Field Experiments

(aka DATASCI W241)

UC Berkeley School of Information

Spring 2015

Instructor: [David Reiley](http://www.davidreiley.com), Research Scientist, Google

(As of Feb 2: Principal Scientist, Pandora)

Co-Instructor: [David Broockman](http://www.davidbroockman.com), PhD student, UC Berkeley

(As of July: Assistant Professor, Stanford Graduate School of Business)

[Course Description](#h.if79zr6x97v8)

[Teaching Philosophy](#h.atstrtmm50yf)

[Preparing for Class](#h.i9g8gvojal71)

[Readings](#h.cam6oyj504o3)

[Schedule](#h.h3naggdr3pww)

[Assignments and Grading](#h.3qooqap6v97l)

[Grading Scale](#h.c7c4gontu3u3)

[Due Dates](#h.gqhop7yaeel0)

[Assignments](#h.1lntqtdn91wz)

# 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

I 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.

I believe firmly in active learning. That is, I believe that the deepest learning occurs when students teach themselves. Therefore, I expect you to do most of your learning through the readings and assignments, both on your own and in cooperation with your classmates. I do not intend to cover all important topics in lecture. Rather, my 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.

For example, when you get stuck on a page of reading you don't understand; don't waste many hours on it, but instead note that you want to ask your instructors about it in class or via an email message. Similarly, when you get stuck on a math problem you can't solve, we can give you a hint. The book can't interact with you, but we can. You will also likely find it valuable to ask questions of each other when studying.

I am excited to try out this new online course format, as I think it fits well with my philosophy of active learning. I intend to emphasize interaction in our live sessions, in order to maximize opportunities to learn from each other. Since this is a new experience for me, I’m sure there is a lot for me to learn about how to take full advantage of the technology. When you have difficulty learning something, or when you see an idea for a way that I can improve the course next time I offer it, I will be grateful for your suggestions.

## Preparing for Class

To prepare for each week’s synchronous session, please first complete the asynchronous session and associated readings. Next complete the assigned reading for synchronous session. You will likely get more out of the reading and if you efer to this [Preparing for Live Sessions document](https://docs.google.com/a/davidreiley.com/document/d/1lE0nknYN16xEB4VgyUJHhwUnebgmigfqcwtLgF4pyYI/) in order to get a preview of the questions we’ll be discussing in class. In particular, for each synchronous-session reading, we have listed specific questions that we hope will guide your reading. We won’t always have time to go over all the questions in class, but if you can answer them for yourself, you will get much more out of the reading. Please come to class prepared to ask questions on anything that you have found confusing, either in the asynchronous content or in the questions we’ve asked to help you prepare for synchronous session.

# Readings

There are three required texts for the course:

* FE: [Field Experiments: Design, Analysis, and Interpretation](http://www.amazon.com/Field-Experiments-Design-Analysis-Interpretation/dp/0393979954), by Alan S. Gerber and Donald P. Green
  + Note: The datasets used in this book can be found at [this Yale website](http://isps.yale.edu/FEDAI). No need ever to type in the data from the tables in the book.
* MHE: [Mostly Harmless Econometrics: An Empiricist’s Companion](http://www.amazon.com/Mostly-Harmless-Econometrics-Empiricists-Companion/dp/0691120358), by Joshua D. Angrist and Jörn-Steffen Pischke (MHE).
* MTGI: [More Than Good Intentions](http://smile.amazon.com/More-Than-Good-Intentions-Improving-ebook/dp/B004BDP02K/), 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.

Additional readings, including newspaper articles and academic journal articles, will be available online. We also have collected a set of [further readings](https://docs.google.com/a/davidreiley.com/document/d/1IMsGTHmklhvetfJJfEm9dhoFM7bvb-YOkN_6mAM8kFM/) for those who want to explore interesting examples of experiments that go beyond the course.

Some readings will be assigned for asynchronous lecture session, while others will be assigned for synchronous session. We attempt to tell you during the asychronous content when we want you to do each reading; often we will ask you to read a section of content before continuing to the next section of asynchronous material. We recommend that you start the aynchronous learning before starting the reading, as we will usually give instructions during the async content about exactly what we want you to read, and when.

The asynchronous sessions and related readings should be completed before the relevant week’s synchronous session, because we will use class time to discuss and answer questions on that material. We also want you to complete the relevant reading before each synchronous class, so that we can have a good discussion of it.

