

PL SC 502: “Statistical Methods for Political Research”

Fall 2016

Professor Christopher Zorn
Department of Political Science
Pennsylvania State University
E-mail: zorn@psu.edu
T-Th 11:15 a.m. - 12:30 p.m.
[IST Building](#), Room 201A

Course Description

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are three main goals of the course: to teach students to read and understand quantitative analyses in published and unpublished work, to provide them with the skills necessary to begin conducting their own quantitative analyses, and to lay the foundation for future courses in quantitative methods. This means learning the basics of data collection, organization, and management; measurement; data visualization and display; univariate, bivariate, and multivariate descriptive statistics; sampling; statistical inference (including essentials of probability and distribution theory); and univariate, bivariate, and multivariate hypothesis testing.

There are no formal prerequisites for this class, mathematical or otherwise, and the only expectation is that students are familiar with high-school level algebra and calculus. That said, much of the material in this course is somewhat technical. While I have chosen readings that present the material as clearly and with as little jargon as possible, most of it will still require several readings to fully comprehend.

This syllabus is designed to provide an overview to the course. Note that all course materials (including this syllabus, slides, notes, data, computer code, homework exercises, etc.) will be available on the course Github repo, at <https://github.com/PrisonRodeo/PLSC502-2016-git>. Clickable links are printed in [Penn State blue](#).

Course Readings

Required Text/Materials

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

Additional readings as necessary, all of which will be available on the course github repo and/or through JSTOR.

Strongly Recommended

Either:

Verzani, John. 2014. *Using R For Introductory Statistics*, 2nd Ed. Boca Raton, FL: Chapman & Hall.

or:

Acock, Alan C. 2016. *A Gentle Introduction to Stata*, 5th Ed. Stata Press.

A Few Other Useful/Recommended Readings

Math/Statistics Books

Abelson, Robert P. 1995. *Statistics as Principled Argument*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Altman, Micah, Jeff Gill and Michael McDonald. 2003. *Statistical Computing for the Social Scientist*. New York: Wiley.

Bohrnstedt, George W., and David Knoke. 2004. *Statistics for Social Data Analysis*, 4th Ed. New York: Wadsworth. (Excerpts below are from the third (1994) edition).

Cleveland, William S. 1985. *The Elements of Graphing Data*. Monterey, CA: Wadsworth.

Cuzzort, R. P., and James S. Vrettos. 1996. *The Elementary Forms of Statistical Reason*. New York: St. Martins.

DeGroot, Morris H., and Mark J. Schervish. 2002. *Probability and Statistics*, 3rd Ed. New York: Addison-Wesley.

Fox, John. 2008. *Applied Regression Analysis and Generalized Linear Models*, 2nd Ed. Newbury Park, CA: Sage Publications.

Evans, Merran, Nicholas Hastings and Brian Peacock. 2000. *Statistical Distributions*, 3rd Ed. New York: Wiley.

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

Larsen, Richard J., and Morris L. Marx. 2006. *An Introduction to Mathematical Statistics and its Applications*. 5th ed. Upper Saddle River, NJ: Duxbury.

Tukey, J. W. 1977. *Exploratory Data Analysis*. Reading, MA: Addison-Wesley.

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

Wonnacott, Thomas H., and Ronald J. Wonnacott. 1990. *Introductory Statistics*, 5th ed. New York: John Wiley and Sons.

Books on Software

Crawley, Michael J. 2005. *Statistics: An Introduction Using R*. New York: Wiley.

Daalgaard, Peter. 2002. *Introductory Statistics With R*. New York: Springer.

Everitt, Brian S., and Torsten Hothorn. 2006. *A Handbook of Statistical Analyses Using R*. Boca Raton, FL: Chapman & Hall.

Kohler, Ulrich, and Frauke Kreuter. 2005. *Data Analysis Using Stata*. Stata Press.

Maindonald, John, and John Braun. 2003. *Data Analysis and Graphics Using R: An Example-Based Approach*. New York: Cambridge University Press.

Mitchell, Michael N. 2008. *A Visual Guide to Stata Graphics*, 2nd Ed. Stata Press.

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

Newton, H. Joseph, and Nicolas J. Cox, Eds. 2006. *Thirty-Three Stata Tips*. Stata Press.

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. They also offer a summer program in quantitative (and other) methods, which we'll discuss a bit later in the year.

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.

