



Department of Information Technology		LP: <b>IT18015</b> Rev. No: <b>00</b>
B.E/B.Tech/M.E/M.Tech : <b>Information Technology</b> Regulation: <b>2018</b>		Date: <b>19-06-2020</b>
PG Specialisation : NA		
Sub. Code / Sub. Name : <b>IT18015 – Statistical Analysis using R Programming</b>		
Unit : <b>I</b>		

**Unit Syllabus: INTRODUCTION****9**

Introduction to R-Basic Syntax-data Types-variables-Operators-Decision Making-Loops-Functions-Strings-Vectors-Lists-Matrices-Arrays-Factors-Data Frames-Packages-Data Reshaping.

**Objective:**

Students will learn about the basic syntax of various elements of R Programming, data structures, packages and reshaping data.

Session No *	Topics to be covered	Ref	Teaching Aids
1	Introduction to R Basic Syntax - Data Types- Numeric, Integer, Complex, Logical and Character, Data structures, variables	1- Ch 2; Pg.23-28 2- Ch 1; Pg.3-10	Online
2	Operators – Arithmetic, Relational, Logical, Assignment and Miscellaneous Operators	1- Ch 2; Pg.40-41	Online
3	Decision Making-if, if-else, if else ladder, switch, Loops – for, while, do while, repeat, break, continue	Internet	Online
4	Functions- Built-in functions, user defined function, recursion, Strings	1- Ch 2; Pg.41-42, 161-168	Online
5	Vectors-Creation, Accessing, Modifying, Recycling, Sorting, Comparing Vectors, Lists- Creation, Listing, Merging	1- Ch 2; Pg.28-39 2- Ch 2; Pg. 23-24	Online
6	Matrices – creation, Accessing, cbind, rbind, computations, Arrays – Creation, Accessing, Manipulation	1- Ch 5; Pg.98-99 2- Ch 2; Pg.24-27	Online
7	Factors – creating factors, Changing factor levels, order of levels, Categorical Variables, Data Frames – Creation, Extraction, Expanding, Displaying	1- Ch 5; Pg.100-107 2- Ch 2; Pg.27-31	Online
8	Packages- Installing and Loading packages	1- Ch 3; Pg.55-60 2- Ch 1; Pg.14-16	Online
9	Data Reshaping – Joining columns and rows, Merging data frames, Transposing, Melting and casting	1- Ch 6; Pg.147-158	Online

**Content beyond syllabus covered (if any):**

\* Session duration: 50 minutes



Sub. Code / Sub. Name: **IT18015 – Statistical Analysis using R Programming**

Unit : II

**Unit Syllabus : DATASET AND GRAPHICS**

**9**

Input and Output-Entering Data from the Keyboard-CSV file-Excel File-Binary File-XML file-JSON file-Web Data-Database-Graphics-Pie Charts-Bar Charts-Box Plots-Dot plots-Histograms-Line Graphs-Scatter plots-Kernel density plots-Writing plot to a file-Changing graphical parameters.

Objective:

Students will acquire Knowledge on working with various file formats and visually representing data using various plots.

Session No *	Topics to be covered	Ref	Teaching Aids
10	Input and Output-print, cat, Entering Data from the Keyboard- readline, scan, edit, reading tabular data files.	1- Ch 4; Pg.71-80 2- Ch 2; Pg. 33-35	Online
11	CSV file – reading from csv file, working with csv files, writing into csv file, Excel File- reading from excel file, working with excel file, writing into excel file	1- Ch 4; Pg.80-83 2- Ch 2; Pg. 35-37	Online
12	Binary File – Writing into binary file, Reading from binary file, XML file – Reading from XML file, Extract XML data, XML to list, XML to data frame	2- Ch 2; Pg. 37-38	Online
13	JSON file – Read JSON file, Convert to data frame, Writing JSON objects into JSON file	Internet	Online
14	Web data – fread, read.csv, <b>Web scraping</b> , Database – Connecting to database, Accessing data, Writing data to database	1- Ch 4; Pg.89-91 2- Ch 2; Pg. 39-41	Online
15	Graphics-Pie Charts – Creation, Adding Title, colors, Legends, 3D Pie, Bar Chart – Stacked Bar Chart, Group Bar Chart, Box Plot- Creation, Box plot with Notch	1- Ch 10; Pg.236-248	Online
16	Dot Plot- Groups, Histogram, Line Graph-Multiple lines	1- Ch 10; Pg.248-252	Online
17	Scatter plots, Kernel density plots, Writing plot to a file	1- Ch 10; Pg.223-235	Online
18	Changing graphical parameters – Set a graphical parameter using par, Text and Symbol size, Plotting Symbols, Lines, Colors, Fonts, Margins and Graph	1- Ch 10; Pg.264-265 2- Ch 3; Pg. 49-59	Online

