Linear Regression Techniques in Comparative Politics
POSC 269, Sec. 1001, Spring 2007, W 1:00-4:00 p.m., Prof. Barrington, office: 452 WWP/478 WWP
Office Phone: 288-5983/288-5234, e-mail: Lowell.Barrington@mu.edu; off. hrs: MW 10-12, Th 12-2.

Overview. This course is designed to give you an introduction to linear statistical analysis, focusing extensively on OLS regression. By the end of the semester, you will be able to estimate an OLS regression model, understand the math underlying the OLS regression coefficient, and have a basic understanding of the strengths and limitations of OLS versus other quantitative techniques commonly used by political scientists.

We will begin the semester by considering the question of quantitative research in comparative politics. While some comparativists conduct statistical analysis of survey data and comparative political economists often conduct time series analysis, comparative politics remains a field in which sophisticated quantitative techniques are the exception rather than the rule. We will then spend two weeks on the basics of statistical relationships between two variables, including correlation. The bulk of the semester will examine the technique of OLS regression. The course will end with a brief discussion of the "next level" of statistical analysis, beyond OLS, that you will be expected to understand if you continue on in a good Ph.D. program.

Requirements. Course requirements include attending class every week, participating in class discussions, completing the required short assignments on time, and completing a research paper by the end of the semester that involves the use of OLS regression. In addition to the regular class discussions, there will be a number of "lab" sessions that will involve you working on a computer to prepare a data set for the estimation of an OLS regression model. We will be using SPSS for these activities. While less sophisticated than other software packages like STATA, SPSS is more than adequate for OLS estimations and is quite user-friendly.

Although the paper must use OLS, it will also include the basics of a quality research paper. Following the introduction, you will have an extensive literature review that identifies a gap in the existing literature that your paper addresses. There must also be a meaningful methods section laying out and justifying the methods that you use. There must be a results and discussion section Note: I take plagiarism on research papers *very* seriously; do not put yourself in a position to find out. While you are encouraged to work together this semester, work that you portray as your own must indeed be your own. Academic dishonesty also includes plagiarizing from yourself, by lifting key text from other papers you have or are turning in for other graduate courses.

There are only two books that you are required to purchase: (1) Eric A. Hanushek and John E. Jackson, *Statistical Methods for Social Scientists* (San Diego: Academic Press/Hardcourt Brace Jovanovich, 1977); (2) Leo H. Kahane, *Regression Basics* (Thousand Oaks; Sage, 2001). Hanushek and Jackson is a bit expensive, but it remains a classic social science statistics text. If you take more advanced statistics courses in a Ph.D. program down the road, you will be grateful to have this book on hand. It is a near-perfect balance of text, formulae, and examples. The Kahane book is less math intensive but also does a solid job of covering the basics of linear regression.

There will also be a few selections taken from Peter Kennedy, *A Guide to Econometrics*, 4th ed. (Blackwell Publishing, 1998). The Marquette library has an e-book copy of the 4th edition. However, the format of the book makes using it as e-book rather awkward. In addition, when one of the students in the class is accessing the e-book no one else can (it is effectively "checked out" during the time someone is looking at it). Thus, you may want to find paper version of the Kennedy text, which should only be around $30.

If you understand the basics of linear algebra (or you want to start to try making sense of it), you might also consider purchasing the following book: Michael P. Allen, *Understanding Regression Analysis* (New York: Plenum Press, 1997). Allen's book covers the same general topics covered in this course, but he approaches the math underlying regression almost exclusively through linear algebra. If you have not had linear algebra before, it will take a while to "click." Once it does, however, you will have a much better understanding of how regression estimates are generated. The linear algebra approach is typically not introduced until the second statistics course of a Ph.D. program's methods sequence, and we will thus not cover it in detail this semester – although it does sneak into a few of the weekly readings in places. A few of Allen's short chapters that do not involve much matrix algebra are assigned as reserve readings.

For fun, but also because it has moments of being quite informative, think about tracking down Larry Gonick and Woollcott Smith's *Cartoon Guide to Statistics*. It should not be more than $15 or so online.

Finally, since we will be doing "lab" work using SPSS, you may want to pick up one of the several books that cover estimating regression models using SPSS. In addition to the SPSS handbook for a recent version of SPSS, these include Babbie, Halley, and Zaino's *Adventures in Social Research: Data Analysis Using SPSS 11.0/11.5 for Windows* (Thousand Oaks, CA: Pine Forge Press, 2003) and Robert Ho, *Handbook of Univariate and Multivariate Data Analysis and Interpretation with SPSS*. There are a couple chapters from the Ho book on reserve.

