

As of September 9, 2009. Subject to change.

WWS 507B: Quantitative Analysis

Woodrow Wilson School
Princeton University
Fall 2009

Instructor: Sam Schulhofer-Wohl, ssschulho@princeton.edu
Preceptor: Penka Kovacheva, pkovache@princeton.edu

All course materials will be posted on the class Web page on Blackboard
(blackboard.princeton.edu).

1 Class times and office hours

Lectures: 9 to 10:30 a.m. Mondays and Wednesdays in Robertson 016. There will be no class Sept. 28 because of Yom Kippur. There will be a makeup class from 6:30 to 8 p.m. Oct. 1 in Robertson 016.

Precepts: 7 to 8:30 p.m. Tuesdays and Thursdays in Robertson 023. The Thursday meeting and the *following* Tuesday meeting (for example, Sept. 24 and Sept. 29) will usually cover the same material, with some exceptions listed on the course outline; attend whichever meeting you prefer. There will be no precept Sept. 22 or Oct. 1. There will be a makeup session to replace the Oct. 1 precept at a time to be announced.

Sam's office hours: 2 to 4 p.m. Wednesdays in 363 Wallace, or by appointment. (Please make appointments through wass.princeton.edu; e-mail me if you cannot find an acceptable time in the online system.)

Penka's office hours: 4 to 6 p.m. Tuesdays in B10 Fisher, or by appointment.

Penka and I are here to help you. We encourage you to come to office hours whenever you have questions.

2 Course description

This course is an introduction to empirical methods commonly used in analyzing public policy. We will discuss how social scientists employ statistical tools to answer policy questions.

Example: Does sending nonviolent criminals to prison (instead of, say, putting them on probation) create a stigma that will reduce their chances of finding legitimate work when they are released? Can we answer this question by taking a national survey of former inmates and people who have never been to prison, and comparing their unemployment rates?

Objectives: By the end of the course, you should be able to perform basic empirical analyses and critically evaluate and explain basic empirical work done by other people.

Prerequisites: This course requires no background in statistics. It does, however, require *fluency in high school mathematics through the level of precalculus*, as well as basic familiarity with calculus concepts. Students who are proficient in calculus may prefer WWS 507C. Please see me or Professor Taryn Dinkelman, the instructor for WWS 507C, if you have questions about which course is appropriate for you.

3 Grading and policies

Final exam: 35%. Date and time TBA. Closed book, but you may bring one page of notes in your own handwriting.

Midterm: 20%. Take-home: You may complete the exam during any continuous four-hour period between Oct. 29 and Nov. 9; due at the beginning of class Nov. 9. Closed book, closed notes.

Projects: 30%. Project 1: distributed Oct. 28, due at the beginning of class Nov. 23. Project 2: distributed Dec. 7, due at 5 p.m. Jan. 20. You may work with classmates on the statistical analysis but must write up your own reports.

Problem sets: 15%. Approximately eight, distributed most Thursday evenings and due at 5 p.m. the following Thursday in the course mailbox in the basement of Fisher Hall. You may work with classmates but must write up your own answers.

I will not accept late work except in cases of documented medical or family emergency. If you require accommodations for a disability, please see me as soon as possible.

4 Readings

There are two required textbooks:

- Freedman, David, Robert Pisani and Roger Purves, 2007, *Statistics*, fourth edition, W.W. Norton; denoted FPP on the reading list.
- Stock, James H., and Mark W. Watson, 2007, *Introduction to Econometrics*, second edition, Addison-Wesley; denoted SW on the reading list.

Several alternative books are on reserve in Stokes Library if you do not find FPP and SW helpful: Ashenfelter/Levine/Zimmerman, *Statistics and Econometrics*; De Veaux/Velleman/Bock, *Stats: Data and Models*; and Moore/McCabe/Craig, *Introduction to the Practice of Statistics*.

Darrell Huff's classic *How to Lie With Statistics*, which gives an entertaining overview of some important topics in the course, is also on reserve. I encourage you to read it early in the semester.