**Content beyond syllabus covered (if any): Web Scraping**

\* Session duration: 50 mins



Sub. Code / Sub. Name: **IT18015 – Statistical Analysis using R Programming**

Unit : III

**Unit Syllabus : PROBABILITY**

**9**

Introduction-Sample Space -Events-Counting Methods-Conditional probability -Independent Events-Bayes Rule-Random Variables-Probability distribution-Discrete and continuous Distribution-Multivariate Distribution.

Objective:

Students will learn about the basics of probability, probability distributions and implementing concepts of probability in R Programming.

Session No *	Topics to be covered	Ref	Teaching Aids
19	Introduction-Deterministic and Random, Sample Space –Sampling from Urns, Sampling from Cards	1- Ch 8; Pg.177-179 3- Ch 4; Pg.65-70	Online
20	Events – Independent, Dependent, Mutually Exclusive, Functions for finding subset, Set Union, Intersection and Difference	3- Ch 4; Pg.70-75	Online
21	Counting Methods – Multiplication Principle, Ordered Samples, Unordered Samples	3- Ch 4; Pg.84-87	Online
22	Conditional Probability- Properties and Rules	3- Ch 4; Pg.87-95	Online
23	Independent Events – Independent, Repeated Experiments	3- Ch 4; Pg.95-98	Online
24	Bayes Rule – Bayes Theorem	3- Ch 4; Pg.98-101	Online
25	Random Variables – Discrete Random Variable, Continuous Random Variable, Marginal Distribution	1- Ch 8; Pg.183-185 3- Ch 4; Pg.102-104	Online
26	Probability Distribution – Discrete Distribution – Probability Mass Function, Mean, Variance, Standard Deviation	1- Ch 8; Pg.185-187 3- Ch 5; Pg.107-111	Online
27	Continuous Distribution – Probability Density Function, Multivariate Distribution – Joint and Marginal Probability Distribution	1- Ch 8; Pg.187-188 3- Ch 6; Pg.137-142 3- Ch 7; Pg.157-163	Online
<b>Content beyond syllabus covered (if any):</b>			

\* Session duration: 50 mins



Sub. Code / Sub. Name: **IT18015 – Statistical Analysis using R Programming**

Unit : IV

**Unit Syllabus : STATISTICS**

**9**

Regression-Linear-Multiple-Logistic-Poisson-Analysis of Covariance-Time Series Analysis-Nonlinear Least Square-Decision Tree-Random Forest-Survival Analysis-t-Test-Chi Square Test,ANOVA.

**Objective:**

Students will acquire knowledge on creating regression models, Analysing Time Series data and performing various hypothesis testing.

