Name of the Department: Department of Statistics

Course name: Computation of Statistics Using R-Software

Course Coordinator name & Contact number: Mr. P. S. Chougule (9822680411, 7083633933)

Duration: 20 days

Course Fee: Rs. 500/-

Eligibility: It's suitable for undergraduates, graduates and researchers from any field that uses statistical computing.

Minimum intake Capacity: 20

Objectives of the Course:

This course aims to provide a practical introduction to the R programming language.

- ➢ In this course you will learn how to program in R and how to use R for effective data analysis.
- ➢ This course covers practical issues in statistical computing which includes programming in R, reading data into R, accessing R packages, writing R functions,
- ➢ R code for in statistical data analysis will provide working examples and running summary statistics and visualizations and simulations form various distributions

Learning Outcomes : By the end of the course students you shall be confident and equipped with all the knowledge required to perform analytical activities in R. Specifically,

- ➢ A new way of thinking
- ➢ Download and Install R
- ➢ A new language for speaking and reading (vectors, data frames, functions, objects, etc.

- ➢ A new syntax for writing, e.g. c(), print(), cat(), sort(), require(), subset() for data analysis and presentation .
- ➢ Understand the concepts of objects and assignment
- ➢ Construct tables and figures
- ➢ Load a script file, run lines from it, edit and save the script file.
- ➢ Load a Workspace containing an R dataframe, edit the dataset, and save the Workspace.
- ➢ Variable types in R
- ➢ Understand the different data types in R
- ➢ Understand the different data structures in R
- ➢ Understanding Assignment operations
- ➢ create loops to solve different types of problems
- ➢ Use R for descriptive statistics
- ➢ Simulation, code profiling
- ➢ Perform appropriate statistical tests using R
- ➢ Expand R by installing R packages
- ➢ Create their own customized function
- ➢ Online R Learning resources

Syllabus:

1. Fundamentals of R:

   1.1 Introduction to R, features of R, Installation of R, starting and ending R session getting help in R, R commands and case sensitivity.

   1.2 Data types: Logical, numeric and complex

   1.3 Vectors and vector arithmetic a) Creation of vectors using function C, assign, seq, rep b) Arithmetic operation on vectors using operators+,-,*, /, ^ c) Numerical functions: log10,log,sort,max,min,unique,range,length,var,prod,sum,summary,fivenum etc. d) Accesing vectors. e) Alternative ways to create vector by scan function.

   1.4 Data frame: creation using data frame, subset and transform commands

   1.5 Resident data sets: Accession and summary

   1.6 Graphics using R: a) High level plotting functions b) Low level plotting functions c) Interactive graphic functions

1.7 Using R as calculator The following Statistical Methods using "R"

2. Sampling Methods: Drawing sample from a population using SRSWR, SRSWOR, Stratified random sampling, Systematic sampling.

3. Diagrams: Simple bar diagram, subdivided bar diagram, multiple bar diagram, Pie diagram, steam and leaf chart.

4. Graphs: Box plot, rod or spike plot, histogram (both equal and unequal class intervals), frequency polygon, ogive curves, empirical distribution function.

5. Measures of central Tendency: Computation of following measures for all types of data. Mean, mode, median, quartiles, Deciles, Percentiles, Geometric mean, Harmonic mean.

6. Measures of dispersion: computation of following measures for all types of data. Range, Quartile Deviation, Variance, Standard Deviation, Coefficient of Variation, Mean Deviation, Mean Squared Deviation.

7. Measures of Skewness and Kurtosis: Bowleys coefficient and Karl Pearson's coefficient of Skewness.

8. Moments: Computation of Raw and central moments, measures of Skewness and kurtosis

9. Probability Distribution: Simulation of random experiment, Hyper geometric distribution, computation of probabilities, model sampling, Fitting of Binomial, Poisson, Normal and Lognormal distribution: computational of probabilities, model sampling from log-normal, Weibull, Logistic distribution

10. Correlation and Regression: Fitting of lines of regression, computation of correlation coefficient, fitting of second curve.

11. Multiple Regression: Fitting of regression plane for trivariate data

12. Testing of Hypothesis: Large sample test for mean and proportions, t-test, variance test, Chi-square test, Computation of probabilities of type I and type II errors and power of the test

13. Analysis of Variance: One way, two way, Analysis of completely randomized Design (CRD), Analysis of randomized block design (RBD)

14. Data Mining: k-nearest neighbor techniques for classification, k-means techniques for clustering.

15. Statistical Quality Control: CUSUM chart, Exponentially weighted moving average charts (EWMA-chart)