Observational Studies Spring 2019 Syllabus

Thomas E. Love, Ph.D.

Version 2019-01-06 11:32:41

Contents

Key Information

This is the Spring 2019 syllabus page for PQHS / CRSP 500: Observational Studies, taught by Professor Thomas Love. The course is given on Thursdays from 8:30 to 11 AM, in Wolstein Research Building room 1217.

Course Home Page

The course home page, with links to everything else you'll need, is at https://github.com/THOMASELOVE/500-2019.

Course Text

As the semester progresses, we are going to read a book together. The book we're reading is

- Paul Rosenbaum's *Observation and Experiment: An Introduction to Causal Inference*, published in 2017 by Harvard University Press. The ISBN-13 number is 978-0674975576. The book is available electronically or in hardcover from Amazon and other retailers. The schedule for reading the book is part of the course schedule

Many additional sources will be presented and referenced in class, and those are gathered at the texts page of our web site.

Getting Help!

To get help for anything related to the course, email Dr. Love at `thomas.love@case.edu`.

- Dr. Love is available on Tuesdays and Thursdays at CWRU, by appointment. To make an appointment, email him at `thomas.love@case.edu`. He is usually available for 15 minutes before and 30 minutes after class for drop-in conversations. His office is Wood WG-82 L.
  - If you have any special concerns about the course, need special accommodations or any other issues for Dr. Love, please email or speak with him before or after class.

Chapter 1

Course Description

PQHS 500 (cross-listed as CRSP 500 and formerly known as EPBI 500) is a one-semester advanced course in modern design and analysis of observational studies, usually in the context of comparative effectiveness research, with a special emphasis on issues that arise in clinical and health services research.

An observational study is an empirical investigation of treatments, policies or exposures and the effects that they cause, but it differs from an experiment in that the investigator cannot control the assignment of treatments to subjects. This course is designed to introduce design, data collection and analysis methods appropriate for scientists engaged in observational studies, and will prepare students to design and interpret their own studies, as well as those of others in their field. Technical formalities will be minimized, and the presentations will focus on the practical application of methods and strategies.

1.1 Prerequisites

People take this course with a wide range of backgrounds and a common interest in using data effectively in research related to biology or medicine. All CWRU students who feel up to it are welcome, regardless of their field of study or prior experience. Students with a working knowledge of R, multiple regression, and some familiarity with logistic regression, should be well prepared.

I strongly encourage people to first complete the PQHS 431-432 sequence, or at least to be substantially familiar with the use of R and R Studio (as well as, ideally, R Markdown) on the coding side, and with multivariate regression modeling and in particular logistic regression on the statistical side. Those who are unsure if they are well prepared for the course should contact the instructor for some advice.

1.2 Everything is on the Web

https://github.com/THOMASELOVE/500-2018 is the place to go for everything related to this course. Please visit any time you need something. I update the web site frequently. You'll find links there related to:

- Your homework **Assignments**
- **Data and Code** I will provide
- my in-class presentation **Slides**
- instructions and hints related to the **Project**
- various outside **Texts**
- a detailed **Schedule** of classes and deadlines

Chapter 2

Dr. Love

Thomas E. Love, Ph.D.

- Professor of Medicine, Population and Quantitative Health Sciences, CWRU
- Director of Biostatistics and Evaluation, Center for Health Care Research & Policy, MetroHealth Medical Center
- Chief Data Scientist, Better Health Partnership
- Track Lead for Health Care Analytics, MS in Biostatistics, Department of Population and Quantitative Health Sciences, CWRU
- Fellow, American Statistical Association

2.1 Email

- Email to get help with the course: **431-help at case dot edu** (seen by Professor Love and the TAs)
- Thomas dot Love at case dot edu (for matters related to grades or individual concerns)
- Dr. Love is hard to reach by phone. Email is always the best way to reach him.

2.2 Offices

- Wood WG-82L on the ground floor of the Wood building (Tuesdays and Thursdays)
- Rammelkamp R-229A at MetroHealth Medical Center (Wednesdays and Fridays)

Dr. Love is generally available for a few minutes before and 30 minutes after class, otherwise by appointment on Tuesdays and Thursdays (send him an email to schedule an appointment.)