Session No *	Topics to be covered	Ref	Teaching Aids
28	Regression – Predictor Variable, Response variable, Linear – Regression Equation, Steps to Establish Regression in R, Visualize Regression Graphically	1- Ch 11; Pg.269-270 2- Ch 13; Pg. 314-315	Online
29	Multiple Regression – Create Model, Predict new Values, Logistic Regression – Create Regression Model, Predict	1- Ch 11; Pg.271-272 2- Ch 13; Pg. 315-320	Online
30	Poisson Regression – Create Regression Model, Predict, Analysis of Covariance	2- Ch 13; Pg. 324-329	Online
31	Time Series Analysis – Plotting Time Series, Representing Time Series Data, Subsetting a Time Series, Components for Time Series Analysis, ARIMA	1- Ch 14; Pg.355-394	Online
32	Nonlinear Least Square – Creating Model, Predict	Internet	Online
33	Decision Tree – Types, Steps in ID3 Algorithm, Attribute Selection Measures, Classification using ID3 Algorithm, Predict	Internet	Online
34	Random Forest – Bootstrapping, Out of Bag dataset, Creating Random Forest, Predict	Internet	Online
35	Survival Analysis-t-Test – One sample t-test, Two Sample t-test, Paired Sample t-test	1- Ch 9; Pg.203-206 3- Ch 10; Pg.224-228	Online
36	Chi Square Test, ANOVA – One way ANOVA, Two way ANOVA	1- Ch 11; Pg.302-303 2- Ch 9; Pg. 220-236	Online

**Content beyond syllabus covered (if any):**

\* Session duration: 50 mins



Sub. Code / Sub. Name: **IT18015 – Statistical Analysis using R Programming**

Unit : V

**Unit Syllabus : ADVANCED METHODS**

**9**

Advanced methods for missing data-Steps in dealing with missing data-Identifying missing values-Exploring missing value patterns-Understanding the sources and impact of missing data-rational Approaches for dealing with incomplete data-Lit wise deletion-Multiple Imputation-Advanced Graphics-Lattice Package-ggPlot2 Package-Interactive graphs.

Objective:

Students will learn about Identifying missing data, methods for handling missing data and advanced graphics packages.

Session No *	Topics to be covered	Ref	Teaching Aids
37	Advanced methods for missing data - Steps in dealing with missing data-classification system, Methods for handling incomplete data	2- Ch 15; Pg.352-355	Online
38	Identifying missing values – NA, NaN, Infinite, Complete. cases	2- Ch 15; Pg.355-356	Online
39	Exploring missing value patterns – Tabulating Missing Values, Visualizing missing data, Using correlation to explore missing values	2- Ch 15; Pg.356-361	Online
40	Understanding the sources and impact of missing data	2- Ch 15; Pg.362-363	Online
41	Rational Approaches for dealing with incomplete data-Lit wise deletion-Multiple Imputation, <b>Other approaches- Pair wise deletion, Simple Imputation</b>	2- Ch 15; Pg.363-371	Online
42	Advanced Graphics – Graphic system in R	2- Ch 16; Pg.373-375	Online
43	Lattice Package – Graph types and corresponding functions, Conditioning variables, Panel function, Grouping variables, Graphical parameters, Page arrangement	2- Ch 16; Pg.375-389	Online
44	ggplot2 package – qplot options	2- Ch 16; Pg.390-394	Online
45	Interactive Graphs – Identifying points, playwith, latticist, iplots, rggobi	2- Ch 16; Pg.394-399	Online

**Content beyond syllabus covered (if any): Pair wise deletion, Simple Imputation**

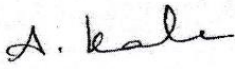


\* Session duration: 50 mins



Sub Code / Sub Name: IT18015 – Statistical Analysis using R Programming

**REFERENCES:**

1. Paul Teetor, "R Cookbook", O'reily, 2011.
2. Robert I Kabacoff, "R in Action: Data Analysis and Graphics with R", Manning Publications, 2015.
3. G. Jay Kerns, "Introduction to Probability and Statistics Using R", First Edition, 2010.

	Prepared by	Approved by
Signature		 
Name	Ms. A.KALA	Dr. V.VIDHYA
Designation	ASSISTANT PROFESSOR	HoD-INT
Date	19-06-2020	19-06-2020
Remarks *:		
Remarks *:		

\* If the same lesson plan is followed in the subsequent semester/year it should be mentioned and signed by the Faculty and the HOD

The same lesson plan is followed in the academic year 2021-22 (odd semester).

A. Kala  
22/7/21