

Statistics 201a
Lecture: MW 2:00pm - 3:15pm, Room: Geology Building 6704
Discussion Section: F 3pm-3:50pm, Room: Mathematical Sciences 5128

Instructor: Erin Hartman, Mathematical Sciences 8105H, ekhartman@ucla.edu

Office Hours: 10:30 AM - 11:30 AM Tuesday (8105H Mathematical Sciences)

TA: Mitchell Hill, mkhill@ucla.edu

Office Hours: Thursdays 4pm - 5pm (8141 Mathematical Sciences)

Course Objectives: This course is intended as an introduction to research design and analysis. We will focus on experimental design, sampling design, and regression analysis. This course will emphasize the relationship of these statistical tools for drawing causal inferences, however prediction and description will also be covered. To achieve this we will focus on:

- The principles of statistical inference
- The difference between design-based inference and model-based inference
- The building blocks of causal inference
- Simple random sampling, stratified and clustered sampling, and survey weighting
- Randomized experiments, blocking, and permutation inference
- Regression analysis: bivariate, regression discontinuity design, multiple regression
- Causal mediation and instrumental variables

Prerequisites: This course assumes that you have a foundation in probability and statistics and matrix algebra. Additionally, students should feel comfortable coding in R.

Course website: The course website will be hosted at:

<https://ccle.ucla.edu/course/view/18F-STATS201A-1>

This will be the main source for all course materials, including problem sets, lecture notes, and readings. In addition to lectures and office hours, *all questions about lectures, problem sets, and other course materials* should be posted to the discussion forums on the course website. This allows all students to benefit from the discussion, and to help each other understand the materials. If you have a question, chances are one of your colleagues does too. All non-personal questions should be posted to the discussion forum, and both students and instructors are encouraged to participate in the discussion. *Both students and the instructor are encouraged to participate in answering questions and helping each other on the forum.*

Textbooks: There is no single required textbook for this course. Some readings may be assigned for specific lectures or problems sets, and will be noted at that time. The main resource will be the lecture notes, which will also be provided on the course website. The following two texts will be very helpful, and I would suggest acquiring them:

- Guido W. Imbens and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, 2015.

This book can be downloaded for free through the library.

- Peter M. Aronow and Benjamin T. Miller. *Foundations of Agnostic Statistics*. *Unpublished manuscript*.. 2017. This book is being provided in preprint at the generosity of the authors. It is available on the course website. **DO NOT SHARE THE MANUSCRIPT.**

Some chapters from the following books may be helpful:

- Morris H. DeGroot and Mark J. Schervish. *Probability and Statistics*. Addison Wesley, Boston, 3rd edition, 2002.
- David A. Freedman. *Statistical Models: Theory and Practice*. Cambridge University Press, Cambridge, 2nd edition, 2009.
- Rice, John. *Mathematical Statistics and Data Analysis*. Nelson Education, 2006.
- Sarndal, Carl-Erik and Sixten Lundstrom. *Estimation in Surveys with Nonresponse*. Wiley. 2005.
- Thomson, Steven K. *Sampling*. Wiley. 2012.
- VanDerWeele, Tyler. *Explanation in Causal Inference*. Oxford University Press, New York. 2015.

For a background on coding in R:

- Garrett Golemund and Hadley Wickham. *R for Data Science*. O'Reilly. 2017. <http://r4ds.had.co.nz>
- Venables, W.N and Brian D. Ripley. *Modern Applied Statistics with S*. Springer-Verlag, New York. 4th edition. 2010.
- Monogan, Jamie. *Poilitical Analysis using R*. Springer, New York. 2015.

Software: One of the main lessons that students can learn is how to clearly present research results. To that end, we will be coding in the programming language R, which is widely used in statistics, the social sciences, computer science, data science, and many other fields. Homeworks will be submitted in Rmd, a tool that can be used to create easily reproducible code (and research papers!).

Requirements and Evaluation

Grading in this class will be based on the components described below. **Late work will not be accepted without prior permission.** Makeup exams will not be given, and students who miss exams will receive a score of 0 absent extraordinary circumstances.

Participation - 10%

Participation will be based on the level of engagement in class and on the CCLE discussion board. The CCLE discussion board is intended as a way for students to help each other learn, and participation is strongly encouraged.

