

ADVANCED QUANTITATIVE METHODS

READING LIST

Jan 2002

Advanced Methods Committee:

- \$ Ken Bollen - bollen@unc.edu
- \$ Guang Guo - gguo@email.unc.edu
- \$ François Nielsen (Chair) - francois_nielsen@unc.edu

OUTLINE

1. LINEAR REGRESSION
2. LOGISTIC/PROBIT & RELATED REGRESSION METHODS
3. STRUCTURAL EQUATION MODELS
4. LONGITUDINAL DATA ANALYSIS
5. MULTILEVEL MODELS
6. LOG-LINEAR MODELS
7. OTHER TOPICS

1. LINEAR REGRESSION

Topics

- Classical Multiple Regression
- Extensions of Multiple Regression
 - Dummy (Indicator) Variables
 - Interaction Effects
 - Functional Form and Transformations
- Outliers and Influential Cases
 - Diagnosics
 - Robust estimation
- Assumption Violations
 - Multicollinearity
 - Heteroscedasticity
 - Autocorrelation
 - $\text{COV}(\mathbf{x}, \mathbf{u}) \neq 0$
 - $E(\mathbf{u}) \neq 0$
 - Specification Tests

Suggested Readings

- Allison, Paul. 1999. Multiple Regression: A Primer. Thousand Oaks, CA: Pine Forge Press.
- Belsley, David A., Edwin Kuh, and Roy E. Welsch. 1980. Regression Diagnostics: Identifying Influential Data and Sources of Collinearity. New York: Wiley.
- Bollen, Kenneth A. and Robert W. Jackman. 1990 [1985]. "Regression Diagnostics: An Expository Treatment of Outliers and Influential Cases." Pp. 257-291 in Modern Methods of Data Analysis, edited by John Fox and J. Scott Long. (Reprinted with revisions from Sociological Methods and Research 13:510-542.)
- Draper, Norman R., and Harry Smith. 1998. Applied Regression Analysis. (3rd ed.) New York: Wiley.
- Fox, John. 1991. Regression Diagnostics. Newbury Park, CA: Sage.
- Fox, John and J. Scott Long, eds. 1990. Modern Methods of Data Analysis. Newbury Park, CA: Sage. [Articles on many of the above topics.]
- Greene, William. 2000. Econometric Analysis. (4th ed.) Upper Saddle River, NJ: Prentice Hall.
- Hadi, Ali S. 1992. "Identifying Multiple Outliers in Multivariate Data." Journal of the Royal Statistical Society - Series B, 54 (3):761-771.
- Hadi, Ali S. 1994. "A Modification of a Method for the Detection of Outliers in Multivariate Samples." Journal of the Royal Statistical Society - Series B, 56 (2):393-396.

- Jaccard, James, Robert Turrisi, and Choi K. Wan. 1980. Interaction Effects in Multiple Regression. Newbury Park, CA: Sage.
- Judge, George G., William E. Griffith, R. Carter Hill and Tsoung-Chao Lee. 1985. The Theory and Practice of Econometrics. (2nd ed.) New York: Wiley.
- Kmenta, Jan. [1986] 1997. Elements of Econometrics. Ann Arbor, MI: University of Michigan Press.
- Koopmans, Lambert H. 1987. Introduction to Contemporary Statistical Methods. (2nd edition.) Boston: Duxbury.
- Neter, John, Michael H. Kutner, Christopher J. Nachtsheim and William Wasserman. 1996. Applied Linear Statistical Models. 4th edition. Burr Ridge, IL: Irwin. ISBN 0-256-11736-5. [Used as main text in SOCI209 by Nielsen.]
- Neter, John, Michael H. Kutner, Christopher J. Nachtsheim and William Wasserman. 1996. Applied Linear Regression Models. 3d edition. Burr Ridge, IL: Irwin. ISBN 0-256-08601-X. [This is a subset - 15 first chapters - of Applied Linear Statistical Models (4th edition.), see previous reference.]
- Velleman, Paul F. and Hoaglin, David C. 1981. Applications, Basics, and Computing of Exploratory Data Analysis. Boston: Duxbury.
- Wilkinson, Leland, Grant Blank, and Christian Gruber. 1996. Desktop Data Analysis with SYSTAT. Upper Saddle River, NJ: Prentice Hall.
- Williams, Valerie S. L., Lyle V. Jones, and John W. Tukey. 1999. "Controlling Errors in Multiple Comparisons, with Examples from State-to-State Differences in Educational Achievement." Journal of Educational and Behavioral Statistics 24, No 1 (Spring):42-69.

