

Political Science 503
Quantitative Methods: Linear Regression and Extensions
Spring 2010
Tuesdays/Thursdays 9:00-10:15 AM
Rosenkranz 06

Thad Dunning
Associate Professor of Political Science
420 Rosenkranz
thad.dunning@yale.edu
Office Hours: Thursday 10:15-11:45 or by appt.
Teaching Fellow: Mario Chacón (Office hours TBA)

This course provides an extensive introduction to the use of linear regression in the social sciences. We begin by considering the algebraic properties of bivariate and multivariate regression. We then discuss the linear regression model, where observed data are viewed as realizations of random variables. Finally, we consider causal inference. Model specification is a central area of concern. We will discuss techniques for addressing limited breakdowns in modeling assumptions, especially instrumental-variables regression and other techniques that address problems of measurement error, reciprocal causation, and nonlinearities. Time series and pooled time-series-cross-sectional models are also covered. The role of research design is given special emphasis. While it is important to master technique, and this course will help you do so, it is even more important to understand the core assumptions.

The course assumes that students have command of the material covered in PLSC 500, including basic knowledge of probability and linear regression. Matrix algebra and calculus are helpful; however, we will be covering those topics as we go.

The URL for the course website is <http://www.thaddunning.com/teaching/plsc-503>. I will post lecture slides for the week to the course website in advance of each Tuesday's class. Viewing or printing the slides out in advance may help you focus on the concepts rather than on taking extensive notes. Don Green, who has taught this course for many years, has wonderful lecture notes available online, and you should consult these as well (<http://vote.research.yale.edu/pls503.html>).

Working exercises is crucial to mastering the material. Problem sets will be assigned most weeks; these will combine theoretical or conceptual problems and derivations with computer exercises. I will ask you to form groups to work on the problem sets (but you should try to work all the problems on your own before meeting with your group). There will also be a take-home midterm and a take-home final exam, to be taken individually. Mario Chacón, who has served as teaching fellow for this course before, and I will be available to help inside and outside of class.

Readings (Required):

Angrist, Joshua D. and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*.

Freedman, David A. 2009. *Statistical Models: Theory and Practice*. New York: Cambridge University Press, 2nd edition.

Freedman, David, Robert Pisani, and Roger Purves. 2007. *Statistics*. New York: W.W. Norton & Company, Fourth Edition.

Wooldridge, Jeffrey M. 2009. *Introductory Econometrics: A Modern Approach*. South-Western Cengage Learning, Fourth edition.

Recommended:

Freedman, David and David Lane. 1981. *Mathematical Methods in Statistics: A Workbook*. New York: W.W. Norton & Company.

Freedman, David A. (David Collier, Jasjeet S. Sekhon, and Philip B. Stark, eds.) 2009. *Statistical Models and Causal Inference: A Dialogue with the Social Sciences*. New York: Cambridge University Press.

Simon, Carl P. and Lawrence Blume. 1994. *Mathematics for Economists*. W.W. Norton & Company.

The four required books are available at the Yale bookstore, as are two of the three recommended books. You should buy all the books if you can; their value to you over the years will more than compensate the hefty initial investment. It may be substantially cheaper to buy an earlier edition of the Wooldridge book, and that should be fine. You could also purchase Freedman (2005, 1st edition) rather than Freedman (2009, 2nd edition), if there are price differentials.

Freedman and Lane (1981) is now out of print, but when last I checked, Amazon had many used (and cheap) copies. The book might seem a little quaint, but it is quite helpful, and I recommend you do your best to try to get it.

Simon & Blume (1994) is on the list because it provides an introduction to matrix algebra (e.g., Chapters 6-11); it may also be useful to have on your shelf for other purposes. If you prefer to have another matrix algebra book, see Freedman (2009) for references.

Section: There will be a two-hour discussion section led by Mario Chacón on Friday, a time TBA. This is a chance for you to reinforce your understanding of the lecture material and, especially, to learn to use the statistical packages. Mario will be leading several sessions on basics in Stata.

Statistical Packages: You can use a statistical package of your choosing to work the computer exercises on the problem sets. Many people seem to like Stata, which is what I most commonly use these days, and this is what Mario will be teaching. Others can work

in R if they prefer. Matlab is a nice option (though few political scientists seem to use it); the lab exercises in Freedman (2009) are written using Matlab.

Problem sets: Problem sets will be handed out most Tuesdays. You should try to work out all the answers before meeting with your group. Each group's problem set is due in class the Tuesday after it is handed out. No late problem sets will be accepted.

