

STATISTICS 704 SYLLABUS (Draft)
Data Analysis I Fall 2015

John M. Grego
TTh 1:15-2:30 LC 201A
Office Hrs: MW 10:30-12

LC216A
777-5070
grego@stat.sc.edu

Text *Applied Linear Statistical Models 5e* by Michael Kutner, Christopher Nachtsheim, John Neter and William Li. Softcover international versions of this text are available online and suitable for this course.

Disabilities If you qualify for accommodations because of a disability, please submit a letter to me from the Office of Student Disability Services in a timely manner so that your needs can be addressed. The Office of Student Disability Services determines accommodations based on documented disabilities. Contact: 777-6142, LeConte 112A; [http : //www.sa.sc.edu/sds](http://www.sa.sc.edu/sds)

Learning Outcomes Students should be able to

- Analyze one and two-sample data sets using established parametric and nonparametric test procedures
- Derive results for the simple linear regression model
- Derive results for the matrix formulation of the multiple linear regression model
- Diagnose multicollinearity and validity of model assumptions in regression models and take appropriate corrective measures
- Analyze regression data sets and prepare written and oral presentations on the analysis
- Build multivariate regression models using model selection and validation techniques
- Analyze regression data using approaches other than linear models, including generalized additive models, robust regression and quantile regression

Attendance Though attendance is never a problem with graduate students, I would like to note that my policy corresponds to the policy stated in the student handbook: If you miss more than 10% of your classes (> 3 classes), the teacher may choose an appropriate penalty. I will deduct 2% from your final class average for each additional day that you miss after the third absence; excused and unexcused absences both count as absences.

Grading Grades will be weighted in the following way:

In-class exam	100 points
Homework/Classwork	100 points
Project	100 points
Final Exam	100 points
<hr/>	
Total	400 points

The project can be undertaken with a partner (or partners) and will consist of a proposal, a final draft and oral presentation. I use the project to enhance (or reinforce) several skills you will need in your future (or current) career: written and oral communication, practical

problem-solving and teamwork. The oral presentation should provide useful practice for professional presentations.

The exam administered during the semester and the final exams are in-class exams because exposure to in-class exams helps prepare students for the qualifying examination at the end of their first year.

You are encouraged to discuss homework and class assignments with your classmates and me, but all such assignments must be written independently. Do not copy any part of another student's work or computer code. Incidences of cheating and academic dishonesty will be punished to the full extent allowed under university regulations.

We will have “in-class” activities that will help in your development as a statistics professional. After each colloquium, we will discuss the colloquium topic (briefly—colloquia are often difficult for first-year students to understand). For portions of the course in which the text is exemplary, we will have lectures that emphasize active learning. Before some of our scheduled lectures, you will have thoroughly read the day's material, and have prepared and uploaded a write-up for an on-line assessment of the material (these assignments will be posted in Blackboard a week ahead of their due date). We will discuss the assignment at the start of class on the due date; grades will be based on preparation and participation. We will have activities that take place during class; these will be built into the lecture notes and range from quick responses to those that take more time and are more in the spirit of the work taking place in a flipped classroom.

Computing and Technology I will maintain a class web page that will be used to post lecture notes, homeworks, tests, and other course-related materials. The URL for the class web page is people.stat.sc.edu/grego/courses/stat704. All assignments will be posted on Blackboard (though assignment materials will be on the website). Grading will be entirely electronic and handled through Blackboard. I will use the computer/LCD projection system extensively in class for presentations and demonstrations of computer software.

We will be using two computer packages throughout the course. Proficiency with SAS is a course objective, and familiarity with R is a professional requirement. SAS is available on office desktops, for rent through University Technology Services, or for free through SAS On-Demand as Enterprise Guide. R can be downloaded for free at <https://cran.r-project.org>. All code used in class will be available on the course webpage; extensive coding examples are available on previous instructors' websites (Profs. Hitchcock and Hanson) as well.

Date	Assignment/Topic	Graded Work
8/20	Random variables, expectation, densities for common distributions	
8/25	Normal data inferences, two sample problem, quantile-quantile plots	CE 1
8/27	Normal data continued	CE 2
9/1	One- and two-sample nonparametric tests	
9/3	Ch 1. Simple linear regression	HW 1 due, CE 3
9/8	Ch 1 and 2. Residuals, normal errors, inferences for slope, the line, and predictions.	
9/10	Ch 2. ANOVA table, general linear test, correlation.	CE 4
9/15	Ch 5. Correlation, matrices, two-sample problem and simple linear regression using matrices	HW 2 due
9/17	Multivariate normal distribution.	CE 5
9/22	Ch 6. Multiple regression in matrix terms	
9/24	Ch 6. Checking assumptions, diagnostics	CE 6
9/29	Ch 3 and 6. Transformations	HW 3 due
10/1	Ch 7. Extra sums of squares, multicollinearity	CE 7
10/6		Class cancelled (flooding) CE 8
10/8		Class cancelled (flooding)
10/13	Ch 8. Polynomial and categorical predictors, interactions	
10/15	Ch 9. Model selection and validation	CE 9 (in Gambrell 003)
10/20		Exam 1 (delayed due to flooding)
10/22	Fall Break	
10/27	Ch 9.	
10/29	Ch 9.	CE 10
11/3	Ch 10. Diagnostics	HW 4 due
11/5		
11/10	Ch 10.	CE 11 (in Gambrell 003)
11/12	Ch 10. Weighted least squares and ridge regression	
11/17	Ch 11. Robust and quantile regression	HW 5 due
11/19	Generalized additive models	
11/24		Project Draft due
11/26	Thanksgiving	
12/1	Ch 11. Bootstrap and Local Regression	
12/3	Oral presentations	

The final exam will be held on Thursday, December 10 at 12:30 PM in Gambrell 003