Finally, a few (other) good places to go looking for social science data include:

- [WorldMapper](#).
- [3stages.org](#).
- McGraw-Hill "[Research Resources](#)" (see also their section on [political science](#)).
- The Federal government's very own <http://www.data.gov/>.

Grading

Grading will be based on ten more-or-less weekly homework exercises (50 points each) and a final paper/poster presentation (500 points). In most instances, exercises will be due seven days from being assigned (that is, they will be assigned on Tuesday and due the following Tuesday). Homework exercises will generally involve analysis and discussion/interpretation of actual data, using statistical computer software (see below). Feel free to work on the assignments in groups of two or three, but you must write up all assignments individually. Details for the homework assignments and the final project will be announced in class.

Also, note that homework exercises and the final paper should be submitted electronically, as PDF files. Note that **only PDF files will be accepted**, without exception. If you do not know how to create a PDF file, go learn.S

Software, Statistical and Otherwise

You are welcome to make use of whatever statistical software you choose to complete the homework exercises, so long as the manner by which your results are generated and conclusions reached are transparent. However, due to the limits of instructor and TA time and patience, we will support only two software packages. Both are available on the machines in the political science computing labs.

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 14.0, but previous versions (back to v. 10, at least)

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
```

Stata newbies may want to check out:

Getting Started with Stata for Windows, Release 14. 2016. 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 effectively the GNU version of the **S** language; as such, it is free (both as in speech and as in beer) and open source; the current version is 2.9.2. **R** is an *object-oriented* language; unlike **Stata** (and most other statistical packages), it operates mostly by assigning values to objects in the workspace. In the notes, 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. I cannot begin to list all the **R**-related resources available on the web; for newbies, however, it might be useful to check out the [Introduction to R](#), [this page](#) in 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](#).

Finally: Don’t use **R**; use [RStudio](#). It is an IDE for **R** that makes using **R** (roughly) 17,311 times easier and more intuitive. It also integrates with number of other valuable, **R**-connected things like [rmarkdown](#), [Shiny](#), [knitr](#), etc.

Other Considerations

In no particular order:

- Your instructor does not have a formally-stated preference for either **Stata** or **R**. My recommendation would be to learn to use both, as each has its strengths and weaknesses. **Stata** has a far flatter learning curve than **R**, which means students tend to gravitate toward it given a choice. But **R** is far more flexible and powerful, and will likely be more useful to you in the long run. Class examples will be conducted exclusively in **R**.

- Learn to use L^AT_EX, 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](#).
- If you insist on using Microsoft Word (or any other WYSIWYG program) for writing assignments, papers, etc., **do not under any circumstances cut and paste graphs from Stata and R into those programs**. Save whatever figures you want to use as .png, .tif, or .jpg files, and import them into the software.

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. 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 of 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.

Other violations include, but are not limited to, any attempt to gain an unfair advantage in regard to an examination, such as tampering with a graded exam or claiming another's work to be one's own. Violations shall also consist of obtaining or attempting to obtain, previous to any examinations, copies of the examination papers or the questions to appear thereon, or to obtain any illegal knowledge of these questions. Lying to the instructor or purposely misleading any Penn State administrator shall also constitute a violation of academic integrity.

