STAT4110/7110: Statistical Software and Data Analysis
Fall Semester, 2012
Time: 10:00-10:50 am, MWF
Location: MDLBH 7
Instructor: Athanasios Christou Micheas
Office: 134G Middlebush Building
Office Hours: 12:10-12:50, MWF
(Also by appointment)
Phone: 884-8828
e-mail: amicheas@stat.missouri.edu
course web site:
http://www.stat.missouri.edu/~amicheas/stat7110/index.html

Prerequisites: At least one course from the statistics department at the 200 level or above.

Text:

DiIorio, SAS Applications Programming
Spector, An Introduction to S and S-Plus

References:
SAS Language Reference: Concepts, Version 8
SAS Language Reference: Dictionary, Version 8
SAS Procedures Guide. Version 8, Volumes 1 and 2
SAS/STAT User’s Guide. Version 8, Volumes 1, 2 and 3
Online documentation:
http://system.missouri.edu/itss/lpp/sas/ver8/sashtml/onldoc.htm
http://www.sas.com
Venables and Ripley, Modern Applied Statistics with S-PLUS
Krause and Olson, Statistics and Computing: The basics of S and S-Plus

Students with disabilities: If you have special needs as addressed by the Americans with Disabilities Act (ADA) and need assistance, please notify the Office of Disability Services, A038 Brady Commons, 882-4696 or the course instructor immediately. Reasonable efforts will be made to accommodate your special needs.

Course Requirements: Regular homework assignments, a midterm on SAS topics, and a final exam, both taken in class. The homework will count approximately 60-70%, and the midterm and final 30-40%.

- Midterm: Friday, October 19
- FINAL EXAM: TBA
**Homework:** Homework assignments will be made and collected throughout the semester. You may work together on these assignments, but you must formulate and write your own solutions. No credit will be given for answers without supporting work.

- All homework is due at the beginning of class.
- Late homework will not be accepted, except for acts of God, e.g. medical emergencies or death in the family.
- On SAS assignments, you must turn in all input programs and SAS output.
- Annotate your output (either with a word processor or by hand) so that the grader can determine that you have accomplished the goals of the assignment.

**Objectives of the course:**

The course illustrates the use of two statistical packages, SAS and R (S-PLUS).

SAS is a statistics software that is widely used in industry, although there are other popular statistics packages that are easier to use for PCs and Macs. (SPSS, Statistical Package for the Social Sciences, is a notable popular package.) The class focuses on aspects of statistical programming with SAS not covered in other statistics department courses. We will initially discuss techniques of data base management and data manipulation. Other SAS capabilities including the graphics package and the interactive data visualization package, PROC INSIGHT, will be discussed in addition to the basic techniques for one and two sample problems, analysis of variance, linear regression, and categorical data.

R is a freeware statistical package that is essentially the free analog of S-PLUS which was an increasingly popular statistical programming language. There are minor differences between the two languages (syntax wise etc). Of course S-PLUS is not free... Both are object-oriented languages with a number of modern statistical capabilities not generally available elsewhere. R/S-PLUS is easily extensible; new capabilities are readily added, either from public domain packages or by writing your own programs. In addition, R/S-PLUS has perhaps the most powerful graphics capabilities of any statistics package since the main routines are very amenable to changes.
Topics to be covered:

SAS

• language basics: variable types and names, syntax
• dataset structure
• input: list, informats, pointer control, iterated formats
• basic procedures: print, means, univariate, sort, chart
• normal probability plotting
• SAS functions including random number generation, quantiles, percentiles
• DO statement
• arrays
• SAS datasets—permanent, libraries
• merge and update
• proc format; recoding
• plotting: proc plot, proc gplot
• statistical methods: proc means, ttest, reg, freq, anova, glm
• Introduction to macros (time permitted)

R/S-Plus

• primitives: vectors, list, matrix, data.frame
• matrix operations
• descriptive statistics: histograms, boxplots, stem plots
• functions including random number generation, quantiles and percentiles of common distributions
• basic methods including parametric and nonparametric one and two sample problems, contingency tables, regression and anova
• loops
• writing programs
• linked plots, scatterplot matrix
IMPORTANT

- Students need to obtain an account (pawprint) with the MU server, bengal.missouri.edu. This is done when you setup your e-mail account. You can use this account to access SAS on the linux system.

- If you cannot login, in order to be able to access SAS during the lectures, you will need to inform the instructor of your bengal account (pawprint) so that we can add it to the lab list. So please send your pawprint names to amicheas@stat.missouri.edu.

- SAS is available in PC labs and on a linux enviroment through bengal. PC SAS (or a Macintosh version) can be obtained from the IATS Outpost. R is freely available on the internet. R is essentially a public domain version of S-PLUS, that is gaining widespread popularity. The R package can be downloaded from http://www.r-project.org.

- For an available list of computing sites offered on campus, please go to http://iatservices.missouri.edu/computing-sites. There should be PC versions of SAS installed on the machines of these labs. You’ll need your pawprint again to login.

Honesty: Academic honesty is fundamental to the activities and principles of a university. All members of the academic community must be confident that each person’s work has been responsibly and honorably acquired, developed and presented. Any effort to gain an advantage not given to all students is dishonest whether or not the effort is successful. The academic community regards academic dishonesty as an extremely serious matter, with serious consequences that range from probation to expulsion. When in doubt about plagiarism, paraphrasing, quoting, or collaboration, consult the course instructor.

Cell Phones: It is your responsibility as a student of this university, to have your cell phones, ipods, kindles etc CLOSED before entering the classroom. If a cell phone rings, you text, or you use electronic devices during a lecture or exam, you will suffer an automatic 10% reduction of your final grade. Only one warning will be given. Using cell phones in class is not proper university conduct.

Computer: During the lecture or exams you may have the SAS or R windows open, as well as a web browser directed to the class
website ONLY. The time at the computer lab is reserved for learning and following along the lecture. Opening web browser email, or other websites than the class website will result in an automatic 10% reduction of your final grade. Only one warning will be given.

**IATS Short Course:**
http://iatservices.missouri.edu/training/
All courses available for PC and MAC, SAS Part I, II, and III. Upon completion of these courses, the student will be able to:
1. Create and combine datasets
2. Sort datasets
3. Utilize data analysis techniques
**PREREQUISITES:** A Bengal/Tiger ID and basic UNIX knowledge is required prior to registering for the course.
**Calendar:** https://iatservices.missouri.edu/training/calendar.html