Syllabus
209.3 Quantitative Methods I

Professors: Kevin Quinn
GSI: Ashley Rubin
Fall Semester 2010

Class Room
2240 Piedmont
8:30am - 11:50am
Wednesday

Office
Kevin Quinn
Simon 490
Phone: (510) 642-2485
Email: `kquinn@law.berkeley.edu`

Ashley Rubin
Email: `poenny@gmail.com`

Preliminaries

Overview and Course Goals
The goal of this course is to provide students with enough background in probability and statistics so that they can successfully:

- evaluate basic quantitative empirical research in law and social science
- *begin* to conduct their own empirical research
- take more advanced quantitative methods courses to further develop their skills.

The course focuses on the general *concepts* that underly statistical inference rather than on specific *techniques*. The hope is that students who successfully complete the course will be able to think clearly about a wide range of substantive problems. That said, ***this course is a starting point***. It is simply not possible to cover what the typical empirical researcher should know about probability and statistics in one course. Students who plan to do empirical research should take several additional methodology courses.

Prerequisites
A willingness to work hard and learn by doing.

Class Requirements
Students are expected to complete the required reading for each week before the week's lecture. Don't be fooled by the fact that the reading has relatively little in the way of Greek letters and mathematical symbols. The required reading for most weeks will deal very difficult concepts and should be read carefully. Make a note of questions that arise during the reading. If your questions aren't answered during the lecture please ask an instructor to clarify– either during the section meeting, office hours, or via email.

There will be approximately 10 problem sets handed out over the course of the semester. These will be graded on a pass/fail basis and will make up 50% of the final grade. Students are encouraged to work on the problem sets in small groups of say 3-4 students. ***However, each student's written answers must be his/her own work.*** A take-home final examination will make up 40% of the final grade. Class participation will account for the last 10% of the final grade.

We will not give incompletes in this course.

Computation
The primary package that we will be using in this course is R. You can download R from `http://www.r-project.org/`.

Learning to use R may be frustrating at first. Nevertheless, I encourage you to stick with it. A moderate initial investment of your time will pay large rewards later. There are several good online references for R. These can be found at the R Project website listed above. The "Introduction to R" is especially well written. You can find this at: `http://cran.r-project.org/manuals.html`. In addition, the Verzani book (see below) is a good resource for learning and using R.

Office Hours and Availability
Professor Quinn will hold office hours on Thursdays from 9:30 am to 12:15 pm. Ashley Rubin will hold office hours from ??? to ??? on ???.

We will also try to answer questions by email whenever possible. If the topic of the question is relevant to the class as a whole and others can learn from the question and response, we will forward the question and answer to the rest of the class. *If you would like identifying information stripped out of your email or you do not want your question forwarded please say so explicitly in your message.*

Course Website
The course website is availabe via bSpace.

Required Books

David Freedman, Robert Pisani, and Roger Purves. 2007. *Statistics.* 4th Edition. New York: W.W. Norton & Company.

John Verzani. 2005. *Using R for Introductory Statistics.* Boca Raton: Chapman & Hall/CRC.

We'll refer to Freedman, Pisani, and Purves as "FPP".

Optional Books

The following books are optional but may prove useful to students looking for a more comprehensive treatment of some of the course topics. The books are listed in roughly increasing order of difficulty / sophistication.

Christopher Achen. 1982. *Interpreting and Using Regression.* Thousand Oaks, CA: Sage.

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

William Cleveland. 1993. *Visualizing Data.* Summit, NJ: Hobart Press.

Robert S. Pindyck and Daniel L. Rubinfeld. 1998. *Econometric Models and Economic Forecasts.* 4th Edition. McGraw Hill.

John E. Freund and Benjamin M. Perles. 2007. *Modern Elementary Statistics.* 12th Edition. Prentice Hall.

Morris H. DeGroot and Mark J. Schervish. 2002. *Probability and Statistics.* 3rd Edition. Addison Wesley.

Joshua D. Angrist and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics.* Princeton: Princeton University Press.

George Casella and Roger L. Berger. 2001. *Statistical Inference*. 2nd Edition. Duxbury Press.

