Multivariate Modeling with Stata and R

Workshops in Social Science Research
Concordia University
May 2016

Instructors:

Harold Clarke, Ashbel Smith Professor, School of Economic, Political and Policy Sciences, University of Texas at Dallas; email: clarke475@msn.com

Guy D. Whitten, Director, Program in Scientific Political Methodology, Professor and Cornerstone Fellow, Department of Political Science, Texas A&M University; email: g-whitten@tamu.edu

Course Description:

This workshop teaches participants how to use the popular software packages Stata and R to conduct theoretically interesting and practically useful analyses of social, economic and political data. Throughout the course, the emphases will be on how to conduct and present sound analyses and strategies for learning new techniques. We begin with a review/crash course in the linear multiple regression model, and then move other important multivariate models including binomial logit and probit models, the multinomial logit model and the mixed logit model. The workshop then considers the interpretation of interaction effects in linear and nonlinear models as well as multilevel models that analyze how socio-economic and political contexts influence individual behavior. Time series and pooled time series methods that investigate how factors such as policy interventions and socio-economic conditions affect dynamic outcomes also are considered. Additional topics, such as the analysis of spatial statistical models will be covered based on student interests and time availability. Assuming only familiarity with introductory statistics, the workshop emphasizes practical approaches to analyzing individual- and aggregate-level social science data. Participants are invited to bring their own data sets to daily lab sessions.

Recommended Texts

Box-Steffensmeier, Janet, John Freeman, Matthew Hitt and Jon Pevehouse. 2014. *Time Series Analysis for the Social Sciences*. New York: Cambridge University Press.

Fox, John and Harvey Sanford Weisberg. *An R Companion to Applied Regression*. 2nd Edition. Thousand Oaks, CA: Sage Publications, 2011.

Paul Kellstedt and Guy Whitten. *The Fundamentals of Political Science Research*. 2nd edition. New York: Cambridge University Press, 2013.

Long, J. Scott and Jeremy Freese. 2014. *Regression Models for Categorical Dependent Variables Using Stata*. 3rd edition. College Station, TX: Stata Press.

Monogan, Jamie. 2016. *Political Analysis Using R*. Springer.

Other Readings

To assist students wishing to learn more about various topics after the course is completed, a number of additional readings are suggested. A list of references is provided at the end of the syllabus.

Data Sets: Students are encouraged to bring their own data sets to analyze in daily labs. The instructors also will provide several data sets for use in lab sessions. Some of these data sets are based on national surveys of political attitudes and voting in recent American, British and Canadian general elections, whereas others track the aggregate dynamics of public attitudes towards important issues such as the economy, health care and immigration, as well support for political parties, presidents and prime ministers.

Topics

1. Introduction to Multivariate Modeling
    - How This Course Will Run
    - Some Rules of the Road for Conducting Statistical Analyses
    - Taking Command of Mathematical Notation
    - Getting to Know Your Data
    - Using Stata
    - Using R
    - Specialized Software Alternatives - Eviews, NLOGIT, RATS
    - Readings:
        - Acock (2014)
        - Fox and Weisberg (2011), chs. 1, 2, 3
        - Long and Freese (2014), ch. 2

2. Multiple Regression Models: Reviewing Basics
    - Introduction
    - Assumptions - Everybody Makes Them!
    - Introduction to Regression Simulation
    - Hypothesis Tests
    - Goodness of Fit
    - Model Selection Criteria
    - Hypothesis Tests

- Readings:
  - Kellstedt and Whitten, chs. 8, 9 and 12
  - Fair (1988)
  - Fox and Weisberg (2011), ch. 4

3. Multiple Regression Models: Specification, Diagnostics, Fix-Ups, Interpretation
   - Dummy Independent Variables
   - Interaction Effects
   - Functional Form
   - Multicollinearity
   - Heteroskedasticity
   - Autocorrelation
   - Readings:
     - Kellstedt and Whitten, ch. 10
     - King, Tomz and Wittenberg (2000)
     - Tomz, Wittenberg and King, (2003)
     - Long and Freese (2014), chs. 3, 4
     - Brambor, Clark and Golder (2006)
     - Clarke, Elliott and Stewart (2016a)

4. Dichotomous and Ordinal Dependent Variables
   - Linear Probability Model
   - Binomial Logit and Probit Models
   - Ordinal Logit and Probit Models
   - Readings:
     - Kellstedt and Whitten, pp. 247-55
     - Long and Freese (2014), chs. 5, 6, 7
     - Fox and Weisberg (20111), ch. 5

5. Unordered Multi-Category Dependent Variables
   - Multinomial Logit Model
   - Multinomial Probit Model
   - Mixed Logit Model
   - Readings:
     - Whitten and Palmer (1996)
     - Glasgow (2001)
     - Hensher, Rose and Green (2005)
     - Clarke et al. (2009), ch. 5
     - Long and Freeze (2014), ch. 8
     - Train (2003)

6. Multilevel Models
   - Nests Everywhere!
   - Multilevel Regression Models
   - Multilevel Discrete Choice Models
   - Cross-Classified Multilevel Models

