

SOCI 30005: QUANTITATIVE DATA ANALYSIS

University of Chicago
Spring Quarter 2016

[Last Updated: March 28, 2016]

LECTURE TIME AND PLACE: Tuesday/Thursday 10:30 a.m. - 11:50 a.m., Haskell Hall-315

DISCUSSION TIME AND PLACE: Friday 1-3 pm; Wieboldt Hall Computing Lab (3rd floor)

INSTRUCTOR: Prof. Xi Song

Office: Social Science Building 422;

Cell phone: (773)834-0335;

E-mail: xisong@uchicago.edu

Office Hours: By appt.

TEACHING ASSISTANT:

Danya Lagos

Office: TBA;

Email: danya@uchicago.edu

Office Hours: By appt;

Joshua Mausolf

Office: TBA;

E-mail: jmausolf@uchicago.edu

Office Hours: By appt;

COURSE OVERVIEW

This is a one-quarter course in how to do theoretically informed quantitative social research with rigorous statistical techniques. The course concentrates on data analysis, and the way one links theory and data. By the end of the course you should have a good idea of how to make sociological sense out of a body of quantitative data. Toward this end, we will cover a variety of techniques, including tabular analysis, regression analysis, regression diagnostics, missing data, factor analysis and scale construction, measurement error, fixed and random effects models, propensity score matching, and related topics. This is not only a statistics course, but also a course that teaches you procedures to draw substantive conclusions about how the social world works. We will also discuss how to organize your work and present your results. We focus on the quantitative analysis of data from probability samples of well-defined populations. Data collection procedures will be essentially ignored — that is, mentioned only in discussions of data-analytic issues.

PREREQUISITES

A prior statistics course — SOCI 30004, or the equivalent — is required.

COURSE ASSIGNMENTS AND EVALUATION

1) Weekly Assignments (~60% of your final grade)

Exercises are due in class the week after they are assigned. There are 8 assignments in the quarter. They will be read and returned the following week. It is important that you do each set of weekly exercises completely and on time; late exercises will not be accepted. If for some reason you do not complete your assignment on time, I encourage you to complete it on your own, but we will not accept it for credit. To compensate for this strict policy, I will drop the lowest grade you receive on an assignment when we tabulate your overall grade. After the first few weeks you will be doing analysis using a major U.S. national sample survey (NORC's General Social Survey). As the quarter progresses, however, for most of the assignments you will be able to substitute data of your own, focusing on topics that interest you and/or that pertain to your term paper.

2) Term Paper (and Term Paper Proposal) (~40% of your final grade)

Your final term paper will be due at the end of the quarter. Late papers will not be accepted.

The course will culminate in a term paper on a topic of your choosing in which you will carry out a quantitative analysis of some substantive issue using the technical and analytic skills developed by doing the exercises. It is not uncommon for course term papers to lead to or revise master's papers or chapters of Ph.D. dissertations and/or publications.

With instructor prior approval, you may write co-authored papers with no more than two authors. Both authors must be students in the class. In the case of co-authorship, the term paper proposal should detail what each author plans to contribute to the project, and the final term paper must include a separate document detailing what each author contributed.

More information on this project will be distributed over the quarters.

COURSE MATERIALS

1) *Calculator*

2) *Stata Software, Version 14*

Starting in the 3rd week of the first quarter all assignments will require doing data analysis of one or more sample surveys, using the statistical package Stata. Stata is a fast and efficient package that includes most of the statistical procedures of interest to social scientists. The 2nd lecture will include an introduction to Stata.

You may carry out your computing work in one of several ways:

- (a) using one of the PC's in the Social Sciences Computing (SSC) Laboratories;
- (b) accessing Stata via public computers in the UChicago network.
- (c) accessing to Stata will be through the UChicago VLab or any on campus labs: <http://academictech.uchicago.edu/vlab>
- (d) [recommended] purchasing and installing Stata on your own computer or some computer to which you have individual access.

To purchase Stata, you have several options. Users of sample survey data sets who do not intend to use Stata as their primary statistical package could use a perpetual license for Stata/IC, which is available under the UChicago "GradPlan". If you expect to work with very large data sets, such as census files, or are a committed Stata user, you should purchase a perpetual license for Stata/SE. You can go to a still more expensive option if you have a multiple-processor computer: Stata/MP. Differences in these versions are summarized below.