# Schedule

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Week | Topics | Read during Async | Read before Live Session (This syllabus takes precedence over ISVC) | Assignment due the day before the following week’s live session |
| 1 | The importance of experimentation  - Reverse causality  - Sample selection | [NYTimes HRT article](http://www.nytimes.com/2007/09/16/magazine/16epidemiology-t.html); FE 1 | [Feynman](http://neurotheory.columbia.edu/~ken/cargo_cult.html); [three news articles](https://docs.google.com/a/davidreiley.com/document/d/1DSuupY-xbNJykxpO3tRA_n7_auozF5gS3gQrC3vk9z8/) | [Essay 1](https://docs.google.com/a/davidreiley.com/document/d/1DSuupY-xbNJykxpO3tRA_n7_auozF5gS3gQrC3vk9z8/) (then, read your assigned peers’ essays for class discussion) |
| 2 | Comparing apples to apples  - Randomization and independence  - Potential outcomes | FE 2; [Lewis and Reiley](http://www.davidreiley.com/papers/OnlineAdsOfflineSales.pdf) [through section III.B] | [Karlan and Appel](http://smile.amazon.com/More-Than-Good-Intentions-Improving/dp/0452297567) book: focus on chapters 1, 5, 8, 9. | [PS1](https://docs.google.com/a/davidreiley.com/document/d/1dcdxit4-0csaMmLm3Auv7hy7v5TO438AOcA4jQ5F8Ts/); Upload revised [Essay 1](https://docs.google.com/a/davidreiley.com/document/d/1DSuupY-xbNJykxpO3tRA_n7_auozF5gS3gQrC3vk9z8/) |
| 3 | Quantifying uncertainty  - Sampling distributions  - Randomization inference  - p-values  - Statistical power  - Confidence intervals | FE 3.0, 3.1, 3.4 | [Lewis and Rao](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2367103) [sections 1, 3.1, 3.2, 4.1, 4.2] | [Essay 2](https://docs.google.com/a/davidreiley.com/document/d/1pyfGMUrcGZbTpbO3vwXTz9v0yf-fLVC8e_2FHNncSm4/)  (then, read your assigned peers’ essays for class discussion) |
| 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 | [PS2](https://docs.google.com/a/davidreiley.com/document/d/1fflmvAZ8rK27H3jESu5ri-EGVPpz548Ro96okm5kvUY/); Upload revised Essay 2 |
| 5 | Covariates and regression  - Diagnostic: randomization check  - Review of multivariate regression  - Covariates can increase precision  - Omitted-variable bias without randomization | MHE 2,  MHE 3.4.3, FE 4.3,  FE 4.1-4.2,  MHE 3.1.4,  MHE 3.2.1 | [Ayres *et al.* (Opower)](http://www.nber.org/papers/w15386.pdf) | Vote on project proposals. See instructions in Essay 2 Forum |
| 6 | Regression; Multi-factor experiments | MHE 3.2.2-3, MHE 1, FE 9.3 | Skim [List and Lucking-Reiley](http://karlan.yale.edu/fieldexperiments/papers/00301.pdf) | [CITI Course on human subjects](https://docs.google.com/a/davidreiley.com/document/d/1fsO9y3c2XB9jE17b4QrbWk7C3V_dEAHAtHAQPR0KyfA/edit#) |
| 7 | Heterogeneous treatment effects  - Dangers of fishing expeditions  - Committing in advance | FE 9 | [Johnson, Lewis, and Reiley](http://www.davidreiley.com/papers/LocationLocationLocation.pdf) (Sections 1, 2, 3.1, 4.3);  [Goodson](http://www.qubitproducts.com/sites/default/files/pdf/mostwinningabtestresultsareillusory_0.pdf) | [PS3](https://docs.google.com/a/davidreiley.com/document/d/1DzmBSpLkKbV3Nce5jiJf_JL39Hkniz5_PDpmVouzCIs/) |
| 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](http://www.campaignfreedom.org/doclib/20110131_GerberandGreen2005.pdf); [Johnson, Lewis, and Reiley](http://www.davidreiley.com/papers/LocationLocationLocation.pdf) (Sections 3.2-4.1, 5) | [Project progress report](https://docs.google.com/a/davidreiley.com/document/d/1tDm17SZo7PZXfatwesygEbB2JbBB11usFVKSRKbJMX0/edit#) |
| 9 | Spillovers | FE 8 | [Miguel and Kremer](http://elsa.berkeley.edu/~emiguel/pdfs/miguel_worms.pdf) (Sections 1-3,8-9); [Blake and Coey](http://www.scu.edu/business/economics/upload/YellowPad-TomBlake.pdf) (Sections 2 and 3) |  |
| 10 | Common problems; Diagnostics; The long term view | FE 11.3 | [DiNardo and Pischke](http://www.jstor.org/stable/2951283) (skim); [Simonsohn *et al.*](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2256237) (skim) | [PS4](https://docs.google.com/a/davidreiley.com/document/d/1qKHFFf0lGK3lESn9wNuK1MEJpnsrkbR4V0k68TMS0i8/) |
|  | SPRING BREAK |  |  |  |
| 11 | Causality from observational data  - Natural experiments (IV)  - Difference in difference  - Regression discontinuity | Optional: MHE 4.1, MHE 5, MHE 6 | [incinerator synopsis (DID)](http://espin086.wordpress.com/2010/08/08/difference-in-difference-estimation-garbage-incinerators-and-home-prices/); [Washington 2008 (natural experiment)](http://economics.yale.edu/sites/default/files/files/Faculty/washington/socialization.pdf) (skim); [Lalive (RD)](http://www.econstor.eu/bitstream/10419/25810/1/517022656.PDF) (skim) | [Peer Evaluations 1](https://docs.google.com/a/davidreiley.com/document/d/1rFZtRrfE3aGYDa2ZZm6DvlteeqIYwsrsQ0Xg8nIdkxw/edit#) |
| 12 | Additional topics:  - (Differential) Attrition  - Mediation  - Generalization of Results | FE 7, 10 | [Allcott and Rogers](https://files.nyu.edu/ha32/public/research/Allcott%20and%20Rogers%202014%20AER%20-%20The%20Short-Run%20and%20Long-Run%20Effects%20of%20Behavioral%20Interventions.pdf) |  |
| 13 | Examples of experimental design | FE 12 | [Sherman et al.](http://www.tandfonline.com/doi/abs/10.1080/07418829500096281) | Data for final project |
| 14 | Final thoughts  - Observation versus experiment  - Prediction versus inference  - Attempts to fix observational data (propensity scores, matching)  - How experiments have changed the world! |  | [Freedman: “Shoe Leather”](http://cooley.libarts.wsu.edu/schwartj/pdf/Freedman_statistical%20model.pdf) | [PS5](https://docs.google.com/document/d/1yzmvi04CJLErE2I_zopPqBoAQiR6YhyhwWP8p0PaTJM/edit?usp=sharing) |
| 15 | Final project presentations;  [Peer evaluations 2](https://docs.google.com/a/davidreiley.com/document/d/18uBEmLfxPS5SM4JqzMUsqNVzwIF1gfx8sggejBZMT20/);  Extra-credit assignment (suggestions for improving course) |  |  |  |

