Computational Data Analysis Using Software

February 16, 2016

1 Aim

In this course you will learn how to program in R and how to use R for effective data analysis. You will learn how to install and configure software necessary for a statistical programming environment and describe generic programming language concepts as they are implemented in a high-level statistical language. The course covers practical issues in statistical computing which includes programming in R, reading data into R, accessing R packages, writing R functions, debugging, profiling R code, and organizing and commenting R code.

Further, this course is intended for students, to bridge the gap between statistical textbooks and empirical research. In this intermediate role, it does not cover the detailed expositions of a professional textbook. Instead, it deomonstrates how to accomplish a wide vaiety of statistical tasks by means of software and realworld dataset. The purpose of this course is twofold: (1) to make student familiar with econometric theory and methods; (2) to master basic skill of empirical research. In the end, students will gain experience in experimental design, data analysis, computer statistical software, and interpretation of results.

There are few requirements for this course. You will need some basic statistical knowledge, particularly of the linear model and its extensions.

- Prerequisites

  - An introductory course in probability and statistics
  - An introductory course in econometrics

2 Class policy

Attendance and Lateness: Students should attend each class and be present when each class begins. Being 15 minutes late is considered absent from the class. Excused absences for special circumstances (e.g., sickness) may be arranged in advance and will not influence the attendance grade. There is no any formal seating chart.

Homework: Homework assignments may contain some combination of R programs and written solutions to questions. The grade will be based on the functional correctness of your code, your adherence to the homework-specific requirements. Homework will no longer be accepted after deadline.

Academic Dishonesty and Misconduct: Students should not only gain knowledge and skills, but also build their character. Particularly, public managers should equip themselves with high and strict professional standards and ethics. All students should complete their own work and be evaluated based upon that work. Be honest with yourself and the instructor; simply say that you know what you know and you don't know what you don't know.

Course Feedback: Given diversity in their backgrounds, students are always encouraged to make comments and suggestions during this term in order to improve this course. Any form of communication (e.g., walk-in, phone, email, etc.) will do.

3   Topic outline

Week 1: Introduction to R	Week 8: R programming
Week 2: Getting and Cleaning Data (1)	Week 9: Regression Models (1)
Week 3: Getting and Cleaning Data (2)	Week 10: Regression Models (2)
Week 4: Exploratory Data Analysis (Basic Graphics)	Week 11: Bootstrapping and resampling
Week 5: Exploratory Data Analysis (Basic Statistics)	Week 12: Financial data analysis using R
Week 6: Advanced Graphics	Week 13: Reproducible Research
Week 7: Creating functions	Week 14: Writing faster R code, Debugging in R

Table 1: Course schedule

4   Grading

The final grade is based on midterm exam, final exam (written examination), homework assignments, and class attendance and participation. Their weights are:

- Final exam (written examination) 40%

- Midterm exam (project and presentation) 40%

- Homework assignment and Class attendance 20%

- Extra credit for outstanding participation and performance (up to 10%)

* Students who missed 30 percent (or more) of class MAY NOT pass this course.

5   Some reading materials

- This class recommends following books:

    - Basic:

        * An Introduction to R

* R for Beginners
* R cookbook
* R in Action
* R in a Nutshell

- Advanced:

  * Statistics with R
  * The R book
  * Applied Econometrics with R

- Time Series analysis

  * Time Series Analysis and Its Applications: with R example
  * Analysis of financial time series
  * Analysis of Integrated and Cointegrated Time Series with R

- Financial analysis

  * Advanced Topics in Analysis of Economic and Financial Data Using R
  * Simulation Inference Stochastic Differential Equations: with R examples
  * Modern Actuarial Risk Theory Using R

- Statistical analysis

  * The element of statistical learning (http://statweb.stanford.edu/~tibs/ElemStatLearn/)

6 Others information

- Teacher: Haifeng Xu

- Office hours: Friday, 10:30–12:00 (Need to make an appointment)

- Lecture Hours: See the timetable

- Homework E-mail address: stata_homework@163.com

- Lecture note will be upload to: http://l.xmu.edu.cn