**STOR 664: COURSE DESCRIPTION: Fall 2009**

**Instructor**

Richard L. Smith, Hanes 303   
Office Phone 962-2660; Home Phone 408-8126; Email: rls (at) email.unc.edu   
Web Page: http://www.unc.edu/~rls/s664/s664.html

Once on the web page you can click "Course Web Page" for updated information about the course. The "Data Page" is used to store all data sets which are used in the text or set as class exercises.

**Please check the course web page often, as important information will be placed on it.**

**Class time and place**

Mondays and Wednesdays 2:00 to 3:15 pm, Hanes 130

**Office Hours**

Feel free to come by my office any time, or approach me after class, or email or call for an appointment. However, there will be an official office hour as well, when anyone can drop in. Day and time to be announced, after consultation with the class.

**Grader**

Jiang Chen, Hanes B05. Phone: 919-843-8571. Email: jiangc(at)email.unc.edu

**Course Text**

The course text will be the draft version of **Linear Regression** by R.L. Smith and K.D.S. Young. This will be available as a course pack through Student Stores. There will be a charge, likely to be around $25.

**Chapter Headings**

**Chapter 1: Air pollution and public health: A case study for regression analysis.**   
This introductory chapter discusses a major public policy issue where the use of regression analysis has featured heavily. It illustrates some of the techniques which we will be discussing in detail later in the course, and also describes some of the pitfalls associated with the use of regression to solve substantive scientific problems.

**Chapter 2: Simple linear regression.**   
For most of you, much of this material will be revision, covering the simple case of one y variable and one x variable. However, we also discuss some more subtle features, such as simultaneous confidence intervals, inverse regression or calibration, and tests for autocorrelation.

**Chapter 3: Multiple regression.**   
Matrix formulation and solutions. Confidence and prediction intervals, and hypothesis tests. Simultaneous estimation. Power of the F test. Examples. Theoretical background.

**Chapter 4: Diagnostics for influential observations.**   
This chapter is concerned with the effect of outliers among either the x or y values. The hat matrix. Diagnostics for influence: DFFITS, DFBETAS, Cook's statistic, COVRATIO. Graphical methods. Examples.

**Chapter 5: Diagnostics for model selection.**   
Multicollinearity. Variable selection. Transformations. Applications.

**Chapter 6: Miscellaneous topics in regression.**   
Weighted and Generalized least squares. Response surface methodology. Nonlinear regression.

**Chapter 7: Analysis of Designed Experiments.**   
One-way and two-way analysis of variance, latin squares, factorial designs.

**Chapter 8: An introduction to generalized linear models.**   
Extensions of linear regression: logistic regression, Poisson regression.

**Computing**

The course includes an extensive practical computing component. The main software packages I intend to use are SAS and R. It will be necessary for you to learn some of both of these packages. The course does not provide formal instruction in either language, but the text contains extensive examples which should provide sufficient information for you to do the exercises that require them.

R is a completely free package. If you do not have it already, you should download it from [**http://cran.r-project.org**](http://cran.r-project.org/)

To use SAS, there are three options (that I am aware of):

1. Many of the departmental machines have SAS already installed, to use in a Windows environment. If you can get as far as opening up SAS and getting to the editor screen, you have enough to begin using SAS for this class. (Which version of SAS? I currently have 9.1 installed, but the version doesn't really matter, because all the SAS programs we are going to use have been part of SAS for a very long time.)

2. If you want to install SAS on your own machine, go to http://help.unc.edu/5546 and follow the instructions. You will need you onyen and password and there is a formal registration procedure, but it's free to all UNC students.

3. You can also access SAS directly on Emerald via a Unix shell. See http://help.unc.edu/4176.

Although I will be providing handouts or weblinks for all the features of SAS and R which will actually be used for the exercises, you may want to gain some familiarity with these packages for yourself. An excellent introduction to SAS is is **The Little SAS Book** by Delwiche and Slaughter, available through the Campus Store. This is written for beginners, but it will take you as far as PROC REG and PROC ANOVA (Chapter 7) which is plenty to give you the flavor of how the package works. In addition, there are various web-based guides - I have included in link to Bob Derr's guide in my home web page and you can also access SAS documentation through the university's Information Technology webpages.

For R, there are by now many introductory guides - one reference is **Introductory Statistics with R** by Peter Dalgaard (Springer Verlag, published 2002) but there are many others.

**Assignments and Exams**

Homeworks consisting of both theoretical and computational exercises will be set, at approximately two-week intervals. There will be a midterm and a final exam. Provisional distribution of marks: 30% for homework assignments, 30% for the midterm, 40% for the final exam.

**Class policies**

Attendance at all classes is expected; if you expect to be away for an extended period you should discuss it with me in advance. There is no need to ask permission if you have to miss an occasional class for personal reasons (e.g. doctor’s appointment).

The Honor Code is in place for all assessed work in this class, and you should familiarize yourself with its provisions if you are not already familiar with them. With homeworks, I allow and encourage students to work together, but the work you hand in must be your own – direct copying from one student to another is not permitted. For exams, no consultation of any form is permitted except that you are always allowed to ask questions directly to the instructor.

If for any reason you think you might miss an exam, I need to know that as soon as the situation becomes apparent to you.

The final exam is scheduled for Monday December 14, beginning at 4pm. The midterm is tentatively scheduled for Monday, October 19, in class, but please let me know if you have a conflict for that time.

**Further reading**

Other references that may be helpful include the following:

Atkinson, A.C. (1985), Plots, transformations, and regression. Oxford : Oxford University Press. QA278.2 .A85 1985   
Cook, R.D. and Weisberg, S. (1982), Residuals and influence in regression. New York : Chapman and Hall. QA278.2 .C665 1982   
Cook, R.D. and Weisberg, S. (1999), Applied regression including computing and graphics. New York : Wiley. QA278.2 .C6617 1999   
Dean, A. and Voss, D. (1999), Design and analysis of experiments. New York : Springer. QA279 .D43 1999   
Draper, N.R. and Smith, H. (1998), Applied Regression Analysis (Third Edition). New York: Wiley. QA278.2 .D7 1998   
McCullagh, P. and Nelder, J.A. (1989), Generalized linear models. London : Chapman and Hall. QA276 .M38 1989   
Neter, Kutner, Nachtsheim and Wasserman (1996), Applied Linear Statistical Models. Fourth Edition: Irwin, Chicago. QA278.2 .A66 1996   
Rawlings, J.O., Pantula, S. and Dickey, D.A. (1998), Applied regression analysis : a research tool. New York : Springer. QA278.2 .R38 1998   
Scheffe, H. (1959), The analysis of variance. New York : Wiley. QA276 .S34   
Weisberg, S. (1985), Applied linear regression. New York : Wiley. QA278.2 .W44 1985