

Class Schedule

Lectures:

- Section 2: Mondays, 2-4pm, in Cheit 210
- Section 3: Mondays, 4-6pm, in Cheit 210

Instructor

Conrad Miller

ccmiller@berkeley.edu (Please include '[UGBA 96]' in the subject)

Office Hours: F 2-4pm, Haas F686

Reader and Connector Associates

<u>Name</u>	<u>Email</u>	<u>Role</u>	<u>Office Hours</u>
Mehmet Seflek	mseflek@berkeley.edu	Reader	Th 2-4pm, F621
Maggy Li	maggyli@berkeley.edu	CA	F 10-11am, DS Nexus (Moffitt)
Tina Teng	tina.teng@berkeley.edu	CA	Th 1-2pm, DS Nexus (Moffitt)
Phoebe Yin	phoebe.yin@berkeley.edu	CA	F 1-2pm, DS Nexus (Moffitt)

Course Description

The objective of this connector course is to provide an understanding of how data and statistical analysis can improve managerial decision-making. Students learn how to ask well-posed questions, find or collect relevant data, and apply appropriate statistical methods to solve problems and make better business and policy decisions. We will explore statistical methods for gleaning insights from economic and social data, with an emphasis on approaches to identifying *causal relationships*. We will discuss how to design and analyze randomized experiments and introduce econometric methods for estimating causal effects in non-experimental data.

To introduce and motivate these methods, we will draw on a variety of business and social science applications, including advertising, management, online marketplaces, labor markets, and education. We will build on the programming tools you develop in Data 8 to analyze real economic and social data.

This course is 2 units. This course, in combination with the Data 8 *Foundations* course, satisfies the statistics prerequisite for admission to Haas.

Prerequisites

For success in this course, concurrent enrollment in or completion of Foundations of Data Science (Data 8/CS8/Statistics 8) and completion of one semester of calculus is *highly recommended*.

Course Objectives

During this course, you will:

- *Develop a comfort with interpreting and analyzing economic and social data.* You will learn to: assess data quality and appropriateness; compute and interpret summary statistics; and construct and interpret figures.
- *Learn standard causal inference methods, including when and how to apply them.* You will learn the importance of randomized experiments, and how to analyze and interpret them. You will also gain a high-level understanding of common approaches to causal inference using non-experimental data, including regression, regression discontinuity, and differences-in-differences.
- *Become a savvy consumer of empirical evidence, particularly evidence underlying causal claims.*

Announcements and Course Materials

We will use *bCourses* for circulating announcements and course materials, including lecture notes, readings, and problem sets.

Discussion

You will use *Piazza* to post questions about the course material, including problem sets and labs. Using *Piazza* will be more effective than emailing course staff with the same question. Be sure to check whether your question has already been answered. I also encourage you to answer posted questions: explaining concepts is a great way to learn and solidify your understanding of the material.

The reader and I will monitor the page and post answers if a correct answer hasn't been posted after some period.

More details on posting etiquette are included on the course *Piazza* page.

Required Textbook

The required text is *Mastering 'Metrics: The Path from Cause to Effect*, by Josh Angrist and Jorn-Steffen Pischke, Princeton University Press, 2015 (ISBN: 9780691152844). This is available at the Cal Store, but feel free to get the book wherever.

There will be additional required and optional background readings, including academic papers and business press articles, posted on *bCourses*.

Software

As a connector course, we will rely on the same Python-based Jupyter infrastructure used in Data 8. Problem sets and labs will be hosted on datahub.berkeley.edu.

Some lectures will include in-class labs. You will need a laptop in class to participate. If you do not have a laptop, you can borrow one from the library. See [here] for details. *Check what lectures feature labs in the Course Schedule below.*

Assignments and Grades

Your grade will be based on four components:

Class Participation (5%)

This class will be difficult to follow if you do not keep up with the material. Prior to each class, you should complete the reading and be engaged in class discussion. *Your participation grade will largely be determined by your participation in in-class polls.*

Note: we will use the *Poll Everywhere* tool for in-class polls. We will circulate details on how to sign up.