Preliminary Schedule

The following is a preliminary schedule of course topics. It is a rough guide to what we will be covering and may well undergo some changes over the semester. There is a lot of material here and we may not cover it all.

August 18: Introduction and Overview

- probability vs. statistics
- probability and statistics in social science and law
- descriptive, predictive, and causal inference
- samples and populations
- large-n vs. small-n research
- experimental vs. observational data
- introduction to computing with R

Required Reading

FPP: Chapters 1 & 2

Verzani: Preface, Chapter 1, Appendix A

August 25: Descriptive Statistics I

- the boxplot
- the histogram
- sample quantiles
- the sample mean
- the sample standard deviation

Required Reading

FPP: Chapters 3 & 4

Verzani: Chapter 2

September 1: Descriptive Statistics II

- plotting bivariate data
- plotting multivariate data
- the conditional sample mean
- the conditional sample standard deviation
- cross-tabulations

Required Reading

FPP: Chapter 7

Verzani: Sections 3.1, 3.2, 3.3, and Chapter 4

Optional Reading

Verzani: Appendix D

Cleveland, William S. and Robert McGill. 1987. "Graphical Perception: The Visual Decoding of Quantitative Information on Graphical Displays of Data." (with discussion) *JRSS A*. 150: 192-229.

Cleveland, William S. and Robert McGill. 1984. "Graphical Perception: Theory, Experimentation, and Application to the Development of Graphical Methods." *JASA*. 79: 531-554

Cleveland, William S.; Persi Diaconis; and Robert McGill. 1982. "Variables on Scatterplots Look More Highly Correlated When the Scales are Increased." *Science*. 216: 1138-1141

Trellis Display homepage at Bell Labs
`http://cm.bell-labs.com/cm/ms/departments/sia/project/trellis/`

September 8: Probability I

- what is probability?
- conditonal probabilities
- statistical independence
- the multiplication rule

Required Reading

FPP: Chapter 13

People v. Collins, 438 P. 2d 33. Available on course website.

September 15: Probability II

- the addition rule
- the binomial formula
- random variables

Required Reading

FPP: Chapters 14 & 15

September 29: Probability III

- discrete distributions
- continuous distributions
- the distribution function
- bivariate distributions
- marginal distributions
- conditional distributions
- statistical independence
- Bayes' theorem

Required Reading

DeGroot and Schervish. 2002. *Probability and Statistics*. 3rd Edition. Chapter 3. (Handout)

Verzani: Section 5.1

Optional Reading

Lindsey A. Foreman, Adrian F. M. Smith, and Ian W. Evett. 1997. "Bayesian Analysis of DNA Profiling Data in Forensic Identification Applications". *Journal of the Royal Statistical Society A*. 160: 429-469.

David H. Kaye. 2009. Commentary, "False, But Highly Persuasive: How Wrong Were the Probability Estimates in *McDaniel v. Brown*?" *Michigan Law Review First Impressions* 108: 1-7. Available at: `http://www.michiganlawreview.org/assets/fi/108/kaye.pdf`

Jeff Strnad. 2007. "Should Legal Empiricists Go Bayesian?". *American Law and Economics Review*. 9: 195-303.

October 6: Probability IV

- expectation of a random variable
- properties of expectations
- the mean and the median
- variance
- covariance and correlation
- conditional expectation
- conditional variance

Required Reading

FPP: Chapters 8 & 9

DeGroot and Schervish. 2002. *Probability and Statistics.* 3rd Edition. Chapter 4. (Handout)

October 13: Chance Variability

- law of large numbers
- standard error
- the normal distribution
- the χ^2 distribution
- central limit theorem
- normal approximations

Required Reading

FPP: Chapters 16, 17, & 18

Verzani: Sections 5.2, 5.3, and Chapter 6

October 20: Sampling

- the *Literary Digest Poll*
- simple random sampling
- survey non-response
- chance error and bias
- accuracy of percentages
- accuracy of averages
- standard errors
- confidence intervals

Required Reading

FPP: Chapters 19, 20, 21 & 23

Verzani: Chapter 7

Optional Reading

Michael J. Saks and Peter David Blanck. 1992. "Justice Improved: The Unrecognized Benefits of Aggregation and Sampling in the Trial of Mass Torts". *Stanford Law Review.* 44: 815-851.

