PL SC 502: Statistical Methods for Political Research

Course Information
Class Time: Tuesdays 2.30-5.30
Place: 103 Pond Lab
Course Website: https://files.nyu.edu/mrg217/public/methods2

Contact Information for Professor
Name: Matt Golder
E-mail: mgolder@psu.edu (preferred method of contact)
Tel: 814-867-4323
Office: 306 Pond Lab
Office Hours: Tuesdays 11-12

Contact Information for Teaching Assistant
Name: Eitan Tzelgov
E-mail: eut113@psu.edu (preferred method of contact)
Office Hours: TBA

Course Objectives

This is the first course in quantitative methods in Penn State's political science Ph.D. program. The course is an introduction to the use of statistics for the social sciences, and political science in particular. There are three main goals of the course: (i) to teach students to read and understand quantitative analyses in published and unpublished work, (ii) to provide them with the skills necessary to begin conducting their own quantitative analyses, and (iii) to lay the foundation for future courses in quantitative methods. The course is divided into three (unequal) parts. The first part of the course provides an introduction to the basics of data collection, organization, and management; measurement; data visualization and display; and univariate, bivariate, and multivariate descriptive statistics. The second part of the course focuses on probability theory; sampling theory; estimation; and hypothesis testing. The third and final part of the course introduces Ordinary Least Squares (OLS) regression analysis.

Although there are no formal prerequisites for this class, the fact that it is an introductory course should not be taken to mean that it will be easy. Rather, the expectation should be that confronting new material will be challenging. This course is intended to provide the foundation for more advanced courses in quantitative methods, and will be fast-paced and rigorous. While I have chosen readings that present the material as clearly and with as little jargon as possible, most of it will require several readings to fully comprehend.

This syllabus is designed to provide an overview to the course. Clickable links are printed in Penn State blue.

Teaching Assistant:

Eitan Tzelgov will be the teaching assistant for this class. He will handle problem sets and is your resource for understanding what we cover together in class. Throughout the semester, he will hold regular office hours and conduct occasional computer lab sessions. Lab sessions will take place on Fridays at 10.00am. Eitan is your primary resource for learning Stata (and R), and LATEX.

Course Requirements

Attendance: Attendance and participation in class is mandatory. Although recommended, attendance at lab sessions is voluntary.

Assignments: There will be weekly homework assignments, which will be a mix of analytical and computer-based problems dealing with the techniques covered in class. These assignments will be made available on Tuesday after class and MUST be handed in to the teaching assistant, Eitan Tzelgov, by the following Monday at 9.00am. Assignments that are handed in after this deadline will not be graded. I assume that you will work together on these assignments and I encourage this. However, I recommend that you do not work in groups larger than three - I do not want the class as a whole to be doing the assignments together! Although you will be working in small groups, each of you should write up your own answer key (don't all use the same sentences etc.) and each of you should run any computer code that is written.

Paper: There will be a final paper in which you apply one or more of the techniques that you have learned to a real social science question. I highly recommend that this paper be written in LATEX. You should start thinking about a possible paper topic as early as possible in the semester and discuss this with me.

Note that with the exception of do-files, log-files, and data files, all homework assignments and the final paper should be submitted as hard (paper) copies. Homework assignments should be placed in Eitan's mailbox. The final paper should be placed in my mailbox. In the exceptional circumstance that you need to submit something electronically, only PDF files will be accepted.

Grading:

Grading is based on the weekly assignments (65%) and the final paper (35%). No incomplete grades will be given unless there is an agreement between the instructor and the student prior to the end of the course. The instructor retains the right to determine legitimate reasons for an incomplete grade.

Required Text/Materials

- Wonnacott, Thomas H. & Ronald J. Wonnacott. 1990. (5th edition). *Introductory Statistics*. New York: John Wiley & Sons. [W&W]

- Gujarati, Damodar N. 2003. *Basic Econometrics*. New York: McGraw-Hill. [G]

Additional readings as necessary, all of which will be available on ANGEL and/or through JSTOR.

Recommended

- Acock, Alan C. 2010. *A Gentle Introduction to Stata*. College Station, TX, Stata Press.

- Verzani, John. 2005. *Using R For Introductory Statistics*. Boca Raton, FL: Chapman & Hall.

Math/Statistics/Econometrics Books

- Agresti, Alan, and Barbara Finlay. 2008. *Statistical Methods for the Social Sciences*, 4th Ed. Upper Saddle River, NJ: Prentice-Hall.

