University of Sri Jayewardenepura

Faculty of Applied Sciences

Department of Statistics

STA 315 2.0 Programming and Data Analysis with R

Type: Core

Duration: 30 lecture hours

Pre-requisites: STA 111 2.0 Elements of Sampling and Descriptive Statistics
STA 112 3.0 Probability and Distribution Theory
STA 121 3.0 Statistical Inference

Course Description:

This course gives students an opportunity to use the public domain and free software R to perform statistical computing. The R language provides a rich environment for working with data, especially for statistical modeling and graphics. The course emphasis is on data manipulation and basic statistical analysis, including exploratory data analysis, classical testing, categorical data analysis, and regression. Students will identify appropriate statistical methods for the data or problems and conduct their own analysis using the R environment. This is a hands-on, project-based course to enable students to develop skills and to solve statistical problems using R.

Course Objective:

The objectives of this course unit are to

- provide opportunity to learn a widely used programming language R
- develop skills of programming
- provide opportunity to learn various Statistical concepts by carrying out simulations

Course contents:

1. **Introduction to R**
   1.1 Why use R?
   1.2 Downloading and Installing R
   1.3 Running the R program
   1.4 Getting Help via the CRAN website and the internet

2. R as a Calculator
2.1 Basic arithmetic
2.2 Storing the results of calculation
2.3 Comments and spacing
2.4 Basic functions
2.5 R's Help function
2.5.1 Known function
2.5.2 Unkown function
2.5.3 Comprehensive help
2.6 Saving Your Work in R

3. Objects, their modes and attributes
3.1 Vectors
3.1.1 Creating vectors
3.1.2 Vector functions used in R
3.1.3 Simplifying vector creation
3.1.4 Indexing
3.2 Matrices
3.2.1 Creating matrices
3.2.2 Indexing matrices
3.2.3 Adding rows and columns to the matrix
3.2.4 Mathematical Operations
3.3 Lists
3.4 Data Frames
3.5 Factors
3.6 Useful functions for working with data objects

4. Basic Data Management
4.1 Missing values
4.1.1 Recoding values to missing
4.1.2 Excluding missing values from analyses
4.2 Sorting data
4.3 Merging datasets
4.3.1 Adding columns
4.3.2 Adding rows
4.4 Subsetting datasets
4.5 Importing data from files

5. Control Structures and Loop Functions
5.1 if, else
5.2 for
5.3 while
5.4 repeat
5.5 break

5.6 next
5.7 return
5.8 Loop functions

6. **Introduction to Graphical Analysis**
   6.1 Basic Graphs
   6.2 Graphical parameters
      6.2.1 Symbols and lines
      6.2.2 Colors
      6.2.3 Text characteristics
      6.2.4 Graph and margin dimensions
   6.3 Adding text, customized axes and legends
   6.4 Combing graphs

7. **Basic Statistics**
   7.1 Probability and distributions
   7.2 Descriptive Statistics

8. **Working with R Packages**
   7.1 Installing a package
   7.2 Loading a package
   7.3 Working with packages

9. **Writing functions in R**

10. **Advanced Data analysis with R**

Learning Outcomes:

By the end of the course unit students should be able to,

- identify the components of the R interface for Windows
- conduct standard arithmetic calculation: both numerical and matrix
- access R help
- load R packages
- import and export data
- write his/her own R functions to analyse data
- carry out analyses using available functions
- use simulation for statistical analysis

Reference Text books:

1. Beginning R, The Statistical Programming Language
   Author: Dr. Mark Gardener
   Publisher: John Wiley & Sons, Inc.

2. R in Action, Data analysis and Graphics with R
   Author: Robert I. Kabacoff
   Publisher: Manning

Method of Assessment:

1. Mid Semester Examination - 20%
2. End Semester Examination - 80%

Lecturer in charge: Ms. Thiyanga Talagala