PSCI 200
Data Analysis I

Fall 2023

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

Prof Curt Signorino
curt.signorino@rochester.edu

TA's	**Bahar Zafer***	**Brianna Jones**	**Yijin (Cro) Qu**	**Sunahra Tanvir**
	bzafer	bjones58	yqu5	stanvir

Office Hours:

Mon	5-6pm	H335	Sunahra
Tues	2-4pm	H303	Prof Curt
Wed	9-10am	H335	Bahar
Thurs	2-3pm	H335	Brianna
Fri	11:15-12:15	H335	Cro

Workshops:

Tues	5-6:30pm	H329	Sunahra
Wed	5:15-6:45pm	H329	Bahar
Fri	1-2:30pm	H112	Brianna & Cro

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 reporters mean when they refer to a poll being accurate to $\pm 3\%$? 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 ($\sim$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 212/213/214, are generally discouraged from taking PSCI 200. Students who have taken the AP exam in Statistics may 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 the professor first.

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 the professor and the head TA 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%

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 the professor. The absence will simply count as one of the five free absences.

Practice sessions. At least once per week, an in-class practice session will require students to enter an answer through Blackboard. In order to receive full credit (100%) for the practice session participation score, students must enter an answer (any answer) for at least 10 of the practice sessions. If a student enters the correct answer, they will receive an extra 10% towards the practice session grade, up to a maximum practice session score of 200%.

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%

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*. This is available in three versions: a webpage (linked above), a pdf that is free to download, or a paperback to purchase. The pdf is donation-ware. You will need to add the pdf to your cart. However, you have the option of downloading it for free – simply move the sliders to zero. Then check out and download the pdf.

- (OIS) David M. Diez, Mine Cetinkaya-Rundel, and Christopher D. Barr. OpenIntro Statistics. 4th ed. This too is available as a donation-ware pdf.

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

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 R and RStudio 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 are available here and on Blackboard (see the "Install R and RStudio" section). Even if you know how to install these files or even if you already

have them installed, 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 as detailed in the aforementioned installation instructions. If you would like to compile 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 "Install R and RStudio" document.

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

. . . . . .

R and Descriptive Statistics
Look at Your Data!

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

3 Introduction to R, pt 2

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

HW 1

datadatabobata 2

4 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

5 Look at your data: two variables at a time

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

Imai QSS 2.2, 3.6

datadatabobata 4

HW 2

. . . . . .

Research Design

6 Operationalization

Operationalization, codebooks, unit of observation, unit of analysis, ecological fallacy

7 Sampling

Sampling, surveys, random sampling, convenience sampling, selection bias, response bias

Imai QSS 3.4.
IMS 2.1

8 Causal effects and experiments

Causality, potential outcomes, simple randomized experiments

Imai QSS 2.1, 2.3–2.7
IMS 2.2–2.4

9 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

10 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

11 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

12 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

13 Expected value and variance

Centrality and spread of a random variable

Imai QSS 6.3.5.
OIS, 3.4.1–3.4.2

datadatabobata 12

HW 5

14 Continuous distributions

Uniform, Normal, PDF, CDF, complement, axioms

Imai QSS 6.3.4
OIS 3.5, 4.1

datadatabobata 13

15 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

datadatabobata 14

HW 6

. . . . . .

Inference: Uncertainty and our Conclusions about a Single Variable

16 Confidence intervals

Large sample mean, large sample proportion, Margin of error, small sample mean, t distribution

Imai QSS 7.1.2–7.1.6
OIS, Ch 5.2, 6.1.2

datadatabobata 15

HW 7

17 Hypothesis tests

Framework, 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

18 Comparing two groups

Difference of means, difference of proportions, significance

Imai QSS 7.2.4–7.2.6
OIS, Ch 6.2, 7.3
datadatabobata 18

HW 8

19 Joint and conditional distributions

Joint distributions, marginal probability, independence, conditional probability

Imai QSS 6.2
OIS 3.2
datadatabobata 19

20 Crosstabs and the test of independence

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

OIS 6.3-6.4

HW 9

21 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

22 Linear regression: hypothesis tests

Interpreting a regression table, significance of the regression

Imai QSS 7.3.4–7.3.5
Verzani 14

HW 10

. . . . . .

23 Final Exam Review

Workshop

24 Final Exam

Tuesday 12/19, 7:15pm-10:15pm, Dewey 2-162

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 additional course-specific academic honesty policies on Blackboard under the Course Academic Honesty tab. If I suspect a student has violated any of these policies, I am required to report the violation. Punchline: don't cheat. If in doubt about what is acceptable behavior concerning completing an exam or homework, just ask me.

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") at the bottom of the Course Home

Page on Blackboard.

Updated: 8/30/23