Compare features

Package	Max. no. of variables	Max. no. of right-hand variables	Max. no. of observations	64-bit version available?	Fastest: designed for parallel processing?	Platforms
Stata/MP	32,767	10,998	unlimited*	Yes	Yes	Windows, Mac (64-bit Intel), or Unix
Stata/SE	32,767	10,998	unlimited*	Yes	No	Windows, Mac, or Unix
Stata/IC	2,047	798	unlimited*	Yes	No	Windows, Mac, or Unix
Small Stata	99	99	1,200	No	No	Windows (32-bit) or Mac (32-bit)

*The maximum number of observations is limited only by the amount of available RAM on your system.

In addition, you may be interested in StatTransfer 13 (available from Stata Corp under the GradPlan), which is an extremely good conversion package that enables you to convert SAS, SPSS, Excel, text or any of a wide number of other file types to Stata files, and vice versa.

If you have a previous version of Stata or of Stat/Transfer, you should check the Stata Corp webpage for upgrade pricing. Although most versions of Stata (10+) should work well for this class, it is preferable to have the most recent version.

3) *Stata Documentation*

It is impossible to become facile at Stata without frequently consulting the documentation. Every copy of Stata ships with complete PDF documentation, including the *Base Reference Manual*, *User's Guide*, *Data-Management Reference Manual*, *Graphics Reference Manual*, and all the programming and specialized statistics manuals. The PDF documentation is linked into the existing interactive help file system. Continue to use the help files to review syntax and options and to see short examples; click on the link in the title of the help file to be taken to the corresponding entry in the PDF documentation.

How to purchase Stata software and documentation: Go to the Stata Corp web site: <http://www.stata.com> click on "Order Stata" or "Upgrade Now," click on "Educational," click on "Place an order" under "Grad Plans," click to get to UChicago, and complete your order.

If you are ordering Stata, do so right away. This will give you time to get and look over the documentation and install and try out the software before the introductory lecture the 4th week of class. If you have questions about the purchase, call Stata Corp and ask to speak to someone about the GradPlan.

4) *Textbooks*

Required

Treiman, Donald. 2009. *Quantitative Data Analysis*. John Wiley & Sons.

Recommended

Hamilton, Lawrence C. 2010. *Statistics with Stata*, Version 12. Stata Press.

Long, J. Scott. 2009. *The Workflow of Data Analysis using Stata*. Stata Press.

Miller, Jane E. 2005. *The Chicago Guide to Writing about Multivariate Analysis*. University of Chicago Press.

Note: Texts are available at the UChicago Bookstore (and online). Additional readings, and the Long readings, will be posted on a week-by-week basis on the course website.

COURSE WEBSITE

The *course chalk website* contains this syllabus, lectures, exercises, links to the data we will use, and the documentation for these data. I urge you to check the website frequently since it will contain the most up-to-date information regarding the course. I also recommend the discussion forum on the website to access your classmates as a resource for questions you may have regarding the course material.

COURSE SESSIONS (SUBJECT TO CHANGE)

Note: Readings marked by * are posted on the course website.

Spring Quarter 2016

Week 1 (March 29): *Course Introduction & Cross-Tabulations and Tables I*

Required Reading: Treiman Ch. 1;

Recommended Reading: Freese 2007*; King 2006*;

(April 1) No Class (PAA)

Week 2 (April 5): *Cross-Tabulations and Tables II*

Required Reading: Treiman Chs. 2-3;

Recommended Reading: Miller 2005 Chs. 5-6*;

(April 7) *Identifying Data and Introduction to Computing*

Required Reading: Treiman Ch. 4;

Assignment: Exercise W02;

Week 3 (April 12, 14): *Descriptive Statistics, Correlation, and Regression I*

Required Reading: Treiman Ch. 5;

Recommended Reading: Miller 2005 Ch, 13*;

Due: Exercise W02;

Assignment: Exercise W03;

Week 4 (April 19, 21): *Regression II*

Required Reading: Treiman Ch. 6;

