SOC542
STATISTICAL METHODS IN SOCIOLOGY II
Rutgers University

Syllabus

Spring 2022

CONTACT AND OFFICE HOURS

Instructor: Thomas Davidson

E-mail: `thomas.davidson@rutgers.edu`

Office hours: Wednesday 11:00-12:00pm, 109 Davison Hall, or by appointment.

Teaching Assistant: Fred Traylor

E-mail: `frederic.traylor@rutgers.edu`

Office hours: Thursday 11:00-12:00pm, Davison Hall or Zoom.

LOGISTICS

Class meetings: Mondays 5:00-8:00 p.m. *Our first meeting will be on Zoom and we will return to in-person classes in Davison Hall from 1/31 onwards.*

Course website: `https://github.com/t-davidson/SOC542-S22`

COURSE DESCRIPTION

This is the second course in a two-semester sequence of graduate-level statistics. The goal of the course is to provide an understanding of the principles and application of statistics to sociological research. The course begins with an overview of the quantitative approach to sociological research and a review of basic statistics and ordinary least squares regression. We then cover multiple regression, followed by generalized linear models for binary, count, and categorical data. Throughout the course, we will consider both frequentist and Bayesian approaches to estimation and will explore various techniques for improving the robustness and validity of statistical analyses. We will pay close attention to the theoretical interpretations of statistical models and emphasize effective and accurate scientific communication.

PREREQUISITES

Students should have taken SOC541 or an equivalent introduction to statistics. The course assumes some basic familiarity with data manipulation and visualization in R and RStudio.

LEARNING GOALS

By the end of the semester, students will:

- Be proficient in preparing datasets, running basic descriptive analyses, and producing informative data visualizations using R.
- Understand the conceptual underpinnings and assumptions of multiple regression and generalized linear models
- Be able to implement and interpret various different forms of regression models
- Be familiar with and proficient in handling interaction effects, non-linear relationships, and violations of assumptions in multiple regression analyses
- Understand and apply frequentist and Bayesian methods for estimation
- Replicate, reassess, and extend quantitative research published in leading sociological journals

ASSESSMENT

1. *Homework assignments* (50%): Homework assignments will be used to assess comprehension of materials covered in class. Assignments will be submitted using Github Classroom. Students can work together on the problem sets but must submit assignments individually.
2. *Replication paper* (40%). Each student will write a replication paper. The objectives of the replication will be to (a) reproduce a finding published in a leading sociological journal, (b) to assess the robustness of the reporting finding to alternative specifications, and (c) to extend the original analysis.
3. *Class presentation* (10%). Each student will present the findings of their replication paper to the class during one of the last two class sessions.

READINGS

There are weekly reading assignments for this course. Students are expected to complete the assigned readings *before* class. Students must purchase copies of the two required texts. Some weeks will also include additional papers published in academic journals. The recommended texts provide useful background material on data analysis and visualization in R. Both are available for free online.

Required

- Gelman, Andrew, Jennifer Hill, and Aki Vehtari. 2020. *Regression and Other Stories*. Cambridge University Press.
- McElreath, Richard. 2020. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. 2nd ed. Chapman and Hall/CRC.

Recommended

- Wickham, Hadley, and Garrett Grolemund. 2016. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. (*R4DS*). O'Reilly Media, Inc. https://r4ds.had.co.nz/
- Healy, Kieran. 2018. *Data Visualization: A Practical Introduction*. Princeton University Press. https://socviz.co/

POLICIES

The Rutgers Sociology Department strives to create an environment that supports and affirms diversity in all manifestations, including race, ethnicity, gender, sexual orientation, religion, age, social class, disability status, region/country of origin, and political orientation. We also celebrate diversity of theoretical and

methodological perspectives among our faculty and students and seek to create an atmosphere of respect and mutual dialogue. We have zero tolerance for violations of these principles and have instituted clear and respectful procedures for responding to such grievances.

Students must abide by the Code of Student Conduct and the university's Academic Integrity Policy at all times, including during lectures and in participation online. Violations of academic integrity will result in disciplinary action.

In accordance with University policy, if you have a documented disability and require accommodations to obtain equal access in this course, please contact me during the first week of classes. Students with disabilities must be registered with the Office of Student Disability Services and must provide verification of their eligibility for such accommodations.

I will also make additional accommodations due to the COVID-19 pandemic. If you or your family are affected in any way that impedes your ability to participate in this course, please contact me as soon as you can so that we can make necessary arrangements.

OUTLINE

Week 1 - January 24 - Statistics review and course overview

Readings

- Gelman, Hill, and Vehtari (GHV) Chapters 1-5
- McElreath Chapter 1
- Imbens, Guido W. 2021. "Statistical Significance, p-Values, and the Reporting of Uncertainty." *Journal of Economic Perspectives* 35 (3): 157–74. https://doi.org/10.1257/jep.35.3.157.

Recommended

- Raftery, Adrian E. 2000. "Statistics in Sociology, 1950–2000." *Journal of the American Statistical Association* 95 (450): 654–61. https://doi.org/10.1080/01621459.2000.10474245.

