PSCI 200: Data Analysis I

Mayya Komisarchik

Spring 2026

Class Location: Goergen Hall 108
Instructor: Mayya Komisarchik (mayya.komisarchik@rochester.edu)
TA: Alice Brocheaux (abrocheu@UR.Rochester.edu)
TA: Yu Xia (yxia24@UR.Rochester.edu)

Class Hours: M/W 2:00-3:15pm
Office Hours: M 3:30-4:30pm
Office Hours: TBD
Office Hours: TBD

Course Description

Do vaccines help prevent disease? Does studying for exams really improve performance? Are people really more likely to go vote if they write down what they plan to do on election day? Are politicians right when they say the economy is improving? How might we even start measuring or evaluating any of these claims? 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.

Class Structure

This course follows the College credit hour policy for four-credit courses. We will meet in person twice a week. This is a lecture course, though we will also work through demonstrations and coding practice together during class sessions. Note that this course also features practice sessions hosted by your TAs. These are not mandatory, but they are designed to help you work through problems on your homework and concepts that will appear on your exams. You are *highly* encouraged to attend these, especially if you find yourself struggling with a specific topic. The instructor and TAs also have office hours and are available to your during those hours if you have questions.

Prerequisites

PSCI 200 is intended to be an introductory course 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. Students who have taken another UR course in statistics, such as ECON 230 or STAT 212/213/214, should not take PSCI 200. Students who have taken the AP exam in statistics may take PSCI 200, especially if you did not use `R` in your AP course.

Grading Policy

Final grades for the course will be based on:

> **Class participation** (10%): Participation can take a variety of forms, including (but not limited to): asking and/or answering questions during lecture, participating actively during in-class practice sessions, or attending TA practice sessions. Attendance will not be explicitly graded to allow students flexibility for illness, travel, or other occasional scheduling conflicts, but coming to class rarely or never is a great way to hurt your score in this category.
>
> **Problem Sets** (50%): Problem sets will be assigned in this course approximately weekly. Problem sets will be posted, submitted, and graded via Blackboard. Accordingly, your problem set answers must be typeset. All assignments are to be completed individually. Be sure to read the PSCI 200 course academic honesty policy concerning HW completion.
>
> **Midterm** (15%): We will have an in-class midterm on **Wednesday, February 25**.
>
> **Final Exam** (25%): We will have a final exam on **Monday, April 29**.

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.

Submitting Assignments

Extensions

The Late Bank

Each of you will have a 48-hour bank of extension time to apply to assignment deadlines in any way you choose. This means that you may, **without penalty**: turn in one assignment 48 hours late, turn in two assignments - each 24 hours late, etc. You do not need to notify me in advance if they plan to use this policy; I will apply this automatically using the timestamps on assignments submitted through Blackboard. This grace period cannot be applied to exams.

Additional Extensions

The late bank is designed to help you manage your schedule by giving everyone some flexibility over deadlines. You may not submit late assignments if you have already used your 48 hours of late time unless one of the following applies: (1) a death in the family (2) an unforeseen medical emergency (3) a public health risk. If any of these circumstances apply, please inform the College Center for Advising Services and request that they reach out to me. You may also contact me directly to discuss the length of your absence and arrange a modified schedule for assignments. I do not require you to provide notes from health care providers or to share other personal information.

If you do not have an approved extension, assignments turned in after the 48-hour grace period will not be evaluated. These will count as zeroes toward the final grade.

R and RStudio

Students are required to complete most problem sets using `R` and RStudio. `R` is the statistical analysis software we will use in this class. RStudio is something called an "integrated development environment", or IDE. This is a user friendly interface that helps you edit, automate, debug, and do many other things you need to do with the `R` code you'll write in this class.

Both `R` and RStudio are free. This is one of the many reasons we like them. 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. Make sure you install `R` first! Detailed, step-by-step installation instructions are available 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.