Recommended Reading: Miller 2005 Ch. 14*; Jann 2007*;

Due: Exercise W03;

Assignment: Exercise W04;

Week 5 (April 26, 28): *Regression III*

Required Reading: Treiman Ch. 7;

Due: Exercise W04;

Assignment: Exercise W05;

Week 6 (May 3, 5): *Regression IV*

Due: Exercise W05;

Assignment: Exercise W06;

Week 7 (May 10, 12): *Regression Diagnostics*

Required Reading: Treiman Ch. 10;

Due: Exercise W06;

Assignment: Exercise W07;

Week 8 (May 17, 19): *Missing Data and Sample Design*

Required Reading: Treiman Chs. 8-9;

Due: Exercise W07;

Assignment: Exercise W08;

Week 9 (May 24, 26): *Factor Analysis and Scale Construction*

Required Reading: Treiman Ch. 11;

Due: Exercise W08;

Assignment: Exercise W09;

Week 10 (May 31): *Causal Inference and Panel Data Methods*

Required Reading: Treiman Ch. 15;

Recommended Reading: Morgan and Winship 2007 Chs. 2, 3, 4, 5, & 6*;

Due: Exercise W09;

Assignment: Work on term papers;

June 10

Due: Term paper;

ADDITIONAL RECOMMENDED READINGS BY TOPIC

The following list of readings offers further elaboration or additional topics beyond the Readings. The list provides you with references to pursue as you advance the research design for your term paper or to elaborate on a week's topic. (Note: There may be newer editions to some of these books.)

Stata

Kohler, Ulrich and Frauke Kreuter. 2012. *Data Analysis Using Stata*. 3rd Edition. College Station: Stata Press.

Mitchell, Michael N. 2008. *A Visual Guide to Stata Graphics, 2nd Edition*. Boston: Brooks/Cole.

Rabe-Hesketh, Sophia, and Brian Everitt. 2007. *A Handbook of Statistical Analyses Using Stata*, 4th edition. Boca Raton, FL: Chapman & Hall/CRC.

Writing Manuscripts and Making Tables

Becker, Howard S. 1986. *Writing for Social Scientists: How to Start and Finish Your Thesis, Book, or Article*. Chicago: University of Chicago Press.

[From Becker's own work with students on editing papers;]

Booth, Wayne, Gregory Colomb, and Joseph Williams. 2008. *The Craft of Writing*. Chicago: University of Chicago Press.

[Describes how to ask questions, make good arguments, make claims, assemble evidence, communicate your results and argument, and connect with readers;]

Ebel, Hans F., Claus Bliefert, and William E. Russey. 2004. *The Art of Scientific Writing*. Wiley.

[For writing about chemistry and related fields, but nevertheless a useful resource.]
Strunk, William and E. B. White. 2009. *Elements of Style*, 50th Anniversary Edition. Pearson and Longman.

[Classic writing manual – a must have;]

Williams, Joseph. 2007. *Style: Lessons in Clarity and Grace*, 9th Edition. New York: Pearson Longman.

[A wonderful small book on effective writing; I highly recommend this one;]

Regression Analysis and Overviews on Quantitative Analysis

Berk, Richard. 2004. *Regression Analysis: A Constructive Critique*. Thousand Oaks: Sage.

Fox, John. 1991. *Regression Diagnostics*. (Sage No. 79)

Fox, John. 1997. *Applied Regression Analysis, Linear Models, and Related Methods*. Thousand Oaks: Sage.

Hamilton, Lawrence C. 1992. *Regression with Graphics: A Second Course in Applied Statistics*. Belmont: Duxbury Press.

[A useful advanced applied statistics text for Stata users;]

Jaccard, James, and Robert Turrisi. 2003. *Interaction Effects in Multiple Regression*. 2nd ed. (Sage No. 72)

Marsh, Lawrence C., and David R. Cormier. 2001. *Spline Regression Models*. (Sage No. 137)

Osborne, Jason W. 2008. *Best Practices in Quantitative Methods*. Los Angeles: Sage Publications.

Logistic Regression and Allied Procedures

Hosmer, David W., and Stanley Lemeshow. 2000. *Applied Logistic Regression*. 2nd Edition. New York: Wiley.