- Readings:
    - Raudenbush and Bryk (2002), chs. 1, 2, 4, 8, 10, 12

7. Time Series Analysis I
    - Time Series Data
    - Specifying Time Series Models
    - Autocorrelation - Just a Nuisance?
    - Autoregressive, Distributed Lag (ADL) Models
    - Readings:
        - Box-Steffensmeier et al. (2014), chs. 1-3
        - Kellstedt and Whitten, pp. 256-69
        - Keele and Kelly (2006)

8. Time Series Analysis II
    - Nonstationarity: Deterministic and Stochastic Trends
    - ARIMA Models
    - Cointegration and Error Correction
    - Fractionally Integrated Models
    - GARCH
    - Readings:
        - Box-Steffensmeier et al. (2014), chs. 5-7
        - Charemza and Deadman (1997)
        - Clarke et al. (2009), ch. 4.
        - Clarke et al. (2016b)
        - Enders (2014)

9. Time Series Cross-Sectional Panel Models
    - To Pool or Not to Pool?
    - Fixes with Fixed Effects?
    - Panacea Corrected Standard Error
    - Interpreting Dynamic TSCS Models
    - Readings:
        - Beck and Katz (1995)
        - Williams and Whitten (2012)
        - Clark and Linzer (2015)

10. Spatial Regression Models
    -Spatial Basics
    -Thinking theoretically about spatial relationships (not just geography)
    -Estimation and interpretation of spatial models
    - Readings:
        - Beck, Gleditsch, and Beardsley (2006)
        - Pluemper and Neumayer (2015)

List of Recommended Readings

Alan Acock. 2014. *A Gentle Introduction to Stata*. 4th Edition. College Station, TX: Stata Press.

Beck, Nathaniel and Jonathan Katz. 1995. "What To Do (and Not To Do) with Time Series Cross-Section Data." *American Political Science Review* 89: 634-47.

Beck, Nathaniel, Kristian Skrede Gleditsch, and Kyle Beardsley. 2006 "Space is More than Geography: Using Spatial Econometrics in the Study of Political Economy." *International Studies Quarterly* 50.1: 27-44.

Brambor, Thomas, William Clark and Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14: 63-82.

Charemza, Wojciech W. and Derek F. Deadman. 1997. *New Directions in Econometric Practice*. 2nd ed. Aldershot: Edward Elgar.

Clark, Tom S., and Drew A. Linzer. "Should I use fixed or random effects?" *Political Science Research and Methods* 3.02 (2015): 399-408.

Clarke, Harold D., Euel Elliott and Marianne C. Stewart. 2016a. "Heuristics, Heterogeneity and Green Choices: Voting on California's Proposition 23." *Political Science Research and Methods*, 4: forthcoming.

Clarke, Harold D., Paul Whiteley, Walter Borges, David Sanders and Marianne C. Stewart. 2016b. "Modeling the Dynamics of Support for a Right-Wing Populist Party: The Case of UKIP." *Journal of Elections, Public Opinion and Parties* 26: 135-54.

Clarke, Harold D., David Sanders, Marianne C. Stewart and Paul Whiteley. 2009. *Performance Politics and the British Voter*. Cambridge: Cambridge University Press.

Enders, Walter. 2014. *Applied Econometric Time-Series*. 4th ed. New York: John Wiley & Sons.

Fair, Ray. 1988. "The Effect of Economic Events on Votes for President: 1984 Update." *Political Behavior* 10: 168-77.

Glasgow, Garett. 2001. "Mixed Logit Models for Multiparty Elections." *Political Analysis* 9: 116-36.

Hensher, David A., John M. Rose and William H. Greene. 2005. *Applied Choice Analysis: A Primer*. Cambridge: Cambridge University Press.

Keele, Luke and Nathan Kelly. 2006. "Dynamic Models for Dynamic Theories: The Ins and Outs of Lagged Dependent Variables." *Political Analysis* 14: 186-205.

King, Gary, Michael Tomz and Jason Wittenberg. 2000. "Making the Most of Statistical Analyses: Improving Interpretation and Presentation." *American Journal of Political Science* 44: 347-61.

Pluemper, Thomas, and Eric Neumayer. 2015. "W." *Political Science Research and Methods.*

Raudenbush, Stephen W. and Anthony Bryk. 2002. *Hierarchical Linear Models: Applications and Data Analysis Methods*. 2nd ed. Thousand Oaks, CA: Sage Publications.

Tomz, Michael, Jason Wittenberg and Gary King. 2003. "Clarify: Software for Interpreting and Presenting Statistical Results." Cambridge, MA: Harvard University, Department of Government.

Train, Ken. 2003. *Discrete Choice Models with Simulation*. Cambridge: Cambridge University Press.

Whitten, Guy and D. and Harvey D. Palmer. 1996. "Heightening Comparativists' Concern for Model Choice: Voting Behavior in Great Britain and the Netherlands." *American Journal of Political Science* 40: 231-60.

Williams, Laron K. and Guy D. Whitten. 2012. "But Wait, There's More! Maximizing Substantive Inferences from TSCS Models." *Journal of Politics* 74: 685-93.