

## STATISTICAL ANALYSIS IN EDUCATIONAL RESEARCH

### EDUC 250A

Tentative Syllabus, Winter, 2005

*(revised 5 January, 2005)*

professor:

**sean f. reardon**

office: 315 cubberley

office hours: by appointment

phone: 736-8517

email: sean.reardon@stanford.edu

MWF 11:00-12:30

class: 108 CERAS

lab: 108 CERAS

teaching assistants:

**liz grant**

office: 229 CERAS

office hours: W 12:30-2:30

email: lizgrant@stanford.edu

**katharine strunk**

office: 225 CERAS

office hours: Th 11:00-1:00

email: kos@stanford.edu

---

### Overview

Statistical analysis can be a powerful tool in understanding social, educational, psychological, and developmental processes. In cases where it is impossible or impractical to collect data on every individual, classroom, teacher, and school of interest, statistical analysis allows us to examine data on a sample of individuals (or classrooms, schools, etc.) in order to infer patterns in a larger population. For example, we might want to examine data on achievement test scores and per-pupil spending for a sample of schools to determine whether there is an association between spending and achievement patterns in the population. Or we might want to examine the association between race/ethnicity and achievement patterns. Moreover, if we find such an association, we might wish to ask additional questions, such as whether race/ethnic differences in achievement patterns can be accounted for by race/ethnic differences in family socioeconomic characteristics or in school quality.

In this course we will learn to answer such questions using regression analysis—a statistical tool that allows us 1) to describe average patterns of association among multiple variables observed in a sample and 2) to make inferences about the patterns of association among these variables in a population. Regression analysis is a powerful statistical method with many variations. Our goal in this course is to develop an understanding of the basic methods, including their limitations, and to develop skill in using regression analysis to answer educational research questions. Finally, because an important part of any analysis is communicating the results to an audience, we will also place considerable emphasis on learning to present (in writing, tables, and figures) the results of regression analyses.

By the end of the quarter, students in this course should be sufficiently skilled in regression analyses that they can critically examine published research using regression and can carefully perform their own analyses.

This is a graduate-level statistics course, so there will be plenty of math. It is a course in the social science application of statistics, however, not a theoretical statistics course. Therefore our math will be in the service of defining statistical models that correspond to the research questions we want to answer, and we will pay little attention to the mathematics by which the programs estimate the parameters in our models. Our concern is with understanding how to use these methods to do good research and with learning to interpret the results they provide.

The course will put a lot of emphasis on conducting statistical analyses using sample data and on communicating the results of these analyses to a general audience of scholars. The software program that we will use for statistical analyses is called Stata (version 8), a general statistics software package similar in capacity to SPSS and SAS. If you have not used Stata before, there are a number of resources available on the web to help you learn (see the document on the course Blackboard site for links).

Assumed background. This course is an introductory/intermediate level applied statistics course, pitched to the level of students who have completed an introductory statistics course, such as EDUC 160. To be prepared for taking this course you should be familiar with basic statistical concepts such as populations, sampling, means, variances, estimates, and statistical inference. A prior familiarity with simple (bivariate) regression is helpful, but we will cover that again in this course. The course relies heavily on algebra; no more sophisticated mathematics are necessary for our purposes. More advanced statistics courses will typically rely on a familiarity with basic matrix algebra; however, we will not use matrix algebra in this course.

Class format. The class will meet in 108 CERAS MWF 11:00-12:30. Two days a week (Monday and Wednesday generally) will consist of lectures on the use of regression analysis to address educational research questions. One day a week (usually Friday) will be a review/lab session run by the teaching assistants. Students are expected to attend all classes, as some supplemental material will be covered on Fridays that is not covered in lectures. The class may be relatively large, but I strongly urge you to ask questions in class. If you have a question, it is likely that others do as well.

---

### **Textbooks:**

There is no required text for this course. However, I have asked the bookstore to stock three textbooks for this course that you may wish to purchase; these books will be useful as a supplement to the class lectures and as reference sources for the future. You need not buy any of the books, and you should definitely not buy all three textbooks—pick only one at most.