In cases of a violation of academic integrity it is the policy of the Department of Political Science to impose appropriate penalties that are consistent with University guidelines. See [here](#) for more details.

Note to students with 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 at: <http://equity.psu.edu/student-disability-resources/>. Instructors should be notified as early in the semester as possible regarding the need for reasonable accommodations.

Course 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. Links are generally to DOIs or to stable PDFs at JSTOR. I won't assign readings from either Verzani (2005) or Acock (2008), but students should consult the relevant parts of those texts for software guidance (depending on whether they are using R or Stata, respectively).

August 23: Introduction, Overview, etc.

- Readings

- *Required:*
 - None.
- *Recommended:*
 - None.

August 25: No Class

August 30: How To Read Tables and Figures

- Readings

- *Required:*
 - Gomez, Brad T., Thomas G. Hansford, and George A. Krause. 2007. "The Republicans Should Pray for Rain: Weather, Turnout, and Voting in U.S. Presidential Elections." *Journal of Politics* 69(3):649-663.
 - Rosh, Robert M. 1987. "Ethnic Cleavage as a Component of Global Military Expenditures." *Journal of Peace Research* 24(1):21-30.
- *Recommended:*
 - Apodaca, Clair, and Michael Stohl. 1999. "United States Human Rights Policy and Foreign Assistance." *International Studies Quarterly* 43(1):185-198
 - Holbrook, Thomas M., and Emily Van Dunk. 1993. "Electoral Competition in the American States." *American Political Science Review* 87(4):955-962
 - Segal, Jeffrey A., and Albert D. Cover. 1989. "Ideological Values and the Votes of U.S. Supreme Court Justices." *American Political Science Review* 83(2):557-565.

September 1: Data: Structure and Measurement

- **Readings**

- *Required:*

- Agresti and Finlay, pp. 1-7, 11-14.
 - Nagler, Jonathan. 1995. “Coding Style and Good Computing Practices.” *The Political Methodologist* 6(2):2-8.

- *Recommended:*

- Jacoby, William G. 1999. “Levels of Measurement and Political Research: An Optimistic View.” *American Journal of Political Science* 43(1):271-301. (Read the first few sections, if you like).

- *Exercise One: Measure, collect, and create some data.*

September 6: Descriptive Statistics: Univariate Graphics

- **Readings**

- *Required:*

- Agresti and Finlay, pp. 31-38.
 - Fox (2008), pp. 28-40.

- *Recommended:*

- Cleveland (1985), pp. 123-153.
 - Jacoby, William G. 2006. “The Dot Plot: A Graphical Display for Labeled Quantitative Values.” *The Political Methodologist* 14(1):6-14.

September 8: Descriptive Statistics: Bivariate and Multivariate Graphics

- **Readings**

- *Required:*

- Fox (2008), pp. 40-49.

- *Recommended:*

- Jacoby, William G. 1997. *Statistical Graphics for Univariate and Bivariate Data*. Thousand Oaks, CA: Sage Publications.
 - Jacoby, William G. 1998. *Statistical Graphics for Visualizing Multivariate Data*. Thousand Oaks, CA: Sage Publications.

- *Exercise Two: Graphing data.*

September 13: Descriptive Statistics: Measures of Central Tendency

- Readings

- *Required:*
 - Agresti and Finlay, pp. 38-46.
- *Recommended:*
 - Groeneveld, Richard A., and Glen Meeden. 1977. “[The Mode, Median, and Mean Inequality.](#)” *The American Statistician* 31(3):120-21.

September 15: Descriptive Statistics: Measures of Dispersion and Higher Moments

- Readings

- *Required:*
 - Agresti and Finlay, pp. 46-55.
 - *Recommended:*
 - Weisberg, Herbert F. 1992. *Central Tendency and Variation*. Newbury Park, CA: Sage Publications.
- *Exercise Three: Central tendency and variation.*

September 20: Introduction to Probability

- Readings

- *Required:*
 - Agresti and Finlay, pp. 73-75.
 - Fox (2008), [Appendix D.](#), pp. 65-68.
- *Recommended:*
 - Rudas, Tamás. 2004. *Probability Theory: A Primer*. Newbury Park, CA: Sage Publications.

September 22: Random Variables

- Readings

- *Required:*
 - Agresti and Finlay, pp. 75-85.

- Fox (2008), [Appendix D.](#), pp. 68-74.
- *Recommended:*
 - None.

September 27: Probability Distributions, I

• Readings

- *Required:*
 - Fox (2008), [Appendix D.](#), pp. 75-78.
- *Recommended:*
 - Evans, Merran, Nicholas Hastings and Brian Peacock. 2000. *Statistical Distributions*, 3rd Ed. New York: Wiley.

