Syllabus for Intro to Statistical Methods PS 812
Fall 2019

Adeline Lo and Dillon Laaker*

Instructor:	Adeline Lo	TA:	Dillon Laaker
Office:	322C North Hall	Office:	TBD
Email:	aylo@wisc.edu	Email:	laaker@wisc.edu
Class:	08:00 - 09:15 AM, MW	Section:	09:55 - 10:45 AM F 216 Ingraham
	586 Van Hise Hall		11:00 - 11:50 AM F 383 Van Hise
OH:	09:30 - 10:30 AM, MW	OH:	09:45 - 11:45 AM T Education Crossroads Cafe
			05:00 - 07:00 PM Th Fair Trade Coffee, State St

1 Overview

Statistical analysis is now an integral part of conducting and consuming research in the social sciences. This class serves as an introduction to the graduate quantitative methods sequence at UW Madison, Political Science. It focuses on foundational probability theory and builds up to linear regression and its application to causal inference. We will also be working on developing programming skills that play an important role in modern research.

1.1 Class goals

This course is the first of a three-semester sequence for PhD students in Political Science. It is primarily designed for graduate students in the social sciences. You will learn the statistical and computational principles necessary to perform modern, flexible and creative analysis of quantitative social data. The goal of the sequence is to move you from being consumers of quantitative research to producers of it. It will require a lot of hard work for all of us to achieve that; however, the class is structured to provide you with the framework

*Last edited: August 30, 2019

to achieve that in combination with hard work and reaching out when appropriate for extra support.

By the end of the semester, you will be able to:

1. Critically read, interpret and replicate the quantitative content of many articles in the quantitative social sciences
2. Conduct, interpret, and communicate results from analysis using multiple regression
3. Explain the limitations of observational data for making causal claims, and begin to useexisting strategies for attempting to make causal claims from observational data
4. Write clean, reusable, and reliable R code
5. Feel empowered working with data

The statistics sequence is designed to get you to a point where you can teach yourself new statistical methods by reading the literature. While we cannot teach you all the statistical techniques that you will need during your career, we can prepare you to teach yourself.

More specifically, PS 812 covers basic probability, univariate inference, linear regression and its applications in causal inference strategies. We will also provide an introduction to statistical programming in R. Upon finishing the course sequence, students should be able to read an original scholarly article describing a new statistical technique, implement it in computer code, estimate the model with relevant data, understand and interpret the results, and explain the results to someone unfamiliar with statistics.

1.2 Class and Section

Instruction for this course is conducted via two avenues: class and section/lab. Class lectures are twice a week and will typically focus on statistical material. Section meets once a week and will typically focus on practical problem solving and/or computational skills. Both are essential to the learning process.

1.3 Prerequisites

Formally, this course assumes familiarity with algebra and the basics of calculus (derivatives and integrals). This means everything covered in your math camp is fair game. If you feel refresher coverage of some of the materials may be helpful, you can refer to Adeline's website (`www.loadeline.com`), under Teaching, which offers crib notes on mathematical concepts directly related to the methods sequence. Informally, and even more importantly, the most essential prerequisite is a willingness to work hard on possibly unfamiliar material. Learning

statistics and programming can be like learning new languages, which require time and dedication. Similar to studying languages, fluency and comfort come from daily practice and consistent effort.

2 Materials

2.1 Computational tools

The best way, and often the only way, to learn about data analysis and new statistical procedures is by doing. We will therefore make extensive use of a flexible (open-source and free) statistical software program called R, RStudio, and a number of companion packages. Problem sets and the takehome midterm will be completed in R Markdown. You will learn how to program in this class, if you do not know already.

2.2 Readings

The majority part of the course we will be using the following book:

(DS) Degroot, Morris H. and Mark J. Schervish. 2012. *Probability and Statistics*. Boston: Addison-Wesley Fourth Edition.

The latter part of the course will require readings from:

(AP) Angrist, Joshua D. and Jorn-Steffen Pischke. 2009. *Mostly Harmless Econometrics*. Princeton: Princeton University Press.

The R programming language, with a developing environment such as RStudio.

Suggested Readings for R The following three books are available for free online or through the library and are excellent introductions to R in increasing order of difficulty.