Problem sets - 40%

Problem sets will be assigned approximately bi-weekly, with each contributing equally towards this portion of the final grade. Problem sets will contain a mix of analytical and coding problems.

Collaboration Policy: Collaboration is allowed with students currently enrolled in the course. You may not collaborate with students outside of the course. All persons and resources consulted in completion of the problem set must be clearly listed. Questions should be posted to the CCLE discussion forum.

Submission Policy: All submissions should be typed in `RMarkdown`. A printed version of the final write-up should be brought to class. Additionally, the `.Rmd` and final PDF should be uploaded to CCLE before the class period the problem set is due.

Midterm exam - 25%

The midterm exam will be held in class on Wednesday, 11/7 and will cover the material discussed in class up to that point.

Final Exam – 25%

The comprehensive final exam will be held on Tuesday, 12/11 from 11:30am - 2:30pm.

Class policies

Grade Appeals

I am happy to meet with students about grading issues, but only after they have met with the graduate TAs and submitted a grading request in writing. Please meet first with the graduate TAs with any concerns about evaluation.

If you wish to appeal the grading of an exam or assignment, you must return it to the graduate TAs. You must staple to the original graded exam or assignment a note that states which question(s) is (are) to be re-graded and why you believe that your answer deserves more credit. Nothing additional (notes, explanations, etc.) should be written on the original assignment and NO changes or erasures should be made on the original before regrading.

Grade appeals will only be considered if submitted, in writing, up to one week after the assignment is returned.

However, no adjustments will be made to final grades under any circumstances.

Academic Honesty

Cheating and plagiarism will not be tolerated. I strongly encourage you to review the University's policies regarding academic honesty, which you can read [here](#).

In general, if you have any question, please feel free to ask your TA or Professor Hartman. Specific rules for this course:

- The homeworks and in-class work are “open book” and “open notes.” However, you *may not* make use of answer keys or graded assignments provided by students from previous years for either homeworks or in-class assignments.
- You may collaborate with other students to discuss homework problems, but all submitted work must be your own.
- All exams will be “closed book” and no calculators will be permitted. You are to consult *only* with Professor Hartman or a TA during exams.

As required by the University's [regulations](#) all suspected cases of cheating or plagiarism will be referred to the Office of Student Conduct.

Students with disabilities

Students with disabilities enrolled in this course who may need disability-related classroom accommodations are encouraged to make an appointment to see Professor Hartman before the end of the second week of the quarter. All conversations will remain confidential. Please also arrange to have the required documentation sent to Professor Hartman for any accommodations *as soon as possible*.

Students needing academic accommodations based on a disability must contact the Center for Accessible Education (CAE) at (310) 825-1501 or present in person at Murphy Hall A255. As the professionals delegated authority from the campus to determine reasonable disability accommodations, CAE will assess all requested accommodations and communicate appropriately with faculty. In the event that a student has approval for proctoring arrangements during exams, please inform your respective professors and/or Teaching Assistant(s) before date of exam(s). When possible, students should contact the CAE within the first two weeks of the term as reasonable notice is needed to coordinate accommodations. For more information visit www.cae.ucla.edu.

Religious observances

Some students may wish to take part in religious observances that occur during this quarter. If you have a religious observance that conflicts with your participation in the course, please meet with Professor Hartman *before the end of the second week of the quarter* to discuss accommodations.

Gender Discrimination

Title IX prohibits gender discrimination, including sexual harassment, domestic and dating violence, sexual assault, and stalking.

Confidential Resources: Students who have experienced sexual harassment or sexual violence can receive confidential support and advocacy at the CARE Advocacy Office for Sexual and Gender-Based Violence, 1st Floor Wooden Center West, CAREadvocate@caps.ucla.edu, (310) 206-2465. You can also report sexual violence or sexual harassment directly to the University's Title IX Coordinator, located in 2241 Murphy Hall, email: titleix@conet.ucla.edu, phone: (310) 206-3417.

Counseling and Psychological Services (CAPS) also provides confidential counseling to all students and can be reached 24/7 at (310) 825-0768.

