

**The University of Calgary
Faculty of Social Sciences
Department of Sociology**

**Sociology 611.01
Social Statistics: The General Linear Model
Fall, 2008
Course Outline**

Instructor: Dr. Richard Wanner
Office: Social Sciences 1058
Office Hours: Mon., Wed. 11-12:00
or by appointment
E-mail: wanner@ucalgary.ca

Teaching Assistant: Annette Tézli
Office: SS 1005
Office Hours: Tues. 10-12:00 in SS 1203
or by appointment
E-mail: atezli@ucalgary.ca

Course Objectives:

Multiple regression provides the prototype for modeling data in which a single continuous dependent variable is predicted from several continuous independent variables plus an error term that is assumed to be normally distributed. The general linear model extends this basic model in several ways, while retaining a continuous dependent variable. This course is designed to familiarize you with the basic properties of linear models, extend your understanding of multiple regression and correlation and their application in testing theoretically-derived models, and enable you to estimate models when either independent or dependent variables are categorical, when the underlying form of the relationship among the variables is nonlinear, or when more than a single equation is required to capture the causal structure. At the same time, you will learn to use the statistical software package Stata to estimate your models.

Prerequisites:

This course assumes that you have completed one or more undergraduate courses in statistics that provide you with a fairly complete understanding of inferential statistics (null hypothesis testing and confidence intervals) and at least an introduction to cross-tabular techniques, bivariate correlation and regression, and basic analysis of variance.

Required Texts:

Pedhazur, E.J. 1997. *Multiple Regression in Behavioral Research* (Third Edition).
Toronto: Harcourt Brace College Publishers.

Hardy, Melissa A. 1993. *Regression with Dummy Variables*. Thousand Oaks, Calif.:
Sage.

Pampel, Fred C. 2000. *Logistic Regression: A Primer*. Thousand Oaks, Calif.: Sage.

Fox, John. 1991. *Regression Diagnostics*. Thousand Oaks, Calif.: Sage.

Chen, Xiao, Philip B. Ender, Michael Mitchell, and Christine Wells 2006. *Regression with Stata*. www.ats.ucla.edu/stat/stata/webbooks/reg/default.htm.

Additional readings on using Stata and on topics not covered in the required texts will be distributed and the cost covered by the lab fee (see below).

Optional Text:

Frankfort-Nachmias, Chava and Anna Leon-Guerrero. 2006. *Social Statistics for a Diverse Society (4th Edition)*. Thousand Oaks, CA: Pine Forge Press.

This book is currently used in Sociology 311 and 315 and does a good job of reviewing the basic undergraduate curriculum in social statistics, covering all the material mentioned as prerequisites for this course. This would be a useful purchase for those students who have not studied or used statistics for many years, or whose undergraduate preparation was inadequate. The StatSoft web site (see below) is also a good resource for reviewing basic material.

A Note on Software

The main statistical software package to be used in this course is Stata. It is rapidly becoming the standard for sociologists, economists, and other social scientists, as well as biomedical researchers, who are doing serious data analysis. It's main virtue is that, because it is programmable, much of it is designed by users themselves. This means that state-of-the-art statistical procedures are generally available first in Stata. While SPSS, SAS, or R can do all of the procedures discussed in this course, they will not be supported in the lectures or labs. For full credit, all lab exercises must include Stata syntax and output.

Some Web Resources

The following are web pages that have substantial amounts of information about Stata and statistics in general. The UCLA and University of North Carolina pages in particular are important tools for learning to use Stata, but the Stata Corp. page also has some excellent resources. The StatSoft web page provides a virtual stats text that is useful for reviewing the basics of statistics and for reference purposes.

www.stata.com

www.ats.ucla.edu/stat/stata/

www.cpc.unc.edu/services/computer/presentations/statatutorial

opr.princeton.edu/computing/software/stata/

www.statsoft.com/textbook/stathome.html

Exercises:

Each unit of the course will be accompanied by either a review exercise involving some hand calculation or a computer exercise applying the Stata program package, available on the Social Sciences server. It is important that you make a serious effort at completing these exercises, since the exams often include questions requiring you to interpret computer output from Stata. The computer exercises will be worth 15 percent of your final mark. These exercises will be "pass-fail", with a pass awarded if all questions are attempted. Incomplete exercises will be awarded a "fail." Specific due dates for the exercises are listed on the lab calendar below. None will be accepted late. Under this system, you will receive an 'A' for 15

percent of your mark by simply submitting the completed exercises on time. Two percent of the final mark will be deducted for each exercise not received on time.

Please note that you will need a basic calculator (i.e. one capable of computing square roots and logarithms) for several of the exercises and exam questions. For the exercises, you could also use the calculator function or a spread sheet (e.g. Excel) on your computer.

Note: A \$20 lab fee, payable to the Department of Sociology, will be charged to cover costs related to copying, printing, and other supplies required for the completion of lab assignments. Your final mark will not be submitted until this fee is paid.

Data Analysis Project

A portion of your grade will be based on a written paper in journal format describing a secondary data analysis involving regression methods of a data set provided to you. Since this is a practice study, you will have some choice of topic, but the format will be prescribed. You will select a topic for the data analysis on the basis of your reading of at least one journal article, book, or book chapter in the field, i.e. the article should motivate the project. Details are available on the project assignment on the course CD.