Problem sets and projects will require you to use the Stata statistical software package, which is available in campus computer labs. We will provide basic Stata training in precepts. If you want to learn more, a useful book is Acock, *A Gentle Introduction to Stata*, which is also on reserve.

We will also read some articles from the press and from research journals. Copies are available through e-reserves on Blackboard.

5 Course outline

Please look at each journal article before the corresponding lecture. You do not need to wade through the technical details of every article. Rather, you should read the introduction to find out what research question the article seeks to answer and what conclusion the article reaches. Then – at minimum – skim the rest of the article to try to understand what data and methods are used to answer the question.

A. Motivation and descriptive statistics.

Lecture 1 (Sept. 21): Experimental and observational studies. FPP 1.1-1.3, 2.1-2.5; optional: SW 1.1-1.3.

- Pager, Devah, 2003, “The Mark of a Criminal Record,” *American Journal of Sociology* 108(5), 937-975.

Lecture 2 (Sept. 23): Describing one variable: histogram, central tendency and dispersion. FPP 3.1-3.4, 4.1-4.2, 4.5-4.6.

Sept. 24 and 29 precepts: summary statistics and histograms in Stata.

Lecture 3 (Sept. 30): Describing relationships between two variables: scatter plots, covariance and correlation. FPP 7.1-7.5 (review on your own), FPP 8.1-8.2, 8.4, 9.1-9.5.

- Yule, G. Udny, 1926, “Why do we Sometimes get Nonsense-Correlations between Time-Series? – A Study in Sampling and the Nature of Time-Series,” *Journal of the Royal Statistical Society* 89(1), 1-63. (Just read Section I.)

Lecture 4 (Oct. 1 makeup): Describing relationships between two variables: linear regression. FPP 10.1-10.3, 10.5, 12.1-12.3; pp. 123-124 of SW 4.3.

- Fortson, Jane G., 2008, “The Gradient in Sub-Saharan Africa: Socioeconomic Status and HIV/AIDS,” *Demography* 45(2), 303-322.

Oct. 1 and 6 precepts: scatterplots, correlations and regressions in Stata.

B. Probability, random sampling and sampling distributions.

Lecture 5 (Oct. 5): Probability theory: unconditional probability, conditional probability, independence. FPP 13.1-13.4, 14.1-14.3; optional: SW 2.1.

Lecture 6 (Oct. 7): Data collection and data quality. FPP 19.1-19.8.

- McNeil, Donald G., Jr., 2007, “U.N. Agency to Say It Overstated Extent of H.I.V. Cases by Millions,” *The New York Times*, Nov. 20, page A1.
- Chandra, Amitabh, 2003, “Is the Convergence in the Racial Wage Gap Illusory?” NBER Working Paper 9476. (Table 5 has the key numbers.)

Oct. 8 and 13 precepts: practice with probability.

Lecture 7 (Oct. 12): Expectation and variance of the sample mean. FPP 16.1-16.4, 17.1-17.2, 20.1-20.2; optional: SW 2.1, 2.3, 3.1.

Lecture 8 (Oct. 14): Distribution of the sample mean: the normal distribution and the central limit theorem. FPP 5.1-5.2, 17.3, 18.1-18.6, 20.3; optional: SW 2.5-2.6.

Oct. 15 and 20 precepts: practice with the normal distribution.

C. Statistical inference.

Lecture 9 (Oct. 19): The precision of an estimate: standard errors and confidence intervals for the mean. FPP 21.1-21.4, 23.1-23.4; SW 3.3.

Lecture 10 (Oct. 21): Hypothesis testing. Hypothesis tests about the mean. FPP 26.1-26.4; SW 3.2.

Oct. 22 and 27 precepts: practice with confidence intervals and hypothesis tests about the mean in Stata.

Lecture 11 (Oct. 26): Comparing two populations. FPP 27.1-27.5; SW 3.4-3.5; Pager 2003.

