STAT 340-3
Introduction to Statistical Computing and Exploratory Data Analysis

Spring 2013
Day Course

Students requiring accommodations as a result of disability must contact the Centre for Students with Disabilities 778-782-3112 or csdo@sfu.ca

Instructor: Dr. Carl Schwarz

Prerequisite:
STAT 285 or STAT 302 or STAT 305 or equivalent.

Textbook:
SAS and R, Data Management, Statistical Analysis, and Graphics, by Ken Kleinman and Nicholas J. Horton, Publisher: CRC Press

Calendar Description:
Statistical computing in R and SAS. Data management: reading, editing and storing statistical data; querying databases with SQL. Data exploration and representation: summarizing data with tables, graphs and other statistical tools. Data simulation: model-based and empirical. The SAS component of the course will give students a good start for writing the SAS programming certification exams. **Quantitative**

Outline:

Part 1. SAS component

1. What is SAS?
   - Downloading and installing
   - Overview of the system
2. Data management in SAS
   a. Data input and structures
      - DATA step
      - Reading specially formatted files
      - Date/time/character formats and manipulations
      - Derived variables
      - Exporting
   b. Data access: from database systems using query languages
   c. Merging and reshaping data
      - sorting/subsetting (set/if/where statements)/ merging/transposing
      - processing using DO LOOPS and SAS arrays
      - modify variable attributes
3. Data exploration and representation in SAS
   - basic procs (print, plot, tabulate, means, univariate, freq)
   - by statement and uses in analysis and simulation
   - output delivery system to extract information from analyses
4. Data simulation in SAS

Part 2 R component

1. What is the R programming environment
   - Downloading and installing
   - Basics of writing R functions
   - Basics of loops/if/while and other control-flow constructs
2. Data management in R
   - Reading and writing data: plain text files and spreadsheets, other file formats
   - Using R to query databases with SQL
   - Merging and re-shaping data
3. Data exploration and representation in R
   - Graphical displays. Customizing and extending these displays for your own research purposes.

- Cross-tabulations and tests of association.

4. Data simulation and resampling in R
    a. Generating data from parametric distributions: uses in evaluating statistical procedures and in understanding classical large-sample results.
    b. Generating data by resampling: introduction to permutation, bootstrapping, cross-validation and their uses.

Grading Scheme:

Assignments – 20%
R Component Exam – 40%
SAS Component Exam– 40%

Grading is subject to change.

Students should be aware that they have certain rights to confidentiality concerning the return of course papers and the posting of marks. Please pay careful attention to the options discussed in class at the beginning of the semester. Students are reminded that Academic Honesty is a cornerstone of the acquisition of knowledge. Scholarly integrity is required of all members of the University. Students are encouraged to review policies pertaining to academic integrity available on Student Services webpage at http://students.sfu.ca/academicintegrity.html

---

Students looking for a Tutor should send an email to stat@sfu.ca with "Tutor Request" in the subject line. Please only include information that you would like forwarded to our tutors mailing list.

Revised October 30, 2012