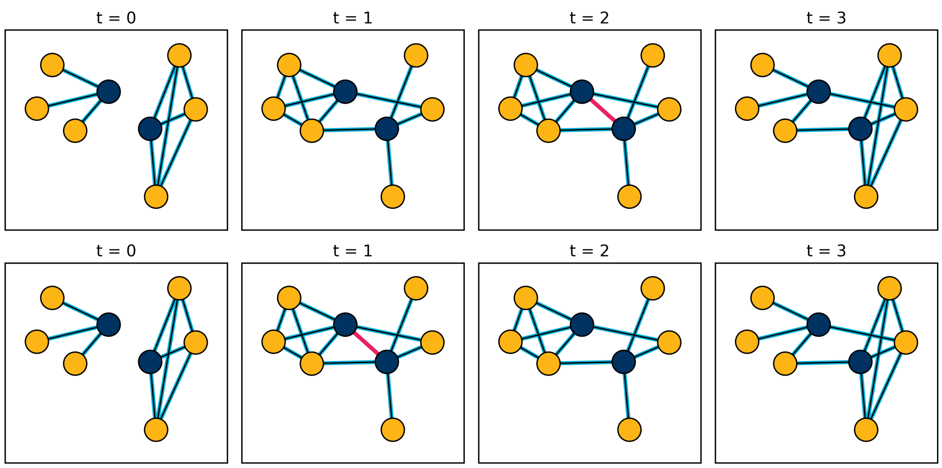
Ec 142 - Applied Econometrics

Department of Economics, UC - Berkeley

Spring 2017

Course Description

This course begins with a review of linear regression. The main properties of regression will be developed within the context of a normed vector space with the projection theorem serving as a unifying tool. This allows for a common treatment of both 'population' and 'sample' regression. The iteration properties of mean and linear regression will receive special emphasis, as will the Frisch-Waugh Theorem. Theoretical development as well as practical application of these results will be covered. We will then briefly discuss the asymptotic approximation of sampling distributions and review basic methods of inference. With these core tools in place, we will turn to an in depth treatment of three 'frontier' topics in econometrics/data science: (i) quantile regression, (ii) survival analysis and (iii) the analysis of social and economic network data.



Course Logistics

Instructor: Bryan Graham, Department of Economics, University of California – Berkeley

Email: `bgraham@econ.berkeley.edu`

Time & Location: Tu/Th 9:30AM to 11AM, 9 Lewis Hall

Office Hours: Thursday 2:00 PM - 4 PM (okay to drop by, but appointment preferred)

Graduate Student Instructor: Seongjoo MIN, e-mail: sjmin711@berkeley.edu

Prerequisites: The minimum requirement for this class is a prior course in econometrics (typically Ec140 or Ec141). A good statistics class in linear models might also suffice. If you have completed Ec140 this also means you have completed an introductory statistics class, the intermediate micro/macro theory sequence (100A/101A & 100B/101B), and (the equivalent of) Math 1A & 1B or Math 16A & 16B. The class will be taught assuming a mastery of the material covered in these courses. A class in linear algebra (e.g., Math 54) and some basic programming experience is also useful, but not required.

Grading: Final grades will equal a weighted average of performance on homework assignments (35%), two midterm exams (40%) and a final project (25%) (in lieu of a final exam). The project will be due at the end of the final exam time slot assigned to this class by the registrar. The midterms are scheduled for **March 23rd, 2017** and **April 27th, 2017**. If you have a known conflict with either of these two dates I suggest dropping the class. There will be 4 to 6 homework assignments (plus ungraded review sheets for both midterm exams). Homeworks are due at 5PM on the assigned due date. They are graded on a ten point scale with one point off per day late. You are free to work in groups but each student must submit an individual write-up and accompanying Jupyter Notebook. Your lowest homework grade will be dropped. With the average of the remaining scores counting toward your final grade. There will be no 'make-up' midterms, although I will drop the lowest of your two mid-term grades.[1] Because exams vary in their difficulty, and cohorts vary in their mastery of the material, I generally apply a "curve" when assigning letter grades. There is no predetermined letter grade distribution.

Textbook: There is no required textbook for this class. Wasserman (2004) is a good, albeit challenging, reference. For a review of basic concepts in probablity, the first few chapters of Mitzenmacher & Upfal (2005) are helpful. Your introductory statistics and Ec140/141 textbooks will also be useful references. While I will occasionally make lecture notes available to students via a course GitHub repository, students should plan on taking *detailed* notes on the material presented during lecture. If you miss class for any reason please be sure to get notes from a classmate. Good note-taking is essential for doing well in this course. I will also assign readings as the course progresses.

Computation: The bulk of class will be devoted to the formal development of the material, albeit with empirical illustrations as well as ample discussions of the various practicalities of implementa-

[1] To incentivize you to take both exams seriously, I will add 10 percentage points to your midterm aggregate if you take both exams and perform comparably on each of them. This means the maximum number of points a student could accumulate toward their final grade is 110.

tion. However I do intend to reserve some class time for actual practice with computation and such exercises will feature prominently in the homework. Computational examples will be done using Python. Python is a widely used general purpose programming language with good functionality for scientific computing. I highly recommend the Anaconda distribution, which is available for download at http://continuum.io/downloads. Some basic tutorials on installing and using Python, with a focus on economic applications, can be found online at `http://quant-econ.net`. You may also wish to install Rodeo, which is an integrated development environment (IDE) tailored to statistics or "data science" applications. Rodeo makes working in Python look and feel similar to working in Stata or MATLAB. Rodeo is also free and available at `https://www.yhat.com/`.