See also: syllabus for SOCI209; sections below on Bootstrap methods, Visual display of quantitative information, Sample selection issues. Many of the methodological issues related to linear regression models are discussed in general statistics and econometric texts, some of which are listed above. For relevant articles see Sociological Methodology (SM) and Sociological Methods & Research (SMR). Several of the Sage papers in the Quantitative Applications in the Social Science Series cover topics on linear regression.

2. LOGISTIC/PROBIT & RELATED REGRESSION METHODS

Topics

Statistical Distributions
 binomial
 multinomial

- normal
- logistic
- Maximum Likelihood Estimation (MLE)
 - steps to find
 - properties
- Dichotomous Dependent Variables
 - Linear Probability Model (LPM)
 - OLS & WLS
 - Probit & Logistic Models
 - Outlier & influence diagnostic
- Ordinal Dependent Variables
 - Probit & Logistic Models
- Counts as Dependent Variables
 - Poisson regression
 - Negative Binomial regression
- Censored Dependent Variables
 - LPM
 - Tobit
- Unordered Dependent Variables
 - Multinomial logit & probit models
 - Independence of irrelevant alternatives problem

Suggested Readings

- Aldrich, John Herbert and Forrest D. Nelson. 1984. Linear Probability, Logit, and Probit Models. Newbury Park: Sage.
- Clogg, Clifford C. and Edward S. Shihadeh. 1994. Statistical Models for Ordinal Variables. Newbury Park: Sage.
- Greene, William. 2000. Econometric Analysis. (4th ed.) Upper Saddle River, NJ: Prentice Hall. [Relevant sections.]
- Kmenta, Jan. [1986] 1997. Elements of Econometrics. Ann Arbor, MI: University of Michigan Press. [Relevant sections.]
- Maddala, G.S. 1983. Limited-Dependent and Qualitative Variables in Econometrics. New York: Cambridge University Press.
- Olzak, Susan. 1989. "Analysis of Events in the Study of Collective Action." Annual Review of Sociology 15:119-141.

Also, see articles from Soci 211 syllabus. Applications in ASR and other journals. Some methodological pieces in SMR and SM. Many more papers are in the econometric literature.

3. STRUCTURAL EQUATION MODELS

Topics

- Matrix Algebra
 - Elementary operations
 - Determinants, inverses, etc.
- Meaning of Causality in Structural Equation Models
- Path Analysis
 - Path Diagrams
 - Decomposition of Effects
- Classical Econometric Models
 - Recursive
 - Nonrecursive
- Consequences of Measurement Error
- Constructing Measurement Models
- Validity and Reliability
- Confirmatory Factor Analysis
- General Structural Equation Models
- Identification of Models
- Estimators
 - 2SLS, OLS, GLS, WLS, ML, ULS, IV
- Model Fit Measures
 - Residuals & Overall fit measures
 - Wald and Lagrangian Multiplier Tests
- Respecification of Models
- Multiple Group Analyses
- Categorical & Discrete Variables

Suggested Readings (not exhaustive)