Semester Grade: The breakdown of your grade will be as follows: (A) the final paper, 50%; (B) the short assignments, 25%; and (C) attendance and general participation, 25%.

CLASS SCHEDULE AND READING ASSIGNMENTS
* = Reading on reserve or D2L; ** = Reading available at the listed website; *** = Reading handed out in class

Part I: Quantitative Analysis and Comparative Politics
Week 1 (Jan. 17): Quantitative analysis in comparative politics.
Readings:
Hanushek and Jackson, ch. 1 (pp. 1-16 only).
***B. Guy Peters, *Comparative Politics: Theory and Method* (NYU Press, 1998), ch. 9.
**Peter Kennedy, *A Guide to Econometrics*, 4th ed. (Blackwell Publishing, 1998), ch. 1. Available online at:
> http://www.blackwellpublishing.com/content/BPL_Images/Content_store/Sample_chapter/9781405115025/Kennedy-001.pdf.
> (For the Kennedy reading, try to read the basic text, general notes, and technical notes for each section before moving to the next section; this will require a bit of jumping around in the chapter).

*Richard Berk, *Regression Analysis: A Constructive Critique* (Thousand Oaks, CA: Sage), chs. 1, 2, and and 11.

Question for consideration: Why have comparativists been reluctant to engage in quantitative research? Is their reluctance justified?

Question for consideration: Berk makes some rather strong claims against OLS regression (and, hence, against offering a course like this one). Is he right?

Week 2 (Jan. 24): Basic statistics involving one and two variables.
Readings:
Hanushek and Jackson, ch. 1 (pp. 16-23).
*David Freedman, Robert Pisani, and Roger Purves, *Statistics*, 1st ed. (New York: W. W. Norton, 1978), ch. 8.
*Christopher Achen, "Measuring Representation: Perils of the Correlation Coefficient," *American Journal of Political Science*, vol. 21 (1977): 805-815.
*Lowell Barrington, "Russian-speakers in Ukraine and Kazakhstan: 'Nationality,' 'Population,' or Neither?" *Post-Soviet Affairs*, vol. 17, no. 2 (April-June 2001): 129-158.
***Marcus Ethridge, ch. 10 ("Univariate Analysis: Statistics of a Single Variable").
> *Excerpt 13.* Greenstein, "The Benevolent Leader Revisited: Children's Images of Political Leaders in Three Democracies."

***Marcus Ethridge, ch. 11 ("Bivariate Analysis: Statistics of Two Variables").
> *Excerpt 14.* Cameron, Epstein, and O'Halloran, "Do Majority-Minority Districts Maximize Substantive Black Representation in Congress?"

Question for consideration: Why would someone examine the characteristics of just one variable?

Question for consideration: What kinds of data distributions would the standard deviation offer misleading information about the distribution? Can you think of real-world examples of such distributions?

Lab Session #1: The basics of working with variables in SPSS. I will give you a copy of Robert Ho, *Handbook of Univariate and Multivariate Data Analysis and Interpretation with SPSS*, ch. 2, before Week 2. Read it before class.

Part II: The Nuts and Bolts (and Math!) of OLS Regression

Week 3 (Jan. 31): Bivariate and multivariate OLS regression models.
Readings:
Kahane, ch. 1, ch. 2 (pp. 20-30 only).
Hanushek and Jackson, ch. 2.
*David A. Freedman, *Statistical Models: Theory and Practice* , ch. 2.
***Michael P. Allen, *Understanding Regression Analysis* (New York: Plenum Press, 1997), chs. 4-6 (pp. 16-30).
***Ethridge, ch. 12 ("Multivariate Analysis: Statistics of More Than Two Variables").
    Excerpt 15. Oliver, "City Size and Civic Involvement in Metropolitan America."
    Excerpt 16. Mutz, "Mass Media and the Depoliticization of Personal Experience."

Question for consideration: What are the advantages and limitations of bivariate regression analysis?

Question for consideration: What exactly does the unstandardized regression coefficient represent?

Week 4 (Feb. 7): Regression analysis using dummy variables.
Readings:
Hanushek and Jackson, ch. 4 (pp. 101-108).
Kahane, ch. 5 (pp. 83-92 only).
*Lowell W. Barrington, "Examining Rival Theories of Demographic Influences on Political Support: The Power of Regional, Ethnic, and Linguistic Divisions in Ukraine," *European Journal of Political Research*, vol. 41, no. 4 (2002): 455-491.
**Richard Wike and Nilanthi Samaranayake, "Where Terrorism Finds Support in the Muslim World: That May Depend on How You Define It - and Who Are the Targets," Pew Global Attitudes Project report, May 23, 2006, available at: http://pewresearch.org/obdeck/?ObDeckID=26.

