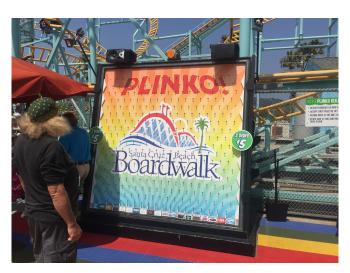# Syllabus

Ec240a - Second Half, Fall 2023

Course Description

After introducing some basic optimization tools, this course begins with an analysis of a basic prediction problem. A decision maker obtains a random sample of covariates (features) and outcomes. She wishes to use her sample to forecast the outcomes of new units on the basis of their covariates. We motivate this problem and provide a canonical representation of it (the K Normal means problem). We use this problem to introduce some elements of (i) statistical decision theory and (ii) modern regression methods.

We then develop some properties of regression functions. The iteration properties of mean and linear regression will receive special emphasis.

Finally, we will develop methods for conducting inference on linear regression coefficients estimated by the method of least squares under random sampling. We will develop two approaches. The first is a nonparametric Bayesian method. The second, frequentist approach, is based on large sample (i.e., asymptotic) approximations. Methods of hypothesis testing and confidence interval construction will be reviewed. If time permits we will introduce some methods for quantile regression analysis.

The central limit theorem at the Santa Cruz Beach Boardwalk (near the Hurricane).

See: https://en.wikipedia.org/wiki/Bean_machine

Instructor: Bryan Graham, 665 Evans Hall, email: `bgraham@econ.berkeley.edu`

Time and Location: Monday and Wednesday, 10:00AM to 12:00PM in Lewis Hall, Room 9

Office Hours: Thursdays 2 to 3:45PM (sign up online here). Office hours will be in my Evans Hall office or, by prior arrangement, on Zoom.

Graduate Student Instructor: Jinglin Yang, e-mail: jinglin.yang@berkeley.edu

Prerequisites: linear algebra, multivariate calculus, basic probability and inference theory.

Course Webpage: Various instructional resources, including occasional lecture notes and Jupyter Notebooks, can be found on GitHub in the following repository

https://github.com/bryangraham/Ec240a

The GSI may make additional resources available on bCourses.

Textbook: There is no mandatory text. Material will be delivered primarily through lecture and assigned papers. Good note taking is essential for successful performance in the class. Nevertheless I do recommend the following book as useful supplement to the material presented in lecture.

1. Wooldridge, Jeffrey M. (2011). _Econometric Analysis of Cross Section and Panel Data, 2nd Ed._ Cambridge, MA: The MIT Press.

This is a useful long term reference for anyone who anticipates undertaking empirical research. Two other useful textbooks are Stachurski (2016) and Hansen (2022). Paul Goldsmith-Pinkham has made Gary Chamberlain's first-year Harvard Ph.D. econometrics Lecture Notes from circa 2010 available online. Gary was my primary advisor when I was a Ph.D. student; you may find these notes helpful as well.

Additional books which you may find helpful include Ferguson (1996), Wasserman (2004), Wasserman (2006) and Manski (2007). Ferguson (1996) is a compact introduction to large sample theory. My treatment of the K Normal means problem draws from Wasserman (2006). Wasserman (2004) is a nice introductory mathematical statistics reference. Manski (2007) provides a textbook treatment of identification with applications of interest to economists. While this course is largely self-contained, I nevertheless encourage you to view course materials as a basic scaffolding which you can use to independently add to on your own. Indeed one goal of the course is to develop some basic literacy in econometrics so that you are better able to learn new material independently. This is not a "cookbook" class.

Grading: Grades for _this half of the course_ will equal a weighted average of homework (40%) and mid-term performance (60%). The mid-term will be held on the last day of class (**November 29th, 2023**). There will be 5 homework assignments (plus a review sheet). Homeworks are due at 5PM on the assigned due date (the GSI may elect to make small modifications to all things homework related). Homeworks are graded on a ten point scale with one point off per day late, for the first

three days, and no additional penalty thereafter. Concretely this means homeworks turned in three or more days late can earn only up to seven points (but as long as you turn it in before the last day of the semester it will be counted).

