Syllabus
209.3 Introductory Statistics

Professor: Kevin Quinn
GSI: Hannah Laqueur
Fall Semester 2013

Class Room
Boalt 130
6:25pm - 8:15pm
Tuesday and Thursday

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

Hannah Laqueur
Email: `hlaqueur@berkeley.edu`

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
- interact with expert witnesses
- 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 for some class sessions has relatively little in the way of Greek letters and mathematical symbols. The required reading for most weeks will deal with very difficult (and subtle) 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.

Also please note that some class sessions will have heavier reading loads than others. You should be continually looking ahead a few weeks in order to plan accordingly.

There will be approximately 10 problem sets handed out over the course of the semester. These will be graded on a 4-point basis (4: excellent, 3: good, 2: fair, 1: poor) 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, we 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`.

We recommend the following process for getting started with R.

1. download and install R-Studio `http://www.rstudio.com/ide/`
2. work through the exercises at: `http://tryr.codeschool.com`

In addition to the free resources above, *Discovering Statistics Using R* by Andy Field, Jeremy Miles, and Zoe Field as well as *Introductory Statistics with R* by Peter Dalgaard are both good texts to learn out of. Please note that these books are not required for this class—in most cases the free online resources should be sufficient for your needs.

Office Hours and Availability
Professor Quinn will hold office hours on Wednesdays from 10:00 am to noon. Hannah Laqueur will hold office hours at a time to be determined.

The section meeting for this course will take place on Fridays from 10:00am to 11:00am in a room to be determined.

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.

Jeffrey M. Wooldridge. 2013. *Introductory Econometrics: A Modern Approach.* 5th Edition. South-Western Cengage Learning.

We'll refer to Freedman, Pisani, and Purves as "FPP" and the Wooldridge book simply as "Wooldridge".

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.

Thursday, August 22: Introduction and Overview

- course goals
- descriptive vs. causal inference

Tuesday, August 27: Motivating Examples

- descriptive vs. causal inference

Required Reading

Elmendorf, Christopher S. and Douglas M. Spencer. 2013. "The Geography of Racial Stereotyping: Evidence and Implications for VRA Preclearance After *Shelby County*." (on course website)

Greiner, D. James, Cassandra Wolos Pattanayak, and Jonathan Philip Hennessy. 2013. "The Limits of Unbundled Legal Assistance: A Randomized Study in a Massachusetts District Court and Prospects for the Future." *Harvard Law Review*. (on course website)

Optional Reading

Greiner, D. James and Cassandra Wolos Pattanayak. 2011. "Randomized Evaluation in Legal Assistance: What Difference Does Representation (Offer and Actual Use) Make?" *Yale Law Journal*. (on course website)

Eppler-Epstein, Steven. 2013. "Response: Passion Caution, and Evolution: The Legal Aid Movement and Empirical Studies of Legal Assistance." *Harvard Law Review Forum*. (on course website)

Thursday, August 29: Lab Session with R

- getting started
- finding help

Tuesday, September 3: Random Sampling as a Basis for Descriptive Inference

- samples and populations
- simple random sampling

Required Reading

FPP, chapter 19

Cobb, Rachael V., D. James Greiner, and Kevin M. Quinn. 2012. "Can Voter ID Laws Be Administered in a Race-Neutral Manner? Evidence from the City of Boston in 2008." *Quarterly Journal of Politial Science*. (on course website).

Thursday, September 5: Random Assignment as a Basis for Causal Inference

- simple randomized controlled trials
- an introduction to the Neyman-Rubin model

Required Reading

FPP, chapter 1 and 2

Holland, Paul. W. 1986. "Statistics and Causal Inference" *JASA*. (on course website) (You should skim this piece—it makes use of concepts and notation that we haven't gotten to yet and so parts of it might seem somewhat mysterious.)

Tuesday, September 10: Probability

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

Required Reading

FPP: Chapter 13

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

Thursday, September 12: Probability

- the addition rule
- the binomial formula
- random variables

Required Reading

FPP: Chapters 14 & 15

Tuesday, September 17: Probability

- 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. (on course website) (Focus on Sections 3.1 to 3.6 and skim 3.7 to 3.9)

Hill, Ray. 2004. "Multiple Sudden Infant Deaths—Coincidence or Beyond Coincidence?". *Paediatric and Perinatal Epidemiology*. (on course website)

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.