Question for consideration: Why in the world does Prof. Barrington like dummy variables so much? What are their main limitations?

Lab Session #2: Using SPSS, estimate an OLS regression model that includes dummy variables.

Week 5 (Feb. 14): The assumptions of the OLS regression coefficient (or why the OLS estimator is "BLUE" on Valentine's Day).
Readings:
Hanushek and Jackson, ch. 3 (pp. 45-56 only).
*Peter Kennedy, *A Guide to Econometrics*, 4th ed. (Blackwell Publishing, 1998), ch. 3.

Question for consideration: Are there any differences between the OLS assumptions presented by H&J and those in Kennedy?

Question for consideration: Which of the main assumptions related to the OLS estimator seems the most likely to be violated when estimating an OLS model?

Part III: Statistical Significance in OLS Estimations
Week 6 (Feb. 21): The statistics underlying statistical significance: Probability and the sampling distribution of the OLS estimator.
Readings:
Hanushek and Jackson, Appendix 1 (pp. 325-344 only), ch. 3 (pp. 56-74 only).
***Michael P. Allen, *Understanding Regression Analysis* (New York: Plenum Press, 1997), chs. 11-12 (pp. 51-60).
*David Freedman, Robert Pisani, and Roger Purves, *Statistics*, 1st ed. (New York: W. W. Norton, 1978), ch. 18.

Question for consideration: Describe the following terms: (1) random variable, (2) probability distribution, (3) confidence interval. What does each have to do with estimating the significance of an OLS coefficient estimate?

Week 7 (Feb. 28): Calculating statistical significance of an individual coefficient.
Readings:
Hanushek and Jackson, ch. 4 (pp. 75-79 only).
*Michael P. Allen, *Understanding Regression Analysis* (New York: Plenum Press, 1997), chs. 13-14.
*David Freedman, Robert Pisani, and Roger Purves, *Statistics*, 1st ed. (New York: W. W. Norton, 1978), ch. 29.

Question for consideration: Look at Hanushek and Jackson's discussion of standardized coefficients (pp. 78-79). Why do they argue that the unstandardized coefficient is superior to the standardized coefficient for a particular variable?

Lab Session #3: In SPSS, estimate an OLS regression model and create a table of the results. Focus on making the statistical significance *and* the substantive significance of each coefficient easy for a reader to understand.

Week 8 (Mar. 7): Calculating statistical significance of sets of coefficients and comparing alternative specifications of a model.
Readings:
Hanushek and Jackson, ch. 5 (pp. 124-131 only).
*Chris Achen, "What Does 'Explained Variance' Explain?: Reply," *Political Analysis*, vol. 2 (1980): 173-184.
*Gary King, "How Not to Lie with Statistics," *American Journal of Political Science*, vol. 30 (1986): 666-687.

Question for consideration: Why is comparing alternative specifications of the same general model using the same data really the only time that the (adjusted) R^2 statistic is especially helpful?

Part IV: Problems in the Estimation of an OLS Linear Regression Model
Week 9 (Mar. 21): Model misspecification.
Readings:
Hanushek and Jackson, ch. 4 (pp. 79-86 only).
*Kennedy, ch. 5.
*H. Michael Crowson, Teresa K. Debacker, Stephen J. Thoma, "The Role of Authoritarianism, Perceived Threat, and Need for Closure or Structure in Predicting Post-9/11 Attitudes and Beliefs," *Journal of Social Psychology*, vol. 146, no. 6 (December 2006): 733-750.
*Kevin Clarke, "The Phantom Menace: Omitted Variable Bias in Econometric Research," *Journal of Conflict Management and Peace Science* (2005): 341-352.

Question for consideration: Although it is rather easy to criticize a work for omitted variable bias, we're going to do it anyway. Are there any glaring omissions in the models in Crowson, Debacker, and Thoma?

Week 10 (Mar. 28): Heteroskedasticity and autocorrelation.
Readings:
Hanushek and Jackson, ch. 6.
Kahane, ch. 6 (**pp. 119-134 only**).
*Kennedy, ch. 8.
*** George Downs and David Rocke, "Interpreting Heteroskedasticity," *American Journal of Political Science*, vol. 23 (November 1979): 816-828.

Question for consideration: How do heteroskedasticity and autocorrelation violate regression assumptions?

Week 11 (Apr. 4): Measurement error.
Readings:
Hanushek and Jackson, ch. 10 (section 10.2 only).
***Kennedy, ch. 9.

