Syllabus for LAW209.3, Introductory Statistics

Overview

This course is designed to introduce students to the use of statistics in the law. On the theoretical side, the course will cover statistical concepts that are important for critically evaluating and undertaking research involving quantitative methods. On the practical side, the course will include instruction in how to compute simple statistical quantities using software. These two approaches enjoy synergies. In my experience, statistics intimidates students up until the point that they try to apply the methods to real-world data—at which point they recognize that the ideas are actually quite simple. A focus for the class will be to review statistical approaches that are commonly encountered in the law.

Requirements

The semester grade will be based on a series of problem sets (10%), a midterm (30%), and a final (60%). Both exams will be open notes and open book. The midterm will be held in class on XXX.

Unless otherwise stated in class, problem sets will be handed out each week on Wednesday and will be due back the following Monday. Most problem sets will be based on computer software exercises. You may work with others on problem sets, but each student is required to turn in a separate problem set. Students are expected to attend and participate in all classes.

Readings

There are two assigned texts for the course. The first is David Freedman, Robert Pisani, and Roger Purves ("FPP") *Statistics* (4th edition), W.W. Norton, 2007. The second is Jeffrey M. Wooldridge ("W"), *Introductory Econometrics: A Modern Approach* (4th edition), Thomson South-Western, 2008. As appropriate, additional readings will be available at http://bspace.berkeley.edu.

The late David Freedman, a long-time Berkeley faculty member, was a giant of the statistics profession, and active as an expert witness for many years on topics as varied as employment discrimination and redistricting. His textbook *Statistics* is probably the clearest exposition of the introductory ideas of statistics available in any language and is likely the most commonly assigned undergraduate textbook on statistics. Possibly even to a fault it strives to communicate the core ideas with a minimum of mathematics. Jeff Wooldridge is a prominent econometrician and is particularly well-known for his devotion to pedagogy and his clarity of exposition. His textbook *Introductory Econometrics* is a staple of second semester undergraduate econometrics courses and sometimes assigned in the first semester of a Ph.D. program sequence in econometrics.¹

To maintain a concrete focus, we will organize our discussion of statistics around substantive discussions, introducing the statistical concepts along the way as needed. See below for an outline of the substantive topics we will cover.

GSI and Problem Sets

The Graduate Student Instructor (GSI) for the course is Ryan Copus. Ryan will run a section and hold office hours. Section will review problem sets and student questions and may additionally amplify concepts from lecture students found confusing.

Class Cancellations

There will be no class on November 11, as I am attending the annual Southern Historical Association meetings. I will post a videotaped lecture.

Statistical Software

"So whats getting ubiquitous and cheap? Data. And what is complementary to data? Analysis. So my recommendation is to take lots of courses about how to manipulate and analyze data: databases, machine learning, econometrics, statistics, visualization, and so on."

-Google Chief Economist, Hal Varian, New York Times, February 25, 2008

An essential part of learning statistics these days is familiarizing yourself with statistical software. Stata and R are the two most commonly used programs in the law school, and either is acceptable in this class. Although you will *not* be tested on your ability to code, R or Stata will be essential for completing problem sets.

You will at times struggle with coding. A simple misplaced comma might cost you an hour or more, particularly if you don't ask for help (read: ask your fellow students and instructors!). Please try to fight your way through it. R and Stata, like all software packages, have notoriously steep learning curves, but a little time and a lot of patience can yield extraordinary benefits.

¹The distinction between statistics and econometrics is slight. Some people say that econometrics is the study of data that result from observational data rather than experimental data. This has a kernel of truth, but for our purposes it does no harm to think of them as being the same field. People also speak of polimetrics, sociometrics, cliometrics, and (my favorite) psychometrics.

To help you choose between R and Stata, I've tried to quickly and informally sum up each package: Stata: A very well put together, professional statistics package. It tends to be used heavily by economists. STATA is generally regarded as more user friendly than R. There are a couple of different flavors of Stata available for purchase through the grad plan:

http://www.stata.com/order/new/edu/gradplans/us-pickup/

You can get away with using Stata/IC if you only are going to use it for class. It costs \$65 for a semester (the cheapest feasible option). If you think you might be interested in doing empirical work yourself, I would strongly suggest getting a perpetual Stata/SE license instead. This is more expensive—\$395. Ask other students which option they found most useful. If you decide to go the "Stata route", then you might go ahead and purchase before class starts to hit the ground running.

R: Free, open source statistics software. To download it, go to

http://www.r-project.org/.

Do not worry about downloading anything now. You will discuss and download software packages in your first section.

Also, for either package, don't bother purchasing documentation. Documentation is available online, and these days a few Goggle searches tells you what the right syntax is. There are many, many extremely clever people using these packages. Some of them apparently have enough time on their hands that they post what they do with them, and the rest of us free ride on their expertise and desire to alert the world.