- Wackerly, Dennis D., William Mendelhall III, and Richard L. Scheaffer. 2008. *Mathematical Statistics with Applications*, 7th Ed. New York: Duxbury.

- DeGroot, Morris H., and Mark J. Schervish. 2002. *Probability and Statistics*. New York: Addison-Wesley.

- Gill, Jeff. 2006. *Essential Mathematics for Political and Social Research*. Cambridge: Cambridge University Press.

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

- Wooldridge, Jeffrey. 2006. *Introductory Econometrics*. USA: Thompson South-Western.

- Greene, William H. 2011. *Econometric Analysis*. Prentice Hall.

- Kennedy, Peter. 2008. *A Guide to Econometrics*. Boston: MIT Press.

- Fox, John. 2008. *Applied Regression Analysis and Generalized Linear Models*. Newbury Park, CA: Sage Publications.

Some Other Useful Resources

The Political Methodology Section of the American Political Science Association was created to provide APSA members with an interest in political methodology with a forum in which to meet and discuss ideas. The section publishes a quarterly newsletter (*The Political Methodologist*), a quarterly journal on political methodology (*Political Analysis*), conducts a discussion list on topics relating to political methodology, and maintains an extensive electronic archive of papers, accessible via their homepage.

Also, the Inter-University Consortium for Political and Social Research (ICPSR), at the University of Michigan, maintains an extensive archive of data in the social and behavioral sciences. Much of it is accessible via their homepage.

Harvard University's Dataverse Project "increases scholarly recognition and distributed control for authors, journals, archives, teachers, and others who produce or organize data; facilitates data access and analysis for researchers and students; and ensures long-term preservation whether or not the data are in the public domain." It's a repository of all sorts of cool and useful stuff for quantitative researchers; check it out.

StatSoft has a very extensive and well-organized web-based electronic textbook which covers a variety of new computational and machine learning methods as well as conventional statistics, and also a very useful glossary

This MIT Open CourseWare course looks to have fairly good lecture notes if you want another take on things.

Software, Statistical and Otherwise

The primary statistical package that will be covered in this class will be Stata . However, we will occasionally employ R as well. You are welcome to use either Stata or R to complete the homework assignments, so long as the manner by which your results are generated and conclusions reached are transparent.

Stata

At the present time, Stata is probably the most widely-used statistical package in the social sciences. It is a powerful tool for data management, analysis, and display, and boasts some of the best manuals and on-line help of any existing software package. Stata is commercial software; the current version of Stata is 12.0, but previous versions can also be used for the class. In the class notes, handouts etc., Stata commands will appear in a fixed-width font and will be preceded by a period ("."):

```
. tab Y X
```

Those new to Stata may want to check out:

- Acock, Alan C. 2010. *A Gentle Introduction to Stata*. College Station, TX, Stata Press.

- Cameron, Colin A. & Pravin K. Trivedi. 2010. *Microeconometrics Using Stata* . College Station, TX, Stata Press.

- Mitchell, Michael N. 2008. *A Visual Guide to Stata Graphics*. College Station, TX, Stata Press.

Beyond this, the Stata homepage is a valuable resource for questions about the Stata statistical software. There are a number of useful Stata references on the web, including Scott Long's page at IU, and an excellent Stata "help page" sponsored by UCLA.

R

R is a statistical environment and high-level programming language for data analysis and display. It is free and open source; the current version is 2.13.1. In the notes and handouts etc., R commands will generally be preceded by a caret (">"):

```
> xtab<-table(Y,X)
```

The Comprehensive R Archive Network (CRAN) is the go-to spot for all things R-related. Those new to R may want to check out the Introduction to R, this page on getting data into R, and the various R "cheat sheets" here, here, and here. Stata users who are interested in learning R should check out the Moving from Stata to R page at the R Project's wiki.

Beyond this, you might also want to check out:

- Adler, Joseph. 2009 *R in a Nutshell: A Desktop Quick Reference*. O'Reilly Publishing.

- Dalgaard, Peter. 2008. *Introductory Statistics with R*. New York: Springer.

- Crawley, Michael J. 2005. *Statistics: An Introduction Using R*. John Wiley & Sons, Ltd.

- Crawley, Michael J. 2007. *The R Book.* John Wiley & Sons, Ltd.

- Murrell, Paul. 2005. *R Graphics*. Boca Raton, FL: Chapman & Hall. (Website is here).

LaTeX

Learn to use LaTeX, now, while you have the time. You will be very, very glad you did. Good places to start are here, here, here, and/or here.

