JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA
KAKINADA – 533 003, Andhra Pradesh, India

DEPARTMENT OF INFORMATION TECHNOLOGY

II Year – II Semester		L	T	P	C
		2	0	2	3
STATISTICS WITH R					

COURSE 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.

COURSE OUTCOMES:
At the end of this course, students will be able to:
• List motivation for learning a programming language
• Access online resources for R and import new function packages into the R workspace
• Import, review, manipulate and summarize data-sets in R
• Explore data-sets to create testable hypotheses and identify appropriate statistical tests
• Perform appropriate statistical tests using R , Create and edit visualizations with R

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

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA
KAKINADA – 533 003, Andhra Pradesh, India

DEPARTMENT OF INFORMATION TECHNOLOGY

UNIT-V:
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,

TEXT BOOKS:
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

EXPERIMENTS:

1) Write a R program to take input from the user (name and age) and display the values. Also print the version of R installation.
2) Write a R program to create a sequence of numbers from 20 to 50 and find the mean of numbers from 20 to 60 and sum of numbers from 51 to 91.
3) Write a R program to create three vectors a,b,c with 3 integers. Combine the three vectors to become a 3×3 matrix where each column represents a vector. Print the content of the matrix.
4) Write a R program to find row and column index of maximum and minimum value in a given matrix.
5) Write a R program to combine three arrays so that the first row of the first array is followed by the first row of the second array and then first row of the third array.
6) Write a R program to create an array using four given columns, three given rows, and two given tables and display the content of the array.
7) Write a R program to create a data frame from four given vectors.
8) Write a R program to find Sum, Mean and Product of a Vector, ignore element like NA or NaN.
9) Write a R program to create a list containing a vector, a matrix and a list and remove the second element.
10) Write a R program to merge two given lists into one list.
11) Write a R program to create an ordered factor from data consisting of the names of months.
12) Plot the density and distribution functions for Normal approximation to the Binomial distribution.
13) Take any dataset, Visualize Tables, charts and plots. Compute visualising Measures of Central Tendency, Variation, and Shape. Box plots, Pareto diagrams. Aslo, find the mean median standard deviation and quantiles of a set of observations.
14) Take any dataset. Calculate the correlation between two variables. Draw the scatter plots. Use the scatter plot to investigate the relationship between two variables.

JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY KAKINADA
KAKINADA – 533 003, Andhra Pradesh, India

DEPARTMENT OF INFORMATION TECHNOLOGY

15) The sales of a company for each year are shown in the table below.

x (year)	2015	2016	2017	2018	2019
y (sales in lakhs)	12	19	29	37	45

a) Find the least square regression line $y = a\,x + b$.

b) Use the least squares regression line as a model to estimate the sales of the company in 2021.

16) Find the least square regression line for the following set of data {(-1 , 0),(0 , 2),(1 , 4),(2 , 5) Plot the given points and the regression line in the same rectangular system of axes