

# Data II: Applied Regression Analysis

SOCY 5021

Spring 2007

## Course Description

Time:	Mon 3:00-5:50pm
Location:	Ketchum 33
Instructor:	Ying Lu
Office:	Ketchum 3E
Office hours:	Mon/Wed 11:00am-12:30pm
Email:	Ying.Lu@colorado.edu

Linear model(linear regression) is one of the most useful tools in applied statistics. The goals of this course are two fold: 1) to provide students an in-depth examination of the linear model, including both topics in simple linear regression and multiple regression, 2) to equip students with practical skills to apply these methods to analyze social science data and to interpret the results. The course will begin with a review of basic statistics and an exposition on the probability theory to prepare students for a better understanding of the linear model. We will then study the simple linear regression model and expand it to allow regression on multiple covariates (multiple regression). We will take a critical perspective on the linear model, spending considerable time evaluating model fit, the influence of outliers, the assumptions of regression, and the consequences of their violation. We will also spend considerable time expanding the model to allow inclusion of nominal and ordinal level covariates, nonlinear relationships, and statistical interactions. Toward the end of the semester, we will begin to examine variations and extensions of the linear model including the models for discrete response(logistic, multinomial and poisson regressions), analysis of two way tables and the analysis of variance.

## Prerequisites

This course is designed for graduate students who have taken SOCY 4061 Social Statistics or other quantitative classes at equivalent level. I will assume that students in the class have solid understanding of the basic statistical concepts and are reasonably comfortable with STATA. If you need to review some of the basic materials, you can refer to books such as,

- McClelland, G. (1999). *Seeing Statistics*. Duxbury Press. The online version [www.seeingstatistics.com/](http://www.seeingstatistics.com/) is available through any CU IP address, or via VPN.
- Hamilton, L. C. (2006). *Statistics with STATA*. Belmont CA: Duxbury.
- *Getting Started with STATA..??*

## Textbooks

It is not easy to find a comprehensive textbook that comes from a truly applied perspective yet covers enough statistical theory of regression analysis at an approachable level of mathematics. For this class, the main source of reference will be the lecture notes which I will make them available on line before each class. Together with the lecture notes, the following three textbooks at different levels(more applied--> more theoretical) are suggested. The [www.amazon.com](http://www.amazon.com) prices are listed. I highly recommend you obtain at least a copy of Allison, and some version of Moore & McCabe.

1. Allison, P. D. (1999). *Multiple Regression: A Primer*. Thousand Oaks, CA: Pine Forge Press. (\$33.99)
2. Moore, D. and McCabe George. (2004). *Introduction to the Practice of Statistics*. 5th Edition, Freeman. (\$84.48, a little pricy but it is a very good desk reference! and there is \$9.95 Cram 101 paperback version of it, also you can get an earlier edition for cheaper price.)
3. Fox, J. (1997). *Applied Regression Analysis, Linear Models, and Related Methods*. Thousand Oaks, CA: Sage. (\$94.30, if you want to do a dissertation with me, get a copy!)

I would also recommend the following very affordable publications by Sage where you can read more about applied regression methods.

*to be added*

## Evaluation

Final grades will be calculated based on the following system:

1. homework. there will be 6 biweekly homework, some of them are data analysis projects using STATA. 30%
2. two closed book mid-term tests(1 hour each). 20% each.
3. one closed book final exam(3 hours). 30%
4. class attendance. It is monitored by occasional popup quizzes. 10% (The grades will not be counted toward your final grade, but if you miss some of them, these 10% will be in jeopardy. And not being able to answer the questions should remind you to spend more time on this class.)

## Auditing Policies

Auditing is in general not encouraged. If you do choose to audit, you will still have to do all the readings and homework(but homework will not be graded, I will try to provide the answer sheet in time). For students who audit this class in preparation of the comprehensive exam, you should pair up to grade each other's work and tests.

## Computation

In this course, we use statistical software, called **STATA**. You should obtain a copy of

- Hamilton, L. C. (2006). *Statistics with STATA*. Belmont CA: Duxbury. (\$65.95, an earlier version is acceptable)

unless you are an extremely competent user of STATA (which I don't think I am one.)

To learn more about STATA, there is a complete set of STATA manual located in 3 Ketchum, but please do not remove them. You can also look for some online help such as joining the STATA listserver <http://www.stata.com/statalist/>

## Tentative course plan

This is just a tentative course plan with evolving reading assignments...in particular, I will add some papers from sociological journal applying these methods.

day	topics	readings
01/22	review of prob. and stat theory	
01/29*	review II, <i>lab/STATA tutorial</i>	
02/05	simple (bivariate) regression—concepts and interpretations	
02/12*	simple regression—estimation, <i>lab</i>	
02/19	multiple regression—concepts, assumptions and interpretations	
02/26*	multiple regression—estimation, <i>lab</i>	
<b>03/05</b>	test 1, multiple regression—dummy variable, nonlinearity and transformation	
03/12	<i>review test 1</i> , statistical interactions and interpretation, <i>lab</i>	
03/19*	inference, assessing model fit, model selection	
03/26	<b>SPRING BREAK!!</b>	
04/02*	multi-collinearity, outliers, heteroscedascity and missing data, <i>lab</i>	
<b>04/09</b>	test 2, analysis of tables	
04/16	<i>review test 2</i> , ANOVA	
04/23*	primer: regressions for binary, ordinal and counts data, <i>lab</i>	
04/20	primer: regression with latent variable	

**00/00** are the test dates, each \* indicates a homework assignment.