[A classic text; authors are biostatisticians;]

Liao, Tim Futing. 1994. *Interpreting Probability Models: Logit, Probit, and Other Generalized Linear Models*. (Sage No. 101)

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

Long, J. Scott, and Jeremy Freese. 2006. *Regression Models for Categorical Dependent Variables Using Stata*, 2nd Edition. College Station: Stata Press.

Menard, Scott. 2001. *Applied Logistic Regression Analysis*. 2nd ed. (Sage No. 106)

[A non-technical introduction to logistic regression;]

Powers, Daniel A., and Yu Xie. 2008. *Statistical Methods for Categorical Data Analysis*, 2nd Edition. Orlando: Academic Press.

Log-linear and Log-multiplicative Analysis

Hout, Michael. 1982. *Mobility Tables*. (Sage No. 31.)

Knoke, David, and Peter Burke. 1980. *Log-Linear Models*. (Sage No. 20.)

Estimation, Statistical Inference, and Related Topics

Good, Phillip I. and James W. Hardin. 2012. *Common Errors in Statistics and How to Avoid Them*. Hoboken: John Wiley & Sons.

Smithson, Michael. 2002. *Confidence Intervals*. (Sage No. 140)

Missing Data

- Allison, Paul D. 2001. *Missing Data*. (Sage No. 136)
- Little, Roderick J. A., and Donald B. Rubin. 2002. *Statistical Analysis with Missing Data*, 2nd ed. New York: John Wiley & Sons.
[The definitive treatment, by the creators of multiple-imputation;]

Sample Selection Bias

- Berk, Richard A. 1983. "An Introduction to Sample Selection Bias in Sociological Data." *American Sociological Review* 48:386-98.
- Breen, Richard. 1996. *Regression Models: Censored, Sample Selected, or Truncated Data*. (Sage No. 111)
- Kang Fu, Vincent, Christopher Winship, and Robert D. Mare. 2004. "Sample Selection Bias Models." Pp. 409-430 in *Handbook of Data Analysis*, Melissa Hardy and Alan Bryman, eds., Sage Publications.
[An update to the citation below by Winship and Mare]
- Winship, Christopher, and Robert D. Mare. 1992. "Models for Sample Selection Bias." *Annual Review of Sociology* 18:327-50.

Basic Econometrics

- Angrist, Joshua and Jorn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton: Princeton University Press.
- Greene, William H. 2008. *Econometric Analysis* 6th Edition. Upper Saddle River, NJ: Prentice Hall.
- Woolridge, Jeffrey M. 2002. *Econometric Analysis of Cross-Section and Panel Data*. Cambridge: The MIT Press.

Causal Inference

- Gangl, Markus. 2010. "Causal Inference in Sociological Research." *Annual Review of Sociology* 36:21-47.
- Manski, Charles. 1995. *Identification Problems in the Social Sciences*. Boston, MA: Harvard University Press.
- Morgan, Stephen, ed. 2013. *Handbook of Causal Analysis for Social Research*. Springer Series.
- Morgan, Stephen, and Christopher Winship. 2014. *Counterfactuals and Causal Inference: Methods and Principles for Social Research, Second Edition*. New York, NY: Cambridge University Press.
- Pearl, Judea. 2000. *Causality: Models, Reasoning, and Inference*. New York, NY: Cambridge University Press.
- Rosenbaum, Paul. 2010. *Observational Studies*. New York: Springer.
- Rubin, Donald. 2006. *Matched Sampling for Causal Effects*. New York, NY: Cambridge University Press.

Fixed and Random Effects Models

- Allison, Paul. 1990. "Change Scores as Dependent Variables in Regression Analysis." *Sociological Methodology* 20:93-114.
- Allison, Paul. 1994. "Using Panel Data to Estimate the Effects of Events." *Sociological Methods and Research* 23:174-199.

Baltagi, Badi H. 2008. *Econometric Analysis of Panel Data 4th Edition*. New York: Wiley.
Halaby, Charles N. 2004. "Panel Models in Sociological Research: Theory into Practice." *Annual Review of Sociology* 30:507-44.