- Matloff, Norman. 2011. *The Art of R Programming: A Tour of Statistical Programming*. No Starch Press.
- Grolemund and Wickham. *R for data science*. (available online)
- Wickham, Hadley. *Advanced R* (available free online)

3 Assignments

There are four types of assignments in this course:

1. **Preparing for class and section**: Often for your classes and sections there will be some reading you must do before class. We expect you to come to both class and sections 100% prepared; we do not assign much reading, but we assume you have read

it. Your participation in class, section and on discussions on Canvas as both a learner and a support to your colleagues is a part of your assessment for the semester and good practice towards creating a positive learning environment!

2. **Weekly problem sets**: learning statistics and programming takes consistent practice. The problem sets are described below.

3. **Midterm exams**: there will be two non-cumulative midterm exams during the semester.

4. **Final poster project**: you will work in small groups to complete a final poster project presentation at the close of the semester amongst your colleagues as well as the your colleagues from Alex Tahk's PS 818 MLE course.

3.1 Preparing for Class and Section/Participation

There are readings for each of the modules and topics within them through the semester; the first part of the semester will heavily feature mathematical notation. We recognize that becoming familiar and comfortable with this type of language can be challenging the first time you do it – and so it will be tempting to skip math that seems particularly daunting– don't do this! The math is often the meat of the statistical work, and part of your learning goals for this course is to become more accustomed and comfortable with reading and understanding mathematical notation. Read carefully and go line by line to make sure you understand. Read the required readings and any others that might evoke your interests as we progress through the semester. Engage with the readings: take notes, write down your questions, impressions and confusions, talk with your classmates and TA, and post questions/answers on Canvas. Actively assisting your classmates in class, section or Canvas will constitute 5% of your final grade.

3.2 Problem sets

Learning by watching and not doing is hard when learning how to use any new tool, so in order to do more and practice with regularity you will have homework on a weekly basis. The assignments will be a mix of analytic problems, computer simulations and data analysis.

Assignments must be completed in R Markdown, which allows you to show both your answers and the code you used to arrive at them. You will need to submit both the .rmd file and the knitted html/pdf. Don't worry if you don't know R Markdown, we will show you how it works in section with more detailed instructions before the first assignment is due.

Each week's homework will be made available on Canvas starting Wednesday immediately after class and is due Wednesday the following week (7 days later) immediately **before** class. Solutions will be made available directly after class through Canvas. Working through the problem sets include looking at the solutions key so please remember to do this portion!

Problem sets are graded out of 0-5 points. We reserve the right to add bonus points for aesthetics including presentable graphs, clear code, nice formatting and well-written answers.

There will be 10 problem sets in total, constituting 50% of your grade. Your problem set with the lowest grade will be dropped when calculating your final grade. Late problem sets drop a grade by 1 point (out of the total 5) each late day, with a maximum of 3 late days, after which we will not accept the problem set anymore. We do not want to hold up the class and will not wait for everyone to submit their problem sets in order to post the solutions key. If you are turning in your problem set late, you are on your honor to *not look at the solutions key* before submitting your work.

Code Conventions: Throughout the course, students will receive feedback on their code from the professor, the TA, and other students. Therefore, consistent code conventions are critical. Good coding style is an important way to increase the readability of your code (even by a future you!). We strongly recommend you follow the code conventions developed by Hadley Wickham and implemented in the package lintr, which is built into R Studio.

Collaboration Policy: Unless otherwise stated, we encourage students to work together on the assignments, but you should write your own solutions (this includes code). That is, no copy-and-paste from other people's code. You would not copy-and-paste from someone's paper, and you should treat code the same way. However, we strongly suggest that you make a solo effort at all the problems before consulting others.

3.3 Midterm exams

There will be two midterm exams. The first will be in-class, closed-book. The second midterm exam will be take home – posted immediately after class and then due two days after by midnight. It is "open-book" in the sense that you can use the slides, your notes, books, and internet resources to answer the questions. However, the exam must be completed *by yourself*. It will be the approximate length of a problem set (although we caution that it might take longer if you are used to collaborating on the problem sets). We encourage you to start early.

3.4 Final poster project

You will work in groups to ask a social science question of a dataset that will culminate in a final poster project which will be presented at the end of the semester. We will provide more details on the project later in the semester.