- Bentler, Peter M. 1989. EQS Program Manual. Los Angeles: BMDP.
- Blalock, Hubert M. 1985. Causal Models in the Social Sciences. Chicago: Aldine.
- Bollen, Kenneth A. 1989. Structural Equations with Latent Variables. New York: Wiley.
- Duncan, Otis D. 1975. Introduction to Structural Equation Models. New York: Academic Press.
- Hayduk, Leslie A. 1987. Structural Equation Modeling with LISREL. Baltimore, MD: Johns Hopkins.
- Jöreskog, Karl G. and Dag Sörbom. 1989. LISREL 7: A Guide to the Program and Applications. Chicago: SPSS.
- Kenny, David A. 1979. Correlation and Causality. New York: Wiley.
- Long, J. Scott. 1983. Confirmatory Factor Analysis. Beverly Hills, CA: Sage.

Long, J. Scott. 1983. Covariance Structure Models. Beverly Hills, CA: Sage.

Also, see the bibliographies in the above sources for articles, see Sociological Methods & Research, Sociological Methodology, Psychometrika, Psychological Bulletin, Multivariate Behavioral Research, British Journal of Mathematical and Statistical Psychology, and Applied Psychological Measurement for recent developments, and see main sociology journals, several psychology journals, and Journal of Marketing Research for application examples.

4. LONGITUDINAL DATA ANALYSIS

Topics

- Processes generating duration data
- Basic definitions
- Continuous time vs discrete time
- Duration dependence
- Estimation:
 - maximum likelihood
 - partial likelihood
- Repeated events
- Multiple states
- Time-dependent covariates
- Unobserved heterogeneity
- Interdependencies between events
- Dynamic models (other than hazard rate)
- Pooling of time-series & cross-sectional data

Suggested Readings (not exhaustive)

Models for events

Allison, Paul D. 1984. Event History Analysis. Beverly Hills, CA: Sage

Carroll, Glenn R. 1983. "Dynamic Analysis of Discrete Dependent Variables." Quality & Quantity 17:425-60.

Cox, David R. and D. Oakes. 1984. Analysis of Survival Data. New York: Chapman and Hall.

Guo, Guang. 1993. "Event History Analysis for Left-Truncated Data." Pp. 217-243 in Peter V. Marsden, ed., Sociological Methodology. Cambridge, MA: Blackwell.

Guo, Guang and Germán Rodríguez. 1992. "Estimating a Multivariate Proportionate Hazard Model for Clustered Data Using the EM Algorithm, with an Application to Child Survival

- in Guatemala." Journal of the American Statistical Association 87:969-976.
- Heckman, James and Burt Singer. 1986. "Econometric Analysis of Longitudinal Data." Pp. 1689-1763 in Z. Griliches and M. D. Intriligator. eds., Handbook of Econometrics. Amsterdam: North-Holland.
- Hutchison, D. 1988. "Event History and Survival Analysis in the Social Sciences I." Quality & Quantity 22:203-219.
- Hutchison, D. 1988. "Event History and Survival Analysis in the Social Sciences II." Quality & Quantity 22:255-278.
- Kalbfleisch, J. D. and Ross L. Prentice. 1980. The Statistical Analysis of Failure Time Data. New York: Wiley.
- Tuma, Nancy B. and Michael T. Hannan. 1984. Social Dynamics. New York: Academic.
- Yamaguchi, Kazuo. 1991. Event History Analysis. Newbury Park, CA: Sage.
- Models for continuous variables**
- Alderson, Arthur S. and François Nielsen. 2002. "Globalization and the Great U-Turn: Income Inequality Trends in 16 OECD Countries." American Journal of Sociology. Forthcoming in March issue.
- Baltagi, Badi H. 1995. Econometric Analysis of Panel Data. New York: Wiley.
- Beck, Nathaniel and Jonathan N. Katz. 1995. "What To Do (and Not To Do) with Time-Series Cross-Section Data." American Political Science Review 89:634-47. [Note that they deal with time-series dominant data sets, not short time series!]
- Greene, William H. 1990. Econometric Analysis. New York: MacMillan. [Chapter 16, especially "longitudinal data" section pp. 480-505.]
- Greene, William H. 1992. LIMDEP User's Guide. New York: Econometric Software. [Chapter 29, "Fixed and Random Effects Linear Models."]
- Hsiao, Cheng. 1986. Analysis of Panel Data. New York: Cambridge.
- Judge, George G., William E. Griffith, R. Carter Hill and Tsoung-Chao Lee. 1980. The Theory and Practice of Econometrics. New York: Wiley. [Chapter 8, pp. 325-373; there is a newer 1985 edition]
- Kmenta, Jan. [1986] 1997. Elements of Econometrics. Ann Arbor, MI: University of Michigan Press. [Material on pooling time-series & cross-sectional data and on time-series analysis.]

