PS 812 Intro to Statistical Methods

Fall 2020 Syllabus *

Instructor:	Adeline Lo	TA:	Dillon Laaker
Office:	Zoom	Office:	Zoom
Email:	aylo@wisc.edu	Email:	laaker@wisc.edu
Contact method:	Slack	Contact method:	Slack
Class:	08:00 - 09:15 AM, TTh SCIENCE 180/Zoom	Section:	9:55AM - 11:35AM, F Zoom
OH:	09:30 - 10:30 AM TTh	OH:	3:30-5:30 PM, Th

1 Course Details

Course subject, Number and Title Political Science 812 Introduction to Statistical Methods

Credits 3 credits.

Instructional modality face-to-face (with virtual access)

Meeting university credit hours This class meets for two, 75-minute class periods each week over the fall semester and carries the expectation that students will work on course learning activities (reading, writing, problem sets, studying, etc) for about 3 hours out of the classroom for every class period. The syllabus includes more information about meeting times and expectations for student work.

*Last edited: September 1, 2020

2 Course 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.

2.1 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 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 use existing 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. **Beyond the methods sequence, we encourage you to participate in the broader statistical life at UW Madison, including the Models and Data Workshop, `https://sites.google.com/view/madwisc`.**

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.

2.2 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.

2.3 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.

3 Materials

3.1 Virtual tools

Given the exceptional times in which we're operating this course, more interaction will inevitably be taking place online. Our main two methods of communicating information and interacting throughout will be via the `Canvas` course website and our class `Slack` workspace. The course website will contain all course related materials and will be the main forum from which assessments will be handed out and submitted. All other correspondence within and between class will be conducted on `Slack`; this includes posing questions to the class and instructing team both in and out of class/sections as well as among yourselves as you work through the materials. We are using `Slack` for a few important reasons: a) it is a central and organized platform with which to quickly communicate broadly (@ an entire channel) and in specific subgroups (@ a set of individuals), b) it keeps track of topics and can allow for searchable discussion, c) it allows for ease in sharing of files and `code` snippets, the latter of which will likely be very useful throughout the semester.

1. `Canvas`: `https://canvas.wisc.edu/courses/##`
2. `Slack`: workspace titled "Fall 2020 PS 812" (reach out to Adeline/Dillon if you have not yet accessed this)

3.2 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.

3.3 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)

4 Weekly Sections with Dillon

Section for this course is designed to teach you how to implement the statistical tests/tools learned in lecture and to further develop the intuition behind these concepts. It will be common to leave lecture understanding the theory, but not knowing how to fully implement it in R. Section exists to bridge the gap between the theory and the application of statistics.

Section will be held every Friday between 9:55am and 11:35am. We will use a combination of Zoom and Slack to communicate and share code. Each section will typically begin with Q/A and/or reviewing material that the class is struggling to comprehend. The remainder of section will be dedicated to mini-problem sets completed in small groups. These exercises are essential to develop your coding skills and to further understand the intuition behind concepts taught in lecture. They also will provide guidance for future problem sets.

The material covered for the week will be made available on Canvas the day before section (Thursday). Similar to problem sets, all assignments in section must be completed in R Markdown. You will submit both the .rmd file and the knitted html/pdf to canvas at the end of each section. While these are not officially graded, we do look over them and they are part of your participation grade for the course.

5 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 `Slack` 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 exam**: there will be one take-home midterm exam during the semester.

4. **Final replication project**: you will work in small groups to complete a final replication presentation at the close of the semester in pairs or threes. More details will be provided on this in the middle of the semester.

5.1 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 Slack. Actively assisting your classmates in class, section or Slack will constitute 5% of your final grade.

5.2 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. **You should have already successfully downloaded `R`/RStudio and knitted a practice html/pdf with your instructors from math bootcamp. If you haven't, please do so ASAP and check in with Adeline or Dillon if you cannot reliably knit out a file before the end of the first week of classes.**

Each week's homework will be made available on Canvas immediately after a class and is due in exactly one 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 11 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 RStudio.

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.

5.3 Midterm exam

There will be one take home midterm exam. You will receive 48 hours to complete it. 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.

5.4 Final replication project

You will work in groups of two or three to work on a replication project and present it 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, 60%
- Midterm exam: 20%
- Final replication project: 15%

6 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.