Labs (10%)

Some lectures will include *labs*, which will provide a hands-on introduction to topics and methods discussed in lecture. Labs will be graded for completion. Your lowest lab score will be dropped in the calculation of your overall grade.

Problem Sets (50%)

Problem sets are an essential part of the learning that you will do in this course. There will be 10 weekly problem sets, which will build on the concepts we discuss in class and the data programming skills you learn in this course and Data 8. Your lowest problem set score will be dropped in the calculation of your overall grade.

You are welcome to discuss problem sets with classmates. However, write up your answers in your own words and don't share your completed work.

Final Exam (35%)

The final exam will be held on December 13th, 3-6pm (Section 2), or December 13th, 8-11am (Section 3). It will cover material from the entire semester. There will be more details about the final announced later in the semester.

Submitting Assignments

Late submissions of labs and problem sets will generally not be accepted, unless you have relevant DSP accommodations. Under *very limited circumstances*, I will grant extensions for extended illnesses and other emergencies. Otherwise, you can 'use' your dropped problem set.

You will submit assignments using OKpy. *Make sure that you have submitted assignments successfully.* You can submit your assignments multiple times. We will grade the latest submission that meets the deadline, unless you 'flag' a different submission on OKpy.

Electronic Devices

Studies have shown electronic devices often distract both you and those around you from learning (for more on those studies, reading this article). Please keep phones on silent and refrain from using laptops, tablets, smartphones and other devices *unless for approved purposes*. These include labs and answering poll questions.

Class Etiquette

Make sure to be in your seat and ready by the time the lecture begins. If you know that you will have to leave early for some reason, get a seat near the door.

Grade Dispute Policy

In the event that an individual requests that a problem set or exam be re-graded, at least 24 hours after and within 1 week of it being returned, class members can submit a petition in writing. The petition should clearly document the specific issue you have with the grading, be well reasoned, and be word-processed. In the case of an exam, please turn in the written petition and the exam at the same time. Please do not write on the exam, but rather on a separate piece of paper. We reserve the right to re-grade the entire problem set or exam.

Honor Code

As members of the UC-Berkeley community, our expectation is that you will adhere rigorously to the UC-Berkeley Honor Code. Anyone caught cheating on a problem set or exam in this course will receive a failing grade in the course and will also be reported to the University Center for Student Conduct. In order to guarantee that you are not suspected of cheating, please keep your eyes on your own materials and do not converse with others during exams. For further information see: www.asuc.org/honorcode/.

DSP

If you are a DSP student, I am happy to make accommodations as necessary. Please feel free to contact me with any requests or concerns.

UGBA 96 Data and Decisions Schedule

<u>#</u>	<u>Date</u>	<u>Topic</u>	<u>Reading</u>	<u>Lab</u>	<u>Due</u>
1	8/27	Course Introduction	(α)Blake et al. (2015)		
	9/3	<i>No Meeting: Labor Day</i>			
2	9/10	Data Collection	(*)Crawford (2013) (*)SticiGui Ch. 24		
3	9/17	Descriptive Statistics	(*)SticiGui Ch. 4 (α)Chetty et al. (2017)	✓	PS1
4	9/24	The Potential Outcomes Framework	(*)MM Ch. 1		PS2
5	10/1	Randomized Experiments	(α)Duflo and Saez (2003) (α)Bloom et al. (2015)	✓	PS3
6	10/8	Noncompliance	(*)MM Ch. 3.1 and Ch. 3.2 (α)Fowlie et al. (2017)		PS4
7	10/15	Inference for Experiments	(*)Review D8 17.2, 17.3	✓	PS5
8	10/22	Observational Studies and Matching	(*)MM Ch. 2.1	✓	
9	10/29	Omitted Variable Bias	(α)Gordon et al. (2017) (α)Carroll (2018)		PS6
10	11/5	Introduction to Regression	(*)MM Ch. 2.2 (*)MM Ch. 2.3 (α)Arceneaux et al. (2006)	✓	PS7
	11/12	<i>No Meeting: Veteran's Day</i>			
11	11/19	Regression Discontinuity	(*)MM Ch. 4 (α)Hoekstra (2009) (α)Anderson and Magruder (2012)		PS8
12	11/26	Differences in Differences	(*)MM Ch. 5		PS9
	12/3	No Meeting: RRR			
	12/13	Final Exam			