Course schedule and readings

N.B. I have organized the material into 13 weeks (with one slide presentation for each week) but don't know exactly how long it will take to cover each topic. We may adjust the timing as we go, if time allows. Most likely, we will need to devote additional time to the time series and time-series-cross-sectional models after the break, which may necessitate dropping other topics.

The first few weeks are likely to be heavy going. It is extremely important to keep up with the reading in the first four weeks and to get on top of the technical material; the rest of the course will prove difficult if you do not do this.

Jan. 12-14: The mechanics of bivariate regression

- Freedman, Pisani, and Purves (2007), Chapters 8-12.
- Freedman and Lane (1981), all, but especially pp. 48-86.
- Freedman (2009), Chapter 2.

Jan. 19-21: The mechanics of multiple regression

- Freedman (2009), Sections 3.1-3.3 and 3.5 (Remember to work the exercises!).
- Read up on matrix algebra (e.g. Simon and Blume 1994, Chapters 6-11).

Jan. 26-28: The linear regression model: statistical inference and hypothesis testing

- Freedman (2009), Chapter 4 and Chapter 5, sections 5.6 and 5.8.
- Freedman, Pisani and Purves (2007), Chapters 5, 13-14, 16-18, 20-21.

Feb. 2-4: The linear regression model: causal inference

- Freedman (2009), Chapter 6.
- Holland, Paul W. 1986. "Statistics and Causal Inference." *Journal of the American Statistical Association* 81 (396): 945-960.
- Angrist and Pischke (2009), Chapter 2.

Feb. 9-11: Instrumental variables: the effect of treatment on compliers

- Freedman, David. 2006. "Statistical Models for Causation: What Inferential Leverage Do They Provide?" *Evaluation Review* 30: 691-713 (read especially sections 1-4 and 10).
- Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. 1996. "Identification of Causal Effects Using Instrumental Variables." *Journal of the American Statistical Association* 91 (434): 444-455.

Feb. 16-18: Instrumental variables: the IVLS regression model

- Freedman (2009), Chapter 9.
- Angrist and Pischke (2009), Chapter 4.
- Sovey, Allison J. and Donald P. Green. 2009. "Instrumental Variables Estimation in Political Science: A Readers' Guide." Mimeo.
- Dunning, Thad. 2008. "Model Specification in Instrumental-Variables Regression." *Political Analysis* 16 (3): 290-302.

Feb. 23-25: Design-based inference: RDD and alternatives

- Freedman (2009), Chapter 1
- Freedman, Pisani, and Purves (2007), Chapter 27 and A-32 to A-36, especially notes 10, 11, 14, and 21.
- Angrist and Pischke (2009), Chapter 6.
- Dunning, Thad, "Improving Causal Inference: Strengths and Limitations of Natural Experiments." *Political Research Quarterly*, 61 (2): 282-293.

March 2-4: Matching as an analogue to regression

- Angrist and Pischke (2009), Section 3.3
- Sekhon, Jasjeet S. 2009. "Opiates for the Matches: Matching Methods for Causal Inference." *Annual Review of Political Science* 12: 487-508.
- Arceneaux, Kevin, Alan S. Gerber, and Donald P. Green. 2006. Comparing Experimental and Matching Methods using a Large-Scale Voter Mobilization Experiment. *Political Analysis* 14: 1-36.

Midterm Exam

March 9, 11, 16, 18 – Spring Break

March 23-25: GLS (and its alternatives) and the bootstrap

- Freedman (2009), Chapter 5 (sections 5.1-5.5 and 5.7) and Chapter 8
- Beck, Neal and Jonathan Katz. 1995. "What to Do (and Not To Do) With Time-Series Cross-Section Data." *American Political Science Review* 89 (3): 634-647.
- Recommended: Freedman, David and Stephen Peters. 1984. "Bootstrapping a Regression Equation: Some Empirical Results." *Journal of the American Statistical Association* 79: 97-106.

April 6-8: Time-series models

- Wooldridge (2009), Chapters 10-12

March 30-April 1: Time-series cross-sectional models

- Angrist and Pischke (2009), Chapter 5
- Wooldridge (2009), Chapters 13-14.

April 13-15: Covariate adjustment with strong designs

- Freedman, David A. 2008. “On regression adjustments to experimental data.” *Advances in Applied Mathematics* 40: 180–193.
- Recommended: Freedman, David A. 2008. “On regression adjustments in experiments with several treatments.” *Annals of Applied Statistics* 2: 176–96
- Green, Donald. 2009. “Regression Adjustments to Experimental Data: Do David Freedman’s Concerns Apply to Political Science?” Mimeo.

April 20-22: Review and wrap-up

Final Exam