STATS C160/C260

UCLA Department of Statistics

Causal Inference for Health Data

Course Information

Professor:	Drago Plecko
Lectures:	Tuesday and Thursday, 11:00am - 12:15am, Mathematical Sciences 5200
Office Hours:	Thursday 10:15am to 11:15am. In-person. Also feel free to post a question on the Bruin Learn Discussion board any time.
Office Location:	Mathematical Sciences 8105E, or zoom at this link

Course description: Lecture, three hours; Preparation: basics of probability theory and regression modelling. Requisites: STATS 100A, STATS 101A, STATS 20 (or Stats 21 or PIC 16A); STATS 102A is recommended but not required, especially for students with high proficiency in python. Designed for graduate students and advanced undergraduate students seeking training in the theory and application of causal inference, with a specific focus on health data. P/NP or letter grading.

Discussion

We will use the Bruin Learn Discussion board for Q&A, discussions on topics, FAQ,

etc. Please use it freely.

Motivation and Synopsis

Statistics C160/C260 is a course in causal inference with a focus on *healthcare data*. It will cover the foundations of causal inference (Structural Causal Models, Causal Diagrams, and Pearl's Causal Hierarchy), causal effect identification theory, and estimation of causal effects. Further advanced topics include heterogeneity of causal effects, variation analysis using counterfactual reasoning, and sensitivity analysis for common data pitfalls such as unobserved confounding, missing data, and measurement error. The course offers multiple case-studies, where the goal is to translate the learned content into practical real-world data analyses using the R Programming Language.

This course is designed for graduate students majoring in Statistics, Computer Science, or Computational Medicine, and advanced undergraduate students who are planning to attend graduate school.

This course is most appropriate for students seeking to learn about causal inference with a focus on real-world health data applications. The goal is to equip students with the foundations of causal inference and familiarize them with software tools necessary to undertake causal inference analyses on health data.

The course has two different perspectives, which makes it suitable for different groups of students:

(A) It focuses on the conceptual underpinnings of modern causal inference, making the course appealing to statistics or computer science graduate or undergraduate students.

(B) It involves the analysis of real-world healthcare data, with implementations using open-source software. This aspect of the course may be appealing to graduate students in medical sciences, or undergraduate students in computational medicine.

Prerequisites

> Stats 100A and Stats 101A, and Stats 20 (or Stats 21 or PIC 16A), while Stats 102A is recommended. However, students with high proficiency in python may complete the course successfully without any background in R.

Course Materials & Resources

There is no specific textbook that the course will follow exactly. Instead, all the slides used in the course will be shared with students in advance, and these slides will often include pointers to academic papers or resources. Several resources are proposed for additional learning (or reading in parallel) on the course-related topics.

Recommended Books

Less Technical but Informative

[**PM**] Pearl, Judea and Mackenzie, Dana

The Book of Why: The New Science of Cause and Effect (2018).
Basic Books

Available in UCLA Library, see here.

More Technical Textbook on Causal Inference

[**P**] Pearl, Judea

Causality: Models, Reasoning and Inference (2009).
Cambridge University Press

Available online in UCLA Library, see here.

Advanced Textbook on Causal Inference

[**B**] Bareinboim, Elias.

Causal Artificial Intelligence (2025).

Available online freely at this link.

Background on R

[**W**] Wickham, Hadley.

Advanced R. (2014).
Chapman and Hall/CRC.

Required and available online as a ebook from the library, free for UCLA students. http://www.crcnetbase.com/doi/book/10.1201/b17487

Students must be connected to the UCLA network to obtain their free download. Students who would like to download the textbook off-campus may do so by connecting the the UCLA network via VPN https://www.bol.ucla.edu/services/vpn/all.html.

Syllabus of the Course

The syllabus of the course is presented through a weekly schedule. Later topics may not be reached or fully covered (Weeks 9 & 10), and in case of need one of the two topics (Sensitivity for Measurement Error / AI in Health) will be dropped.

Week	Contents by Lecture
1	**L1:** Introduction & Motivation, Pearl's Causal Hierarchy **L2:** Structural Causal Models & Causal Diagrams
2	**L3:** Testable implications & d-separation **L4:** Identification of Causal Effects
3	**L5:** Confounding & Back-Door Criterion **L6:** Estimation of Causal Effects (Regression-based, Propensity, Mixed)
4	**L7:** Case Study – Comparing Estimators in Practice **L8:** Beyond Back-Door & Causal Calculus
5	**L9:** Counterfactuals – Part I **L10:** Counterfactuals – Part II
6	**L11:** Heterogeneity of Causal Effects (Oxygen Therapy in ICU) **L12:** ***Midterm Exam***

