PSCI 205
Data Analysis II

Spring 2023
Tues/Thurs 9:40-10:55, Dewey 2110E

Prof Curt Signorino
303 Harkness Hall
Office Hours: Tues 11:30-1:30
curt.signorino@rochester.edu

TA's:
Ben Gien
Office hours: Wed noon-1 (in person)
bgien@u.rochester.edu

Fred Liu
Office hours: Thurs 1-2pm (via zoom)
sliu85@u.rochester.edu

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 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.

PREREQUISITES: Students should have taken a course (such as PSC 200, ECO 230, STT 212, STT 213, or STT 214) that introduces them to probability, hypothesis tests, confidence intervals, and bivariate regression. Students should have a basic familiarity with R. Calculus and matrix algebra are not required. If you need to refresh yourself concerning R and/or introductory statistics, I have uploaded review material to Blackboard.

COURSE MEETING & CREDITS. This course follows the College credit hour policy for four-credit courses. We will meet in person twice a week (Tues & Thurs) in Dewey 2110E for 75-minute classes. These classes will consist of lectures and hands-on practice sessions. Additionally, every week a separate workshop will be held (day and time TBD). 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. Some workshop attendance is required. Remaining credit hours (or portions thereof) are fulfilled through independent reading and completion of the homework assignments.

GRADING: Course grades will be based on a series of homeworks (60%), a final exam (30%), class attendance (5%), workshop attendance (3%), 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.

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.

Workshop attendance. Attendance will also be taken at workshops. Students must attend at least five (5) workshops to receive full workshop participation credit. If not, that part of the grade will be calculated as the percent of workshops (out of five) that the student attended.

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%.

Examples:

1. Student A scores 90% for the HW score, 89% on the final exam, attends all but 5 (out of 25) lectures, attends 8 workshops, and enters 12 practice sessions answers, with three of those being correct answers.

   Class attendance score = 100%
   Workshop attendance score = 100%
   Practice session score = 100% + 30% = 130%

   Course score = .6(90) + .3(89) + .05(100) + .03(100) + .02(130) = 91.3%

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

   Class attendance score = (15/20)*100 = 75%
   Workshop attendance score = (2/5)*100 = 40%
   Practice session score = 40%

   Course score = .6(92) + .3(62) + .05(75) + .03(40) + .02(40) = 79.5%

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

   Course score = .6(100) + .3(100) + .05(100) + .03(100) + .02(200) = 102%

In order to receive a passing grade for the course, students must take the final exam and complete at least four of the homeworks.

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.

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.

  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, 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: We will use R and RStudio for our statistical analysis. R is open source and free. There are versions for Mac OSX, Windows, and Linux. You can download it from https://cran.r-project.org/. Additionally, we will use RStudio as a graphical interface for R. RStudio is free for students to download. Instructions for downloading R and RStudio are provided on Blackboard under the Prerequisite R Review folder on the Course Home Page.

COURSE SCHEDULE

1. Course Introduction

Workshop: R, RStudio, & RNotebook

R vs 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.

---

2. 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

3. Bivariate Regression II: Inference

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

Reading: Wooldridge Ch 2.5

HW 1

---

4. **Multiple Regression I: Estimation & Inference**
   It's a plane!, research hypotheses, estimating and interpreting coefficients and standard errors

   Reading: Wooldridge Ch 3.2

5. **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

   HW 2

6. **Multiple Regression III: Multicollinearity, Nested Models**
   Correlated regressors, F test for nested models, reexamining Black voter registration and global warming

   Reading: Wooldridge 3.4a

---

7. **Dummy Variables**
   Binary regressors, factor variables, omitted/baseline category, fixed effects

   Reading: Wooldridge Ch 7.1-7.4b

   HW 3

---

8. **Interactions & Plotting Fitted Values**
   Interacted regressors, plotting fitted values for more complicated models

   Reading: Wooldridge Ch 6.2c, Ch 7.4a.

---

9. **Polynomial & Log Transformations**
   Polynomial regression, log-linear & linear-log regression models

   Reading: Wooldridge Ch 2.4, Ch 6.2.

   HW 4

---

10. **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.

---

11. Maximum Likelihood Estimation
Intuition, one parameter model, multiple parameters, Normal, Binomial

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

HW 5

---

12. Binary Data I
Logistic regression, probit, nonlinear E(Y|X), glm(), interpretation.

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

13. Binary Data II
Hypothesis tests: individual coefficients, test of model, nested models; AIC, BIC

HW 6

---

14. 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."

15. Survival Models
Exponential, weibull, hazard, survival function, censoring

Reading: TBD

HW 7

---

16. 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…"

---

17. 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*.

HW 8

18. 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*.

19. Difference in Difference Design
Natural experiment, 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*.

HW 9

20. Regression Discontinuity Design
Natural experiment, effect of Head Start on children's futures

Reading: Ludwig & Miller. 2007. "Does Head Start Improve Children's Life Chances?" *QJE*.

FINAL EXAM: The final exam will be 3 hours, held in-class during finals week. The exact date and location will be announced later in the semester.

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 at the bottom of the Course Academic Honesty 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 – please use good judgment concerning whether to attend class. If you do attend class, please mask-up for the safety of everyone else. If you think you are going miss more than two classes, email me as soon as possible, so we can coordinate on how to handle the missed lectures and HW's.

Updated: 1/11/23