PSCI 200
Data Analysis I

Fall 2025

Tu/Th 12:30-1:45, Dewey 2-162

Prof Curt Signorino
curt.signorino@rochester.edu
Office hours: Tues 2-4pm, Harkness 303

TA's	**Adam Roberts***	**Shiyou Li**	**Ryan Lin**	**Alex Trauben**
	arober48	sli121	rlin30	atrauben

Office Hours:

Mon	1-2pm	Alex	Harkness 335
Wed	11am-12	Adam	H335
Thurs	10am-11	Shiyou	H335
Fri	2-3pm	Ryan	H335

Workshops:

Mon	5:30-7pm	Ryan	TBD
Tues	4-5:30pm	Alex	TBD
Wed	3:15-4:45pm	Adam	TBD
Thurs	2-3:30pm	Shiyou	TBD

Course Description

How do we empirically evaluate the claims politicians make? How do we determine whether theories of political behavior are supported by evidence? What do pollsters mean when they refer to a poll having a 3% margin of error? In this course, students are introduced to data analysis and statistical inference relevant to political science research in particular and to the social sciences more broadly. Topics will include descriptive statistics, surveys, experiments, probability, confidence intervals, hypothesis tests, correlation, and regression analysis. Data analysis will be conducted using R and RStudio. Students should bring a laptop to class with R and RStudio installed.

Course Meeting and Credits. This course follows the College credit hour policy for four-credit courses. We will meet in person twice a week: Tues & Thurs, 12:30-1:45pm. Attendance for the Tues/Thurs classes is required. Classes will be a mix of lecture and practice sessions. Practice sessions will consist of short (∼8–15min) sessions where students have a chance to practice or implement the techniques just presented in lecture.

Each week, the TA's will also hold workshops. Attending the workshops is not required but is *highly* recommended. During the workshops, students will have the opportunity to work on problems very similar to those on the current or upcoming homework, but with the

assistance of the TA's. The workshop schedule is shown on the first page of this syllabus. The remaining credit hour is fulfilled through independent reading and completion of the homeworks.

Prerequisites: PSCI 200 is intended for students with little to no experience in statistics or for those who want to improve their understanding of introductory material before proceeding to more advanced courses. Calculus and matrix algebra are not required and will not be used during the semester. Students who have taken another UR course in statistics, such as ECON 230 or STAT 180/190/212/213/214, are generally discouraged from taking PSCI 200. Students who have taken the AP exam in Statistics are encouraged to take PSCI 200, especially if R was not used in the AP course. If you have taken a similar "intro to statistics" course and would like to take PSCI 200, please consult with me first.

Readings

Students are responsible for keeping up with the reading each week. Whenever possible, I will post to Blackboard pdf's of any readings or lecture notes.

Required:

- (QSS) Kosuke Imai, *Quantitative Social Science: An Introduction*. Princeton. Available in paperback for a reasonable price.

  Do NOT purchase the Imai & Williams book with "An introduction in tidyverse" in the title.

- (IMS) Mine Çetinkaya-Rundel and Johanna Hardin, *Introduction to Modern Statistics*. 2nd ed. You can download this using the link. Note that it is donation-ware. Simply enter a purchase price of $0 to download it for free. Alternatively, you can find it in the "Texts" folder on Blackboard.

- (OIS) David M. Diez, Mine Cetinkaya-Rundel, and Christopher D. Barr. OpenIntro Statistics. 4th ed.

- John Verzani, *SimpleR: Using R for Introductory Statistics*.

Note that all but the Imai text are available as free downloads and can be found in the "Texts" folder on Blackboard.

Laptops, R, and RStudio

Students are required to complete most homeworks using R and RStudio. R is the main statistics program. RStudio is a user-friendly interface with many other features. Both are free. If you have never installed either or if your current installation is over six months old, you should install the most recent version of each on your laptop.

Step-by-step installation instructions for Mac OSX and MS Win are available here and on Blackboard. Even if you know how to install these programs, please work through the instructions and verification. A fresh install using the provided instructions will help avoid most of the problems we regularly see at the start of the semester.