You will need to submit your problem sets via Blackboard in .pdf. Unfortunately, `R` and RStudio don't automatically produce nicely-formatted write-ups for your homework! There are, however, several ways to convert what you produce in `R` to nicely formatted documents for submission:

- Type all of your answers and descriptions into your preferred word-processing software (e.g. Word, Google Doc, etc.). `R` and RStudio give you options to export graphics and tables (as .pdf files and html, respectively), so you can just export those, paste them into your document manually, save the final version as a .pdf, and submit.
- You may use LaTeX to integrate your writing and your `R` output. LaTeX is another free typesetting software with its own language that lets you include tables and graphics exported from `R` and update them automatically if you store them as specific files. Learning this approach is purely optional. We will not cover this explicitly in class but the instructor and the TAs will be able to provide support if you choose this approach.
- You may use RMarkdown to integrate your writing and your `R` output. RMarkdown is another language that combines, or "knits" together blocks of written responses and `R` code to create formatted documents. This also has its own language and can be edited in the

RStudio environment. This is also purely optional. The instructor at the TAs will be able to provide support if you choose this approach. We will post RMarkdown versions of problem sets when relevant for those of you how would like to use RMarkdown, but this is not required.

Academic Honesty

I encourage all students to discuss readings and assignments with classmates. However, you must do all of your written work independently; you may not collaboratively do problem sets and submit either group solutions or identical copies of solutions across group members. All course assignments and in-class activities must be performed in accordance with the University of Rochester's Academic Honesty Policy (**https://www.rochester.edu/college/honesty/**). Please note that the university requires me to report possible violations of this policy to the Board on Academic Honesty.

Resources

If there are accommodations of any sort that would make the class work better for you, please come talk to me about it.

If you think you may need to seek accommodations due to a disability, the Office of Disability Resources (**http://www.rochester.edu/college/disability/index.html**) can help you figure out your options. Speaking with faculty about disability-related accommodations is strictly confidential; I encourage you to come see me with concerns or requests. Please note that if your accommodation requires testing in a separate location, you must reach out to the Office of Disability Resources *well in advance* of your exam to schedule testing.

Other resources that may prove helpful during your time at the University of Rochester:

- CARE (**https://www.rochester.edu/care/about.html**).
- The Writing & Speaking Center (**http://writing.rochester.edu/index.html**)
- Tutoring (**http://www.rochester.edu/college/cetl/undergraduate/tutoring.html**)
- University Counseling Center (**http://www.rochester.edu/uhs/ucc/**)
- David T. Kearns Center (**https://www.rochester.edu/college/kearnscenter/**)
- Office of Minority Student Affairs (**http://www.rochester.edu/College/OMSA/**)

Readings

Most readings for this course are available online free of charge. We will be reading excerpts from the following books in class. The first title below can be purchased via Amazon or other online retailers, but assigned excerpts will also be made available to you via the online course reading repository linked below.

- (QSS) Imai, Kosuke. *Quantitative Social Science: An Introduction*. Princeton. 2017. Assigned readings made available here.

- (IMS) Çetinkaya-Rundel, Mine and Johanna Hardin. *Introduction to Modern Statistics*. 2024. Available here.

- (OIS) Diez, David, Mine Çetinkaya-Rundel, and Christopher D. Barr. *OpenIntro Statistics*. 4th Edition. Available here.

- (Verzani) Verzani, John. *SimpleR: Using R for Introductory Statistics*. 2002. Available here.

Class Schedule and Readings

This syllabus is tentative and subject to change. Please check the course syllabus link periodically for updates.

I. Descriptive Statistics

Wednesday, 01/21 Course Introduction

Welcome!

