W241: Experiments and Causality

UC Berkeley School of Information, Summer 2017

Instructors: Dr. D. Alex Hughes
Dr. Daniel Hedblom

Office Hours: Monday: 5:30 - 6:30 (Daniel)
Tuesday: 5:30 - 6:30 (Alex)
Thursday: 5:30 - 6:30 (Daniel)
Friday: 1:00 - 2:00 (Alex)
By Appointment

Lead Instructor: Dr. David Reiley, Principal Scientist, Pandora

NOTE: UPDATES TO THE COURSE MATERIAL WILL OCCUR IN THE GITHUB VERSION OF THIS SYLLABUS!

Course Description
Teaching Philosophy
    Preparing for Class
Readings
Schedule
Assignments and Grading
    Grading Scale
    Due Dates
    Assignments

Course Description

This course introduces students to experimentation in the social sciences. This topic has increased considerably in importance since 1995, as researchers have learned to think creatively about how to generate data in more scientific ways, and developments in information technology has facilitated the development of better data gathering. Key to this area of inquiry is the insight that correlation does not necessarily imply causality. In this course, we learn how to use experiments to establish causal effects, and how to be appropriately skeptical of findings from observational data.

Our goals for each student in the course are:

- Become skeptical about claims of causality. When faced with a piece of research on observational data, you should be able to tell stories that illustrate possible flaws in the conclusions.
- Understand why experimentation (generating one's own data by doing deliberate interventions) solves the basic causal-inference problem. You should be able to describe several examples of successful experiments and what makes you feel confident about their results.
- Appreciate the difference between laboratory experiments and field experiments.
- Appreciate how information systems and websites can be designed to make experimentation easy in the modern online world.
- Understand how to quantify uncertainty, using confidence intervals and statistical power calculations.
- Understand why control groups and placebos are both important.
- Design, implement, and analyze your own field experiment.
- Appreciate a few examples of what can go wrong in experiments. Examples include administrative glitches that undo random assignment, inability to fully control the treatment (and failure to take this inability into account), and spillovers between subjects.

Teaching Philosophy

We want to make this course interesting and thought-provoking, and one from which you will remember some important lessons even after the final exam is over.

We believe firmly in active learning. That is, we believe that the deepest learning occurs when students teach themselves. Therefore, we expect you to do most of your learning through the readings and assignments, both on your own and in cooperation with your classmates. We do not intend to cover all important topics in lecture. Rather, our job in this course is to guide the learning by choosing readings and exercises for you, and to coach you through this learning process in a way that maximizes understanding with as little frustration as possible.

Preparing for Class

To prepare for each week's synchronous session, please complete the asynchronous session and associated readings. Each instructor will also provide guidance on topics that will be covered in the synchronous sessions, so that you may better focus your prep time.

Readings

In addition to journal articles and papers linked from the syllabus, there are three required texts for the course:

- *FE*: Field Experiments: Design, Analysis, and Interpretation, by Alan S. Gerber and Donald P. Green
  - Note: The datasets used in this book can be found at this Yale website. No need ever to type in the data from the tables in the book.
- *MM*: Mastering Metrics, by Joshua Angrist and Jörn-Steffen Pischke. (One supplmental reading will come from these authors' more advanced textbook, Mostly Harmless Econometrics: An Empiricist's Companion, which we used to use in this course until fall 2016.)
- *MTGI*: More Than Good Intentions, by Dean Karlan and Jacob Appel. This is a popular-press book rather than a textbook; it introduces us to many examples of valuable experiments in development economics.

In addition to the required readings, we also maintain a list of Suggestions for Further Reading, which we hope will be useful to you after the course is over. This list contains a number of interesting and useful books and articles that we didn't have room to cover during the course.

Schedule -- NOTE UPDATES ARE IN THE GITHUB SYLLABUS. WE RECOMMEND YOU DO NOT USE THIS!

Week	Topics	Read during Async	Read before Live Session (This syllabus takes precedence over ISVC)	Assignment due the day of the live session
1	*The importance of experimentation:* - Reverse causality - Sample selection	NYTimes HRT article; *FE* 1	Feynman; three news articles	
2	*Comparing apples to apples:*	*FE* 2; Lewis and Reiley	Karlan and Appel Chapters 1, 5, 8, 9.	Essay 1

	- Randomization and independence - Potential outcomes	[abstract and first 4 paragraphs; §1; §2 A&B]	(Other chapters optional, to read at your leisure.)	(then, read assigned peers' essays for class discussion)
3	*Quantifying uncertainty:* - Sampling distributions - Randomization inference - p-values - Statistical power - Confidence intervals	*FE* 3.0, 3.1, 3.4	Lewis and Rao [sections 1, 3.1, 3.2]	PS1 [.Rmd]; PS1 [.pdf]; Upload revised Essay 1
4	*Blocking and clustering* - Blocking can increase power - Clustering can decrease power	*FE* 3.6.1, 4.4, 3.6.2, 4.5	N/A	Essay 2 (then, read assigned peers' essays for class discussion)
5	*Covariates and regression* - Diagnostic: randomization check - Review of multivariate regression - Covariates can increase precision - Omitted-variable bias without randomization	*MM* 1, *FE* 4.3, *MM* 2, *MHE* pp 16-24, *FE* 4.1-4.2, *MHE* 3.1.4	Ayres *et al.* (Opower)	PS2 [.Rmd]; PS2 [.html]; Upload revised Essay 2
6	*Regression; Multi-factor experiments*	*MM* 6.1, *MM* pp 95-97, *FE* 9.3.3, *FE* 9.4	Skim List and Lucking-Reiley	- Vote on project proposals (See instructions in Essay 2 Forum)
7	*Heterogeneous treatment effects* - Dangers of fishing expeditions - Committing in advance	*FE* 9	Johnson, Lewis, and Reiley (Sections 1, 2, 3.1, 4.3); Goodson	