Good books for learning Python, with some coverage of statistical applications, are Guttag (2013) and McKinney (2013). The former is an excellent introduction to computer science as well as Python, the latter is heavily focused on the pandas module.

The code I will provide will execute properly in Python 2.7. Nevertheless I recommend installing Python 3.5. This is the latest Python release.

Graphviz is a free graph visualization program that is also useful (`http://www.graphviz.org/`).

Ec142 is not a Python programming course. Although I will provide hints in class about how to approach computation, including sharing code snippets, I will not "teach" computation. Section will include some assistance with computation and you are always welcome to ask about computation during my office hours. There are a large number of useful resources available for learning Python (including classes at the D-Lab). Consequently, much of the programming aspect of the course *is something you will need to learn outside of class.* I encourage you to seek out the help of classmates and work together in groups. I also ask those of you with strong prior experience in technical computing to help your classmates.

Extensions: Extensions for assignments will not be granted. The penalty for lateness is relatively minor and I also drop the lowest homework grade. Likewise there will be no accommodations for missed exams.

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 special 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.

Additional notes: 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. Do feel free to chat with me immediately before class. For longer questions please make use of my office hours. This is time specifically allocated for your use; please come by! I look forward to getting to know all of you.

Course Outline

(Additional references to be added)

	Date	Topic	Readings
Week 1	Tu 1/17	Introduction/Python	Shen (2014)
	Th 1/19	Projection/Regression #1	
Week 2	Tu 1/24	Projection/Regression #2	
	Th 1/26	Projection/Regression #3	Card & Krueger (1996)
Week 3	Tu 1/31	Projection/Regression #4	Case & Paxson (2008)
	Th 2/2	Large sample analysis #1	Wasserman (2004, Ch. 5)
Week 4	Tu 2/7	Large sample analysis #2	
	Th 2/9	Testing/Confidence Intervals	
Week 5	Tu 2/14	Quantiles #1	Mood et al. (1974, Ch. 11.3)
	Th 2/16	Quantiles #2	Koenker & Hallock (2001)
Week 6	Tu 2/21	Quantile Regression #1	Chamberlain (1994)
	Th 2/23	Quantile Regression #2	
Week 7	Tu 2/28	Quantile Regression #3	
	Th 3/2	Quantile Regression #4	
Week 8	Tu 3/7	Survival Analysis #1	Singer & Willett (2003, Chs. 9 - 12)
	Th 3/9	Survival Analysis #2	Efron & Hastie (2016, Ch. 9)
Week 9	Tu 3/14	Survival Analysis #3	Ashenfelter & Card (2002)
	Th 3/16	Survival Analysis #4	
Week 10	Tu 3/21	Catch-up/Review	
	Th 3/23	**Midterm #1**	
	Tu 3/28	*Spring Recess*	
	Th 3/30		
Week 11	Tu 4/4	Networks/Graphs #1	Graham (2015)
	Th 4/6	Networks/Graphs #2	Apicella et al. (2012)
Week 12	Tu 4/11	Networks/Graphs #3	Holland & Leinhardt (1976)
	Th 4/13	Networks/Graphs #4	Bhattacharya & Bickel (2015)
Week 13	Tu 4/18	Networks/Graphs #5	Graham (2014)
	Th 4/20	Networks/Graphs #6	
Week 14	Tu 4/25	Catch-up/Review	
	Th 4/27	**Midterm #2**	

References

Apicella, C. L., Marlowe, F. W., Fowler, J. H., & Christakis, N. A. (2012). Social networks and cooperation in hunter-gatherers,. *Nature*, 481(7382), 497 – 501.

Ashenfelter, O. & Card, D. (2002). Did the elimination of mandatory retirement affect faculty retirement? *American Economic Review*, 92(4), 957 – 980.

Bhattacharya, S. & Bickel, P. J. (2015). Subsampling bootstrap of count features of networks. *Annals of Statistics*, 43(6), 2384 – 2411.

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.

Case, A. & Paxson, C. (2008). Stature and status: height, ability, and labor market outcomes. *Journal of Political Economy*, 116(3), 499 – 532.

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.

Efron, B. & Hastie, T. (2016). *Computer Age Statistical Inference*. Cambridge: Cambridge University Press.

Graham, B. S. (2014). *An econometric model of link formation with degree heterogeneity*. NBER Working Paper 20341, National Bureau of Economic Research.

Graham, B. S. (2015). Methods of identification in social networks. *Annual Review of Economics*, 7(1), 465–485.

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

Holland, P. W. & Leinhardt, S. (1976). Local structure in social networks. *Sociological Methodology*, 7, 1 – 45.

Koenker, R. & Hallock, K. F. (2001). Quantile regression. *Journal of Economic Perspectives*, 15(4), 143 – 156.

McKinney, W. (2013). *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.

Shen, H. (2014). Interactive notebooks: sharing the code. *Nature*, 515(7525), 151–152.

Singer, J. D. & Willett, J. B. (2003). *Applied Longitudinal Data Analysis*. Oxford: Oxford University Press.

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