# 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 1 is due by midnight Pacific time on the day before your Week 2 class session.

## Assignments

Here are the different graded components of the course, together with their weights in the final grade. In general, please plan to :

* **Problem Sets**. A series of problem sets, mostly drawn from GG, many requiring programming or analysis in R. - 50%
  + Note: Due to resource constraints, we will not be grading every single problem. After each problem set, we will choose a random 40% of problems to grade, and we will give you solutions to all of the problems. We want you to have incentive to do every problem on the problem set, because we feel they are important for your learning.
  + 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.
* [**Essay 1**](https://docs.google.com/a/davidreiley.com/document/d/1DSuupY-xbNJykxpO3tRA_n7_auozF5gS3gQrC3vk9z8/)**.** Find an observational study and critique it. (2-page paper) - 10%
* [**Essay 2**](https://docs.google.com/a/davidreiley.com/document/d/1pyfGMUrcGZbTpbO3vwXTz9v0yf-fLVC8e_2FHNncSm4/)**.** 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) - 10%
* **Class experiment.** In a team of 5 students, carry out a pilot experiment that measures a causal effect of interest. - 30%
  + 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](http://berkeley.qualtrics.com). Then, [this tutorial](http://brentcurdy.net/qualtrics-tutorials/link/) 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.
* **Optional extra credit assignment: Final reflections.** During final-exam week, reflect on what you learned in this course and how you can make use of it in your future work. (At least one page, but more if you feel inspired.) Tell us concrete ways that the course has caused you to think differently about research. Provide concrete suggestions for improvement for the course. If you are interested in this option, please consider opening a document early in the course, so that you can make detailed suggestions for small changes you might want to make on specific content segments or problem-set questions. Your grade on this assignment will boost your final average by 0 to 3 percentage points. If we don’t learn anything from you, you’ll get zero extra-credit points. If we learn a ton from you, you’ll get +3 percentage points.