

## **DSC4217 Advanced Analytics with R**

**Lecturers :** Dr. Liu Qizhang

**Session**: Semester 2, 2017/2018

### **Description**

We are now at the era of big data. Data and algorithms dominate the day. Competitive advantage, for more and more enterprises, is obtained via data analytics and idea sharing in the current fast-paced, data-intensive, and open-source business environment. The capability of understanding data, digging out valuable insights from data, and thus making right managerial decisions accordingly has gradually become an essential skill that business graduates must master in order to excel in their career.

This course prepares students with fundamental knowledge of using R, a powerful complete analytical environment, to organize, visualize, and analyze data. It is, however, not a pure programming course. It will focus on case studies from the most commonly used tasks that a business analyst will face on a day-to-day basis and how R could be used to handle such cases.

The topics covered in the course include foundations of R programming, data visualization, regression-based data modelling, classification, and cluster analysis.

### **Objectives**

At the end of the course, students should be able to

- Master basic R programming techniques to model and analyse data.
- Self-explore R packages for analytics needs.
- Understand some common analytics methodologies used in current business environment.

### **Course Outline**

#### **1) Foundations of R Programming**

- a) Data types
- b) Vectors, lists and matrices
- c) Reading and organizing data
- d) Functions
- e) Control structure

#### **2) Exploring and Discovering Data**

- a) Basic data summaries and visualization
- b) Time series graph

c) Spatial graph

**3) Simulation with R**

- a) Introduction to Simulation
- b) Simulation Framework in R

**4) Regression-based Data Modeling**

- a) Basics of data modeling
- b) Making data model more flexible
- c) Making data model more selective

**5) Classification**

- a) Introduction to Logistic Regression
- b) Interpretation of Logistic Regression results with R
- c) Classification with Decision Tree

**6) Cluster Analysis**

- a) Common steps in cluster analysis
- b) Hierarchical cluster analysis
- c) Partitioning cluster analysis
- d) Avoiding nonexistent clusters

**Reading List**

Compulsory reading:  
Course Package

Supplementary reading:

- “R in Action – Data analysis and graphics with R”, Robert I. Kabacoff, Manning
- “Data Mining and Business Analytics with R”, Johannes Ledolter, Wiley.
- “Marketing Data Science”, Thomas W. Miller, Pearson.
- “Business Analytics for Managers”, Wolfgang Jank, Springer.

**Prerequisites**

DSC1007 and DSC2008.

**Weightage of Assessment**

Continuous Assessment :

Class Participation	20%
Group Project	40%
Test 1	20%
Test 2	20%