The textbooks are written at different levels of mathematical and statistical complexity, so you should choose a text that you find readable. If you think this may be the most advanced statistics course you take, or if you find it difficult to read a textbook packed with equations, then you might

choose either the Chatterjee, Hadi, and Price (CHP) text or the Kleinbaum, Kupper, Muller, and Nizam (KKMN) text. Neither of these rely on matrix algebra. The Kutner, Nachsteim, and Neter (KNN) text is an excellent advanced text; this text relies heavily on matrix algebra (but has a nice chapter reviewing matrix algebra for those who are rusty). This is a text that will be a valuable resource in the future to students who plan to continue on in more advanced applied statistics courses.

Chatterjee, S., Hadi, A, & Price, B. (2000). *Regression Analysis by Example* (3rd Ed.). Wiley.

Kleinbaum, D, Kupper, L., Muller, K, & Nizam, A. (1998). *Applied Regression Analysis and Multivariable Methods* (3rd Ed.). Duxbury Press.

Kutner, M., Nachsteim, C.J., & Neter, J. (2004). *Applied Linear Regression Models* (4th Ed.). McGraw-Hill.

In addition, lecture note, handouts, data sets, and other course materials will be posted on the Blackboard site for this course.

### Assignments and Grades

Grading will be based on 4 homework assignments (15 points each) and one 40-point final exam/assignment, for a total of 100 available points. Late assignments will lose one point per day late, except in the case of a demonstrated personal emergency. Problem sets will be due the following dates:

Problem Set #1	distributed January 12	due January 19, 2005
Problem Set #2	distributed January 31	due February 7, 2005
Problem Set #3	distributed February 12	due February 23, 2005
Problem Set #4	distributed February 28	due March 7, 2005
Final Exam	distributed March 7	due March 16, 2005

Students have the option of taking this course for a grade or for credit/no credit. As noted in the SUSE graduate student handbook, “students are strongly encouraged to ask for grades in all courses where they are available.” For students who take the course for a grade, the grading scale is as follows:

A+	97 to 100
A	93 to 96
A-	90 to 92
B+	86 to 89
B	83 to 85
B-	80 to 82
C+	75 to 79
C	70 to 74
F	0 to 69

A grade of C is required for students who take the course for credit/no credit to receive credit for the course.

Getting help. In addition to the Friday review/lab sessions, each teaching assistant will hold office hours two hours a week.

### Academic Integrity:

Academic integrity is the pursuit of scholarly activity free from fraud and deception and is an educational objective of this institution. Academic dishonesty includes, but is not limited to, cheating, plagiarizing, fabricating information, facilitating acts of academic dishonesty by others, and submitting the work of another person or work previously used without informing the instructor. The Honor Code, outlining the general expectations pertaining to Academic Integrity applicable to this course are published in the Graduate Student Handbook available at:

<http://honorcode.stanford.edu>. Students are expected to conform to the highest standards of academic integrity in this course — meaning, essentially, don't lie; don't cheat; don't pass off someone else's work as your own.

It is, nonetheless, acceptable (and encouraged) to work together on problem sets and the final exam. Answers to the problem sets, however, should be written entirely on your own and in your own words. Do not submit assignments that duplicate word-for-word or paraphrase another student's assignment.

### Approximate Course Schedule

Topic	Approximate Dates	Suggested Supplemental Reading		
		CHP	KKMN	KNN
1 Overview, Review of Descriptive Statistics	1/5	ch. 1	ch. 1-3	---
2 Simple Linear Regression: The Basic Foundation	1/10-1/14	ch. 2 ch. 4.2	ch. 4, 5.1- 5.5, 6, 7	ch. 1
3 Simple Linear Regression: Inference and Hypothesis Testing	1/19-1/24	ch. 2	ch. 5.6-5.10, ch. 6.	ch. 2 ch. 4
4 Simple Linear Regression: Detection and Correction of Model Violations	1/26-1/31	ch. 4, 6, 7	ch. 12	ch. 3 ch. 11.1
5 Multiple Regression: Extending the Basic Model	2/2-2/11	ch. 3.	ch. 8, 9	ch. 5 ch. 6
6 Multiple Regression: Multicollinearity and Relationships Among Variables	2/14-2/18	ch. 9	ch 10, 11.4	ch. 7.6, 10.5, 11
7 Multiple Regression: Categorical Variables and Interactions	2/23-3/4	ch. 5	ch. 11, 14	ch. 8
8 Logistic Regression	3/7-3/11	ch. 12	ch. 23	ch. 14