Beyond this, you might also want to check out:

- The Not So Short Introduction to LaTeX2$_\epsilon$

- Mittelbach, Frank & Michel Goosens. 2004. *The LaTeXCompanion*. New York: Addison-Wesley.

- Kopka, Helmut & Patrick W. Daly, 2003. *Guide to LaTeX*. New York: Addison-Wesley.

- Goosens, Michel, Frank Mittelbach, Sebastian Rahtz, Denis Roegel, & Herbert Voß. 2007. *The LaTeXGraphics Companion.* New York: Addison-Wesley.

- Lamport, Leslie. 1994. *LaTeX: A Document Preparation System*. New York: Addison-Wesley.

Academic Dishonesty

The Department of Political Science, along with the College of the Liberal Arts and the University, takes violations of academic dishonesty seriously. Observing basic honesty in one's work, words, ideas, and actions is a principle to which all members of the community are required to subscribe.

All course work by students is to be done on an individual basis unless an instructor clearly states that an alternative is acceptable. Any reference materials used in the preparation of any assignment must be explicitly cited. Students uncertain about proper citation are responsible for checking with their instructor.

In an examination setting, unless the instructor gives explicit prior instructions to the contrary, whether the examination is in class or take home, violations of academic integrity shall consist but are not limited to any attempt to receive assistance from written or printed aids, or from any person or papers or electronic devices, or of any attempt to give assistance, whether the one so doing has completed his or her own work or not.

Lying to the instructor or purposely misleading any Penn State administrator shall also constitute a violation of academic integrity.

In cases of any violation of academic integrity it is the policy of the Department of Political Science to follow procedures established by the College of the Liberal Arts. More information on academic integrity and procedures followed for violation can be found here.

Disabilities

Penn State welcomes students with disabilities into the University's educational programs. If you have a disability-related need for reasonable academic adjustments in this course, contact the Office for Disability Services. For further information regarding policies, rights and responsibilities please visit the Office for Disability Services (ODS) Web site. Instructors should be notified as early in the semester as possible regarding the need for reasonable accommodations.

Acknowledgements

Much of the material for this course has built up over time and has been influenced by numerous scholars. Particular thanks go to Sona Golder, Brad Gomez, Jonathan Nagler, Phil Schrodt, and Chris Zorn for sharing their notes.

Tentative Schedule

Readings should be completed prior to coming to class on the assigned day. Note that we will not, in general, hew closely (or at all) to the readings themselves, other than topically. The course schedule should be treated as tentative and flexible. It may be the case that it takes us more or less time for a particular topic than I have allotted. We will adapt accordingly.

Week 1: August 23 – Introduction, Overview, LaTeX etc.

- **Readings: none**

- *Exercise One: LaTeX.*

Week 2: August 30 – Data Structure, Measurement, Coding Basics

- **Readings**
  - *Required*:
    - W & W, Chapter 1.
    - Clark, William, Matt Golder, & Sona Golder. 2009. *Principles of Comparative Politics*. Washington D.C.: CQ Press. Chapters 2, 5.
    - Nagler, Jonathan. 1995. "Coding Style and Good Computing Practices." *The Political Methodologist* 6 :2-8.
    - King, Gary. 1995. "Replication, Replication." *PS: Political Science and Politics*. 28: 443-499.
  - *Recommended*:
    - Bowers, Jake. 2011. "Six Steps to a Better Relationship with your Future Self." *The Political Methodologist* 18 :2-8.
    - Adcock, Robert & David Collier. 2001. "Measurement Validity: A Shared Standard for Qualitative and Quantitative Research." *American Political Science Review* 95: 529-546.
    - Elkins, Zarchary. 2000. "Gradations of Democracy? Empirical Tests of Alternative Conceptualizations." *American Journal of Political Science* 44: 293-300.
    - Treir, Shawn & Simon Jackman. 2008. "Democracy as a Latent Variable." *American Journal of Political Science* 52: 201-217.
    - Pemstein, Daniel, Stephen A. Meserve, & James Melton. 2010. "Democratic Compromise: A Latent Variable Analysis of Ten Measures of Regime Type." *Political Analysis* 18: 426-449.

- *Exercise Two: Measurement. Collect and create some data.*

- *Lab Session: Stata and R .*

Week 3: September 6 – Visualizing Data. Descriptive Statistics.