Grading

- Participation: 5%
- Problem sets: 10 total, 50%
- Midterm exams: 2 total, each 15%
- Final poster project: 15%

4 How to Learn in this Course

If you find this course challenging, you are not alone. Statistics can be challenging and we cover a lot of ground. I have confidence in your abilities as smart and engaged researchers who can handle it. Below are some details on forms of support that we offer in this class.

Your primarily responsibilities in this class are to *work hard* and *communicate* with us about what you need. You cannot learn if you aren't putting in the time. We also can't help if we don't know there's a problem.

4.1 Resources for Getting Help

Below are a few main sources of support for this class.

1. Class and Section

   We encourage you to be an active participant in class and section. Ask questions if you don't understand something that is happening.

2. Readings and Slides

   If you are studying alone and hit something you don't understand, you should turn to the readings and slides. There will be a fair amount of material in the slides and they are intended to be used and reviewed multiple times, not just seen once during lecture.

3. Canvas

We will be using the Canvas discussion board for communication in this class, but also as a source to post and answer questions about the material. You will not be required to post, but the system is designed to get you help quickly and efficiently from classmates, the TA and the professor. **Unless the question is of a personal nature or completely specific to you, you should not email teaching staff**; instead, you should post your questions on Canvas. We will be monitoring the discussion board, but we encourage you to help your classmates as well. Likely a significant amount of overlap will exist for both things people want to know more about and things people have just figured out.

4. TA office hours

   Dillon is merging office hours with the MLE class and offering four office hours each week. TA office hours are often useful for getting help with new tricky material and problem sets.

5. Instructor office hours

   My office hours are in my office immediately after each class session. If these times are not convenient, I will also be hosting office hours for my undergraduate course 1:00 PM - 2:50 PM, Mondays and Wednesdays as well. If you swing by during the latter hours I will prioritize undergrad questions, but will be happy to answer your questions after.

6. Problem Set Key

   As soon as the problem sets are due, we will post the key. It may be tempting to immediately turn focus towards the next problem set, but if you were uncertain about anything in the problem set, I recommend you check the key to lock down core concepts. Some of the material builds directly on previous concepts!

Class Schedule

Note, this WILL change as we roll through the semester, though no exam dates will change. Please check Canvas regularly for updates.

I. PROBABILITY

Wednesday, September 4 to Monday, September 9

Topic: Probability
Reading: DS 1.5-10

Wednesday, September 11 to Monday, September 16

Topic: Conditional probability
Reading: DS 2.1-3

Wednesday, September 18 to Monday, September 23

Topic: Random variables (discrete, continuous)
Reading: DS 3.1-3.3, 3.8, 5.2, 5.6

Wednesday, September 25, Monday, September 30, & Wednesday, October 2

Topic: Expectation, Variance, Covariance and Correlation, Conditional Expectation
Reading: DS 4.1-4.3,4.6-4.7

Monday, October 7, Wednesday October 9, & Monday, October 12

Topic: Normal distributions, Law of Large Numbers, Central Limit Theorem
Reading: DS 5.6, 6.1-6.3

Wednesday, October 14

Midterm 1: in class, closed book.

II. STATISTICAL INFERENCE

Monday, October 21 to Wednesday, October 23

Topic: Estimation (intro to statistical inference, MLE)
Reading: DS 7.1, 7.2, 7.4-7.9

Monday, October 28 to Wednesday, October 30

Topic: Finish estimation, confidence intervals
Reading: DS 8.5

Monday, November 4 to Wednesday, November 6

Topic: Hypothesis testing
Reading: DS 9.1-9.6

III. Regression

Monday, November 11 to Wednesday, November 13

Topic: Linear regression
Reading: DS 11.1-4

Monday, November 18

Topic: Multiple linear regression
Reading: DS 11.5

Wednesday, November 20

Topic: Simulations, Bootstrap
Reading: DS 12.1-3, 12.6 Take home Midterm 2 (on materials in Statistical Inference and Regression, only)

