

STAT 704, Data Analysis I -- Fall 2008

Instructor:

David Hitchcock, assistant professor of statistics
209A LeConte College
Phone: 777-5346
Email: hitchcock@stat.sc.edu
Course Web Page: <http://www.stat.sc.edu/~hitchcock/stat704.html>

Classes:

Meeting Times:
Tue-Thu 12:30-1:45 p.m., LeConte College, Room 201A

Office Hours:

Mon. 10:00-11:00 a.m., Tues. 11:00 a.m.-12:00 noon, Wed. 10:00-11:00 a.m.,
Thurs. 11:00 a.m.-12:00 noon, Fri. 1:30 p.m.-2:30 p.m.
Please feel free to make appointments to see me at other times.

Textbook:

Applied Linear Statistical Models, 5th edition, by Kutner, Nachtsheim, Neter and Li.

Purpose: To provide an overview of normal-model inference and a strong background in regression modeling.

Prerequisite/Corequisite: Prerequisite is the completion of an elementary statistics course or consent of department. Corequisite is STAT 712.

Official Course Description: Primarily for graduate students in statistics and the mathematical sciences. Probability concepts, inferences for normal parameters, regression, correlation, use of computer statistical packages.

Course Outline: Random Variables and Moments; Important Probability Distributions. Single-Sample Inference; Two-Sample Inference; Nonparametric Alternatives. Simple Linear Regression and Correlation; Multiple Regression. Nonlinear Regression; Logistic and Count Regression; Nonparametric Regression.

Homework:

Homework exercises from the textbook (and possibly other sources) will be assigned on the course web page. Due dates are given on the course web page. Late homework will be penalized. Please write up homework papers neatly and clearly.

Each student's homework must be done independently. You may ask each other informal questions about the homework, but everyone is to do his/her own work. If homework is found to be copied, all students involved will receive a 0. Of course, you may always ask me questions about the homework. [To be clearer, students can ask each other informal ORAL questions about homework, but cannot look at or copy each other's homework papers. All submitted homework must be their own work.]

Project:

A project involving the analysis of real data using methods learned in this class will be due near the end of the semester. The first part of the project will be a preliminary proposal with a data description and the other part will be the final written report. More information will be given out later.

Exams:

There will be two in-class midterm exams and a final exam on Thursday, December 11 at 2:00 p.m. Exams may not normally be made up, except in extreme circumstances, for which written documentation of excuse (doctor's note, funeral notice, etc.) is required. If you suspect you may miss an exam day, it is important to contact me well in advance of the test date.

Grading:

The course grade will be based on homework/quizzes (15%), 2 midterm exams (24% each), and a final exam (27%) and a data analysis project (10%). A course average of 90-100 will result in an A, 87-89 a B+, 80-86 a B, etc.

Learning Outcomes:

The successful students will learn important principles of normal-model inference and methods for data analysis, especially regression methods. Successful students will be able to interpret and clearly communicate the results of common analyses.

Computing:

Some problems in this course involve significant computations, and for these, we will learn to use the software packages SAS and R. You will have an account on the MATHSTAT domain. Currently the computers in LC 124 (and some in PSC 102 and LC 303A) have SAS and R. Student copies of SAS for home use are also available for purchase; for information, call Barbara Koski (803-777-4657) at USC Computer Services (corner of Blossom and Sumter). R is a free, open-source statistical programming language. Details about how to download R for free onto your home computer are posted on the course web page.

It is not assumed that you have much previous experience with SAS or R. In many industries and jobs, SAS is the standard statistical computing package used, and this course will introduce you to some of the most common SAS procedures. R is an extremely useful statistical programming language that has become widely used in recent years.

Tentative Course Schedule: Tues-Thurs, August 21 through December 4, except:

No class (Fall Break): October 9 (Thursday)

No class (Thanksgiving): November 27 (Thursday)

Tuesday, September 30 (tentative): Exam 1

Tuesday, November 4 (tentative): Exam 2

Thursday, December 11 (2:00 p.m.): final exam

** Homework Due Dates will be posted on the course web page with each homework assignment.