

**SPEA V507 - Spring 2012**

MW, 2:30-3:45; SPEA 277

Class Number: 19496

Lab: M, 7:00-8:30 pm, SPEA 277/LIB 402

TA: Collin Drat, SPEA 412, 855-8963

TA Office Hrs: TR 12:00-2:00, F 2:00-4:00/5:00

TA E-Mail and Cell: [cdrat@umail.iu.edu](mailto:cdrat@umail.iu.edu), (309) 531-1087

Madison Busker's Office Hrs: TR 10:00-12:00, F 11:00-1:00/2:00

**Instructor: Barry Rubin**

Office: SPEA 329

Phone: (SPEA) 855-4556

(Cell) 327-3852

E-Mail: [rubin@indiana.edu](mailto:rubin@indiana.edu)

Office Hrs: MW 1:00-2:15,

or by appointment

## Data Analysis and Modeling for Public Affairs

### Course Description and Objectives

V507 provides students of public and environmental affairs and related disciplines with a detailed, intermediate-level perspective on statistical concepts and techniques for analyzing and modeling complex systems. The course content includes estimating the parameters of such models based on existing data, testing hypotheses about these systems, and forecasting. The context of the course is the application of these techniques to problems and policies in public and environmental affairs. Multivariate regression analysis is one of the primary tools for statistical modeling for purposes of policy analysis, program evaluation, simulation of systems, and general forecasting. Thus, most of the course is devoted to single equation regression models and the extension of these models to a variety of situations. A prerequisite for the class is a graduate-level, introductory statistics course that includes coverage of the simple (two-variable) regression model and an introduction to multivariate regression.

The primary objectives for the course are to enable students to:

- understand how regression techniques can be used for data analysis and applied problem solving;
- apply regression techniques to such problems using the appropriate technology;
- understand the assumptions required to use regression analysis, the impact of violating these assumptions, and how to correct for such violations;
- model and forecast the behavior of systems using regression tools;
- comprehend and critically analyze the results of research employing these tools as presented in academic/professional journals, public/non-profit sector reports, and policy analyses; and
- explore several alternative multivariate analysis and forecasting techniques.

The major topics that will be covered during the semester are:

- the general (multivariate) linear model;
- summary statistics, hypothesis testing, and implicitly linear models;
- multicollinearity and the use of dummy variables;
- violations of the general linear model assumptions and errors of specification;
- logit models; and
- an introduction to simultaneous equation systems.

A lecture and dialog format, along with in-class exercises and demonstrations, will be used for the Monday and Wednesday class sessions. There is also a lab session scheduled for 7:00 to 8:30 p.m. on Mondays in SPEA 277 or LIB 402, beginning the first week of class. The lab sessions will be used for SAS workshops, administration of exams, distribution and explanation of graded homework exercises and exams, current homework help sessions, review of general concepts, and pre-exam review sessions. Other than the SAS workshops, exams, and project presentation sessions, attendance at these lab sessions will not be required (although regular attendance is advised for most students.) The lab sessions will generally be conducted by the TA.

## Course Requirements and Grading Criteria

There will be two examinations, each equally weighted. These exams are scheduled for **February 27** and **May 2**. Each exam will combine multiple choice questions, short answer essays, and problems. The exams will be comprehensive to the extent that concepts covered in previous sections of the course are required to understand those covered later.

Six homework exercises will be assigned during the semester. These will utilize computer software (primarily SAS Version 9.2) to apply the regression techniques covered in the lecture and readings. Students are required to work in groups of two on these homework assignments, with assignments submitted by both students collectively. However, each student should be able to successfully utilize the problem-solving approaches used on the homework assignments. Late homework assignments will be accepted with a 10 percent penalty per day late, up to three days (including weekends). Assignments are due in the TA's mailbox by 5:00 p.m. on the date specified. It is strongly encouraged that students complete homework assignments ahead of the due date to allow time for proofreading. Homework is expected to be word-processed and of professional, academic quality.

A major project will be required which applies multivariate regression analysis to a data set that is identified and compiled by students. Detailed information on the project and the associated presentation will be provided in early March. Students are required to work in groups of three on this project.

Students must be familiar with desktop computing software, including spreadsheet software, email, and Oncourse. Email will serve as an important means of communication for the class and Oncourse will be used for posting class information and data sets. Each student must also have a USB drive with at least 1 Gb capacity for storing class materials.

The following grading criteria will be used:

Midterm Exam	20%
Final Exam	20%
Data Set Analysis Project/Presentation	25%
Homework Exercises	30%
Class Discussion/Participation	5%

## Text, Lecture Notes, and Reading Assignments

The required text for the class is:

Damodar Gujarati and Dawn Porter, 5th ed., *Basic Econometrics*, (New York: McGraw-Hill/Irwin, 2009).

Students will also need to purchase a copy of the lecture notes for the class from the IU or TIS Bookstore. These notes should be brought to class every day. Additional readings identified in the syllabus will be on available via Oncourse.

Note that the course syllabus is somewhat flexible. The exact dates on which topics will be covered depend on progress made throughout the semester. As a result, these dates may be altered.

**SPEA V507: DATA ANALYSIS AND MODELING  
TOPICS AND READING ASSIGNMENTS - SPRING 2012**

**Date   Topic and Reading Assignments**

**SECTION 1: The General (Multivariate) Linear Model**

Jan 9   Introduction to the Course

Readings:   G&P, scan\review pp. 34-48, 55-85, 97-102, and 107-134.