IV. Causal inference

Monday, November 25

Topic: Causal framework and the experimental ideal
Reading: AP 1-2

Wednesday, November 27

Topic: Regression with observational data in the causal framework
Reading: AP 3

Monday, December 2 to Wednesday, December 4

Topic: Instruments, Regression discontinuity
Reading: AP 4, 6

Monday, December 9 to Wednesday, December 11

Topic: Fixed effects, diff-in-diff
Reading: AP 5, 6

Either afternoon Wednesday December 11 or December 12 poster presentation: TBD

Problem set & exam schedule

1. Problem Set 1 (Probability): given Mon 9/9, due Mon 9/16
2. Problem Set 2 (Conditional probability): given Mon 9/16, due Mon 9/23
3. Problem Set 3 (Random variables): given Mon 9/23, due Mon 9/30
4. Problem Set 4 (Expectation): given Wed 10/2, due Wed 10/9

5. NO problem set for (Normal distributions, LLN, CLT) – practice problems provided to prepare for in-class midterm Wed 10/14
6. Wed 10/14: in class midterm 1
7. Problem Set 5 (Estimation/MLE): given Wed 10/23, due Wed 10/30
8. Problem Set 6 (Confidence intervals/Hypothesis testing): given Wed 11/6, due Wed 11/13
9. Problem Set 7 (Linear regression): given Wed 11/13, due Wed 11/20
10. Wed 11/20: take-home midterm 2 (on materials in Statistical Inference and Regression, only) due Friday November 22 end of night, midnight.
11. Problem Set 8 (Multiple linear regression, simulations): given Fri 11/22, due Fri 11/29 (to account for take-home midterm, this set will be shorter)
12. Problem Set 9 (Causal framework and experiments): given Mon 11/25, due Mon 12/2
13. Problem Set 10 (IV, RD): given Mon 12/2, due Mon 12/9
14. Poster session Wed or Thurs 12/11 TBD

Acknowledgements

This course was developed on the shoulders of giants, in some cases borrowing directly from materials developed by the amazing methods community in political science, economics, statistics and sociology. I am extremely grateful to everyone who has contributed directly, or indirectly. Lecture slides and related circulated materials should have appropriate citations – please send me an email if you believe they are incorrectly citing or lacking in citation rigour. Individuals include but are not limited to: Matt Blackwell, Dalton Conley, Adam Glynn, Justin Grimmer, Jens Hainmueller, Erin Hartman, Chad Hazlett, Kosuke Imai, Gary King, Dean Knox, Kevin Quinn, Matt Salganik, Brandon Stewart, and Teppei Yamamoto.

All errors that remain are my own.

ACADEMIC INTEGRITY

By enrolling in this course, each student assumes the responsibilities of an active participant in UW-Madison's community of scholars in which everyone's academic work and behavior are held to the highest academic integrity standards. Academic misconduct compromises

the integrity of the university. Cheating, fabrication, plagiarism, unauthorized collaboration, and helping others commit these acts are examples of academic misconduct, which can result in disciplinary action. This includes but is not limited to failure on the assignment/course, disciplinary probation, or suspension. Substantial or repeated cases of misconduct will be forwarded to the Office of Student Conduct & Community Standards for additional review. For more information, refer to `https://conduct.students.wisc.edu/academic-integrity/`.

ACCOMMODATIONS FOR STUDENTS WITH DISABILITIES

The University of Wisconsin-Madison supports the right of all enrolled students to a full and equal educational opportunity. The Americans with Disabilities Act (ADA), Wisconsin State Statute (36.12), and UW-Madison policy (Faculty Document 1071) require that students with disabilities be reasonably accommodated in instruction and campus life. Reasonable accommodations for students with disabilities is a shared faculty and student responsibility. Students are expected to inform faculty [me] of their need for instructional accommodations by the end of the third week of the semester, or as soon as possible after a disability has been incurred or recognized. Faculty [I], will work either directly with the student [you] or in coordination with the McBurney Center to identify and provide reasonable instructional accommodations. Disability information, including instructional accommodations as part of a student's educational record, is confidential and protected under FERPA. `http://mcburney.wisc.edu/facstaffother/faculty/syllabus.php`.

DIVERSITY & INCLUSION

Diversity is a source of strength, creativity, and innovation for UW-Madison. We value the contributions of each person and respect the profound ways their identity, culture, background, experience, status, abilities, and opinion enrich the university community. We commit ourselves to the pursuit of excellence in teaching, research, outreach, and diversity as inextricably linked goals. The University of Wisconsin-Madison fulfills its public mission by creating a welcoming and inclusive community for people from every background – people who as students, faculty, and staff serve Wisconsin and the world. `https://diversity.wisc.edu/`