANOVA.

UNIT-VI:

Linear Models, Simple Linear Regression, -Multiple Regression Generalized Linear Models, Logistic Regression, - Poisson Regression- other Generalized Linear Models-Survival Analysis, Nonlinear Models, Splines- Decision- Random Forests

Course plan:

S.No.	Topics to be covered	Teaching aids used LCD/BB	Textbook/ Reference
I	**UNIT-I**		
1.1	Introduction , How to run R	BB	T1-1-6 Pages
1.2	R Sessions and Functions	BB	T1-7, 22 Page
1.3	Basic Math	BB/LCD	T1-32 Page
1.4	Variables	BB/LCD	T1-46 Page
1.5	Data Types	BB/LCD	T1-56 Page
1.6	Vectors	BB/LCD	T1-18 Page
1.7	Conclusion	BB	
1.8	Advanced Data Structures	BB/LCD	T1-26-31 Page
1.9	Data Frames	BB	T1-26-31 Page
1.10	Lists	BB	T1-26-31 Page
1.11	Matrices	BB	T1-26-31 Page
1.12	Arrays	BB	T1-26-31 Page
1.13	Classes	BB	TB/Internet
II	**UNIT-II**		
2.1	R Programming Structures	BB/LCD	T1-87-99 Page
2.2	Control Statements	BB/LCD	T1-87-99 Page
2.3	Loops	BB/LCD	T1-87-99 Page
2.4	Looping Over Non-vector Sets, If-Else	BB/LCD	T1-87-99 Page
2.5	Arithmetic and Boolean Operators and values	BB/LCD	T1-63-70 Page
2.6	Default Values for Argument, Return Values	BB/LCD	T1-119 Page
2.7	Deciding Whether to explicitly call return- Returning Complex Objects	BB/LCD	T1-119-122 Page
2.8	Functions are Objective	BB	T1-117-118 Page
2.9	No Pointers in R	BB	T1-117-118 Page
2.10	Recursion	BB/LCD	T1-123 page
2.11	A Quick sort Implementation-Extended Example: A Binary Search Tree.	BB/LCD	T1-124-125 Page
III	**UNIT-III**		
3.1	Doing Math and Simulation in R,	BB	T1-163 Page
3.2	Math Function	BB	T1-164 Page
3.3	Extended Example Calculating Probability-Cumulative Sums and Products-Minima and Maxima- Calculus	BB	T1-164 Page
3.4	Functions for Statistical Distribution,	BB/LCD	T1-169 Page

6.4	Logistic Regression	BB/LCD	T1-326 Page
6.5	Poisson Regression	BB/LCD	T1-333 Page
6.6	other Generalized Linear Models	BB/LCD	T1-334 Page
6.7	Survival Analysis	BB/LCD	T1-335 Page
6.8	Nonlinear Models	BB/LCD	T1-337 Page
6.9	Splines- Decision-	BB/LCD	T1-334 Page/Internet
6.10	Random Forests	BB/LCD	T1-334 Page/Internet

Evaluation Scheme:

Test	Marks
Internal Test-1	15
Internal Test-2	15
Internal Marks	Average of 80% of best internal and 20% of other internal test (15)
Assignment-1	5
Assignment-2	5
Assignment Marks	2 Assignments (10)
External exam	70
Total	100

Notices:

Answers for questions in internal and external examinations will be available in the Department Library of Information Technology. Any circulars related to course will be displayed in notice boards of Department of Information Technology.

Method of Evaluation:

Continuous Assessment Examination: Yes / <u>No</u>

Assignments: <u>Yes</u> / No

Questions in class room: <u>Yes</u> / No

Quiz as per University Norms: <u>Yes</u> / No

*<u>Others</u>***:**   Make the students to solve the problems on the board
(Please Specify)

Students are being taken to lab and hands-on experience provided to them and guiding to analyze data sets and preparing statistical and visualization reports.

On *completion of the course the student shall be able to:*

- Model different types of graphs and charts by importing, reviewing, manipulating data sets in R.
- Interpret and compare the outputs of statistical functions, and various graphs and charts created using R software.

List out any new topic(s) or any innovation you would like to introduce in teaching the subject in this semester:

- Working on real-world data sets for statistical analysis and visualization.

*Guidelines to study the subject***:**

1. Students are expected to know and understand the fundamentals of programming
2. Students are expected to know and understand the fundamentals of mathematics and probability concepts.
3. Students are advised to observe the real-world scenarios where large data are being generate and need to learn how to preprocess the data and apply statistical and graphics functions on the data sets.

Expected date of completion of the course:

Unit Number: 1   30th JUN 2019

Unit Number: 2   20th JUL 2019

Unit Number: 3   10th AUG 2019

Unit Number: 4   30th AUG 2019

Unit Number: 5   18th SEPT 2019

Unit Number: 6   31st SEPT 2019