Week 2 - January 30 - Linear regression with a single predictor

Readings

- GHV 6-7
- Freese, Jeremy, and David Peterson. 2017. "Replication in Social Science." *Annual Review of Sociology* 43: 147–65. https://doi.org/10.1146/annurev-soc-060116-053450

Week 3 - February 7 - Frequentist and Bayesian approaches to estimation

Readings

- McElreath 2-3
- GHV 8-9

Recommended

- Gelman, Andrew. 2014. "How Bayesian Analysis Cracked the Red-State, Blue-State Problem." *Statistical Science* 29 (1). https://doi.org/10.1214/13-STS458.
- Kruschke, John K., and Torrin M. Liddell. 2018. "The Bayesian New Statistics: Hypothesis Testing, Estimation, Meta-Analysis, and Power Analysis from a Bayesian Perspective." *Psychonomic Bulletin & Review* 25 (1): 178–206. https://doi.org/10.3758/s13423-016-1221-4.

- Lynch, Scott M., and Bryce Bartlett. 2019. "Bayesian Statistics in Sociology: Past, Present, and Future." *Annual Review of Sociology* 45 (1): 47–68. https://doi.org/10.1146/annurev-soc-073018-022457.
- Lundberg, Ian, Rebecca Johnson, and Brandon M Stewart. 2021. "What Is Your Estimand? Defining the Target Quantity Connects Statistical Evidence to Theory." *American Sociological Review* 86 (3): 532–65. https://doi.org/10.1177/00031224211004187.

Week 4 - February 14 - Multiple regression

Readings

- GHV 10-11 (selections)
- McElreath 4 (selections)

Week 5 - February 21 - Dummy, categorical, and non-linear variables

Readings

- GHV 10-12 (selections)
- McElreath 5 (selections)

Week 6 - February 28 - Interactions

Readings

- GHV 10 (selections)
- McElreath 8

Week 7 - March 7 - Model checking and missing data

Readings

- GHV 11, 17.3-17.8
- McElreath 6-7 (selections)

Recommended

- Watts, Duncan J. 2014. "Common Sense and Sociological Explanations." *American Journal of Sociology* 120 (2): 313–51. https://doi.org/10.1086/678271.
- Young, Cristobal, and Katherine Holsteen. 2017. "Model Uncertainty and Robustness: A Computational Framework for Multimodel Analysis." *Sociological Methods & Research* 46 (1): 3–40. https://doi.org/10.1177/0049124115610347.
- Slez, Adam. 2017. "The Difference Between Instability and Uncertainty: Comment on Young and Holsteen (2017)." *Sociological Methods & Research* 48 (2): 400–430. https://doi.org/10.1177/0049124117729704.
- Muñoz, John, and Cristobal Young. 2018. "We Ran 9 Billion Regressions: Eliminating False Positives through Computational Model Robustness." *Sociological Methodology* 48 (1): 1–33. https://doi.org/10.1177/0081175018777988.
- Molina, Mario, and Filiz Garip. 2019. "Machine Learning for Sociology." *Annual Review of Sociology* 45: 27–45. https://doi.org/10.1146/annurev-soc-073117-041106.

SPRING BREAK - No class

Week 8 - March 21 - GLMs I: Binary outcomes and logistic regression

Readings

- GHV 13
- McElreath 10-11 (selections)

Week 9 - March 28 - GLMs II: Logistic regression and marginal effects

Readings

- GHV 14
- Long, J. Scott, and Sarah A. Mustillo. 2018. "Using Predictions and Marginal Effects to Compare Groups in Regression Models for Binary Outcomes." *Sociological Methods & Research* 50 (3): 1284–1320. https://doi.org/10.1177/0049124118799374.
- Mize, Trenton D., Long Doan, and J. Scott Long. 2019. "A General Framework for Comparing Predictions and Marginal Effects across Models." *Sociological Methodology* 49 (1): 152–89. https://doi.org/10.1177/0081175019852763.

Recommended

- Mize, Trenton. 2019. "Best Practices for Estimating, Interpreting, and Presenting Nonlinear Interaction Effects." *Sociological Science* 6: 81–117. https://doi.org/10.15195/v6.a4.

Week 10 - April 4 - GLMs III: Count outcomes and overdispersion

Readings

- GHV 15 (selections)
- McElreath 11-12 (selections)

Week 11 - April 11 - GLMs IV: Categorical and ordered outcomes

Readings

- GHV 15 (selections)
- McElreath 12 (selections)

Week 12 - April 18 - Clustered data

Readings

- McElreath 13-14 (selections)
- Bell, Andrew, Malcolm Fairbrother, and Kelvyn Jones. 2019. "Fixed and Random Effects Models: Making an Informed Choice." *Quality & Quantity* 53 (2): 1051–74. https://doi.org/10.1007/s11135-018-0802-x.
- TBD

Week 13 - April 25 - Causal inference using observational data

Readings

- McElreath 6, 14 (selections)
- TBD

Week 14 - May 2 - Student presentations