PSCI 205
Data Analysis II

Spring 2026

Tues/Thurs 9:40-10:55, Dewey 2-110E

Prof Curt Signorino
Office hours: Tues 11-1, Harkness 303
curt.signorino@rochester.edu

TA's	Adam Roberts*	Jacob Cohen	Lillian Good
	arober48@ur.rochester.edu	jcohen43@u.rochester.edu	lgood5@u.rochester.edu

Office Hours				Workshops			
Mon	3:30 – 4:30pm	H335	Lillian	Mon	6:15 – 7:45pm	TBD	Jacob
Wed	3:30 – 4:30pm	H335	Jacob	Wed	10 – 11:30am	TBD	Adam
Thurs	11am – 12pm	H335	Adam	Thurs	5:30 – 7pm	TBD	Lillian

COURSE DESCRIPTION: This course builds on PSCI 200, Data Analysis I, taking the linear regression model as its starting point. We will explore various statistical techniques for analyzing a world of data that is relevant to political science in particular, and to the social sciences and business more broadly. In addition to the linear regression model, we will examine regression models for binary data, durations, counts, and censored outcomes, among others. Students will also be introduced to methods and research designs from the causal inference literature, such as experiments, matching, diff-in-diff, and regression discontinuity designs. The models and techniques will be applied to topics such as international conflict, civil war onset, parliamentary cabinet survival, international sanctions, campaign contributions, and voting. Students will be taught how to (1) frame research hypotheses, (2) analyze data using the appropriate statistical model, and (3) interpret and present their results. Statistical analysis will be conducted using R/RStudio.

PREREQUISITES: Students should have taken a course (such as PSCI 200, ECON 230, or STAT 180/190/212/213/214) that introduces them to probability, hypothesis tests, confidence intervals, and bivariate regression. Students should have a basic familiarity with R and RStudio. Calculus and matrix algebra are not required. If you need to refresh yourself concerning R and/or introductory statistics, I have uploaded PSCI 200 review material to Blackboard. If you have only taken AP Statistics, without any use of R/RStudio, I strongly recommend that you take one of the above introductory courses before taking this course.

COURSE MEETING & CREDITS: This course follows the College credit hour policy for four-credit courses. We will meet in person twice a week for 75-minute classes. These classes will consist of lectures and hands-on practice sessions. Additionally, every week separate workshops will be held. Workshops may supplement the material presented in lecture. However, they are primarily intended to provide students additional opportunities to work on the data analysis skills relevant to the HW assigned that week and to receive help or instruction from the teaching assistants. Workshop attendance is not mandatory but is *highly* encouraged. Remaining credit hours (or portions thereof) are fulfilled through independent reading and completion of the homework assignments.

READINGS: There is no perfect text for this course. Instead, I will assign readings from various texts and articles and supplement those with lecture notes. Except for the Wooldridge text, all texts and articles will be available as pdf's on Blackboard. Texts used for this course will include

- John Verzani. *SimpleR: Using R for Introductory Statistics*. This is an open source pdf that introduces students to using R for statistics. There is little to no math. It focuses on the mechanics of data analysis, hypothesis testing, and linear regression using R. Detailed answers to exercises are provided here.

  If you feel rusty concerning R, please work through pages 1-24 before the end of the first week of class.

- G. Jay Kerns. *Introduction to Probability and Statistics using R*. 3rd ed. The topics overlap quite a bit with Verzani. However, Kerns is much more mathematical, including the use of calculus. The open source pdf is available as part of R's IPSUR package.

- David M. Diez, Christopher D. Barr, and Mine Cetinkaya-Rundel. *OpenIntro Statistics*. 3rd ed. For some, this will be a more user-friendly version of Kerns, without any calculus or more advanced math.

- Jeffrey M. Wooldridge. *Introductory Econometrics*. 7th ed. More advanced than *OpenIntro Statistics*. Does not demonstrate with R. Does not use calculus or linear algebra in main text, but does provide some proofs. The 5th and 6th editions are perfectly fine. However, you may have to determine which chapters in the 5th and 6th editions correspond to the readings shown in the list of topics that follow.

- Marco R. Steenbergen. 2008. *Discrete Choice Models for Political Analysis*. Advanced Political Methodology Lecture Notes. (pdf on Blackboard)

STATISTICAL SOFTWARE: 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. 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 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 available on Blackboard as well.

LAPTOPS, CHROMEBOOKS, & TABLETS. You will need a laptop for this course, preferably running Mac OSX, Windows, or Linux operating systems. If you haven't updated your operating system in the last year, you may need to do so before installing the most recent versions of R/RStudio.