September 29: Probability Distributions, II

• Readings

- *Required:*
 - Fox (2008), [Appendix D.](#), pp. 79-86.
- *Recommended:*
 - None.
- *Exercise Four: Distributions.*

October 4: Randomization and Sampling

• Readings

- *Required:*
 - Agresti and Finlay, pp. 15-25.
- *Recommended:*
 - Kaltom, Graham. 1983. *Introduction to Survey Sampling*. Newbury Park, CA: Sage Publications.
 - Thompson, Steven K. 2002. *Sampling*, 2nd Ed. New York: Wiley.

October 6: Sampling Distributions

- Readings

- *Required:*
 - Agresti and Finlay, pp. 58-61, 85-99.
- *Recommended:*
 - None.

- *Exercise Five: Sampling.*

October 11: Estimation and Estimators

- Readings

- *Required:*
 - Agresti and Finlay, pp. 107-110.
 - Fox (2008), [Appendix D.](#), pp. 89-92.
- *Recommended:*
 - Fisher, R. A. 1922. “[On the Mathematical Foundations of Theoretical Statistics.](#)” *Philosophical Transactions of the Royal Society of London, A*, 222:309-368.

October 13: Statistical Inference: Confidence Intervals

- Readings

- *Required:*
 - Agresti and Finlay, pp. 109-123.
- *Recommended:*
 - Pomeranz, Janet Bellcourt. 1982. “[Confidence in Confidence Intervals.](#)” *Mathematics Magazine* 55(1):12-18.

October 18: Statistical Inference: Hypothesis Testing

• Readings

- *Required:*
 - Agresti and Finlay, pp. 147-174.
- *Recommended:*
 - Gill, Jeff. 1999. “The Insignificance of Null Hypothesis Significance Testing.” *Political Research Quarterly* 52(3):647-674.
- *Exercise Six: Statistical inference.*

October 20: Two-Group Comparisons, I

• Readings

- *Required:*
 - Agresti and Finlay, pp. 183-193.
- *Recommended:*
 - None.

October 25: Two-Group Comparisons, II

• Readings

- *Required:*
 - Agresti and Finlay, pp. 193-209.
- *Recommended:*
 - Wood, Graham R., and David J. Saville. 2002 . “A New Angle on the t -Test.” *The Statistician* 51(1):99-104.
- *Exercise Seven: Conduct two-group comparisons.*

October 27: Measures of Association: Nominal Variables

• Readings

- *Required:*
 - Agresti and Finlay, pp. 221-229.
- *Recommended:*
 - None.

November 1: Measures of Association: Binary Variables

- **Readings**

- *Required:*
 - Agresti and Finlay, pp. 229-239.
- *Recommended:*
 - Bohrnstedt and Knoke (1994), pp. 155-181.
- *Exercise Eight: Measures of association.*

November 3: Measures of Association: Ordinal Variables

- **Readings**

- *Required:*
 - Agresti and Finlay, pp. 239-246.
- *Recommended:*
 - None.

November 8: Measures of Association: Interval/Ratio Variables

- **Readings**

- *Required:*
 - Agresti and Finlay, pp. 255-259, 269-273..
- *Recommended:*
 - Bollen, Kenneth A., and Kenney H. Barb. 1981. “[Pearson’s \$r\$ and Coarsely Categorized Measures.](#)” *American Sociological Review* 46(2):232-239.
- *Exercise Nine: Measures of association, redux.*

November 10: Linear Regression, I

- **Readings**

- *Required:*
 - Agresti and Finlay, pp. 259-268.
- *Recommended:*
 - Bohrnstedt and Knoke (1994), pp. 191-204.

November 15: Linear Regression, II

- **Readings**

- *Required:*

- Agresti and Finlay, pp. 273-279.

- *Recommended:*

- Kahane, Leo H. 2001. *Regression Basics*. Thousand Oaks, CA: Sage Publications.

- *Exercise Ten: Linear least-squares regression.*

November 17: Linear Regression, III

- **Readings**

- *Required:*

- Agresti and Finlay, pp. 280-289.

- *Recommended:*

- Kahane, Leo H. 2001. *Regression Basics*. Thousand Oaks, CA: Sage Publications.

- *Exercise Ten: Linear least-squares regression.*

November 22 & 24: No Class – Thanksgiving Break

November 29: Nonlinearity and Data Transformations

- **Readings**

- *Required:*

- Fox (2008), pp. 50-68.

- *Recommended:*

- Bohrnstedt and Knoke (1994), pp. 321-332.

December 1: Bayesian Inference: A Very Brief Introduction

• Readings

- *Required:*
 - Fox (2008), [Appendix D.](#), pp. 101-106.
- *Recommended:*
 - Gill, Jeff. 2007. *Bayesian Methods: A Social and Behavioral Sciences Approach*, 2nd Ed. London: Chapman & Hall.

December 6: Multivariate Statistics: A Preview

• Readings

- *Required:*
 - Agresti and Finlay, pp. 301-314.
- *Recommended:*
 - None.

December 8: Catch-up, Wrap-Up, and Review

• Readings

- *Required:*
 - Tufte, Edward R. 1969. “[Improving Data Analysis in Political Science.](#)” *World Politics* 21(4):641-54.
- *Recommended:*
 - None.