Laurens Walker and John Monahan. 1998. "Sampling Damages". *Iowa Law Review.* 83: 545-568.

October 27: Tests of Significance

- logic of frequentist hypothesis testing
- the z-test
- the t-test

Required Reading

FPP: Chapters 26 & 27

Verzani: Chapter 8

Optional Reading

Barry Brosi and Eric Biber. 2009. "Statistical Inference, Type II Error, and Decision-Making Under the U.S. Endangered Species Act". *Frontiers in Ecology and Environment.* Available at: `http://www.law.berkeley.edu/php-programs/faculty/facultyPubsPDF.php?facID=6482&pubID=1`.

November 3: Simple Linear Regression

- regression as conditional expectation
- least squares as a fitting method
- RMS error for regression
- interpretation of regression estimates

Required Reading

FPP: Chapters 10, 11, & 12; go back and skim Chapters 8 & 9

Verzani: Section 10.1

Optional Reading

Epstein, Lee and Carol Mershon. 1996. "Measuring Political Preferences." *American Journal of Political Science*, 40: 261-294.

Krehbiel, Keith. 1997. "Restrictive Rules Reconsidered." *American Journal of Political Science*, 41: 919-944.

November 10: Multiple Linear Regression

- model fitting
- descriptive inference
- predictive inference
- causal inference
- selecting rhs variables

Required Reading

Daniel L. Rubinfeld. 2000. "Reference Guide on Multiple Regression." in *Reference Manual on Scientific Evidence*. 2nd Edition. Federal Judicial Center. pp. 179-227. Available at: `http://www.fjc.gov/public/pdf.nsf/lookup/sciman00.pdf/$file/sciman00.pdf`

Verzani: Section 10.3

Optional Reading

Daniel L. Rubinfeld. 1985. "Econometrics in the Courtroom". *Columbia Law Review*. 85: 1065-1078.

Daniel L. Rubinfeld and Peter O. Steiner. "Quantitative Methods in Antitrust Litigation". *Law and Contemporary Problems*. 46: 69-141.

David Hyman, Bernard Black, Kathryn Zeiler, Charles Silver, and William Sage. 2007. "Do Defendants Pay What Juries Award? Post-Verdict Haircuts in Texas Medical Malpractice Cases, 1988-2003". *Journal of Empirical Legal Studies*. 4: 3-68.

Gelman, Andrew and Gary King. 1990. "Estimating the Incumbency Advantage without Bias." *American Journal of Political Science*. 34: 1142-1164.

Wilkerson, John D. 1999. " 'Killer' Amendments in Congress." *The American Political Science Review*, Vol. 93: 535-552.

Tsebelis, George. 1999. "Veto Players and Law Production in Parliamentary Democracies: An Empirical Analysis." *The American Political Science Review*, 93: 591-608.

November 17: Hypothesis Testing, Assessing Model Adequacy, and Practical Data Analysis (Not All the Same Thing)

- z-test and t-test
- χ^2-test and F-test
- leverage and influential data points
- assessing linearity
- assessing normality

Required Reading

FPP: Chapter 29

Verzani: Section 10.2

Optional Reading

Tatem, Andrew J; Carlos A. Guerra; Peter M. Atkinson; and Simon I. Hay. 2004. "Momentous Sprint at the 2156 Olympics." *Nature* 431 (30 September): 525.

November 24: Regression Analysis and Causal Inference

- causality and counterfactuals
- the need for causal assumptions
- using regression to estimate average treatment effects

Required Reading

Paul W. Holland. 1986. "Statistics and Causal Inference (with discussion)". *Journal of the American Statistical Association*. 81: 945-970.

Guido W. Imbens. 2004. "Nonparametric Estimation of Average Treatment Effects Under Exogeneity: A Review". *The Review of Economics and Statistics*. 86: 4-29.