PSCI 200
Data Analysis I

Fall 2022

Tu/Th 2-3:15, Morey 525

Prof Curt Signorino
curt.signorino@rochester.edu
Office Hours: H303, Tu & Th 3:30-4:30

TAs	**Neha Iyer**	**Payge Vukelic**	**Jessica Yao**
Office Hours (H335)	Mon 3-4	Fri 10-11	Wed 1-2
Email	niyer3	pvukelic	yyao32

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. 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.

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, 2-3:15pm. The Tues/Thurs sessions will be a mix of lecture, practice sessions, and labs. Practice sessions will consist of short ($\sim$10–15min) sessions where students have a chance to practice or implement the techniques just presented in lecture. Labs will provide the opportunity for students to begin working on new homework assignments. 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. If you have taken a similar course and would like to take PSCI 200, please consult with me first.

Grading

Course grades will be based on a series of homeworks (65%), a midterm exam (10%), a final exam (20%), and class participation/attendance (5%). Attendance is required and will be taken each class using Qwickly.

Homeworks. Homeworks will typically be handed out (via Blackboard) in class, at which point a short "lab" will be held. The lab allows students to read over the HW, download any data, begin working on the HW, and ask me questions about the HW before taking it home to complete it. Homeworks will normally be due by the start of lecture one week after they are handed out. 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 HWs 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 new HWs.

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 – contact 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 me 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 assignments will be penalized (as a percentage of total points possible) as follows:

Lateness	**Penalty**
Up to 10 hrs	5%
10 hrs to 24 hrs ($\leq$ 1 day)	10%
24 hrs to 48 hrs ($\leq$ 2 days)	20%
48 hrs to 72 hrs ($\leq$ 3 days)	30%
$>$ 3 days but $\leq$ 4 days	40%
$>$ 4 days but $\leq$ 5 days	50%
$>$ 5 days but $\leq$ 6 days	60%
$>$ 6 days but $\leq$ 7 days	70%
$>$ 7 days	100%

Examples: Suppose a HW has a total of 20 points possible and, without the penalty, you score 18 out of 20 points.

1. You turn in the HW two hours after class ends. Since the HW was late by two hours, you receive a 5% penalty or 1 point Your HW score is $18 - 1 = 17$ out of 20.

2. You turn in the HW by noon the day after the lecture in which it was due. Since the HW was late by more than 10 hours but less than 24 hours, you receive a 10% penalty or 2 points. Your HW score is $18 - 2 = 16$ out of 20.

3. You turn in the HW three and a half days after it was due. Since the HW was more than 3 days late but less than 4 days late, you receive a 40% penalty or 8 points. Your HW score is $18 - 8 = 10$ out of 20.

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.
- (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, RStudio, and LaTeX

Students are required to complete homeworks in 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. You should install the most recent versions of both on your laptop. LaTeX is a text formatting language. You will not have to learn LaTeX for this course. However, RStudio will use LaTeX to turn your HW R code into a nicely formatted pdf.

Step-by-step instructions are available on Blackboard in the folder "Installing R, RStudio, and LaTeX." Even if you know how to install these files or even if you already have them installed, please work through the instructions. A fresh install based on the provided instructions will help avoid most of the problems I regularly see at the start of the semester.

Course Outline

Course topics do not have dates assigned. Students are responsible for keeping up with the readings, lectures, labs, and HWs 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

Install R, RStudio, & LaTeX

. . . . . .

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

3 Introduction to R, pt 2

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

HW 1

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

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

Interactive examples on logical operations: datadatabobata 2

Interactive examples on covariance & correlation: datadatabobata 16

HW 2

. . . . . .

Research Design

6 Sampling and measurement

Operationalization, codebooks, unit of observation, sampling, surveys

Imai QSS 3.4.
IMS 2.1

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, pt 1

Least squares, bivariate linear regression, residuals, lm

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

10 Fitting a line, pt 2

Experimental data, observational data

Imai QSS 4.3.1

HW 4

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

12 Midterm Exam Review

In class

13 Midterm Exam

In class. Exam will cover topics 1–10

Probability: The Mathematics of Uncertainty

14 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

Interactive examples: datadatabobata 6

15 Expected value and variance

Centrality and spread of a random variable

Imai QSS 6.3.5.
OIS, 3.4.1–3.4.2

Interactive examples: datadatabobata 7

HW 5

16 Continuous distributions

Uniform, Normal, PDF, CDF, complement, axioms

Imai QSS 6.3.4
OIS 3.5, 4.1

Interactive examples: datadatabobata 9

17 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

Interactive examples: datadatabobata 10

HW 6

. . . . . .

Inference: Uncertainty and our Conclusions about a Single Variable

18 Confidence intervals, pt 1

Large sample mean, large sample proportion

Imai QSS 7.1.2–7.1.4
OIS, Ch 5.2, 6.1.2

Interactive examples: datadatabobata 11

19 Confidence intervals, pt 2

Margin of error, small sample mean, t distribution

Imai QSS 7.1.5–7.1.6

HW 7

20 Hypothesis tests, pt 1

Framework, large samples: proportion

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

Interactive examples: datadatabobata 12

21 Hypothesis tests, pt 2

Mean, large sample, small sample, type I/II error, p-values

Interactive examples: datadatabobata 13

. . . . . .

Inference: Uncertainty about whether Two Variables are Related

22 Comparing two groups

Difference of means, difference of proportions, significance

Imai QSS 7.2.4–7.2.6
OIS, Ch 6.2, 7.3
Interactive examples: datadatabobata 14

HW 8

23 Joint and conditional distributions

Joint distributions, marginal probability, independence, conditional probability

Imai QSS 6.2
OIS 3.2
Interactive examples: datadatabobata 8

24 Crosstabs and the test of independence

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

OIS 6.3-6.4

HW 9

25 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

26 Linear regression: hypothesis tests

Interpreting a regression table, significance of the regression

Imai QSS 7.3.4–7.3.5
Verzani 14

HW 10

. . . . . .

27 Final Exam Review

In class

28 Final Exam

12/21, 12:30-3:30, Morey 525

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.

COVID-19. Please maintain an awareness of the University's covid tier level and employ appropriate masking and social distancing. The University covid resource center can be found here: www.rochester.edu/coronavirus-update.

If you become ill – with covid, the flu, or anything else – and need to miss class, please notify me by email as soon as possible.

Class Recording. I may need to record lectures for students who cannot attend in person. Students should simply assume all lectures are being recorded, regardless of whether I announce it at the start of class or not.

Updated: 8/26/22