Econ 309/QR 309/Soc 319: Causal Inference

Tuesday, Friday 1:30-2:40, PNE 139

Spring 2014 Syllabus

Instructor: Cassandra Pattanayak
cpattanayak@wellesley.edu
Clapp 238

Office Hours: Tue, 2:45-4 and Wed 2:15-4, unless announced otherwise; or by appointment.

My office, Clapp 238, is in the back of the main floor of Clapp Library, off of Brackett Reading Room, behind the Sanger Room.

My experience is that the concepts in this course are best explained in person. I strongly encourage you to take advantage of office hours, and I look forward to discussing your questions and reactions to the material.

Description: This course focuses on statistical methods for causal inference, with an emphasis on how to frame a causal (rather than associative) research question and design a study to address that question. What implicit assumptions underlie claims of discrimination? Why do we believe that smoking causes lung cancer? We will cover both randomized experiments – the history of randomization, principles for experimental design, and the non-parametric foundations of randomization-based inference – and methods for drawing causal conclusions from non-randomized studies, such as propensity score matching. Students will develop the expertise necessary to assess the credibility of causal claims and the ability to design and analyze studies that lead to causal inferences. A variety of examples will be used, from economics, psychology, sociology, political science, medicine, and beyond.

Goals: After this course, you should be able to:
- Frame a question in terms of the causal model
- Design and analyze randomized experiments and observational studies to uncover causal effects
- Evaluate the methods used in causal studies encountered in your future coursework, your career, or the media
- Use the statistical software R

Prerequisites: Any one of ECON 203, SOC 290, POL 199, PSYC 305, or a Psychology 300-level R course; or permission of the instructor.

Distribution: Social and Behavioral Analysis.

Course Website: I will post assignments and readings on the google site.

Email: You will be automatically subscribed to the course google group (econ-309-01-qr-309-01-soc-319-01-sp14@wellesley.edu). I will use the google group for announcements, and you should feel free to use it to communicate with each other.

You can also email me personally. I will make every attempt to answer your emails within 24 hours. It is not always possible for me to answer last-minute questions just before a deadline.

Textbook: The primary textbook for this class is *Causal Inference in Statistics and Social Sciences*, by Guido W. Imbens and Donald B. Rubin, which is in the final draft stage. The pdf will be posted on the course website. Additional journal articles will also be posted. Students do not need to purchase any reading materials.

Computing: This course involves learning the statistical software R. R is popular among statisticians and other researchers because it is free, downloadable, open source, and powerful. No matter where you are after college, R will always be available.

No previous experience with R is necessary. However, you should expect that it will take time to familiarize yourself with R, and I expect to answer lots of R questions – don't start an R assignment the day it is due! Computing questions (like conceptual questions) are usually easier to answer in person than by email, so plan ahead and attend office hours.

Resources:

- *OpenIntro Statistics*, Diez, Barr, and Cetinkaya-Rundel

  http://www.openintro.org/stat/textbook.php

  This is a free, downloadable introductory statistics textbook that may be helpful if we refer to statistical concepts that you'd like to brush up on or have never seen.

- R is downloadable from www.r-project.org (or, google the letter "R"). Instructions for downloading R and R resources will be posted on the course website.

Course Requirements and Grading:

15% - Project (due Tue, May 20)

20% - Midterm I (Fri, Feb 28).

25% - Midterm II (Fri, Apr 25).

36% - Assignments (six of them, spread over semester – deadlines in outline below)

4% - Participation

Policies:

- Assignments should be submitted electronically by 10 pm on the due date, unless specified otherwise. Submission instructions will be posted. Typically, you will be required to submit your R code in addition to your write-up.

- You should also submit a hard copy of your write-up shortly after the deadline (e.g., at class the next day). You should not print out hard copies of R code.

- I will deduct 25% of the possible points from graded assignments for each day of lateness: if the deadline is 10 pm on Thu, then assignments received by 10 pm Fri can score no higher than 75%; assignments received by 10 pm Sat can score no higher than 50%; assignments received by 10 pm Sun can score no higher than 25%; and assignments received after 10 pm Sun will receive no credit.