Substantive Topics

As noted, to keep things lively, we will use substantive topics to motivate the introduction of statistical concepts. Substantive topics we will touch upon:

- 1. Social progress and inequality: What has happened to real wages in the U.S. over time? What about the top of the wage distribution? What about the bottom? Do people bear income and wealth risk cross-sectionally or intertemporally? How quickly is China's economy growing?
- 2. Health: Why are some people healthier than others? Cuba has an infant mortality much better than one would expect based on income. Why? What is the connection between smoking and lung cancer? What is behind the rise in obesity in the U.S.?
- 3. Group inequality: Some groups within society shoulder particular burdens or enjoy particular fruits corresponding to wealth, health, and justice. Some of these differences are stable over time and others change quickly.
- 4. Financial markets: How do they operate? Can we use them to predict the results of the upcoming presidential election?
- 5. Crime: It goes up, it goes down. Does anybody know why? Is it about to go up again? What ever happened to the "super predators"?

Draft Outline of Lecture Schedule: Substantive Topics

- Weeks 1, 2: Measuring progress and measuring inequality. Changes in the wage distribution over time in the U.S. and globally. The role of supply and demand factors and the role of institutions. Free agency in professional sports and executive compensation. Implications for fiduciary duties of members of boards of directors and compensation committees.
- Week 3: More on measurement. Application to health over time in the U.S. and globally. A 1st conversation regarding causality: is there a link between infant mortality and GDP? What do we mean by that? Is it a question of prediction or causality? What is meant by either notion?
- Week 4: Causality and interventions. Education as an intervention to minimize inequality. An introduction to quasi-experimental research designs: twin differences in education and wages. Efforts to improve birthweight as a means of improving survival. Twin differences in birthweight and infant mortality.
- Week 5: Prediction from data. Principles of forecasting. The 2012 presidential election and prediction markets (e.g., http://www.intrade.com and http://tippie.uiowa.edu/iem/). Arbitrage. Event studies: what happened to the prediction market for Obama after the June 2012 health care ruling?
- Weeks 6, 7: Inequality in wages between groups and its evolution over time. Implications for employment discrimination litigation. Statistical tests used in employment discrimination. Discussion of Piva v. Xerox and related cases. Are log wages distributed normally? Does it matter?
- Week 8: Employment discrimination litigation and financial markets: Xerox, Duke Power, and Wal-Mart. Event studies again. William Fisher (2005), Does the Efficient Market Theory Help us Do Justice in a Time of Madness?, Emory Law Journal.
- Week 9: Discussion of financial markets and 10(b)5 suits. Event studies. Regression and the capital asset pricing model (CAPM). Are stock returns distributed normally? Does it matter?
- Week 10: Regression analysis and estimation of damages in antitrust litigation. Discussion of the role of expert witnesses in litigation. Influential observations. The specter of specification searching. Burdens of proof. Interpretation of regression results.
- Week 11: Back to financial markets: the 2010 Flash Crash and the role of high frequency traders. The SEC's mandate that stock and commodity exchanges change from fractional quotes (i.e., 1/16ths of a dollar) to decimal quotes (i.e., pennies). The ongoing discussion regarding subdecimalization. Another quasi-experimental research design: the regression discontinuity design.
- Week 12: Back to inequality: inequality in criminal justice outcomes between groups and its evolution over time. Interpretation of coefficients on dummy variables and coefficients on interactions between dummy variables. Discussion of *McCleskey v. Kemp*.
- Week 13: Trends in crime. Causes of crime: demographics, poverty, education, sentencing policy, police, lead, abortion. Measurement error in regression. How to read tables from sophisticated analyses. A third quasi-experimental research design: instrumental variables.
- Week 14: Analyzing Supreme Court decisions quantitatively. Predicting votes. Connections to item response theory. Priest Klein hypothesis. Publication rates of federal district court and appeals court decisions.
- Week 15: Explosion of data on the internet and implications for the future of empirical work. Privacy.

Draft Outline of Lecture Schedule: Statistical Topics

What follows is a list of core statistical concepts, coupled with a list of the weeks during which they will come up.

- 1. multiple regression
 - Weeks 9, 10, 11, 12, 13
- 2. differences in means and proportions; connections between differencing schemes and regression Weeks 4, 6, 7, 12, 13
- 3. categorical data Weeks 12, 13, 14
- 4. maximum likelihood estimation Weeks 9, 12, 13, 14
- 5. probability distributions Weeks 1, 2, 3, 6, 7, 11
- 6. sampling distributions, confidence intervals, and tests of significance Weeks 1, 2, 3, 4, 6, 7, 10
- 7. sampling schemes Weeks 1, 2, 3, 6, 7, 15
- 8. selection bias Weeks 6, 7, 11, 14, 15