ST.JOSEPH'S COLLEGEFORWOMEN (A),VISAKHAPATNAM VII SEMESTER STATISTICS TIME: 4Hrs/Week ST 7205(3) DATA ANALYSIS USING 'R' Max. Marks:100 SYLLABUS

Objectives:

CO1: To expose the students to R- statistical programming language developed by scientists

that has open source libraries for statistics, machine learning, and data science. It is

free software environment for statistical computing and graphics supported by the R

Foundation for Statistical Computing.

CO2: The student will learn a large coherent and integrated collection of statistical tools

available in R. Further, the student will be familiar with graphical facilities for data analysis available in R.

CO3: R is widely used by statisticians, data scientists and major corporations like Google,

Airbnb, Facebook etc. for data analysis.

CO4: R is well-developed, simple and effective programming language for data handling

and storage facility.

CO5: R is data analysis software includes a wide variety of functions, such as data manipulation, statistical modeling, and graphics. One really big advantage of R, however, is its extensibility. Developers can easily write their own software and distribute it in the form of add-on packages.

Learing Outcomes:

After learning the course the student

LO1: Will be able to handle the data analysis using the R-statistical tools and can also

perform graphical presentation of the data.

LO2: Will be able to write own R-scripts for handling the data analysis in their own way.

LO3: Has better job potentiality to acquire opportunities as data analyst in software companies, clinical trial data analysis companies and etc.

COURSE:

UNIT - I

Essentials of R - language – Expressions and objects, assignments, creating vectors, vectorized arithmetic, creating matrices, operations on matrices, lists, data frame creation, indexing, sorting and conditional selection with examples.

UNIT-II

Programming using conditional statements and loops (flow control: if (), for () and while () loop), data editor, reading data from text files.

UNIT-III

Learn how to load data, plot a graph: bar-plot, pie-chart, and box plot, stem-leaf, histograms (equal class intervals and unequal class intervals), frequency polygon, ogives with graphical summaries of data.

UNIT-IV

Random number generation and sampling procedures. Application problems based on fitting of suitable distribution, Q-Q plot, Multiple Regression.

UNIT-V

Basics of statistical inference in order to understand hypothesis testing, compute p-values and confidence intervals. Simple analysis and manage statistical analysis projects, import data. **References:**

- 1. Braun, W.J., and Murdoch, D.J.(2007). A First Course in Statistical Programming with R. Cambridge University Press. New York.
- 2. Dalgaard, P.(2008). Introductory Statistics with R. 2 nd Ed Springer.
- 3. Gardener, M.(2012): Beginning R: The Statistical Programming Language, Wiley Publications