- No assignments will be dropped when your grade is calculated.

- Extensions will be granted only in exceptional circumstances, such as serious illness or a family emergency, or to accommodate special circumstances as described below. **If you think you may miss a deadline because of travel, interviews, your senior thesis, or other scheduling conflicts, submit the assignment early.**

- Midterms will be held only on the specified dates. There will be no make-ups. The midterms will be closed-book. You may bring two two-sided pages of notes (on 8.5" x 11" paper) and a calculator.

Honor Code and Collaboration Policy: The Honor Code will be strictly enforced. Sources must be cited in written assignments. Unless specified otherwise, all assignments and the project should be submitted individually. However, you are encouraged to (orally) discuss the assignments with your classmates. Each student must write up solutions separately. Be sure that you have worked through each problem yourself and that the answers you submit are the results of your own efforts. For assignments and the project, you also may not share another student's computer code, submit output from another student's computer session, or allow another student to share your code or output. A good rule of thumb: if a fellow student asks you if you would like to discuss a problem on an assignment, you are encouraged to say "yes"; if a fellow student asks to see your answer to a problem or code, the answer is "no." **You are expected to explicitly acknowledge collaborators by writing their names at the top of your assignments.**

Laptop/Phone Policy: I expect that you will actively engage in class meetings. This means that laptops are used only for note-taking or class exercises and phones are away. My experience is that it is easier to take notes by hand in a quantitative class, and screens can be distracting to others. Please let me know if you plan to take notes electronically.

Accommodations: If you have documentation from Disability Services or anticipate conflicts with the course due to religious observance or a Wellesley-sponsored activity (such as an athletic team), please let me know early in the semester so that arrangements can be made.

Course Outline (subject to change) and Deadlines

		Topic	Readings from Imbens and Rubin	Due
Week 1	Tue, Jan 28 Fri, Jan 31	Introduction to the causal framework	Chapter 1 (Chapter 2)	
Week 2	Tue, Feb 4 Fri, Feb 7	Assignment mechanisms	Chapter 3	Asst 1 (Thu, Feb 6)
Week 3	Tue, Feb 11 Fri, Feb 14	Classical randomized experiments Fisher's approach	Chapter 4 Chapter 5	
Week 4	Tue, Feb 18 Fri, Feb 21	Fisher's approach Neyman's approach	Chapter 5 Chapter 6	Asst 2 (Thu, Feb 20)
Week 5	Tue, Feb 25 Fri, Feb 28	Neyman's approach **Midterm I**	Chapter 6	**Midterm I (Fri, Feb 28)**
Week 6	Tue, Mar 4 Fri, Mar 7	Imputation-based inference	Chapter 8	
Week 7	Tue, Mar 11 Fri, Mar 14	Regression in randomized experiments	Chapter 7	Asst 3 (Thu, Mar 13)
Week 8	Tue, Mar 18	Power and related considerations	(Chapter 9, Chapter 10)	
Spring Break	*No class:* *Fri, Mar 21* *Tue, Mar 25* *Fri, Mar 28*			
Week 9	Tue, Apr 1 Fri, Apr 4	Observational studies Propensity scores	At this point, we will focus on other readings, though you may still find chapters 12-21 a helpful resource	Asst 4 (Thu, Apr 3)
Week 10	Tue, Apr 8 Fri, Apr 11	Matching and subclassification, application to health example		Project proposal (Thu, Apr 10)
Week 11	Tue, Apr 15 Fri, Apr 18	Theory underlying propensity score methods		Asst 5 (Thu, Apr 18)
(Mon schedule)	*No class:* *Tue, Apr 22*			
Week 12	Fri, Apr 25	**Midterm II**		**Midterm II (Fri, Apr 25)**
Week 13	Tue, Apr 29 Fri, May 2	Regression in observational studies Instrumental variables	Chapter 23 Chapter 24	Project update (Thu, May 1)
Week 14	Tue, May 6 Fri, May 9	Principal stratification		Asst 6 (Thu, May 8)
Finals Period				**Project (Tue, May 20)**