- **Readings**

  - *Required*:
    - W & W, Chapter 2.
    - Fox, John. 2008. *Applied Regression Analysis and Generalized Linear Models*, 2nd Ed. Newbury Park, CA: Sage Publications. pp. 28-49.
    - Kastellec, Jonathan P. & Eduardo L. Leoni. 2007. "Using Graphs Instead of Tables in Political Science." *Perspective on Politics* 5: 755-771.

  - *Recommended*:
    - Cleveland, William S. 1985. *The Elements of Graphing Data*. Monterey, CA: Wadsworth.
    - Tukey, J. W. 1977. *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.
    - Tufte, Edward R. 2001. *The Visual Display of Quantitative Information*.
    - Tufte, Edward R. 1990. *Envisioning Information*.
    - Tufte, Edward R. 1997. *Visual Explanations: Images and Quantities, Evidence and Narrative*.
    - Tufte, Edward R. 2006. *Beautiful Evidence*.
    - Jacoby, William G. 2006. "The Dot Plot: A Graphical Display for Labeled Quantitative Values." *The Political Methodologist* 14: 6-14.
    - Weisberg, Herbert F. 1992. *Central Tendency and Variation*. Newbury Park, CA: Sage Publications.

- *Exercise Three: Central tendency and variation. Graphing data.*

- *Lab Session:* Stata *and* R *.*

Week 4: September 13 – Basic Probability Theory: Probability concepts, set theory, conditional probability and independence, Bayes' Theorem.

- **Readings**

  - *Required*:
    - W & W, Chapter 3.
    - Fox, John. 2008. *Applied Regression Analysis and Generalized Linear Models*, 2nd Ed. Newbury Park, CA: Sage Publications. Appendix D, pp. 68-74.
    - Sekhon, Jasjeet S. 2004. "Quality Meets Quantity: Case Studies, Conditional Probability, and Counterfactuals." *Perspectives on Politics* 2: 281-293.

  - *Recommended*:
    - Rudas, Tamás. 2004. *Probability Theory: A Primer*. Newbury Park, CA: Sage Publications.

- *Exercise Three: Basic Probability.*

Week 5: September 20 – Probability Distributions I: Probability distributions, mathematical expectations, variance.

- **Readings**
    - *Required*:
        - W & W, Chapters 4-5.
        - Fox, John. 2008. *Applied Regression Analysis and Generalized Linear Models*, 2nd Ed. Newbury Park, CA: Sage Publications. Appendix D, pp. 75-78.
    - *Recommended*:
        - Evans, Merran, Nicholas Hastings & Brian Peacock. 2000. *Statistical Distributions*. New York: John Wiley & Sons, Inc.

- *Exercise Four: Mathematical expectations and variance. Independence.*

- *Lab Session: Integrals.*

Week 6: September 27 – Probability Distributions II: Binomial, normal, and other distributions, simulation.

- **Readings**
    - *Required*:
        - W & W, Chapter 4.
        - Fox, John. 2008. *Applied Regression Analysis and Generalized Linear Models*, 2nd Ed. Newbury Park, CA: Sage Publications. Appendix D, pp. 79-86.
    - *Recommended*:
        - None

- *Exercise Five: Probability distributions. Simulating probability distributions.*

Week 7: October 4 – Sampling: Sampling schemes, sampling distributions, Central Limit Theorem.

- **Readings**
    - *Required*:
        - W & W, Chapter 6.
        - Fox, John. 2008. *Applied Regression Analysis and Generalized Linear Models*, 2nd Ed. Newbury Park, CA: Sage Publications. Appendix D, pp. 79-86.
    - *Recommended*:
        - Kaltom, Graham. 1983. *Introduction to Survey Sampling*. Newbury Park, CA: Sage Publications.

· Thompson, Steven K. 2002. *Sampling*. New York: Wiley.

- *Exercise Six: Sampling and sampling distributions.*

Week 8: October 11 – Estimation: Point estimators, interval estimators and confidence intervals.

- **Readings**
  - *Required*:
    - · W & W, Chapters 7-8.
    - · Fox, John. 2008. *Applied Regression Analysis and Generalized Linear Models*, 2nd Ed. Newbury Park, CA: Sage Publications. Appendix D, pp. 89-92.
  - *Recommended*:
    - · Pomeranz, Janet Bellcourt. 1982. "Confidence in Confidence Intervals." *Mathematics Magazine* 55(1):12-18.

- *Exercise Six: Estimation.*

Week 9: October 18 – Hypothesis Testing: Elements of a statistical test, Type I and Type II errors, large and small sample tests, p-values, substantive and statistical significance.