Rosenfeld, Rachel A. and François Nielsen. 1984. "Inequality and Careers: A Dynamic Model of Socioeconomic Achievement." Sociological Methods and Research 12:279-321.

Also see ASR, AJS, and SF for recent applications, see SM, SMR, and biometrics literature for recent developments, and see bibliographies of above readings for additional material.

5. MULTILEVEL MODELS

Suggested readings

Bryk, Anthony S. and Stephen W. Raudenbush. 1992. Hierarchical Linear Models: Applications and Data Analysis Methods. Newbury Park, CA: Sage.

Goldstein, Harvey. 1995. Multilevel Statistical Models. 2nd ed. London: Arnold; New York: Halstead. [An earlier edition was published in 1987 as Multilevel Models in Educational and Social Research. London: Griffin.]

Guo, Guang and Hongxin Zhao. 2000. "Multilevel Modeling for Binary Data." Annual Review of Sociology 26:441-462.

Mason, William et al. 1983. "Contextual Analysis through the Multilevel Linear Model." Pp. 72-103 in Sociological Methodology 1983-84. San Francisco: Jossey-Bass.

6. LOG-LINEAR MODELS

Topics

- Measures of association
- Hierarchical models
- Models of quasi-independence
- Other models incorporating patterns of associations
- "Dependent" variables
- Ordered categorical data
- Category counts

Suggested Readings (not exhaustive)

Agresti, Alan. 1990. Categorical Data Analysis. New York: Wiley.

Clogg, Clifford C. 1982. "Using Association Models in Sociological Research: Some Examples." American Journal of Sociology 88: 113-134.

Clogg, Clifford C. and Edward S. Shihadeh. 1994. Statistical Models for Ordinal Variables. Newbury Park, CA: Sage.

Clogg, Clifford C. and Scott R. Eliason. 1988. "Some Common Problems in Log-Linear Analysis." In Common Problems/Proper Solutions, edited by J. Scott Long. Newbury Park, CA: Sage.

Fienberg, Stephen E. 1980. The Analysis of Cross-classified Categorical Data. (2nd ed.) Cambridge, MA: MIT Press.

Fingleton, Bernard. 1984. Models of Category Counts. New York: Cambridge University Press.

Hout, Michael. 1983. Mobility Tables. Beverly Hills, CA: Sage.

Knoke, David and Peter J. Burke. 1980. Log-linear Models. Beverly Hills, CA: Sage.

Long, J. Scott. 1997. Regression Models for Categorical and Limited Dependent Variables. Thousand Oaks, CA: Sage.

Also, see bibliographies of above sources for additional references, see ASR, AJS, or SF for some applications, and see SM, SMR, the biometrics, and statistical journals for recent developments.

7. OTHER ISSUES & TECHNIQUES

The list of topics and readings for this section is the most difficult to compose. The above references provide some guidelines but the best way to learn what other quantitative techniques are applied in sociology is to become familiar with SM and SMR. Occasionally a quantitative methodology piece is published in ASR, AJS, or SF.

Topics

- Epistemological issues
- Visual display of quantitative information
- Bootstrap methods
- Monte Carlo simulations
- Factor analysis
- Sample selection issues
- Other topics

Fox, John and J. Scott Long (eds). 1990. Modern Methods of Data Analysis. Newbury Park, CA: Sage. [This collection has articles on several of the above topics.]