You are free, indeed encouraged, to work in groups but each student must submit an individual write-up and accompanying Jupyter Notebook (when required; see below). Please write the names of any study partners on the top of your homework. Your lowest homework grade will be dropped, with the average of the remaining scores counting toward your final grade. I will add 5 points to homework aggregates for students who make serious efforts to complete *all* five problem sets. *Concretely this means that students may amass up to 45 homework points; it also means that if you only do four problem sets you can earn no more than 40 out of 45 homework points.* Problem Set 5 is formally the course final assessment.

The due dates for the five problem sets are:

Problem Set	Due Date
1	October 27th
2	November 9th (Thursday)
3	November 22nd (Wednesday)
4	December 1st
5	December 15th (assignments accepted after this date)

Computation: All computational work should be completed in Python. Python is a widely used general purpose programming language with good functionality for scientific computing. There are lots of ways of accessing Python (EML, online at `https://datahub.berkeley.edu/`). For those wishing to manage a Python environment on their personal computer, the Anaconda distribution, which is available for download at https://www.anaconda.com/distribution/, is a convenient way to get started. Some basic tutorials on installing and using Python, with a focus on economic applications, can be found online at `https://quantecon.org/`. Good books for learning Python, with some coverage of statistical applications, are Guttag (2013), VanderPlas (2017), and McKinney (2017).

The code I will provide will execute properly in Python 3.6, which is (close to) the latest Python release. Python is also available on the EML workstations (which are remotely accessible). There are a large number of useful resources available for learning Python (including classes at the D-Lab).

While issues of computation may arise from time to time during lecture, I will not teach Python programming. *This is something you will need to learn outside of class.* I do not expect this to be easy. I ask that those students with strong backgrounds in technical computing to assist classmates with less experience. I am happy to answer programming questions during office hours.

Extensions: Extensions for assignments will not be granted. The penalty for lateness is relatively minor and I also drop the lowest homework grade. I am mindful that you may find the workload during the first semester of the Ph.D. program challenging. The goal is not to create a miserable

experience, but rather to introduce you to a variety of tools that will be of continuing value as an economist. Part of the "trick" of getting through the core Ph.D. coursework is to not let work pile up. There will be times where you may not complete all your work to the standards you are perhaps used to. This is normal. Do the best you can on your problem sets and turn them in on time. This allows you to not fall behind. Adopt a growth mindset. Perfection is neither expected nor optimal; try to enjoy the challenge. Think of the problem sets as puzzles if that helps. Conditional of meeting a basic proficiency standard, your course grade is not particularly important. Work hard, but do not worry.

Accommodations: Any students requiring academic accommodations should request a 'Letter of Accommodation' from the Disabled Students Program at http://dsp.berkeley.edu/ *immediately.* I will make a good faith effort to accommodate any specific needs conditional on certification. Please plan well in advance as I may not be able accommodate last minute requests.

Academic Integrity: Please read the Center for Student Conduct's statement on Academic Integrity at http://sa.berkeley.edu/conduct/integrity. I take issues of intellectual honest *very* seriously.

Cooperation/Civility: I remember graduate school as a period of immense intellectual excitement, punctuated by periods of equally intense frustration and stress. My classmates were an important source of intellectual support and encouragement. Please be open to helping one another learn the material. Don't be afraid to ask classmates for help and, if asked for help, be generous and gracious in providing it. Everyone will learn more in this class if they work together. *Class is not a tournament.* We are here to learn together. I also encourage each of you to familiarize yourself with Berkeley's Principles of Community (available here).

E-mail and office hours: I prefer to avoid having substantive communications by e-mail. Please limit e-mail use to short yes/no queries. I am unlikely to read or respond to a long/complex e-mail. Please make use of my office hours. This is time specifically allocated for your use; please come by (virtually!) I look forward to getting to know all of you. You can sign-up for office hour slots online here.