8	*Incomplete control over treatment delivery* - One-sided non-compliance - Encouragement designs - Downstream experiments - CACE vs. ATE - Attenuation bias	*FE* 5	Gerber and Green 2005; Johnson, Lewis, and Reiley (Sections 3.2-4.1, 5)	PS3 [.Rmd]; PS3 [.html] (this is a long one, allocate time accordingly; likely two weeks)
9	*Spillovers*	*FE* 8 (Alex's Sections: also: Lyft and Uber)	Miguel and Kremer (Sections 1-3,8-9); Blake and Coey (Sections 2 and 3)	Project progress report
10	*Common problems; Diagnostics; The long term view*	*FE* 11.3	DiNardo and Pischke (skim); Simonsohn *et al.* (skim)	
11	*Causality from observational data* - Natural experiments (IV) - Difference in difference - Regression discontinuity	*MM* 3.1, 4.1, 5.1	incinerator synopsis (DID); Washington 2008 (natural experiment) (skim); Lalive (RD) (skim)	PS4 [.Rmd]; PS4 [.html]
12	*Additional topics:* - (Differential) Attrition - Mediation - Generalization of Results	*FE* 7, 10	Allcott and Rogers	Peer Evaluations 1
13	*Examples of experimental design*	*FE* 12	Sherman et al.	

14	Async: *Final thoughts* - Observation versus experiment - Prediction versus inference - Attempts to fix observational data (propensity scores, matching) - How experiments have changed the world!		NYT article on 2014 Montana election experiment and Freedman: "Shoe Leather"	
15	Final project presentations			PS5 [.Rmd]; PS5 [.html]
16+				Final project report and problem set 5

Assignments and Grading

Grading Scale

We intend to use the following grading scale when grading assignments in this course:

- A+: [97.5,100]
- A: [92.5, 97.5)
- A-: [90. 92.5)
- B+: [87.5, 90)
- B: [82.5, 87.5)
- B-: [80, 82.5)
- We hope nobody will earn grades below this level, but we will extend this same pattern as far as necessary through the ranges of C (<80), D (<70), and F (<60).

Due Dates

Assignments are due each week the day before the next class session. For example, your assignment for Week 2 is due by midnight Pacific time on the day before your Week 2 class session.

Assignments

Problem Sets (50%, 10% each) A series of problem sets, mostly drawn from *FE*, many requiring programming or analysis in R.

- We encourage you to work together on problem sets, because great learning can come out of helping each other get unstuck. We ask that each person independently prepare his or her own problem-set writeup, to demonstrate that you have thought through the ideas and calculations and can explain them on your own. This includes making sure you run any code yourself and can explain how it works. Collaboration is encouraged, but mere copying will be treated as academic dishonesty.
- At this point, the course has lived for a number of semesters, and we have shared solution sets each semester. We note in particular that struggling with the problems is a key part of the learning in this course. Copying from past solutions constitutes academic dishonesty and will be punished as such; you should know that we have included language in the solutions that will make it clear when something has been merely copied rather than understood.

Essay 1 (10%) Find an observational study and critique it. (2-page paper)

Essay 2 (10%) Experiment proposal. Pose a question and sketch an experiment to answer it. This is a proposal for an experiment that a team of 5 students could carry out during the semester. (4-page paper)

Class experiment (30%) In teams of 3-5 students, carry out a pilot experiment that measures a causal effect of interest.

- The experiment
    - The experiment should involve at least 30 observations per treatment. The data may be collected either online or offline. If the latter, students may choose to divide up the data collection, but be careful to balance the data collection across potentially heterogeneous clusters in different locations.
    - The intention here is for you to learn what it's like to do an experiment in practice, not for you to have the *perfect* design or enough observations that your data will be academically publishable.
    - It's very important to run a pilot experiment with a small number of observations, to help you debug problems in execution, before going ahead to collect all your data.

    - We highly encourage you to collect real field data instead of survey data. However, collecting data via a survey is common given the time constraints of a semester. If you do so, a common solution is to use Qualtrics, to which Berkeley has a license. Register for a free account using your Berkeley login at berkeley.qualtrics.com. Then, this tutorial has good instructions on recruiting subjects to your survey using Mechanical Turk.
- Presentation
    - During one of the final classes, we will ask you to present your findings to your peers for feedback that might help you improve your final paper. Please don't spend time making the presentation pretty; this will not get an explicit grade.
- The final paper
    - The final research report should be about 10 to 20 pages.
    - There is no template or "required sections" - just describe what you did, how you estimated the effect, and the conclusions you will draw from the data. Reviewing some of the academic papers we read this term and the FE chapter on writing a research report may help.
- Peer evaluations
    - At two points during the semester, we will ask you to write short evaluations of your peers and your team as a whole. This is partially to help ensure that we don't have free-rider problems: individuals will potentially have their group grades modified by the results of the final peer evaluations if it becomes clear that some students relied too much on teammates to get the paper done. It is also a useful opportunity to think about your group's strengths and weaknesses, and look for areas of improvement in working together.