If your laptop is a Chromebook-style laptop (with a keyboard), it is possible to install R/RStudio. However, it is a more complicated procedure. Instead, we recommend that you use the cloud version of RStudio available at Posit Cloud (see below concerning tablets).

Although I strongly recommend against it, it is possible to complete this course if you have only a tablet *with a physical keyboard*. A free, cloud (i.e., web browser) version of RStudio is available at Posit Cloud. 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 Mac/Windows/Linux laptop. If you do not own a laptop, you may be able to borrow one for the semester from University IT. Lastly, it is simply not possible to complete this course successfully with a keyboard-less tablet.

COURSE SCHEDULE

1. Course Introduction

2. A Very Quick Introduction to R

R/RStudio, calculations, variables, classes, vectors, matrices, logical operations, data frames, loading data sets, descriptive statistics, tables, plots, help, knitting an R script, RNotebook, NA's

Reading: Verzani, pp. 1-24. IPSUR, Ch 2.

Linear Regression: Estimation & Inference

3. Bivariate Regression I: Estimation

It's a line!, estimating the coefficients, lm(), applied to observational data

Reading: Wooldridge Ch 2.1, 2.3, 2.4

4. Bivariate Regression II: Inference

OLS assumptions, sampling distribution of coefficients, t tests, CI's

Reading: Wooldridge Ch 2.5

5. Multiple Regression I: Estimation & Inference

It's a plane!, research hypotheses, estimating and interpreting coefficients and standard errors

Reading: Wooldridge Ch 3.2

6. Multiple Regression II: Model Fit & F Test of Regression

R^2, F test of the regression, statistical vs substantive significance, Is R^2 evil?

Reading: Wooldridge Ch 4.5-4.6

7. Multiple Regression III: Multicollinearity, Nested Models

Correlated regressors, F test for nested models, reexamining Black voter registration and global warming

Reading: Wooldridge 3.4a

Topics in Multiple Regression

8. Dummy Variables

Binary regressors, factor variables, omitted/baseline category, fixed effects

Reading: Wooldridge Ch 7.1-7.4b

9. Interactions & Plotting Fitted Values

Interacted regressors, plotting fitted values for more complicated models

Reading: Wooldridge Ch 6.2c, Ch 7.4a.

10. Polynomial & Log Transformations

Polynomial regression, log-linear & linear-log regression models

Reading: Wooldridge Ch 2.4, Ch 6.2.

11. Potential Problems & Diagnostics

Checking whether our modeling assumptions are incorrect, linearity, heteroskedasticity, Normality, outliers

Reading: *OpenIntro* Ch 8.3. *IPSUR*, 11.4-11.5. Wooldridge Ch 9.5.

12. Maximum Likelihood Estimation

Intuition, one parameter model, multiple parameters, Normal, Binomial

Reading: Gary King. 1998. *Unifying Political Methodology*. Ch 2 & 4.

13. Binary Data I

Logistic regression, probit, nonlinear E(Y|X), glm(), interpretation.

Reading: *OpenIntro*, Ch 8.4. Steenbergen, Ch 2. Wooldridge Ch 17.1.

14. Binary Data II

Hypothesis tests: individual coefficients, test of model, nested models; AIC, BIC

15. Count Data

Poisson, negative binomial regression

Reading: Wooldridge, Ch 17.3.
Beaujean & Morgan. 2016. "Tutorial on Using Regression Models with Count Outcomes…"
Zeileis et al. "Regression Models for Count Data in R."

16. Survival Models

Exponential, weibull, hazard, survival function, censoring

17. Censored & Truncated Data

Tobit model, Heckman selection, survey nonresponse

Reading: Wooldridge Ch 17.2, Ch 17.4.
Arne Henningsen. "Estimating Censored Regression Models in R using the censReg…"
Sigelman & Zeng. "Analyzing Censored and Sample-Selected Data with Tobit…"

18. Causal Inference & Experiments

Experiments w/clones, potential outcomes, RCT's, internal/external validity, B&M 2004 resumes experiment

Reading: Bertrand & Mullainathan. 2004. "Are Emily and Greg More Employable than Lakisha and Jamal?" *AER*.

19. Matching using Observational Data

Exact matching, propensity scores, MatchIt, voter mobilization data, effect of training programs

Reading: Arceneaux, Gerber, & Green. 2006. "Comparing Experimental and Matching Methods Using a Large-Scale Voter Mobilization Experiment." *Political Analysis*.

20. Difference in Difference Design

Natural experiments, parallel trends assumption, effect of minimum wage laws, evaluating the presence of police

Reading: Card & Krueger, 1994. "Minimum Wages and Employment." *AER*.
Tella & Schargrodsky. 2004. "Do Police Reduce Crime?" *AER*.