Thursday, September 19: Probability

- 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. (on course website) (You can skip Sections 4.4 and 4.9)

Tuesday, September 24: Descriptive Statistics

- sample quantiles
- the sample mean
- the sample standard deviation
- the conditional sample mean
- the conditional sample standard deviation
- correlation

Required Reading

FPP: chapters 4, 5, 8 (skim chapter 9)

Thursday, September 26: Lab Session with R

- working with datasets
- plotting data

Tuesday, October 1: 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

Thursday, October 3: Tests of Significance

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

Required Reading

FPP: chapters 26 & 27

Castaneda v. Partida 430 U.S. 482 (on course website)

Tuesday, October 8: Randomized Experiments

- experiments and causal inference
- the importance of random assignment of treatment
- field experiments

Required Reading

Chapters 1 and 2 of: Gerber, Alan S., and Donald P. Green. 2012. *Field Experiments: Design, Analysis, and Interpretation.* W.W. Norton. (on course website)

Page, Stewart. 1998. "Accepting the Gay Person: Rental Accomodation in the Community." *Journal of Homosexuality.* (on course website)

Sherman, Lawrence W. and Dennis P. Rogan. 1995. "Deterrent Effects of Police Raids on Crack Houses: A Randomized Controlled Experiment." *Justice Quarterly.* (on course website)

Thursday, October 10: Lab Session with R

- building simple Monte Carlo experiments
- basic data analysis

Tuesday, October 15: Simple Linear Regression

- motivation for linear regression
- regression as conditional expectation

Required Reading

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

Wooldridge: chapter 2.1

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.

Thursday, October 17: Simple Linear Regression

- OLS as an estimation method
- calculating OLS estimates

Required Reading

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

Wooldridge: chapter 2.2

Tuesday, October 22: Simple Linear Regression

- properties of the OLS estimator
- hypothesis testing
- confidence intervals

Required Reading

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

Wooldridge: chapter 2.3

Thursday, October 24: Simple Linear Regression

- interpretation

Required Reading

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

Wooldridge: chapter 2.4, 2.5, and 2.6

Tuesday, October 29: Multiple Linear Regression

- motivation for multiple regression

Required Reading

Wooldridge: chapter 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.

Thursday, October 31: Multiple Linear Regression

- the OLS estimator
- calculating OLS estimates

Required Reading

Wooldridge: chapter 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.

Tuesday, November 5: Multiple Linear Regression

- interpretation

Required Reading

Wooldridge: chapter 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.

Thursday, November 7: Multiple Linear Regression

- properties of the OLS estimator

Required Reading

Wooldridge: chapter 3

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. (on course website)

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.

Tuesday, November 12: Multiple Linear Regression

- inference for multiple regression

Required Reading

Wooldridge: chapter 4

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.

Thursday, November 14: Multiple Linear Regression

- introduction to asymptotics

Required Reading

Wooldridge: chapter 5

Tuesday, November 19: Regression and Causality

- importance of research design
- measured and unmeasured confounding
- credible causal inference

Required Reading

Angrist, Joshua D. and Jörn-Steffen Pischke. 2010. "The Credibility Revolution in Empirical Economics: How Better Research Design is Taking the Con out of Econometrics." *Journal of Economic Perspectives.* (on course website)

Ho, Daniel E. and Donald B. Rubin. 2011. "Credible Causal Inference in Empirical Legal Studies." *Annual Review of Law and Social Science.* (on course website)

Thursday, November 21: Regression and Causality

- importance of research design
- measured and unmeasured confounding
- when does an OLS estimate have a bona fide causal interpretation?

Required Reading

Hainmueller, Jens and Dominik Hangartner. 2013. "Does Direct Democracy Hurt Immmigrant Minorities? Evidence from Naturalization Decisions in Switzerland" (on course website)

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

Tuesday, November 26: Partial Identification

- how to make credible inferences when your design is not strong
- Manski bounds

Required Reading

Manski, Charles F. and Daniel S. Nagin. 1998. "Bounding Disagreements about Treatment Effects: A Case Study of Sentencing and Recidivism." *Sociological Methodology* (on course website)

Manski, Charles F. and John V. Pepper. 2013. "Deterrence and the Death Penalty: Partial Identification Analysis Using Repeated Cross Sections." *Journal of Quantitative Criminology*. (on course website)