6.1 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; you can do this either by raising a hand (in-person or virtually) or posting a question on the Slack #lectures channel.

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. Slack

   We will be using Slack 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. **If a question is of a personal nature or completely specific to you, you should use a direct channel with the teaching staff on Slack; otherwise instead, you should post your questions on a general Slack channel**. We will be monitoring the discussion channels, 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's office hours are on Thursdays from 3:30 to 5:30pm and will be held virtually via Zoom. If these times are not convenient, please do not hesitate to set-up an appointment. TA office hours are often useful for getting help with new tricky material and problem sets.

5. Instructor office hours

   Adeline's office hours are on Zoom immediately after each class session. If these times are not convenient, Adeline will also be hosting office hours for my other graduate course, on Thursday afternoons. If you swing by during the latter hours she will prioritize the other course's students 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/Slack regularly for updates.

I. PROBABILITY

Thurs 9/3, Tues 9/8

Topic: Probability
Reading: DS 1.5-10

Tues 9/8, Thurs 9/10

Topic: Conditional probability
Reading: DS 2.1-3
Assignment: Problem Set 1 (Probability) given T 9/8, due T 9/15

Tues 9/15, Thurs 9/17

Topic: Random variables (discrete, continuous)
Reading: DS 3.1-3.3, 3.8, 5.2, 5.6
Assignment: Problem Set 2 (Conditional probability) given T 9/15, due T 9/22

Tues 9/22, Thurs 9/24, & Tues 9/29

Topic: Expectation, Variance, Covariance and Correlation, Conditional Expectation
Reading: DS 4.1-4.3,4.6-4.7
Assignment: Problem Set 3 (Random variables) given T 9/22, due T 9/29

Thurs 10/1, Tues 10/6

Topic: Normal distributions, Law of Large Numbers
Reading: DS 4.4, 5.6, 6.1-6.3
Assignment: Problem Set 4 (Expectation) given Th 10/1, due Th 10/8

Thurs 10/8, Tues 10/13

Topic: Finish Law of Large Numbers, Central Limit Theorem
Reading: DS 4.4, 5.6, 6.1-6.3
Assignment: Problem Set 5 (LLN/CLT) given Th 10/8, due Th 10/15

Below updated 10/22/2020:

II. STATISTICAL INFERENCE

Thurs 10/15

Topic: Estimation (intro to statistical inference, MLE)
Reading: DS 7.1, 7.2, 7.4-7.9
Assignment: Problem Set 6 (Estimation/MLE) given T 10/15, due T 10/22

Tues 10/20, Thurs 10/22

Topic: Confidence intervals, Start hypothesis testing
Reading: DS 8.5
Assignment: Problem Set 7 (Confidence intervals) given Th 10/22, due Th 10/29

Tues 10/27

Topic: Hypothesis testing
Reading: DS 9.1-9.6
Assignment: Problem Set 8 (Hypothesis testing) given Th 10/29, due Th 11/5

III. REGRESSION

Thurs 10/29, Tues 11/3

Topic: Linear regression
Reading: DS 11.1-4
Assignment: Problem Set 9 (Regression) given **TUES 11/3, due TUES 11/10**

Thurs 11/5, Tues 11/10, Thurs 11/12

Topic: Multiple linear regression
Reading: DS 11.5

IV. CAUSAL INFERENCE

Tues 11/17

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

MIDTERM EXAM

Thurs 11/19

NO CLASS. MIDTERM IN PROGRESS.

Thurs 11/19 – Sat 11/21

Take home midterm examination. Handed out Thurs 11/19 at 9 am Central. Submission deadline Saturday 11/21 9 am Central.

IV. CAUSAL INFERENCE (CONT'D)

Tues 11/24

Topic: Regression with observational data in the causal framework
Reading: AP 3
Assignment: Problem Set 10 (Causal framework) given T 11/24, due T 12/1

Thurs 11/26

Thanksgiving break; no classes.

Tues 12/1, Thurs 12/3

Topic: Instruments, Regression discontinuity
Reading: AP 4, 6
Assignment: Problem Set 11 (IV, RD) given Th 12/1, due Th 12/8

Tues 12/8

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

LAST CLASS

Thurs 12/10

Topic: Last day of class, presentations.
Assignment: Group presentations and submission of `.rmd` and knitted files.