Lecture 12 (Oct. 28): Bayesian inference. Handouts.

- Lempert, Richard, 1991, “Some Caveats Concerning DNA as Criminal Identification Evidence: With Thanks to the Reverend Bayes,” *Cardozo Law Review* 13(2-3), 303-341.

Oct. 29 precept (Thursday only): midterm review.

Fall break and midterm

Nov. 10 precept (Tuesday only): advice on Project 1.

D. Causal and statistical inference in simple linear regression.

Lecture 13 (Nov. 9): Assumptions of the classical linear regression model. Unbiasedness and consistency. SW 4.4-4.5.

Lecture 14 (Nov. 11): Statistical inference in simple linear regression. SW 5.1-5.4.

Nov. 12 and 17 precepts: practice with interpreting simple linear regressions.

Lecture 15 (Nov. 16): Endogeneity: omitted variable bias and measurement error. FPP 10.4; SW 6.1 and pp. 319-321 of 9.2.

- Case, Anne, and Christina Paxson, 2008, “Stature and Status: Height, Ability, and Labor Market Outcomes,” *Journal of Political Economy* 116(3), 499-532.

E. Solving the endogeneity problem with multiple regression.

Lecture 16 (Nov. 18): Solution 1: multiple regression. Interpreting multiple regression coefficients. SW 6.2-6.3; pp. 212-213 of FPP 12.3; Case/Paxson 2008.

Nov. 19 and 24 precepts: practice with omitted variables problems.

Lecture 17 (Nov. 23): Mechanics of multiple regression. SW 6.4-6.7.

Lecture 18 (Nov. 25): Nonlinear relationships: logs and quadratics. SW 8.1-8.2.

- Krueger, Alan B., 1993, “How Computers Have Changed the Wage Structure: Evidence from Microdata, 1984-1989,” *Quarterly Journal of Economics* 108(1), 33-60.

Nov. 26 and Dec. 1 precepts: estimating and interpreting multiple regressions in Stata.

Lecture 19 (Nov. 30): Qualitative data: categorical variables, dummy variables and interactions. SW 5.3, 8.3; Krueger 1993.

Lecture 20 (Dec. 2): Inference in multiple regression. SW 7.1-7.3.

- Duflo, Esther, 2003, “Grandmothers and Granddaughters: Old-Age Pensions and Intrahousehold Allocation in South Africa,” *World Bank Economic Review* 17(1), 1-25.

Dec. 3 and 8 precepts: practice with multiple regression.

Lecture 21 (Dec. 7): Interpreting and assessing regression results. SW 9.1-9.2.

- DiNardo, John E., and Jorn-Steffen Pischke, 1997, “The Returns to Computer Use Revisited: Have Pencils Changed the Wage Structure Too?” *Quarterly Journal of Economics* 112(1), 291-303.

F. Additional methods for solving the endogeneity problem.

Lecture 22 (Dec. 9): Solution 2: instrumental variables. SW 12.1, 12.3-12.5.

- Angrist, Joshua D., 1990, “Lifetime Earnings and the Vietnam Era Draft Lottery: Evidence from Social Security Administrative Records,” *American Economic Review* 80(3), 313-336.

Dec. 10 precept (Thursday only; optional): instrumental variables in Stata.

Dec. 15 precept (Tuesday only): advice on Project 2.

Lecture 23 (Dec. 14): Solution 3: fixed effects. SW 10.1-10.2.

- Levitt, Steven D., 1994, “Using Repeat Challengers to Estimate the Effect of Campaign Spending on Election Outcomes in the U.S. House,” *Journal of Political Economy* 102(4), 777-798.

Lecture 24 (Dec. 16): Solution 4: difference in differences. SW 13.1-13.3; optional: SW 13.4-13.7.

- Card, David, 1990, “The Impact of the Mariel Boatlift on the Miami Labor Market,” *Industrial and Labor Relations Review* 43(2), 245-257.

Dec. 17 precept (Thursday only; optional): panel data in Stata.