Course Outline

Date	Topic	Readings/Notes
M 10/16	**Projection Theorem**	Stachurski (2016, Ch. 2)
W 10/18	**Probability Distributions, Bayes' Rule**	Mitzenmacher & Upfal (2005, Chs. 1-2) Stachurski (2016, Ch. 4)
M 10/23	**Conditional Expectation Functions**	Wooldridge (2010, Ch. 2) Stachurski (2016, Ch. 5)
W 10/25	**K-Normal Means**	Efron (2004) Wasserman (2006, Ch. 7)
M 10/30	**K-Normal Means**	Wasserman (2006, Ch. 7) Stein (1981)
W 11/1	**Linear Regression**	Wooldridge (2010, Chs. 2, 4) Card (1995); Card & Krueger (1996)
M 11/6	**Bayesian Bootstrap**	Chamberlain & Imbens (2003)
W 11/8	**Large Sample Theory for OLS**	Wooldridge (2010, Chs. 3-4)
M 11/13	**Applications of OLS**	Deaton (1989); Altonji & Pierret (2001) Ashenfelter et al. (2006); Goldberger (1984)
W 11/15	**Quantiles**	Mood et al. (1974)
M 11/20	**Quantile Regression** *(Class will be held on Zoom)*	Chamberlain (1994)
W 11/22	*No Class*	*Thanksgiving recess*
M 11/27	**In Class Review Session**	
W 11/29	**2nd Midterm Exam**	Good luck!

References

Altonji, J. G. & Pierret, C. R. (2001). Employer learning and statistical discrimination. *Quarterly Journal of Economics*, 116(1), 313 – 350.

Ashenfelter, O., Ashmore, D., Baker, J. B., Gleason, S., & Hosken, D. S. (2006). Empirical methods in merger analysis: econometric analysis of pricing in ftc v. staples. *International Journal of the Economics of Business*, 13(2), 265 – 279.

Card, D. (1995). Earnings, schooling, and ability revisted. *Research in Labor Economics*, 14(23 - 48).

Card, D. & Krueger, A. (1996). *Does Money Matter?*, chapter Labor market effects of school quality: theory and evidence, (pp. 97 – 140). Brookings Institution Press: Washington D.C.

Chamberlain, G. (1994). *Advances in Econometrics: Sixth World Congress*, volume 2, chapter Quantile regression, censoring, and the structure of wages, (pp. 171 – 209). Cambridge University Press: Cambridge.

Chamberlain, G. & Imbens, G. W. (2003). Nonparametric applications of bayesian inference. *Journal of Business and Economic Statistics*, 21(1), 12 – 18.

Deaton, A. (1989). Rice prices and income distribution in thailand: a non-parametric analysis. *Economic Journal*, 99(395), 1 – 37.

Efron, B. (2004). The estimation of prediction error. *Journal of the American Statistical Association*, 99(467), 619 – 632.

Ferguson, T. S. (1996). *A Course in Large Sample Theory*. London: Chapman & Hall.

Goldberger, A. S. (1984). Reverse regression and salary discrimination. *Journal of Human Resources*, 19(3), 293 – 318.

Guttag, J. V. (2013). *Introduction to Computation and Programming Using Python*. Cambridge, MA: The MIT Press.

Hansen, B. (2022). *Econometrics*. Princeton, NJ: Princeton University Press.

Manski, C. F. (2007). *Identification for Prediction and Decision*. Cambridge, MA: Harvard University Press.

McKinney, W. (2017). *Python for Data Analysis*. Cambridge: O'Reilly.

Mitzenmacher, M. & Upfal, E. (2005). *Probability and Computing*. Cambridge: Cambridge University Press.

Mood, A. M., Graybill, F. A., & Boes, D. C. (1974). *Introduction to the Theory of Statistics*. New York: McGraw-Hill Book Company, 3rd edition.

Stachurski, J. (2016). *A Primer in Econometric Theory*. Cambridge, MA: The MIT Press.

Stein, C. M. (1981). Estimation of the mean of a multivariate normal distribution. *Annals of Statistics*, 9(6), 1135 – 1151.

VanderPlas, J. (2017). *Python Data Science Handbook*. Boston: O'Reilly.

Wasserman, L. (2004). *All of Statistics*. New York: Springer.

Wasserman, L. (2006). *All of Nonparametric Statistics*. New York: Springer.

Wooldridge, J. M. (2010). *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press, 2nd edition.