STATISTICS CM120AB/C217AB
APPLIED REGRESION ANALYSIS

Fall 2003, Winter 2004
Monday/Wednesday: 9:30 AM to 10:45 AM
Math Sciences 5203

Richard Berk
206-9544
berk@stat.ucla.edu

This is not your parents' applied regression course. Since the first major textbook on regression, written in the 1950's, most treatments have pretty much covered the same ground. What has varied is the applications used to illustrate the material and some of the ways the material has been explained. This is especially true of textbooks written for particular academic disciplines such as economics, engineering, sociology, or public health. Meanwhile, statisticians have been working away so that now "modern regression analysis" looks quite different from conventional methods. In this course, we will cover modern regression analysis. It is both far more powerful and far more accessible than conventional regression analysis. It is also far more cautious.

Over the next two quarters, we will cover the generalized linear model (e.g., linear regression, analysis of variance, logistic regression, Poisson regression, etc.), along with regression smoothers and regression diagnostics. Near the end of the course we will touch on statistical inference using the bootstrap and related techniques and then spend several weeks on sampling. In each case, the emphasis will be on applications.

This is a transition year for the course. The last several years, we have used as our text *Applied Regression Including Computing and Graphics*, R. Dennis Cook and Sanford Weisberg, John Wiley, 1999. The book comes with free software (*ARC*) that you download. Documentation for this software can be found in Appendix A of the text along with instructions on how to obtain it. There are Mac and Wintel version of the software. However, this year Cook and Weisberg is only suggested reading, although it will be essential for students who want to use their software.

This year we are making two changes. First, the preferred software is now R rather than ARC. R is also shareware and is an excellent programming language very similar to S-Plus. And there are a very large number of statistical procedures that can be easily called and used. The reference for R is *Introductory Statistics with R* by Peter Dalgaard (Springer- Ver-

lag, 2002). You can obtain a version of R and supporting documentation from the web. From the Statistics home page (www.stat.ucla.edu) go to "Services and Archives" and then to the "UCLA CRAN Mirror." You will want to download the precompiled binary distribution appropriate for your platform. Also download from the manual at least "An Introduction to R" and "R Installation and Administration." Installation should go easily and smoothly, but allow for Murphy's Law too. Get on this right away and work through Chapters 1-4 in Dalgaard. This material will get you into R and review some elementary statistics.

I plan to still draw heavily on material in Cook and Weisberg. But my notes, available over the web, will cover the essentials, and the new text is my own *Regression Analysis: A Constructive Critique* (Sage Publications, 2003). This change simply provides easy access to material I was already covering in the course in a less systematic manner.

There is a web page for the course. Go to www.stat.ucla.edu, click the icon for "Teaching" and then "Courses" and from there "Stat 217A." There you will find a pull-down menu. You should now find sitting there this course outline under "handouts." There will be class notes over time as well under "handouts." Later you will be able to download the midterm and other assignments.

Grading will be based on a midterm and a final paper using data I provide. You will be asked to write a "research report" based on your analysis of the data. Those reports will generally be between 10 and 20 pages of text (tables and figures extra). You will use R or ARC for all those analyses. (Other software will not have the procedures you need.) The aim is to make this a "hands-on" course, and that will be reflected in the content of the paper. There will also be ungraded homework discussed in class.

The best way to reach me is with e-mail.