HW submissions will need to be compiled (or knit) either to html or to pdf. If you would like to compile your HW results to html output, you need do nothing more than install R and RStudio. If you would like to compile directly to pdf, you will need to install LaTeX. Using LaTeX is purely optional. LaTeX is a text formatting language. You will not have to learn LaTeX for this course. However, RStudio can use LaTeX to turn your HW R code into a nicely formatted pdf. Instructions for installing the tinytex version of LaTeX are included in the R/RStudio installation instructions.

Although we strongly recommend against it, it is possible to complete this course if you have only a tablet with a *physical* keyboard. A cloud (i.e., web browser) version of RStudio is available at Posit Cloud. That said, it has a slightly different interface than the version that will be taught in class; and the professor and TA's cannot provide the same support for that version that they can for the laptop versions of R/RStudio. Students are strongly encouraged to use a laptop. If you do not own a laptop, you may be able to borrow one for the semester from University IT.

Course Outline

Course topics do not have dates assigned. Students are responsible for keeping up with the readings, lectures, and HW's throughout the semester. As a guidepost, at the end of a topic, I will normally indicate in lecture the subject of the next topic.

1 Course Introduction

Course organization. This syllabus. Why you should care about statistics.

2 Introduction to R, pt 1

R as a calculator, variables, assignment, vectors, indices, R scripts, knitting

Imai QSS 1.3
Verzani 1-2 datadatabobata 2

2.5 Introduction to R, pt 2

Working directory, loading data frames, dim, head, help

HW 1 datadatabobata 2

3 Look at your data: one variable at a time

Types of variables, measures of centrality, variance, descriptive statistics, hist, NA's

Imai QSS 3.2–3.3
OIS Ch 2
Verzani 3 datadatabobata 3

4 Look at your data: two variables at a time

Logical operations, subset, diff of means, crosstabs, scatterplots, covariance, correlation

Imai QSS 2.2, 3.6

HW 2 datadatabobata 4

Where does the data come from?

5 Operationalization

Operationalization, observational vs experimental data, level of observation / analysis

6 Sampling

Surveys, random sampling, convenience sampling, selection bias, response bias

Imai QSS 3.4.
IMS 2.1

Experiments and Research Design

7 Causal effects and experiments

Causality, potential outcomes, simple randomized experiments

Imai QSS 2.1, 2.3–2.7
IMS 2.2–2.4

8 Research design and threats to validity

Internal validity, external validity, confounders

James McDavid & Laura Hawthorn. 2005. *Research Designs for Program Evaluations.* Chapter 3. (Sage Proof)(Google Books)
Research Methods Knowledge Base, Section on Design

HW 3

Estimating a Linear Relationship

9 Fitting a line

Least squares, bivariate regression, residuals, lm(), r^2, experimental vs observational data

Imai QSS 4.1–4.2, 4.3.1
IMS 7
OIS, Ch 8.1–8.3

10 Estimating a plane: multiple regression

Partial regression coefficient, dummy variables, R^2

Imai QSS 4.3.2–4.3.3
IMS 8
OIS, Ch 9.1–9.3

Midterm Exam Review

Workshop

Midterm Exam

In class. Exam will cover topics 1–10

Probability: The Mathematics of Uncertainty

11 Discrete distributions

Random variable, PMF, CMF, complement, axioms, Bernoulli, Binomial

Imai QSS 6.1, 6.3.1–6.3.3
OIS, Ch 3.1, 3.3, 3.4, 4.3 datadatabobata 11

12 Expected value and variance

Centrality and spread of a random variable

Imai QSS 6.3.5.
OIS, 3.4.1–3.4.2

HW 5 datadatabobata 12

13 Continuous distributions

Uniform, Normal, PDF, CDF, complement, axioms

Imai QSS 6.3.4
OIS 3.5, 4.1 datadatabobata 13

14 Central limit theorem

The sample mean is a random variable, Law of Large Numbers, CLT

Imai QSS 6.4, 7.1.1
OIS 5.1

HW 6 datadatabobata 14

. . . . . .

Inference: Uncertainty and our Conclusions about a Single Variable

15 Confidence intervals

Large sample mean and proportion, margin of error, small sample mean, t distribution

Imai QSS 7.1.2–7.1.6
OIS, Ch 5.2, 6.1.2

HW 7 datadatabobata 15

16 Hypothesis tests

Large samples: proportion & mean, small sample: mean, type I/II error, p-values

Imai QSS 7.2.1–7.2.3
OIS, Ch 5.3, 6.1.3, 7.1. datadatabobata 16 datadatabobata 17

. . . . . .