Grading:

Your mark for the course will be based upon the computer exercises, three "open-note" examinations, and the data analysis project. Marks will be assigned as follows:

First Exam	Friday, October 24 in SS921	20%
Second Exam	Tuesday, November 25 in SS921	20%
Third Exam	Tuesday, December 9 in SS921	20%
Computer Exercises	Due weekly	15%*
Data Analysis Project	Proposal due Nov. 7	25%
	Final report Due December 16	

*See lab calendar below for details

Tentative Calendar:

Week	Topics, Readings, and Labs
1-2	Administration and description of course Models, model building, and model testing Bivariate correlation and regression <i>Reading:</i> Pedhazur, pp. 1-14; 15-41. <i>Lab:</i> Introduction to Stata, exercises on bivariate correlation and regression <i>Lab reading:</i> Long and Freese pp. 15-73; Chen et al., Chapter 1

Week Topics, Readings, and Labs

- 3-5 Multiple regression and correlation
Confidence intervals and null hypothesis tests
Reading: Pedhazur, pp. 95-111, 156-165; 170-183
Lab: Review exercise and computer exercise on multiple regression and correlation
Lab reading: Annotated regression output at www.ats.ucla.edu/stat/stata/output/reg_output.htm

First Examination

(Tentative date: Friday, October 24)

- 6 Basic matrix concepts and operations
Multiple regression: the matrix approach
Reading: Pedhazur, pp. 983-992 (Appendix A), 135-154
Lab: Computer exercise on matrix algebra and its application in regression analysis
- 7 Regression with categorical independent variables:
 Analysis of variance
 Dummy vs. effect coding
Reading: Hardy pp. 1-26.
Lab: Computer exercise on dummy variable regression
Lab reading: Chen et al. Chapter 3
- 8 Regression with categorical and continuous independent variables:
 Analysis of covariance
 Contextual effects
 Cross-level inference
Reading: Hardy pp. 26-75;
Lab: Computer exercise on analysis of covariance
- 9 Models with a categorical dependent variable: logistic regression
Reading: Pampel pp. 1-54; 74-82
Lab: Computer exercise on logistic regression

Second Exam

(Tentative date: Tuesday, November 25)

- 10 Issues in multiple regression:
 Explanation vs. prediction
 Shrinkage and the adjusted R^2
 The specification problem
 Multicollinearity
 Measurement error
Reading: Pedhazur, pp. 195-198, 207-209, 288-294; Fox, pp. 10-21
Lab: Exercise on above topics; **project proposal due Nov. 7**

Week	Topics, Readings, and Labs
11-12	Issues in multiple regression (continued): Residuals and detecting violations of assumptions Standardized vs. unstandardized coefficients Nonlinearity Homoscedasticity and autocorrelation Discrete dependent variables Robust standard errors <i>Reading:</i> Pedhazur, pp. 319-325, 520-522; Fox, pp. 21-66. <i>Stata User's Guide</i> , 23.11 Obtaining robust variance estimates <i>Lab:</i> Computer exercise on nonlinearity, violating assumptions <i>Lab reading:</i> Chen et al. Chapter 2
13	Introduction to structural equation models: Path analysis <i>Reading:</i> Pedhazur, pp. 765-807 <i>Lab:</i> Computer exercise on path analysis

Third Exam

(Tentative date: December 10)

Tentative Lab Calendar

The following are the important dates for the lab component of the course, though these dates are subject to change if this becomes necessary for pedagogical reasons. Nevertheless, this should give you the ability to gauge your workload at different periods during the semester.

Sept. 12 – First lab: Stata introduction; exercise 1
Sept. 19 – Exercise 1 due; exercise 2
Sept. 26 – Exercise 1 returned; exercise 2 due; exercise 3
Oct. 3 – Exercise 2 returned; exercise 3 due; exercise 4
Oct. 10 – Exercise 3 returned; exercise 4 due; exercise 5
Oct. 17 – Exercise 4 returned ; exercise 5 due; exercises 6 and 7; discuss project proposal
Oct. 24 – First exam (SS921); exercise 6 due
Oct. 31 – Exercise 6 returned; exercise 7 due; exercise 8
Nov. 7 – Exercise 8 due; exercise 7 returned; exercise 9; project proposal due
Nov. 14 – Exercise 8 returned; exercise 9 due; exercise 10
Nov. 21 – Exercise 9 returned; exercise 10 due
Nov. 25 – Second exam (SS921) in lecture
Nov. 28 – Lecture
Dec. 2 – Lecture during hours 1-2; Exercise 11 during hour 3
Dec. 5 – Exercise 11 due; discuss project, review last exam topics

Important Conventions in this Course:

- a. Except in circumstances of illness, domestic affliction, or religious conviction that are fully documented (see U. of C. Calendar, p. 46), late computer assignments will not be accepted.

- b. Intellectual honesty is the cornerstone of the development and acquisition of knowledge and requires that the contribution of others be acknowledged. As a result, cheating or plagiarism on any assignment or examination are regarded as extremely serious academic offenses. Students are advised to consult pp. 18-21 of the Faculty of Graduate Studies Calendar, which present a Statement of Intellectual Honesty and definitions and penalties associated with plagiarism, cheating, and other academic misconduct.

- c. A student who is unable to write one of the mid-term examinations for a fully documented reason listed in (a) above must discuss an alternative course of action with the instructor. At his discretion, the instructor will transfer the percentage weight to another component of the course.

- d. Examinations will be "open-note" exams, meaning that you may bring into the exams any notes you have taken on course materials (formulas, definitions, etc.), but not the course materials themselves.