7	**L13:** Causal Explanations for Health Equity (Race Differentials in ICU) **L14:** Sensitivity to Unobserved Confounding – Part I (Obesity Paradox)
8	**L15:** Sensitivity to Unobserved Confounding – Part II **L16:** Causal Inference with Missing Data – Part I
9	**L17:** Causal Inference with Missing Data – Part II (Marijuana & Mental Health) **L18:** Causal Inference with Measurement Error (Obesity Paradox Revisited)
10	**L19:** Causal Artificial Intelligence Tools for Health Data **L20:** Using AI for Health Data (Opioids & Delirium; Allocating Respirators)

Learning Outcomes

- Students will learn the basics of causal inference, and how to differentiate between observational, interventional, and counterfactual data.
- Students will be able to formulate and model causal questions in the context of healthcare data.
- Students will learn how to perform causal explanations of different health phenomena.
- Students will learn about the key difficulties of handling healthcare data, and about techniques for investigating the impact of data deficiencies.
- Students will be able to write code in R to implement various causal inference techniques.

Course Webpage and Discussion Forum

The course has a webpage through the UCLA Bruin Learn system. The webpage will be continuously updated throughout the course with handouts, homework assignments and solutions. Users sign in to Bruin Learn with their UCLA Logon IDs.

I will be using Bruin Learn Discussion board to provide discussion of issues in class and related questions. For questions that might be of interest to other students, please use it to ask a question rather than solely emailing me. On the Discussion board, other students and the TA can also answer questions. Example of questions are about problems with access to resources, homework or computer questions.

Please regularly consult this classes Bruin Learn home page, the Discussion board and the archive of the Announcements mailing list. It will contain lecture notes, homework, solutions and course information.

Computer Usage and Software

Access to a computer during the course is assumed. Our computer interface to R will be the RStudio IDE, which you can download from Posit. For case studies, Students with a high degree of proficiency in python may choose to write their own code.

Course Requirements and Grades

The course grading is different for undergraduate and graduate students. Specifically:

Undergraduate students (STATS C160)

- 10% Homework (4 assignments, 2.5% each)
- 50% Midterm exam (Thursday, February 12)
- 40% Final project

Graduate students (STATS C260)

- 10% Homework (4 assignments, 2.5% each)
- 30% Midterm project report
- 60% Final project

Homework

There will be four (4) homeworks in total. All homework assignments will be made available from the Bruin Learn *Assignments* page. Students will submit the solution as self-contained PDF files electronically via the Gradescope link on the Bruin Learn home page.

Homework is worth 10% of the course grade (2.5% per assignment). Homework is graded **entirely based on submission**: any reasonable attempt submitted by the deadline receives full credit. None of the homework scores will be dropped. It is your responsibility to verify that your homework assignment successfully uploaded by the deadline.

Students are free to discuss homework problems and solutions. Discussing the contents of

the course with fellow students can be a valuable element of the learning process, and doing so is therefore generally encouraged. However, each student must hand in their own solutions, and the student should, if asked, be able to explain the solutions.

Midterm Exam (Undergraduates)

The midterm exam applies to undergraduate students (STATS C160) only and takes place on **Thursday, February 12**. The exam assesses understanding of causal questions, causal models, identification strategies, and interpretation of results covered in the first part of the course.

Graduate students may attend and take the exam voluntarily. Their performance will not be graded or recorded.

Course Project

The project can take one of two forms. You should choose only one of the two options. Regardless of the form, the requirements are as follows:

- Undergraduate students submit a single final project report, worth 40% of their grade,
- Graduate students submit a midterm report (30% of the grade), and a final project report (60% of the grade).

The **first project form** is to use a proposed template for data analysis.
For the first option, the course will offer three suggestions for data analysis. Specifically, these template suggestions will include (1) the dataset to be used; (2) the type of analysis to be carried out, conveying the general idea. When choosing this approach, students should seek to analyze data by applying methods covered in the class, but also methods beyond the scope of the class.

The **second project** is to undertake your own data analysis.
For the second option, you can select any dataset you find interesting. This is especially encouraged for graduate students, who can aim to choose a topic related to their graduate work or major area. A routine analysis will not be sufficient for the final project report; a good job will show understanding of the problem and possible solutions and techniques to consider.

Midterm Project Report (Graduates Only)

The mid-term report should be thought of as a research proposal. The key goal is to formulate the investigation clearly, i.e., (i) what is the causal query of interest; (ii) what is the assumed causal model? (iii) what are the identification/estimation/sensitivity techniques used, and why?

This way, the student provides a sketch of what they wish to achieve by the end of the quarter. The mid-term report should be used as the starting point for completing the final report; it serves to ground the overall investigation, and obtain feedback about halfway through the quarter. This report is graded based on: (a) clarity of conveying the key causal aspects of the analysis; (b) level of mathematical formalism of the described analysis; (c) quality & depth of the proposed analysis.