[A nice review, focusing on causal inference and fixed effects models;]

Hausman, Jerry A. 1978. "Specification Tests in Econometrics." *Econometrica* 46:1251-1272.

[Hausman tests are frequently used to compare fixed and random effects models; this article introduces the test;]

Propensity Score Matching

Abadie, Alberto, David Drukker, Jane Leber Herr, and Guido Imbens. 2002. "Implementing Matching Estimators for Average Treatment Effects in Stata." *Stata Journal*.

[A nice Stata routine to use for PS matching.]

Becker, Sascha O., and Andrea Ichino. 2002. "Estimation of Average Treatment Effects Based on Propensity Scores." *Stata Journal* 2:358-377.

[Use this routine to generate p-scores. It also contains matching estimators;]

Morgan, Stephen and David Harding. 2006. "Matching Estimators of Causal Effects." *Sociological Methods & Research* 35(1):3-60.

Rosenbaum, Paul R. and Donald B. Rubin. 1984. "Reducing Bias in Observational Studies using Sub-classification on the Propensity Score." *Journal of the American Statistical Association*, 79, 516-524.

Smith, Herbert L. 1997. "Matching with Multiple Controls to Estimate Treatment Effects in Observational Studies." *Sociological Methodology* 27:325-353.

Instrumental Variables

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.

Angrist, Joshua and Alan Krueger. 2001. "Instrumental Variables and the Search for Identification: From Supply and Demand to Natural Experiments." *Journal of Economic Perspectives* 15(4), 69-85.

Heckman, James. 1997. "Instrumental Variables: A Study of Implicit Behavioral Assumptions Used in Making Program Evaluations." *Journal of Human Resources*, 32(3), 441-462.

Imbens, Guido and Joshua Angrist. 1997. "Identification and Estimation of Local Average Treatment Effects." *Econometrica* 62: 467-476.

Event History / Survival Models

Blossfeld, Hans-Peter, Katrin Golsh, and Gotz Rohwer. 2007. *Event History Analysis with Stata*. London: Lawrence Erlbaum Associates Publishers.

Allison, Paul D. 1984. *Event History Analysis: Regression for Longitudinal Event Data*. (Sage No. 46)

Hosmer, D.W. and S. Lemeshow. 1999. *Applied Survival Analysis: Regression Modeling of Time to Event Data*. Wiley Series in Probability and Statistics.

Structural Equation Modeling

- Bollen, Kenneth A. 1989. *Structural Equations with Latent Variables*. New York: John Wiley and Sons.
- Duncan, Otis Dudley. 1975. *Introduction to Structural Equation Models*. New York: Academic Press.
- Jöreskog, Karl G., and Sörbom, Dag. 1996. *LISREL 8: User's Reference Guide*. Chicago: Scientific Software International.
- Loehlin, John C. 2004. *Latent Variable Models: An Introduction to Factor, Path, and Structural Analysis*. 4th Edition. Lawrence Erlbaum Associates.

Multilevel Models

- DiPrete, Thomas and Jerry D. Forristal. 1994. "Multilevel Models: Methods and Substance." *Annual Review of Sociology* 20:331-357.
- Rabe-Hesketh and Anders Skrondal. 2008. *Multilevel and Longitudinal Modeling Using Stata* 2nd Edition, College Station: Stata Press.
- Raudenbush, Stephen and Anthony Bryk. 2002. *Hierarchical Linear Models*. Thousand Oaks: Sage.

Growth Curve and Trajectory Models

- Bollen, Kenneth A. and Patrick J. Curran. 2006. *Latent Curve Models: A Structural Equation Perspective*. New York: Wiley.
- Jones, Bobby L. and Daniel S. Nagin. 2007. "Advances in Group-Based Trajectory Modeling and a SAS Procedure for Estimating Them." *Sociological Methods and Research* 35:542-571.
- Singer, Judith and John B. Willet. 2003. *Applied Longitudinal Data Analysis*. Oxford: Oxford University Press.

Network Analysis

- Scott, John. 1991. *Social Network Analysis: A Handbook*. London: Sage
- Wasserman, Stanley and Katherine Faust. 1994. *Social Network Analysis: Methods and Applications*. Cambridge: Cambridge University Press.