Reading List

Legend:

- (*) Required reading
- (α) Optional reading—application covered in lecture
- MM: *Mastering 'Metrics*, Josh Angrist and Jorn-Steffen Pischke
- D8: *Computational and Inferential Thinking: Foundations for Data Science*, the Data 8 textbook
- SticiGui: online textbook written by Prof. Philip Stark [link]
 - This text may be a useful reference for some portions of the course

Lecture 1, Course Introduction: Prediction versus Causality

(α)Blake, Thomas, Chris Nosko and Steven Tadelis. 2015. "Consumer Heterogeneity and Paid Search Effectiveness: A Large-Scale Field Experiment." *Econometrica*, 83(1): 155-174.

Lecture 2, Data Collection

(*)Crawford, Kate. "The Hidden Biases in Big Data." *Harvard Business Review*, April 1, 2013. Web. Accessed January 2nd, 2018. [link]

(*)SticiGui, Ch. 24 [link]

Lecture 3, Data Exploration

(*)SticiGui, Ch. 4 [link]: *Measures of Location, Spread or Variability, The Range, IQR and SD*

(α)Chetty, Raj, John N. Friedman, Emmanuel Saez, Nicholas Turner and Danny Yagan. 2017. "Mobility Report Cards: The Role of Colleges in Intergenerational Mobility." Working paper. [link] [non-technical summary]

Lecture 4, Potential Outcomes Framework

(*)MM, Ch. 1

Lecture 5, Randomized Experiments

(α)Duflo, Esther and Emmanuel Saez. 2003. "The Role of Information and Social Interactions in Retirement Plan Decisions: Evidence from a Randomized Experiment." *Quarterly Journal of Economics*, 118(3): 815-842.

(α)Bloom, Nicholas, James Liang, John Roberts, and Zhichun Jenny Ying. 2015. "Does Working from Home Work? Evidence from a Chinese Experiment." *Quarterly Journal of Economics*, 130(1): 165-218.

Lecture 6, Noncompliance

(*)MM, Ch. 3.1 and Ch. 3.2

(α)Fowlie, Meredith, Catherine Wolfram, C. Anna Spurlock, Annika Todd, Patrick Baylis, and Peter Cappers. 2017. "Default Effects and Follow-On Behavior: Evidence from an Electricity Pricing Program." Working paper. [link]

Lecture 7, Inference for Experiments

(*)Review D8 sections 12.1 and 12.3 [12.1 link, 12.3 link]

Lecture 8, Observational Studies and Matching

(*)MM, Ch. 2.1

Lecture 9, Omitted Variable Bias

(α)Gordon, Brett R., Florian Zettlemeyer, Neha Bhargava and Dan Chapsky. 2018. "A Comparison of Approaches to Advertising Measurement: Evidence from Big Field Experiments at Facebook." Working paper. [link]

(α)Carroll, Aaron E. "Workplace Wellness Programs Don't Work Well. Why Some Studies Show Otherwise." *New York Times*, August 6, 2018. Web. Accessed August 22, 2018. [link]

Lecture 10, Introduction to Regression

(*)MM, Ch. 2.2

(*)MM, Ch. 2.3

(α)Arceneaux, Kevin, Alan S. Gerber and Donald P. Green. 2006. "Comparing Experimental and Matching Methods Using a Large-Scale Voter Mobilization Experiment." *Political Analysis*, 14: 37-62.

Lecture 11, Regression Discontinuity

(*)MM, Ch. 4

(α)Hoekstra, Mark. 2009. "The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach." *Review of Economics and Statistics*, 91(4): 717-724.

(α)Anderson, Michael and Jeremy Magruder. 2012. "Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online Review Database." *Economic Journal*, 122: 957-989.

Lecture 12, Differences in Differences

(*)MM, Ch. 5