Final Project Report (All Students)
The final report should be concise, at the level so that a fellow student can read it (after they have taken the course). Key ingredients include the description of a causal query, causal model, and causal techniques that are employed (graduate students, who also submit a midterm project report, are allowed to improve this part in the final report). Emphasis is placed on (a) clarity of conveying the key causal aspects of the analysis; (b) level of mathematical formalism of the described analysis; (c) quality & depth of the proposed analysis.
On top of a clear description of the causal analysis, the final report should include (iv) all the analyses carried out, and (v) key implications of the analyses. Additional analyses on semi-synthetic data to illustrate the main points are encouraged. The length of the final report can vary, but there is a strong emphasis on quality over quantity. Concise reports as short as 4-5 pages can obtain maximum marks for undergraduates (these reports often contain high-quality visualizations, and condensed result summaries – think about the plots/visuals you want to include!), while for graduate students reports as concise as 7-8 pages can obtain maximum marks. The final report should also contain a link to a code repository with all the used code, fostering reproducibility of all the results. The key point for the final report is to convey ideas clearly, and highlight the takeaway points from the analysis. Use of generative AI tools such as ChatGPT for writing entire project reports is strictly forbidden. On top of the rubrics (a)-(c) described at the start of the paragraph, the final report is also evaluated based on (d) quality and depth of the undertaken analysis & clarity of the described results; (e) quality of the analysis implications.

Grade rubric.
>90% = A, 85-90% = A-, 80-85%=B, 75-80% = B-, 65-75% = C, 50-65%=D, <50% = F.

Differentiation between Undergraduate and Graduate requirements

The project requirements and expectations will be different for C160 and C260 students. The level of the project is required to be substantially higher for graduate students (C260). C260 students are strongly encouraged to choose and carry out their own data analysis instead of using a proposed template, while C160 students may wish to use a proposed template (although undergraduate students are more than welcome to do their own analysis if time

permits). Furthermore, the level of exploratory detail in project reports is expected to be higher for C260 students compared to C160 students.

Late Submissions

Late submissions will not be accepted.

Comments of suggestions about the course

Suggestions on how to improve the course are welcome at any time, via email or in person. Towards the end of the quarter, an anonymous feedback survey will be carried out. Here, students will also have the opportunity to provide feedback anonymously.

University Policies

Academic Integrity

UCLA is a community of scholars. In this community, all members including faculty, staff and students alike are responsible for maintaining standards of academic honesty. As a student and member of the University community, you are here to get an education and are, therefore, expected to demonstrate integrity in your academic endeavors. You are evaluated on your own merits. Cheating, plagiarism, collaborative work, multiple submissions without the permission of the professor, or other kinds of academic dishonesty are considered unacceptable behavior and will result in formal disciplinary proceedings usually resulting in suspension or dismissal. See the Dean of Students website for more information.

[source: Dean of Students syllabus statement (syllabus)]

Accommodations for Students with Disabilities:

If you are already registered with the Center for Accessible Education (CAE), please request your Letter of Accommodation in the Student Portal. If you are seeking registration with the CAE, please submit your request for accommodations via the CAE website. Students with disabilities requiring academic accommodations should submit their request for accommodations as soon as possible, as it may take up to two weeks to review the request. For more information, please visit the CAE website , visit the CAE at A255 Murphy Hall, or contact us by phone at (310) 825-1501.

[source: Center for Accessible Education (Faculty Questions)]

Resources for Students

UCLA provides resources if you are feeling overwhelmed and need personal and/or academic assistance.

Please see the Red Folder REV2020 web for more information.

Title IX

Advocacy and Confidential Services

Please note that Title IX prohibits gender discrimination, including sexual harassment, domestic and dating violence, sexual assault, and stalking. If you have experienced sexual harassment or sexual violence, you can receive confidential support and advocacy at the CARE Advocacy Office for Sexual and Gender-Based Violence, 205 Covel Commons, Los Angeles, CA, 90095, care@careprogram.ucla.edu, (310) 206-246 5. Counseling and Psychological Services (CAPS) provides confidential counseling to all students and can be reached 24/7 at (310) 825-0768.

Reporting and Non-confidential Services

Your professor is required under the UC Policy on Sexual Violence and Sexual Harassment to inform the Title IX Coordinator should he become aware that you or any other student has experienced sexual violence or sexual harassment. In addition, You can also report sexual violence or sexual harassment directly to the University's Title IX Coordinator, 2255 Murphy Hall, titleix@equity.ucla.edu , (310) 206-3417. Reports to law enforcement can be made to UCPD at (310) 825-1491.