Friday, 1/23 Introduction to R Part I

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

- QSS 1.3
- Verzani 1-2
- databobata 2

Monday, 01/26 Introduction to R Part II

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

- databobata 2

Wednesday, 01/28 Look at Your Data! One Variable at a Time

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

- QSS 3.2-3.3
- OIS 2
- Verzani 3
- databobata 3

Monday, 02/02 Look at Your Data! Two Variables at a Time

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

- QSS 2.2, 3.6
- databobata 4

II. Where Does Your Data Come From?

Wednesday, 02/04 Operationalization

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

Monday, 02/09 Sampling

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

- QSS 3.4
- IMS 2.1

III. Experiments and Research Design

Wednesday, 02/11 Experiments and Causal Effects

Causality, potential outcomes, simple randomized experiments

- QSS 2.1, 2.3-2.7
- IMS 2.2-2.4

Monday, 02/16 Research Design and Threats to Validity

Internal validity, external validity, confounders

- McDavid, James and Laura Hawthorn. 2025. *Research Design for Program Evaluations*. Chapter 3. Available here.
- Research Methods Knowledge Base, Section on Design. Available here.

IV. Estimating a Linear Relationship

Wednesday, 02/18 Fitting a Line

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

- QSS 4.1-4.2, 4.3.1
- IMS 7
- OIS 8.1-8.3

Monday, 02/23 Midterm Exam Review

Wednesday, 02/25 Midterm Exam in Class

V. Probability: The Mathematics of Uncertainty

Monday, 03/02 Discrete Distributions

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

- QSS 6.1, 6.3.1-6.3.3
- OIS 3.1, 3.3, 3.4, 4.3
- databobata 11

Wednesday, 03/04 Expected Value and Variance

Centrality and spread of a random variable

- QSS 6.3.5
- OIS 3.4.1-3.4.2
- databobata 12

Monday, 03/09 No Class. Have a Great Spring Break!

Wednesday, 03/11 No Class. Have a Great Spring Break!

Monday, 03/16 Continuous Distributions

Uniform, Normal, PDF, CDF, complement, axioms

- QSS 6.3.4
- OIS 3.5, 4.1
- databobata 13

Wednesday, 03/18 Central Limit Theorem

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

- QSS 6.4, 7.1.1
- OIS 5.1
- databobata 14

VI. Inference: Uncertainty and our Conclusions about a Single Variable

Monday, 03/23 Confidence Intervals

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

- QSS 7.1.2-7.1.6
- OIS 5.2, 6.1.2
- databobata 15

Wednesday, 03/25 Hypothesis Tests

Framework, large samples: proportion and mean, small sample: mean, type I/II error, p-values

- QSS 7.2.1-7.2.3
- OIS 5.3, 6.1.3, 7.1
- databobata 16
- databobata 17

VII. Inference: Uncertainty about whether Two Variables are Related

Monday, 03/30 Comparing Two Groups

Difference of means, difference of proportions, significance

- QSS 7.2.4-7.2.6
- OIS 6.2, 7.3
- databobata 18

Wednesday, 04/01 Joint and Conditional Distributions

Joint distributions, marginal probability, independence, conditional probability

- QSS 6.2
- OIS 3.2
- databobata 19

Monday, 04/06 Crosstabs and the Test of Conditional Independence

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

- OIS 6.3-6.4

Wednesday, 04/08 Linear Regression: Standard Errors

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

- QSS 7.3.1-7.3.3
- OIS 8.4
- Verzani 13

Monday, 04/13 Linear Regression: Hypothesis Tests

Interpreting a regression table, significance of the regression

- QSS 7.3.4-7.3.5
- Verzani 14

Wednesday, 04/15 Bonus: Regression and Variable Types

Indicators, interaction terms, higher orders

Monday, 04/20 Bonus: Regression Diagnostics Part 1

Best Linear Unbiased Estimator, heteroskedasticity, non-linearity

Wednesday, 04/22 Bonus: Regression Diagnostics Part 2

Non-normality, outliers, multicollinearity

Wednesday, 04/24 Final Exam Review

Monday, 04/29 Final Exam in Class