Statistics with 'R'

CREDIT DISTRIBUTION, ELIGIBILITY AND PRE-REQUISITES OF THE COURSE

Course title & Code	Credits	Credit distribution of the course			Eligibility criteria	Pre-requisite of the course (if any)
		Lecture	Tutorial	Practical/ Practice		
Statistics with R	2	0	0	2	Class XII	Basic course in Statistics

Learning Objectives

The Learning Objectives of this course are as follows:

- To enable students to handle data in the R software thereby helping them to understand meaningful statistical analysis performed on the data.
- To enable students to extract data, and perform basic statistical operations entailing data analysis such as – data cleaning, data visualisation, data summarisation, and regression amongst others.

Learning outcomes

The Learning Outcomes of this course are as follows:

- After studying this course, students will be able to extract and Read data into R, manipulate, and analyse it
- After studying this course, students will be able Tto debug, organize, and comment R code
- After studying this course, students will be able to understand the R environment for downloading, installing, and using packages
- After studying this course, students will be able to do basic programming to write own functions
- After studying this course, students will be able to use loops
- After studying this course, students will be able to create standard and customized graphics
- After studying this course, students will be able to perform basic statistical operations and regression.

SYLLABUS

Unit 1: Data Extraction and Spread Sheet Exploration (12 hours)

Extraction of economics and financial data from Prowessiq, RBI, IMF, World bank or an equivalent financial/economic database. The students should be able to save and export the data to 'R-environment' for further analysis.

Unit 2: Basics of R-language **(28 hours)**

Overview of the R language: Installing R and R Studio : Using R studio, Scripts, Text editors for R, Graphical User Interfaces (GUIs) for R, Creating and storing R workspaces, installing packages and libraries, Mathematical operations.

Data Types in R – Numeric, Integer, Character, Logical, Complex and missing data. Data Structures in R

- Vectors – Creation, Arithmetic operations of Vectors, Vector Sub setting, Sorting and Sequencing functions.
- Matrix and Arrays – Creation, Arithmetic Operations of matrix, Sub setting, Use of Drop Function.
- Factors – Converting a vector into factor, assigning levels and labels, ordered Factor.
- List – Creating a list, accessing elements from a list, adding a new element and eliminating an existing element form the list, converting list to vectors.
- Data Frames – Creation of Data Frame, adding new columns, rows and removing columns, accessing column using the $ sign, importing a data set (important file formats such as csv, txt and spreadsheet), aggregate function and subsetting of dataframes, tapply function, manipulation using dplyr package (select, filter, arrange, mutate and group by function, pipe operator).

Programming Fundamentals: Logical operators, conditional statements (if, else, else if statements in R), While loops, For loops, repeat loops.

Creating functions in R.

Reading data in R (file formats such as csv, txt, and xlsx), Writing data to external files (file formats such as csv, txt, and xlsx), writing a table to a file, print function.

Unit 3: Basic Statistics and Regression **(20 hours)**

Summarizing and exploring data: Descriptive statistics (mean, median, mode, variance, skewness, five-point summary), dealing with missing data in R, Data cleaning (dplyr package, tidyr package and pipe operator), Exploratory Data Analysis; data visualization using inbuilt functions and ggplot2 package (pie chart, bar chart, line chart, histogram, box plot, scatter plot, Normal QQ plot).

Regression analysis using R: Regression vs Correlation, Simple and multiple regression, Ordinary least square, Assumptions of classical normal linear regression model (CNLRM), corrplot package, car package, lmtest package, scatter plot (using plot function and ggplot2 package) to understand the relationship between variables, lm, abline, predict, resid function, interpreting 'summary table' of the regression model, normality of residuals (qqnorm and qqPlot functions), multicollinearity (correlation matrix, corrplot and vif function), autocorrelation (acf plot and Durbin Watson test), heteroscedasticity (graphically, bptest, ncvTest), impact on estimates and inferences in case of violations of assumptions of CNLRM, methods to take care of violations.

Time series data, components of a time series data, additive and multiplicative time series model, ts function, diff function, plot of a time series data, time series data with linear trend; regression analysis using 'lm' function, stationarity in time series (concept only).

Essential/recommended readings

- Gardener, M. (2018), Beginning R: The Statistical Programming Language, Wiley & Sons.
- Sekhar, S.R.M., et al. (2017), Programming with R, Cengage Learning India.
- Wickham, H., et al. (2017), R for Data Science: Import, Tidy, Transform, Visualize, and Model Data, O'Reilly'.
- Field, A., Miles, J and Field (2012), Z. Discovering Statistics using R (Indian Reprint 2022), SAGE
- SimpleR - Using R for Introductory Statistics: John Verzani.
- The R Guide.
- Analysis of Epidemiological Data Using R and Epicalc: Virasakdi Chongsuvivatwong.
- Statistics Using R with Biological Examples: Kim Seefeld and Ernst Linder.
- An Introduction to R: Software for Statistical Modeling & Computing: Petra Kuhnert and Bill Venables.
- Gujarati, D.N. et al (2018), Basic Econometrics, McGraw Hill India, 5th Ed.
- CRAN website: https://cran.r-project.org/
- https://prowessiq.cmie.com,
- https://data.worldbank.org/indicator,
- https://rstudio.com/products/rstudio/download/(Rstudio)
- http://r-statistics.co

Examination scheme and mode:

Evaluation scheme and mode will be as per the guidelines notified by the University of Delhi.