Sociology 604 – Regression Analysis
Prof. Stephen Plank
Spring 2011

Monday (10-10:50am) and Friday (10-11:50am)

This course provides an accessible but in-depth coverage of multiple regression (least squares and alternative estimation procedures) with a focus on sociological problems and software applications. The course is intended for graduate students in the social sciences, though well-prepared undergraduates are also welcome with Professor Plank's permission.

Students should have had (at least) a one-semester introduction to statistics. Some matrix algebra will be used in the course, but reviews or tutorials will be provided as needed. Calculus will be used very little or not at all. Probability theory, including properties of distributions and random variables, will be discussed in some depth.

The course grade will be determined as follows:

Homework sets	*55%*
Final project (take-home exam)	*35%*
Participation	*10%*

Substantive lectures on statistical concepts and the craft of quantitative sociological analysis will generally begin during the Friday session and conclude on Monday. Parts of each session will also be dedicated to instruction in Stata, with both Steve and the students taking the keyboard and offering examples and insights.

Homework will be assigned most weeks, generally to be collected at the start of the Friday session unless otherwise announced. Every time a homework set is due, the next homework set will be distributed (which will help you see where we're going next, and help you manage your time).

The final project will require a comprehensive analysis of one dataset, with accompanying narrative and interpretation. It should be approximately 15-20 pages of "journal-style" text, tables, and figures – plus any needed appendices or supporting materials.

Required text:

Fox, John. (2008). *Applied Regression Analysis and Generalized Linear Models (Second Edition).* Sage Publications.

(See book description at http://socserv.socsci.mcmaster.ca/jfox/Books/Applied-Regression-2E/index.html)

Optional (supplemental) text:

Hamilton, Lawrence C. (2005). *Statistics with Stata (Version 9).* Duxbury Press.
(Consider also Hamilton's edition for Stata Version 10, published in 2008).

Course schedule:

Week One Jan. 31 & Feb. 4		**For Friday, read Chs. 1 & 2.** Introductory lecture and surveying students' background knowledge and learning goals. Lecture on Chs. 1 & 2) Statistics and social science. What is regression analysis? * statistical models and social reality * observation and experiment * populations and samples * naive nonparametric regression * local averaging Using Stata, as well as online datasets and examples.

Week Two
 	
Feb. 11
		Homework from Ch. 2 due on Friday.
		
		***no class on Feb. 7 – Steve out of town ***
		
		Lecture on Ch. 3) Examining data.
		* univariate displays
		  * histograms
		  * nonparametric density estimation
		  * quantile-comparison plots
		  * boxplots
		* plotting bivariate data
		* plotting multivariate data

Week Three Feb. 14 & 18		**For Monday, read Ch. 3.** **Homework from Ch. 3 due on Friday.** Lecture on Ch. 4) Transforming data. * the family of powers and roots * transforming skewness * transforming nonlinearity * transforming nonconstant spread * transforming proportions

Week Four
 	
Feb. 21 & 25
		For Monday, read Ch. 4.
		Homework from Ch. 4 due Friday.
		
		Lecture on Ch. 5) Linear least-squares regression
		* simple regression
		  * least-squares fit
		  * simple correlation
		* multiple regression
		  * two explanatory variables
		  * several explanatory variables
		  * multiple correlation
		  * standardized regression coefficients

Week Five
 	
Feb. 28 & March 4
		For Monday, read Ch. 5.
		Homework from Ch. 5 due Friday.
		
		Lecture on Ch. 6) Statistical inference for regression
		* simple regression
		  * the simple-regression model
		  * properties of the least-squares estimator
		  * confidence intervals and hypothesis tests
		* multiple regression
		  * the multiple-regression model
		  * confidence intervals and hypothesis tests
		* empirical versus structural relations
		* measurement error in explanatory variables

Week Six
 	
March 7 & 11
		For Monday, read Ch. 6.
		Homework from Ch. 6 due Friday.
		
		Lecture on Ch. 7) Dummy-variable regression
		* a dichotomous factor
		* polytomous factors
		* modeling interactions
		  * constructing interaction regressors
		  * the principle of marginality
		  * interactions with polytomous factors
		  * interpreting dummy-regression models with interactions
		  * hypothesis tests for main effects and interactions
		* a caution concerning standardized coefficients
		
		** ALSO, discussion of Analysis of Variance.

Week Seven March 14 & 18		**For Monday, read Ch. 7.** **Homework from Ch. 7 due Friday.** Ch. 11) Diagnostics I: Unusual and influential data * outliers, leverage, and influence * assessing leverage: hat-values * detecting outliers: studentized residuals * measuring influence * numerical cutoffs for diagnostic statistics * joint influence * should unusual data be discarded?
SPRING BREAK		March 21-25

Week Eight
 	
March 28 & April 1
		For Monday, read Ch. 11.
		Homework from Ch. 11 due Friday.
		
		Ch. 12) Diagnostics II: Non-normality, nonconstant error variance, and nonlinearity
		* non-normally distributed errors
		* non-constant error variance
		  * residual plots
		  * weighted-least-squares estimation
		  * correcting OLS standard errors for non-constant variance
		  * how non-constant error variance affects the OLS estimator
		* nonlinearity
		  * component-plus-residual plots
		  * when do component-plus-residual plots work?
		* discrete data
		  * testing for non-linearity

Week Eight (cont'd)		* testing for non-constant error variance * maximum-likelihood methods * Box-Cox transformation of Y * Box-Tidwell transformation of the X's * Non-constant error variance revisited
Week Nine April 4 & 8		**For Monday, read Ch. 12.** **Homework from Ch. 12 due Friday.** Ch. 13) Diagnostics III: Collinearity and its purported remedies * detecting collinearity * principal components * generalized variance inflation * coping with collinearity: no quick fixes * model respecification * variable selection * biased estimation * prior information about the regression coefficiencts * some comparisons
Week Ten April 11 & 15		**For Monday, read Ch. 13.** **Homework from Ch. 13 due Friday.** Pausing to review. Pausing to discuss students' final projects. **Any reading, or memos on final projects, *to be determined*.**
Week Eleven April 18 April 22		Logit and probit models for categorical response variables. * Drawing upon Ch. 14 and other materials. Time-series regression and generalized least squares. * Drawing upon Ch. 16 and other materials. **Any reading, or memos on final projects, *to be determined*.**
Week Twelve April 25 & 29		Missing Data. * Drawing upon Ch. 20 and other materials. **No homework due this week; be working on final project.**
Week Thirteen May 2 & 6		Maximum likelihood estimation, and generalized linear models as a comprehensive framework for modeling. **No homework due this week; be working on final project.**
May 18		**Final Project Due.**