***Frederick G. Conrad and Michael F. Schober, "Clarifying Question Meaning in a Household Telephone Survey," *Public Opinion Quarterly*, vol. 64, no. 1 (Spring 2000): 1-28.
***George D. Gaskell and Daniel B. Wright Colm A O'Muircheartaigh, "Telescoping of Landmark Events: Implications for Survey Research," *Public Opinion Quarterly*, vol. 64, no. 1 (Spring 2000): 77-89.

Question for consideration: Why is measurement error arguably a much greater issue for the independent variables in an OLS model than it is for the dependent variable?

Week 12 (Apr. 11): Multicollinearity.
Readings:
Hanushek and Jackson, ch. 4 (pp. 86-96).
Kahane, ch. 6 (pp. 113-119 only).
***Kennedy, ch. 11.

Recommended Reading:
***Donald E. Farrar and Robert R. Glauber, "Multicollinearity in Regression Analysis: The Problem Revisited," *The Review of Economics and Statistics*, vol. 49, no. 1. (Feb. 1967): 92-107.

Question for consideration: What is multicollinearity, and what problems does it pose for OLS estimations? Is it a violation of one of the regression assumptions?

Week 13 (Apr. 18): Non-linear relationships, data transformations, and interaction terms.
Readings:
Hanushek and Jackson, ch. 4 (pp. 96-101).
Kahane, ch. 5 (pp. 79-83, 92-101 only).
*Bear Braumoeller, "Hypothesis Testing and Multiplicative Interaction Terms," *International Organization* vol. 58, no. 4 (2004): 807-820.
***Thomas Brambor, William Roberts Clark, Matt Golder, "Understanding Interaction Models: Improving Empirical Analyses," *Political Analysis*, vol. 14, no. 1(Winter 2006): 63-82.

Question for consideration: In what situations would it *not* be appropriate to estimate multivariate statistical models?

Part V: Looking Ahead: Quantitative Techniques beyond OLS
Week 14 (Apr. 25): Time series analysis, Logit, Probit, and Bayesian Statistics.
Readings:
Hanushek and Jackson, ch. 7.
*Eric Noreen, "An Empirical Comparison of Probit and OLS Regression Hypothesis Tests," *Journal of Accounting Research*, vol. 26, no. 1 (Spring, 1988): 119-133
*Jeff Gill, "Introduction to the Special Issue," *Political Analysis*, vol. 12, no. 4 (Autumn 2004): 323-337.
*B. Efron, "Why Isn't Everyone a Bayesian?," *The American Statistician*, vol. 40 (1986): 1-5.
*Lowell Barrington, Brian Silver, and Erik Herron The Motherland is Calling: Views of Homeland among Russians in the Near Abroad," *World Politics*, vol. 55, no. 2 (January 2003): 290-313.

Recommended Reading:
*Kennedy, ch. 13, ch. 15, ch. 17.

Question for consideration: Why is OLS (officially) not appropriate in the case of a dichotomous dependent variable? In what ways might the OLS estimator be biased in such a case? Given Noreen's findings, is OLS likely to generate different results from probit in the case of dichotomous dependent variables?

Question for consideration: What is autocorrelation, and why is it a particular problem in time series analysis?

Question for consideration: Why isn't everyone a Bayesian?

<u>**Week 15 (May 2): Structural equations modeling and conclusion**</u>.
Hanushek and Jackson, chs. 8-9.
*Lowell Barrington, "Economic Calculations, Cultural Perceptions, or General Outlooks on Life?: What Shapes Views of the 'Ethnic Other' in Kazakhstan." *Journal of Central Asian Studies*, forthcoming, December 2006.
*Lowell W. Barrington and Erik S. Herron, "Understanding Public Opinion in Post-Communist States: The Effect of Statistical Assumptions on Substantive Results," *Europe-Asia Studies*, vol. 53 (2001): 573-594.
*Larry M. Bartels, "Five Approaches to Model Specification," *The Political Methodologist*, vol. 3 no. 2 (1990): 2-6.

Recommended Reading:
*Kennedy, ch. 10 (sections 10.1-10.4 [basic text, general notes, and technical notes for those sections]).

Question for consideration: Do Barrington and Herron make a compelling case for not mixing together attitudinal and demographic explanatory variables in an OLS model? If one is convinced that demographic and attitudinal variables fall at different stages of the "causal funnel," why would a SEM approach make more sense than OLS?

FINAL PAPER DUE: Thursday, May 10, 5:00 p.m. (in my department mailbox).