Epistemological issues

Blalock, Hubert M. 1984. Basic Dilemmas in the Social Sciences. Beverly Hills, CA: Sage

Blalock, Hubert M. 1982. Conceptualization and Measurement in the Social Sciences. Beverly Hills, CA: Sage.

Duncan, Otis D. 1984. Notes on Social Measurement. New York: Russell Sage Foundation.

Liebertson, Stanley. 1985. Making It Count : The Improvement of Social Research and Theory. Berkeley, CA: University of California Press. [Also look at discussion of Liebertson in Sociological Methodology 1987- 88.)

Visual display of quantitative information

Chambers, John M., William S. Cleveland, B. Kleiner & John A. Tukey. 1983. Graphical Methods for Data Analysis. Boston, MA: Duxbury.

Cleveland, William S. 1994. The Elements of Graphing Data. (Revised edition.) Summit, NJ: Hobart Press.

Tufte, Edward R. 1983. The Visual Display of Quantitative Information. Cheshire, CT: Graphics Press.

Wilkinson, Leland. 1988. SYGRAPH. Evanston, IL: SYSTAT, Inc.: Ch.2: "Cognitive science and graphic design."

Bootstrap methods

Bollen, Kenneth A. and Robert Stine. 1990. "Direct and Indirect Effects: Classical and Bootstrap Estimates of Variability". Pp. 115-140 in Sociological Methodology 1990, edited by Clifford C. Clogg. Cambridge, MA: Basil Blackwell.

Diaconis, Persi and Bradley Efron. 1983. "Computer-Intensive Methods in Statistics." Scientific American 248: 5 (May):116-130.

Efron, Bradley and Robert Tibshirani. 1986. "Bootstrap Methods for Standard Errors, Confidence Intervals, and Other Measures of Statistical Accuracy." Statistical Science 1:54-75.

Efron, Bradley and Robert Tibshirani. 1993. An Introduction to the Bootstrap. New York: Chapman & Hall.

Dynamic models (other than hazard rates)

Doreian, Patrick and Norman P. Hummon. 1976. Modeling Social Processes. New York: Elsevier.

Nielsen, François and Rachel A. Rosenfeld. 1981. "Substantive Interpretations of Differential Equations Models." American Sociological Review 46:159-174.

Petersen, Trond. 1993. "Recent Advances in Longitudinal Methodology." Annual Review of Sociology 19:425-454.

Missing values & sample selection issues

Allison, Paul D. 2002. Missing Data. Sage. (Forthcoming.)

Berk, Richard A. and Subashi C. Ray. 1982. "Selection Biases in Sociological Data." Social Science Research 11:352-398.

Little, Roderick J. A. and Donald B. Rubin. 1987. Statistical Analyses with Missing Data. New York: Wiley.

Little, Roderick J. A. and Donald B. Rubin. 1989. "The Analysis of Social Science Data with Missing Values." Pp. 292-326 in Data Analysis: Modern Methods and New Perspectives, edited by J. Scott Long. Special double issue of Sociological Methods & Research 18, No 2 (November 1989) & 3 (February 1990).

Stolzenberg, Ross M. and D. A. Relles. 1990. "Theory Testing in a World of Constrained Research Design: The Significance of Heckman's Censored Sampling Bias Correction for Nonexperimental Research." Sociological Methods & Research 18:395-415.

Winship, Christopher and Robert Mare. 1992. "Models for Sample Selection Bias." Annual Review of Sociology 18:327-350.

Other topics

Marsden, Peter V. 1990. "Network Data & Measurement." Annual Review of Sociology 16:435-463.

McCullagh, Peter and John A. Nelder. 1983. Generalized Linear Models. New York: Chapman and Hall.

Sørensen, Aage B. 1978. "Mathematical Models in Sociology." Annual Review of Sociology 4:345-371.