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.

Students' Rules, Rights & Responsibilities

During the global COVID-19 pandemic, we must prioritize our collective health and safety to keep ourselves, our campus, and our community safe. As a university community, we must work together to prevent the spread of the virus and to promote the collective health and welfare of our campus and surrounding community. See here for details; UW Madison Badger Pledge.

UW-MADISON Face Covering Guidelines

While on campus all employees and students are required to wear appropriate and properly fitting face coverings while present in any campus building unless working alone in a laboratory or office space. Guidelines can be found here.

Face Coverings During In-person Instruction Statement (COVID-19) Individuals are expected to wear a face covering while inside any university building. Face coverings must be worn correctly (i.e., covering both your mouth and nose) in the building if you are attending class in person. If any student is unable to wear a face-covering, an accommodation may be provided due to disability, medical condition, or other legitimate reason.

Students with disabilities or medical conditions who are unable to wear a face covering should contact the McBurney Disability Resource Center or their Access Consultant if they are already affiliated. Students requesting an accommodation unrelated to disability or medical condition, should contact the Dean of Students Office.

Students who choose not to wear a face covering may not attend in-person classes, unless they are approved for an accommodation or exemption. All other students not wearing a face covering will be asked to put one on or leave the classroom. Students who refuse to wear face coverings appropriately or adhere to other stated requirements will be reported to the Office of Student Conduct and Community Standards and will not be allowed to return to the classroom until they agree to comply with the face covering policy. An instructor may cancel or suspend a course in-person meeting if a person is in the classroom without an approved face covering in position over their nose and mouth and refuses to immediately comply.

Quarantine or isolation due to COVID-19

Student should continually monitor themselves for COVID-19 symptoms and get tested for the virus if they have symptoms or have been in close contact with someone with COVID-19. Student should reach out to instructors as soon as possible if they become ill or need to isolate or quarantine, in order to make alternate plans for how to proceed with the course. Students are strongly encouraged to communicate with their instructor concerning their illness and the anticipated extent of their absence from the course (either in-person or remote). The instructor will work with the student to provide alternative ways to complete the course work.

Course evaluations

Students will be provided with an opportunity to evaluate this course and your learning experience. I will be informally soliciting feedback throughout the semester as we roll through, though we will also have a formal evaluation at the end of the semester. Your feedback is particularly important given the context of delivering the course under COVID constraints and I take these seriously in updating the course and its materials.

Digital Course Evaluation (AEFIS) UW-Madison now uses an online course evaluation survey tool, AEFIS. In most instances, you will receive an official email two weeks prior to the end of the semester when your course evaluation is available. You will receive a link to log into the course evaluation with your NetID where you can complete the evaluation and submit it, anonymously. Your participation is an integral component of this course, and your feedback is important to me. I strongly encourage you to participate in the course evaluation.

Usage of Audio Recorded Lectures Statement

Lecture materials and recordings for PS 812 are protected intellectual property at UW-Madison. Students in this course may use the materials and recordings for their personal use related to participation in this class. Students may also take notes solely for their personal use. If a lecture is not already recorded, you are not authorized to record my lectures without my permission unless you are considered by the university to be a qualified student with a disability requiring accommodation. [Regent Policy Document 4-1] Students may not copy or have lecture materials and recordings outside of class, including posting on internet sites or selling to commercial entities. Students are also prohibited from providing or selling their personal notes to anyone else or being paid for taking notes by any person or commercial firm without the instructor's express written permission. Unauthorized use of these copyrighted lecture materials and recordings constitutes copyright infringement and may be addressed under the university's policies, UWS Chapters 14 and 17, governing student academic and non-academic misconduct.

Academic Calendar & Religious Observances

See here.

Academic Integrity

By virtue of enrollment, each student agrees to uphold the high academic standards of the University of Wisconsin-Madison; academic misconduct is behavior that negatively impacts the integrity of the institution. Cheating, fabrication, plagiarism, unauthorized collaboration, and helping others commit these previously listed acts are examples of misconduct which may result in disciplinary action. Examples of disciplinary action include, but is not limited to, failure on the assignment/course, written reprimand, disciplinary probation, suspension, or expulsion. 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.