# PSC 205
# Data Analysis II

Spring 2020
Tues/Thurs 11:05-12:20, Gavett 208

**Prof. Curtis S. Signorino**
303 Harkness Hall
Office Hours: Wed 1:30-3:30pm
curt.signorino@rochester.edu

**TA: Ziyu Song**
zsong10@u.rochester.edu

**COURSE DESCRIPTION:** This course builds on PSC 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 models for binary data, durations, counts, censoring and truncation, self-selection, discrete choice, and strategic choice, among others. These models 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.

**Course Meeting and Credits**. This course follows the College credit hour policy for four-credit courses. We will meet twice a week (Tues & Thurs) for 1.5 hour sessions. There is no separate designated day for labs. Rather, the normal Tues/Thurs sessions will be a mix of lecture and computer labs. During the labs, students will receive computer instruction, analyze data, discuss past homework problems, and start on new homework problems. The remaining credit hour is fulfilled through independent reading and completion of the homeworks.

**PREREQUISITES:** Students should have taken a course (such as PSC 200, ECO 230, STT 211, STT 212, STT 213, or STT 214) that introduces them to probability, hypothesis tests, confidence intervals, and linear regression. Familiarity with R, calculus, or matrix algebra is not required.

**GRADING:** Course grades will be based on a series of homeworks (65%), a final exam (30%), and class participation/attendance (5%).

Unless otherwise noted, homeworks will generally be due at the start of class, one week after they are handed out. Students must deliver their homework in hardcopy. Late assignments will be penalized one half-grade (e.g., B to B-) for each day they are late. Homeworks more than seven days late will receive a grade of zero. Finally, while you are encouraged to study together and to learn the software together, all assignments are to be completed individually.

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

**STATISTICAL SOFTWARE:** As noted, we will be using R for our statistical analysis. R is open source and free. There are versions for Mac OSX, Windows, and Linux. If you have a laptop or desktop computer, you can download it from https://cran.r-project.org/. Additionally, we will be using RStudio as a graphical interface for R. RStudio is free for students to download.

**COURSE OUTLINE:**

1. **Course Introduction**

2. **Introduction to R**
   - Starting R, calculations, variables, classes, vectors, matrices, logical operations, data.frames, loading data sets, descriptive statistics, tables, plots, help
   - *IPSUR*, Ch 2. Verzani, pp. 1-7.

3. **Describing Data**
   - Categorical, ordinal, and quantitative data. Missing data. Descriptive statistics.
   - *OpenIntro*, Ch 1. *IPSUR*, Ch 3. Verzani, pp. 8-18.

4. **Probability**
   - Random variables, discrete distributions, continuous distributions, expectation, variance, independence, joint distributions, conditional probability
   - *OpenIntro*, Ch 2-3. *IPSUR*, Ch 4-7.

5. **Sampling Distributions, Confidence Intervals, and Hypothesis Tests**
   - Law of large numbers, central limit theorem, estimation
   - *OpenIntro* Ch 4, 5.1-5.4, 6.1-6.2. *IPSUR*, Ch 8-10. Verzani, pp. 59-71.

6. **Correlation and Bivariate Linear Regression**
   - Interpretation, confidence and prediction intervals, outliers, nonlinear relationships
   - *OpenIntro*, Ch 7. *IPSUR*, Ch 11. Verzani, pp. 24-31, 77-83

**7. Multiple Regression**

- Research hypotheses, interpretation, statistical vs substantive significance, dummies, interactions, nested models, $R^2$, polynomials, log transforms, diagnostics
- *OpenIntro*, Ch 8.1-8.3. *IPSUR*, Ch 12. Verzani, pp. 84-89.

**8. Two Important Topics in Linear Regression**

- Heteroskedasticity, fixed effects
- Required reading: TBD

**9. Some Problems to Consider When Analyzing Data**

- Omitted variables, measurement error, endogeneity, selection, missing data
- Required reading: TBD

**10. Maximum Likelihood**

- Intuition, one parameter, multiple parameters, the normal distribution
- Gary King. 1998. *Unifying Political Methodology*. Chapters 2 & 4. pdf's on Blackboard.

**11. Binary Data**

- Logit, probit, research hypotheses part II, derivations, nonlinear E(Y), interpretation, implicit interactions, CI's
- *OpenIntro*, Ch 8.4. Steenbergen, Ch 2.

**12. Count Data**

- Poisson, negative binomial, hurdle models
- Beaujean & Morgan. 2016. "Tutorial on Using Regression Models with Count Outcomes…"
- Zeileis et al. "Regression Models for Count Data in R."

**13. Censoring and Truncation**

- Tobit, truncated normal
- Arne Henningsen. "Estimating Censored Regression Models in R using the censReg Package."
- Sigelman & Zeng. "Analyzing Censored and Sample-Selected Data with Tobit…"

**14. Survival Models**

- Exponential, Weibull, Kaplan Meier, Cox proportional hazard

**15. Ordered Logit/Probit**

- Steenbergen, Ch 3.

**16. Discrete Choice**

- Random utility, multinomial logit, conditional logit
- Steenbergen, Ch 4-5.

**17. Selection Models**

**18. If Time Permits…**

- LASSO, neural nets, parallel computing in R, estimating games

Final Exam (TBD: Finals Week)

## 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. 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 coordinate with me concerning the 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. If I suspect a student has violated the University's academic honesty policies, I am required to initiate the procedures detailed on that webpage. Punchline: don't cheat. If in doubt about what is acceptable behavior concerning completing an exam or homework, please ask me.

**Class Behavior/Courtesy.** You may use the class lab computer or your own laptop to follow lecture notes in class. Sometimes you will also need to execute R code as I execute it in lecture. However, please do not engage in behavior that is rude or disruptive to everyone, such as reading email, checking facebook, browsing the web, etc.

Updated: 1/15/20