- **Readings**
  - *Required*:
    - · W & W, Chapters 9, 17.
    - · Golder, Matt & Gabriella Lloyd. 2011. "Re-evaluating the Relationship between Electoral Rules and Ideological Congruence." Unpublished paper, pp. 1-5.
    - · Gill, Jeff. 1999. "The Insignificance of Null Hypothesis Significance Testing." *Political Research Quarterly* 52(3):647-674.
    - · Achen, Christopher H. 1982. "Interpreting and Using Regression." Newbury Park, CA: Sage. pp. 41-51.
  - *Recommended*:
    - · Wood, Graham R. & David J. Saville. 2002 . "A New Angle on the t-Test." *The Statistician* 51(1):99-104.
    - · Lempert, Richard. 2009. "The Significance of Statistical Significance." *Law and Social Inquiry* 34: 225-249.

- *Exercise Seven: Hypothesis testing.*

Week 10: October 25 – Simple Regression I: PRF and SRF, estimation, numerical properties of OLS, Gauss-Markov Theorem.

- **Readings**
  - *Required*:
    - · Gujarati, Chapters 2-3.
  - *Recommended*:
    - · Berry, William D. 1993. *Understanding Regression Assumptions*. Newbury Park, CA: Sage.
- *Exercise Eight: Simple regression.*

Week 11: November 1 – Simple Regression II: Goodness of fit, regression and normality, statistical inference, prediction.

- **Readings**
  - *Required*:
    - · Gujarati, Chapter 4.
    - · Achen, Christopher. 1977. "Measuring Representation: Perils of the Correlation Coefficient." *American Journal of Political Science* 21: 805-815.
    - · King, Gary. 1991. "How Not to Lie with Statistics: Avoiding Common Mistakes in Quantitative Political Science." *American Journal of Political Science* 30: 666-678.
    - · King, Gary. 1990. "Stochastic Variation: A Comment on Lewis-Beck and Skalaban's 'The R-Squared'." *Political Analysis* 2: 185-200.
  - *Recommended*:
    - · King, Gary. 1991. "Truth is Stranger than Prediction, More Questionable than Causal Inference." *American Journal of Political Science* 35: 1047-1053.
    - · Achen, Christopher H. 1982. *Interpreting and Using Regression*. Newbury Park, CA: Sage.
    - · Fox, John. 1991. *Regression Diagnostics*. Newbury Park, CA: Sage.
- *Exercise Eight: Simple regression and R^2.*

Week 12: November 8 – Multiple Regression I: Frisch-Waugh-Lovell Theorem and partial regression coefficients, omitted variable bias, F tests.

- **Readings**
  - *Required*:
    - · Gujarati, Chapters 6-8, 13 pp. 506-524.

· Beck, Nathaniel. 2011. "Making Regression and Related Output More Helpful to Users." *The Political Methodologist* 18: 4-8.

   ○ *Recommended*:

   · None.

- *Exercise Nine: Multiple regression.*

- *Lab Session: Derivatives.*

Week 13: November 15 – Multiple Regression II: Data transformations, qualitative independent variables, interaction terms.

- **Readings**

   ○ *Required*:

   · Gujarati, Chapter 9.
   · Brambor, Thomas, William Clark, & Matt Golder. 2006. "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis* 14: 63-82.
   · Berry, William D., Matt Golder, & Daniel Milton. 2011. "Improving Tests of Theories Positing Interaction."

   ○ *Recommended*:

   · Kam, Cindy D. & Robert J. Franzese. 2007. *Modeling and Interpreting Interactive Hypotheses in Regression Analysis.* Ann Arbor, MI: University of Michigan Press.
   · Clark, William, Michael Gilligan, & Matt Golder. 2006. "A Simple Multivariate Test for Asymmetric Hypotheses."

- *Exercise Ten: Interaction terms and qualitative independent variables.*

Thanksgiving (November 21-25)

Week 14: November 29 – Multiple Regression III: Measurement error, multicollinearity, heteroskedasticity.

- **Readings**

   ○ *Required*:

   · Gujarati, Chapters 10-11.
   · Golder, Matt & Gabriella Lloyd. 2011. "Re-evaluating the Relationship between Electoral Rules and Ideological Congruence." Unpublished paper.

   ○ *Recommended*:

   · None.

- *Exercise Eleven: Multiple regression.*

Week 15: December 6 – Free week to catch up. Possibly OLS in matrix form, autocorrelation, or nonparametric regression.

- **Readings**
  - *Required*:
    - · None.
  - *Recommended*:
    - · None.