Inference: Uncertainty about whether Two Variables are Related

17 Comparing two groups

Difference of means, difference of proportions, significance

Imai QSS 7.2.4–7.2.6
OIS, Ch 6.2, 7.3

HW 8 datadatabobata 18

18 Joint and conditional distributions

Joint distributions, marginal probability, independence, conditional probability

Imai QSS 6.2
OIS 3.2 datadatabobata 19

19 Crosstabs and the test of independence

Crosstabs, independence, expected and observed frequencies, Chi-square test

OIS 6.3-6.4

HW 9

20 Linear regression: standard errors

Assumptions of the classical linear model, $\hat{\beta}$ is a random variable

Imai QSS 7.3.1–7.3.3
OIS, Ch 8.4
Verzani 13

21 Linear regression: hypothesis tests

Interpreting a regression table, significance of the regression

Imai QSS 7.3.4–7.3.5
Verzani 14

HW 10

Final Exam Review

Workshop

Final Exam

Sunday 12/14, 7:15pm-10:15pm, Dewey 2-162

Course Objectives

In this course, you will develop

- An understanding of descriptive statistics and how to apply them to data.
- An understanding of sampling and why it's so important.
- An understanding of randomized experiments, research design, and threats to validity.
- An understanding of regression estimation.
- A basic understanding of probability and how it applies to statistical inference.

- An understanding of sampling distributions and the central limit theorem.
- An understanding of how to evaluate hypotheses using tools of statistical inference.

Course Learning Outcomes

By the end of this course, you will be able to

- Calculate the sample mean and variance for a variable in a dataset.
- Calculate the correlation between two interval variables.
- Estimate bivariate and multiple regressions.
- Calculate the expected value and variance of a discrete random variable.
- Calculate probabilities for continuous random variables.
- Calculate a 95% confidence interval for a mean or a proportion.
- Conduct an hypothesis test concerning a mean or a proportion.
- Conduct an hypothesis test concerning the difference of two means or two proportions.

Grading

Course grades will be based on a series of homeworks (53%), a midterm exam (15%), a final exam (25%), class attendance (5%), and participation through in-class practice sessions (2%).

Homeworks. Typically, homeworks will be handed out via Blackboard at the end of lecture and due by the start of lecture one week later. Students should submit their homework answers, properly formatted, via Blackboard. Homework grades will also be posted on Blackboard. All assignments are to be completed individually. Be sure to read the PSCI 200 course academic honesty policy concerning HW completion.

Late homework submissions. It is important that students submit their HW's on time. We do our best to provide grades and answer keys in a timely manner. Late HW submissions can hold up that process, in which case the class will not have as much time to review previous HW answers before starting a new HW.

That said, life happens. If you need to attend a major event – e.g., a conference, a job interview, an athletic tournament for UR, etc – email me ahead of time. As long as this is

a one-time occurrence, you will likely be given permission to turn in the current HW after the deadline. Similarly, if you fall ill, email us immediately and we'll try to work something out. In either case, you should expect that the deadline extension will be no more than 7 days after the original due date, usually less.

In all other cases, late submissions will be penalized (as a % of total points) as follows:

Lateness	**Penalty**
Up to 10 hrs	5%
10 hrs to 24 hrs ($\leq$ 1 day)	10%
$>$ 1 day but $\leq$ 2 days	20%
$>$ 2 days but $\leq$ 3 days	30%
$\vdots$	
$>$ 6 days but $\leq$ 7 days	70%
$>$ 7 days	100%

Again, it is important that you contact us as soon as possible concerning a late HW submission. If you delay and the HW answer key is posted before your submission, you will receive a zero for that HW assignment.

Class attendance. Attendance will be taken each class using Qwickly (on Blackboard). A student's class attendance score will be calculated by dropping up to five (5) absences and then calculating the percent of times the student attended class. Examples: Suppose a student misses 3 out of 25 lectures. The class attendance grade will be 100%. If a student misses 7 out of 25 lectures, then their class attendance grade would be (18/20)*100 = 90%.

Because students automatically receive five free absences from lecture, no allowances will be made for students who fail to enter the Qwickly code during class. It will simply count as one of the five absences. Similarly, if a student misses a class due to illness or needs to attend a religious observance, conference, sporting event, etc, there is no need to report those to the TA's or to me. The absence will simply count as one of the five free absences. Please do not ask the TA's to adjust your absence for a particular day. They are not authorized to do so.