2.3 Web

- Web site for this course
- Dr. Love's GitHub name is THOMASELOVE.
- His Twitter handle is @ThomasELove

2.4 A More Complete Biography

Hi. I have at least three different jobs.

- I am a Professor in the Departments of Medicine and Population & Quantitative Health Sciences at Case Western Reserve University. I teach three courses per year there (PQHS 431, 432 and 500) and also lead the Health Care Analytics track of the MS program in Biostatistics.
- I direct Biostatistics and Evaluation at the Center for Health Care Research & Policy, which is a joint venture of CWRU and MetroHealth Medical Center.
- For ten years, I was the (founding) Data Director for Better Health Partnership, an alliance of people who provide, pay for and receive care in Northeast Ohio. I now serve as Chief Data Scientist there.
- I am a Fellow of the American Statistical Association, and have won some awards for my teaching and my research.
- I have been teaching at CWRU since 1994, and have taught every type of CWRU student over the years, especially graduate students in biostatistics, medicine, and management.

In research, I use statistical methods to look at questions in health policy and in particular the provision of health services. I mostly work with observational data, rather than data that emerge from randomized clinical trials, and I have a special interest in working with data from electronic health records.

- You may be interested in a study in Health Affairs showing the impact of a Medicaid-like expansion plan on care and outcomes of poor patients in Cleveland.
- Or you might be interested in our New England Journal of Medicine study of the effect of electronic health records on the care and outcomes of people with diabetes.
- In 2011, James O'Malley and I chaired the Ninth International Conference on Health Policy Statistics, here in Cleveland. Here's a recap. We may chair it again in 2021.
- I've also worked on many projects involving the use of propensity scores to make causal inferences from observational studies, particularly in heart failure.

If you want to see a list of many of my publications, knock yourself out.

I hold degrees from Columbia University in the City of New York and from the University of Pennsylvania. My dissertation advisor was Paul Rosenbaum. I am married to a brilliant woman and we are raising two terrific sons, the elder of whom just finished his first semester of college. I live in Shaker Heights. In spare moments, I do community theater, and have appeared onstage with several local groups. Occasionally I sing in concerts with the Chagrin Valley Studio Orchestra, and I also play golf and try to lose weight.

Chapter 3

Assignments

The course requires you to complete four types of "assignments" all of which are described on the course assignments page, and also listed as part of the Class Schedule.

The four types of assignments are:

1. A course project
2. An "observational studies in action" presentation
3. Seven brief "essays" about chapters in the Rosenbaum (2017) text, and
4. Five "homework" assignments involving using R to perform and present analyses.

3.1 A course project

You will do a small observational study as a capstone project for the course. Your deliverables include:

1. a **proposal** which you will submit for my approval in February (you should expect the need to do a revision or two) and you will need to receive my approval by March 1.
2. an **update** verifying that you have the data and are proceeding appropriately, due April 1.
3. a final 20-minute **presentation** to the class about your results accompanied by an **abstract**, and by R Markdown and HTML files that describe the work you did. The presentations are on April 19 and 26. You will submit your abstract and presentation slides on April 18, and then submit all of the final materials by May 1.

Complete details on the Course Project are available here.

3.2 An "Observational Studies in Action" presentation

During the semester, you will be responsible for presenting the methods and results of an observational study from the literature that uses propensity scores. Your choice of manuscript must be accepted before March 1 by Dr. Love. After you've claimed a study, you'll give a 15-minute presentation of it to the class in late March / early April, and you'll also act as "second reviewer" for one of the studies selected by your colleagues.

Complete details on the OSIA work are available here.

3.3 Essays in reaction to the course text

In most sessions, we will be reading chapters from Paul Rosenbaum's *Observation and Experiment*. In early sessions, we will be reading several chapters at once (see the Schedule for up-to-date details.) Beginning with Class 4, we will slow the pace to one chapter per class, and you will then produce an essay in reaction to each chapter (Chapters 7-13, specifically), to share with the rest of the class.

The essay prompts are available here.

