SEC: MS 306: SOFTWARE FOR DATA EXTRACTION AND ANALYSIS

Course Objectives:

This course is about how to obtain data from financial database and how to use R language for statistical &econometrics applications. The main objective is to develop skills that can help in solving your research problems.

Prerequisites - Basic knowledge of statistics is desired.

Learning Outcomes:

After studying this course, you will be able to-

1. How to obtain data from financial database (Prowess IQ)

2. How to perform data analysis using R

3. Use of R and prowess in research.

Course Content:

Unit I   (4 weeks)

An Introduction to financial database Prowess IQ from CMIE: Creating company set-, creating spreadsheets, use of elements in Ribbons – Company address and identity identicators, business segments and products, Ownership structure and governance indicators, Financial Statements, Stock prices and capital changes, Capex and M&A, indices and index number. Formulating queries and advance queries Student is expected to be able to extract different types of data for an index, an industry and company data Selection of company/s, period to be studied. Data extraction from balance sheet, profit & loss statement and cash flow statements Stock market data- price and volume, BSE/NSE, adjusted prices Saving and exporting data to a spreadsheet for further analysis.

References:

https://www.prowess.com/(PDF Page 1-25)

Unit II   (4 weeks)

Overview of the R language: Generating R code, data structures, creating functions, conditional formatting, looping, list, dictionary, array. Using R studio, Scripts, Text editors for R, Graphical User Interfaces (GUIs) for R, installing packages and libraries, Variable classes (factor, numeric, logical, complex, missing), matrices operations, Data sets included in R packages, Summarizing and exploring data. Data cleaning and mining. Using data from

external files- reading& writing data to external files, Creating and storing R workspaces, Basic exploratory graphics, Mathematical operations.

References:

Sekhar, Kumar and Kasa, Programming with R; Cengage Learning [Chapter 1-8]

Unit III   **(4 weeks)**

Analysis of data Using R: Descriptive Statistic of data. Estimating a Multiple Regression Equation by Ordinary Least Squares, Violations of Classical Assumptions: multicollinearity, heteroscedasticity, autocorrelation and model specification errors, their identification, their impact on parameters; tests related to parameters and impact on the reliability and the validity of inferences in case of violations of Assumptions; methods to take care of violations of assumptions, goodness of fit. Testing of stationarity. Panel data models estimation.

References:

Sekhar, Kumar and Kasa, Programming with R; Cengage Learning [9 and10]

Text Books:

1. Sekhar, Kumar and Kasa, Programming with R; Cengage Learning

Additional Readings:

1. Vishwanathan, Data Analytics with R- A hands on approach; Infivista Inc.
2. Chang, R Graphics Cookbook- Practical Recipes for Visualizing Data; O' Reilly Media

Teaching Learning Process:

Class room lecture, Lab sessions, Workshop, Project Assignments.

Assessment Method:

1. Practical evaluation of 50% marks
2. End term University Exam of 50% marks

Teaching and Learning Process:

This course will be taught using a mix of the following tools:

1. Relevant Case studies
2. Hands on practice on prowess and R

SOFTWARE FOR DATA EXTRACTION AND ANALYSIS

3. Relevant and important articles from academic linked journals in the domain of Management such as Harvard Business Review,

Assessment Method:

- Internal Assessment : 25 marks
- Practical : 25 marks
- Written Theory Exam : 50 marks

Keywords:

Prowess, R- software, Stata, Multiple Regression and Classical Assumptions