Practice sessions. In every lecture, there will be opportunities for hands-on practice of that lecture's concepts. At least once per week, an in-class practice session will be "for credit." For those practice sessions, students will earn 0 points if they do not submit an answer, 1 point if they submit an answer but it is incorrect, and 2 points if they submit a correct answer. The course practice session score will be calculated as the average of the 10 best practice session submissions. For full credit, a student need only submit an answer (any answer) for 10 "for credit" practice sessions. For example, if a student submits 10 practice session answers that are all incorrect, their practice session score will be 100%. If a student only has five submissions, but they're all correct, their practice session score will be 100%.

Notice that students can earn extra credit through practice sessions. If a student submits at least 10 answers that are correct, they'll receive the maximum practice session score of 200%.

Midterm and Final Exams. The midterm exam will be held in class — usually after topics 9 or 10. The final exam will be a 3-hour exam, held in our normal classroom during finals week. The final exam is scheduled by the University. Unless granted an accommodation through the Office of Disability Resources, all students are expected to take the same final exam, at the same time, under the same conditions. Exceptions, including taking the final exam via zoom, cannot be granted based on students' end-of-semester return travel plans. Students are expected to schedule their return travel so that it does not conflict with their course final exams.

Minimal Requirements for Passing Grade: In order to receive a passing grade for the course, students must take the midterm exam, the final exam, and complete at least five of the homeworks. While that does not guarantee a passing grade, students who do not satisfy these minimal requirements will receive a grade of E, regardless of their weighted average score.

Course score grading examples

Course score = .53 (HW) + .15 (midterm) + .25 (final) + .05 (attend) + .02 (practice)

1. Student A scores 90% for the HW score, 92% on the midterm exam, 89% on the final exam, attends all but 5 (out of 25) lectures, and enters 12 practice sessions answers, with three of those being correct answers.

Class attendance score = 100%
Practice session score = 100% + 30% = 130%
Course score = .53 (90) + .15 (92) + .25 (89) + .05 (100) + .02 (130) = 91.35%

2. Student B scores 93% for the HW score, 61% on the midterm exam, 52% on the final exam, misses 10 (out of 25) lectures, and enters only four practice session answers, with one correct answer.

Class attendance score = (15/20)*100 = 75%
Practice session score = 40% + 10% = 50%
Course score = .53 (93) + .15 (61) + .25 (52) + .05 (75) + .02 (50) = 76.19%

3. Student C scores 100% for the HW, 100% on the midterm exam, 100% on the final exam, misses no lectures, and answers 14 practice sessions with all correct answers.

Course score = .53 (100) + .15 (100) + .25 (100) + .05 (100) + .02 (200) = 102%

Other Important Items

Course Organization. The course organization may be adjusted/optimized during the semester according to the pace of learning and the priority of topics. Students are responsible for attending lectures and maintaining an awareness of any changes to the course materials, homework requirements, or exam dates.

Student Disability Accommodation. I am happy to work with any student who requires an accommodation due to a disability. However, I am not authorized to grant any accommodations on my own. It is important that students first contact the Office of Disability Resources. They will discuss any barriers a student is experiencing, explain the process for establishing academic accommodation, and then authorize me to provide a specific level of accommodation. You can reach the Office of Disability Resources at disability@rochester.edu or (585) 276-5075.

Academic Honesty. Students are expected to be familiar with the University's policies on academic honesty. I have provided course-specific academic honesty policies on Blackboard under "Course Academic Honesty." During the first week of class, please review both the University policies and the course policies. You must confirm that you have read and accept these policies by completing the Acceptance of Academic Honesty Policy activity ("test") on the Course Home Page on Blackboard.

Punchline 1: Don't cheat.

Punchline 2: Be careful of how you use an AI/LLM/chatgpt-like app: it may be considered cheating!

If I suspect a student has violated any of the academic honesty policies, I am required to report the violation. Academic honesty violations on HW's usually result in a zero for that HW. Violations on an exam usually result in a zero for the entire exam and a full letter grade reduction in the course score once the final grade is calculated.

If in doubt about what is acceptable behavior concerning completing an exam or homework, just ask the professor.

Updated: 8/25/25