Jan 9   *Lab:* Introductory SAS Workshop, LIB 402, (required for students  
who have no SAS experience)

Jan 11   The General Linear Model and its Assumptions

Readings:   G&P, pp. 188-196.

Jan 16   No Class (Martin Luther King Day)

Jan 16   *Lab:* None

Jan 18   Estimation and Strength of Relationship of the GLM

Readings:   G&P, pp. 196-207.

**Exercise 1 Assigned (Due Fri, Jan. 27)**

Jan 23   Hypothesis Testing

Readings:   G&P, pp. 233-243.

Jan 23   *Lab:* Exercise 1 Help Session, LIB 402

Jan 25   Hypothesis Testing  
Implicitly Linear Models

Readings:   G&P, pp. 243-259 and pp. 159-173.

Jan 27   **Exercise 1 Due by 5:00 p.m.**

Jan 30   Implicitly Linear Models and Variable Transformations

Readings:   G&P, pp. 207-213.

Jan 30 *Lab*: SAS Workshop II, LIB 402, required  
Exercise 1 Returned

Feb 1 Multicollinearity

Readings: G&P, pp. 320-337.

**Exercise 2 Assigned (Due Fri, Feb. 10)**

Feb 6 Multicollinearity

Readings: G&P, pp. 337-351.  
Application Article Assigned in Class

Feb 6 *Lab*: Exercise 2 Help Session, LIB 402

Feb 8 Multicollinearity Application Example  
Dummy Variables

Readings: G&P, pp. 277-290.

**Exercise 3 Assigned (Due Fri, Feb. 17)**

Feb 10 **Exercise 2 Due by 5:00 p.m.**

Feb 13 Dummy Variables

Readings: G&P, 290-305 and 591-602.

Feb 13 *Lab*: Exercise 2 Returned, LIB 402  
Exercise 3 Help Session

Feb 15 Dummy Variable Application Example

Readings: None.

Feb 17 **Exercise 3 Due by 5:00 p.m.**

## **SECTION 2: Violations of Assumptions and Other Problems in the GLM**

Feb 20 Heteroscedasticity: Causes and Impacts

Readings: G&P, pp. 365-376.

Feb 20 *Lab*: Exercise 3 Returned, SPEA 277  
Review for Exam I

Feb 22 Catch-Up or Additional Examples

Readings: None.

- Feb 27 Review for Exam I
- Feb 27 *Lab: Exam I* (covers content through Feb 22), SPEA 277
- Feb 29 Data Set Analysis Project Assigned
- Mar 5 Heteroscedasticity: Impacts and Theoretical Resolution  
Readings: None.
- Mar 5 *Lab: Exam I Returned*, SPEA 277
- Mar 7 Heteroscedasticity: Practical Resolution  
Readings: G&P, pp. 376-401.  
**Exercise 4 Assigned (Due Fri, Mar 23)**
- Mar 10- **Spring Break**  
Mar 18
- Mar 19 Heteroscedasticity: Application Example and In-Class Exercise  
Readings: None.
- Mar 19 *Lab: Exercise 4 Help Session*, LIB 402
- Mar 21 Autocorrelation: Causes and Effects  
Readings: G&P, pp. 412-429.
- Mar 23 **Exercise 4 Due by 5:00 p.m.**
- Mar 26 Autocorrelation: Theoretical and Practical Resolution  
Readings: G&P, pp. 429-440.
- Mar 26 *Lab: Exercise 4 Returned*, SPEA 277  
Project Help Session
- Mar 28 Autocorrelation: Application Example and In-Class Exercise  
Readings: G&P, pp. 440-453.  
**Exercise 5 Assigned (Due Fri, Apr. 6)**
- Apr 2 Errors of Specification and Proxy/Instrumental Variables  
Readings: G&P, pp. 467-486.
- Apr 2 *Lab: Exercise 5 Help Session*, LIB 402

Apr 4 Errors of Specification and Proxy Variables

Readings: Application Article - Barry Rubin, 1985, "Specification of Wage Determination Mechanisms in Urban Labor Markets," *Socio-Economic Planning Sciences*, Vol. 19, No. 6, pp. 387-398.

Apr 6 **Exercise 5 Due by 5:00 p.m.**

### **SECTION 3: Alternative Forms and Extensions of Regression Analysis**

Apr 9 The Linear Probability Model and Introduction to Logit Analysis

Readings: G&P, pp. 541-552.

Apr 9 *Lab*: Project Help Session, LIB 402

Apr 11 The Logit Model

Readings: G&P, pp. 552-566.

**Exercise 6 Assigned (Due Fri, Apr. 20)**

Apr 16 The Logit Model – Application Example  
Autoregressive and Distributed Lag Models

Readings: G&P, pp. 617-629; and Application Article Assigned in Class.

Apr 16 *Lab*: Exercise 6 Help Session, LIB 402

Apr 18 Introduction to Simultaneous Equation Models (time permitting)

Readings: G&P, pp. 673-684 and 718-726.

Apr 20 **Exercise 6 Due by 5:00 p.m.**

Apr 23 Project Presentations  
Overview of Exam II Content

**Data Set Analysis Project Report Due In-Class**

Apr 23 *Lab*: Project Presentations, SPEA 277

Apr 25 Teaching Evaluations  
Project Presentations

Apr 28 Review for Final Exam, time and location to be announced  
or 29

May 2 (Wed) **Final Exam, 12:30-2:30 p.m., SPEA 277**