3.4 Homework Assignments

The five "regular" homework assignments usually require you to do some analyses (using R and R Markdown) on data I will provide to you. All of these occur before the middle of March.

Details on the homework assignments are available here.

3.5 A Note on Feedback

500 is a very different course from 431-432 in terms of the "hands on" assistance that I make available to you as you're working on a homework assignment, in part because I don't have a TA for the course. In particular, I'm not going to review your code to be sure you're going in the right direction, even though I understand some of you have come to expect that from 431-432.

Instead, we discuss the HW assignments in class and I provide a detailed answer sketch after the fact. So if you have questions, please feel free to send them, and I'll answer sometimes, but what I am likely to say to anything I cannot resolve quickly for you (or think is worthy of deeper discussion as a group) is:

1. That sounds like an excellent question to bring up on Thursday in class, and
2. I'm going to let you (and everyone else) flounder a bit between now and our discussion of this HW in class.

3.6 Grading

- If you complete all deliverables on time, and your project and OSIA presentations are solid, you will receive an A in the course.
  - If you mostly meet that standard but don't quite (either because more than one thing is late or because you have substantial project problems that linger), either an A or B is possible.
  - Grades that appear on Canvas will reflect successful and timely completion of assignments, and nothing more.
- This is an advanced graduate school course. I don't anticipate that anyone who makes a concerted effort will fall below the B standard, as no one has in the past.
- If you have global or local concerns, raise them with me directly as soon as possible. Email is likely the best approach.

Chapter 4

A Few Writing/Presenting Tips

Note that some coding and strategy tips specific to the final project are provided here. This material is focused on communication issues.

1. Statistics is a "getting the details right" business - we care deeply about details, and this applies to writing code or complete English sentences.
2. Nothing impresses us as much as a clear and concise argument, presented using well-written English sentences, effective and well-labeled figures and tables.
3. Don't parrot back material that Dr. Love wrote or said. State ideas in your own words. Stating them in other words is, technically, plagiarism.
4. Edit your more adventurous output; don't present everything you know how to do in R, and don't forget that someone is trying to read both your code and your results.
5. Make your work easy to evaluate. In responding to an assignment, be sure to answer the question that was asked, restating it as necessary.
6. Clearly label everything: graphs, tables, your answer to a specific question. Everything. Again, make your work easy to evaluate.
7. Simplify. Emphasize ideas in plain language. Avoid jargon. Use English well.
8. Data are plural. Use "the data **are** ..." rather than "the data *is* ... "
9. A paragraph must contain more than one sentence.
10. Don't switch tenses. If you want to write in the present tense, stick to it throughout.
11. Don't write or say random sample unless you used a random number generator. If you used haphazard sampling or convenience sampling, call it what it is, and indicate whether any problems could have cropped up as a result.
12. Similarly, don't defend a method of data collection because it is random. Most of the time we want to represent some population, and a random sample is just one way to ensure that certain types of biases have a low probability of creeping in.
13. If you want to write that you used $\alpha = 0.05$ as your significance level, then state that your results were obtained using a 95% confidence level, not a 95% confidence interval, unless you are actually interpreting a confidence interval.
14. If you're looking at a p-value, then you should state either:
    - [1] We're using a 95% confidence level.

- [2] We're using a 5% significance level. or
- [3] We're using $\alpha = 0.05$.

- Don't use more than one of these expressions.

15. Refer to all p-values that are less than 0.001 or perhaps less than 0.0001 as $p < 0.001$, rather than, for instance, $p = 0.00000001$ or, worse yet, $p = 0$. In a similar vein, write all p-values that exceed 0.99 as $p > 0.99$ instead of, for instance, $p = 1$.
16. To the extent possible, don't use `computer-ese` to label variables, plots or tables. R and Markdown allow you to change the labels on graphs and tables to meaningful things – do so. Use meaningful abbreviations, as necessary, explaining what they mean on the first usage.
17. Use words that we all know, whenever possible, and provide clear definitions at the first encounter when jargon is mandatory.
18. Often the most useful thing you can do in an analysis is to turn a table into a meaningful graph.
19. When in doubt, err on the side of clearer expression. Clear thinking causes and is demonstrated by clear writing.