Non-confidential Resources: You can also report sexual violence or sexual harassment directly to the University's Title IX Coordinator, 2241 Murphy Hall, titleix@conet.ucla.edu, (310) 206-3417. Reports to law enforcement can be made to UCPD at (310) 825-1491. These offices may be required to pursue an official investigation.

Faculty and TAs are required under the UC Policy on Sexual Violence and Sexual Harassment to inform the Title IX Coordinator – **a non-confidential resources** – should they become aware that you or any other student has experienced sexual violence or sexual harassment.

Counseling and Psychological Services

Graduate school is a very stressful time. There are many resources on campus for students to help with study habits, anxiety, stress, and depression. Students are encouraged to check out the Counseling and Psychological Services (CAPS) center at UCLA (<http://www.counseling.ucla.edu>) for these resources. In addition to counseling and other psychological and mental health services, they provide additional resources such as readings on dealing with stress and anxiety, group counseling sessions, mindfulness trainings, and other behavioral services. Walk-in hours are Monday-Thursday 8am-4:30pm and Friday 9am-4:30pm in John Wooden Center West. **Crisis counseling is also available 24 hours/day at (310) 825-0768.** CAPS services are often covered by UC SHIP.

Additional Resources

Find additional resources [here](#). This includes immigrant students, homeless and food shelters, the office for equity, diversity, and inclusion, healthy living, and the UCLA recreation department.

Development of this course

Learning should not happen in a vacuum. To help ensure the best chance for success for the students of this course, this course draws on the format, syllabus, and materials from similar successful courses at peer institutions. I am incredibly thankful to Kosuke Imai, Jas Sekhon, and Chad Hazlett, and all those who influenced them, for their gracious help.

Tentative Schedule

The following serves as a tentative schedule, with approximately two weeks for each section.

Introduction

- Statistical Inference: Descriptive, Predictive, and Causal
 - Freedman: Chapter 1
- Probability
 - Aronow and Miller: Chapter 1
- Identification, Estimation, Confidence Intervals, and Hypothesis Testing
 - Aronow and Miller: Chapter 2-3
 - Degroot and Schervish: 7.1-7.5, 8.1-8.5, Chapter 9
 - Freedman: Chapter 3
- Design-based vs. Model-based Inference
 - Thompson Chapter 2.

Sampling Design and Surveys

- Sampling: Sampling Frames, Stratified and Clustered Sampling, Multi-stage Sampling
 - Thompson: Chapters 2, 11, 12, 13
- Unequal-probability sampling and Horvitz-Thompson
 - Thompson Chapter 6
- Precision and Sample Size
 - Thompson: Chapter 4
- Total Survey Error
 - Groves (2010)
- Survey Weighting: Raking, Post-stratification, Calibration
 - Lohr: Chapter 8
 - Galton and Cervantes (2003)

Causal Inference and Experimental Design

- Introduction to Causal Inference
 - Imbens and Rubin: Chapters 1-3
- Randomization, Variance, Neyman Paradox

- Imbens and Rubin: Chapter 6
- Randomization Inference
 - Imbens and Rubin: Chapter 5
- Experimental Design: Blocking, Clustering, Paired and Optimal Designs
 - Imbens and Rubin: Chapters 4, 7-9
- Identification vs. Estimation, Partial Identification
 - Aronow and Miller: Chapter 7
- Bias, pre-analysis plans, sample splitting

Linear Regression

- Least Squares: Assumptions, Characterization, Interpretation
 - Aronow and Miller: Chapter 4
 - Freedman: Chapter 2
 - Imbens and Rubin: Chapter 7
 - DeGroot and Schervish: Chapter 11.1-11.4
- Robust standard errors
- Regression Discontinuity

Multiple Regression

- Multiple Regression: Assumptions, Characterization, Gauss-Markov, Interpretation, Diagnostics
 - Hayashi: Chapter 1.1-1.4
 - Freedman: Chapter 4, 5
 - DeGroot and Schervish: Chapter 11.5-11.8
- Weighted Least Squares
 - Hayashi: 1.6
- Threats to validity: Omitted Variable Bias, Measurement Error
- Non-linear CEFS and Overfitting
 - Aronow and Miller: Chapter 4.3
- Regression and Experimental Data
 - Imbens and Rubin: Chapter 7
- Missing data
 - Aronow and Miller: Chapter 6