21. Regression Discontinuity Design

Natural experiments, effect of Head Start on children's futures, effect of election to Parliament on wealth

Reading: Ludwig & Miller. 2007. "Does Head Start Improve Children's Life Chances?" *QJE*.
Eggers & Hainmueller. 2009. "MP's for Sale? Returns to Office in Postwar British Politics." *APSR*.

Final Exam Review
In Workshop

Final Exam
Friday, May 8, 4-7pm, Dewey 2-110E

COURSE OBJECTIVES

In this course, you will develop

- An understanding of multiple regression estimation and inference.
- An understanding of common issues or aspects of multiple regression, such as dummy variables, interaction terms, nonlinear models, and regression diagnostics.
- An understanding of the theory behind maximum likelihood estimation.
- An understanding of maximum likelihood estimation and inference for regressions with binary, count, and duration outcomes.
- An understanding of experiments, causal inference, and research designs that are better enable researchers to make causal claims.

COURSE LEARNING OUTCOMES

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

- Estimate a multiple regression and evaluate hypotheses concerning the coefficients.
- Conduct hypothesis tests concerning multiple linear restrictions of the coefficients in a regression.
- Interpret regression output when using dummy variables with multiple categories.
- Estimate and interpret regressions that contain an interaction term.
- Estimate and interpret regressions that contain polynomial or log terms.
- Evaluate basic regression diagnostics.
- Estimate and interpret regression models with binary, count, and duration outcomes.
- Calculate and plot fitted values for logistic regressions, negative binomial regressions, and weibull regressions.
- Estimate and interpret randomized experiments, diff-in-diff designs, and regression discontinuity designs.

GRADING: Course grades will be based on a series of homeworks (55%), a final exam (38%), class attendance (5%), and participation through practice session answers (2%). A 0-100 grade is calculated for each category. The course score is the weighted average using the above weights.

Homeworks. Homework assignments will be made available via Blackboard and will typically be due 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 205 course academic honesty policy concerning HW completion.

Late Homework Submissions. It is important that students submit 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 new HW's.

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 late	5%
10hrs to 24hrs (≤1day)	10%
24hrs to 48hrs (≤2days)	20%
48hrs to 72hrs (≤3days)	30%
>3days but ≤ 4 days	40%
>4days but ≤ 5 days	50%
>5days but ≤ 6 days	60%
>6days but ≤ 7 days	70%
>7days	100%

Again, it is important that you contact me 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). Class attendance will be calculated by dropping up to five (5) absences and then calculating the percent of times the student attended class. Example: 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, needs to attend a conference, needs to participate in a 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%.

Final Exam: The final exam is scheduled for Friday, May 8, 4–7pm, Dewey 2-110E. 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 final exams.

Course Grade Examples:

IMPORTANT: In order to receive a passing grade for the course, students must take the final exam *and* complete at least four of the homeworks. While that does not guarantee a passing grade, students who do not take the final exam and complete at least four homeworks will receive a grade of E, regardless of their weighted average score.

Assuming a student has taken the final exam and completed at least four homeworks, their course score is the weighted average using the weights shown previously. The following provides examples for three hypothetical students.

1. Student A scores 90% for the HW score, 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 = .55(90) + .38(89) + .05(100) + .02(130) = 90.92%

2. Student B scores 92% for the HW score, 62% on the final exam, misses 10 (out of 25) lectures, and enters only four practice session answers, none being correct.

   Class attendance score = (15/20)*100 = 75%
   Practice session score = 40%

   Course score = .55(92) + .38(62) + .05(75) + .02(40) = 78.71%

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

   Course score = .55(100) + .38(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. This section is not exhaustive. 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. 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 at the bottom of the Course Academic Honesty page on Blackboard.

Use of AI/LLM/Chat-GPT-like resources: Students are encouraged to use any (legal) online resources to *learn* about the concepts in this course, including AI / large-language models (LLMs) – e.g., Chat-GPT. These tools can be extremely helpful in providing summary information on topics, details on specific techniques, and additional examples. That said, the only way to really learn technical material is by repeatedly solving problems without outside help.

HW's and the final exam are tools to assess (1) the extent to which you understand the underlying concepts and techniques and (2) your ability to apply those techniques using R when analyzing data. To that end, students are *not* allowed to use AI/LLM/Chat-GPT-like tools to answer HW or exam questions. Students are not allowed to copy and paste or type in HW or exam questions into AI/LLM/Chat-GPT-like tools. Any R code you use to answer a HW or exam question must have been typed in/written by you alone. Any answers you report should be your independent work – i.e., without the aid of AI/Chat-GPT-like tools.

Again, if in doubt about what is acceptable behavior concerning the use of AI/Chat-GPT-like tools, please ask the professor.

Updated: 1/20/26