SOC542
STATISTICAL METHODS IN SOCIOLOGY II
Rutgers University

Syllabus

Spring 2023

CONTACT AND OFFICE HOURS

Instructor: Thomas Davidson Email: thomas.davidson@rutgers.edu

Office hours: Wednesday 5-6pm, Davison 109*

Teaching Assistant: Fred Traylor Email: frederic.traylor@rutgers.edu

Office hours: Wednesday 4-5pm, Davison 013*$^{+}$

*Office hours will be held in person, email ahead of time to request Zoom. $^{+}$ Email Fred before office hours for confirmation.

LOGISTICS

Class meetings: Mondays 1:00-3:40 p.m.

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

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 and 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, conducting descriptive analyses, and producing informative data summaries and visualizations using R.
- Understand the conceptual underpinnings and assumptions of multiple regression and generalized linear models.
- Understand the key differences between frequentist and Bayesian methods for estimation.
- Confidently implement, interpret, and present common varieties of regression models using R.
- Develop an original research paper using techniques covered in class.

ASSESSMENT

1. *Homework assignments* (40%):Four 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 cannot share solutions and must submit their own answers.
2. *Final paper* (50%). Each student will write a final paper. This should either be an original research paper or a replication and extension of an existing paper. Students will submit a proposal outlining the research question, data, and methodology and a preliminary results section prior to the final submission, each worth 10% of the final grade. The final paper will be worth 30% of the overall grade.
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. Most readings will be from the two required textbooks, but some weeks will also include additional papers published in academic journals or readings from the recommended texts.

Required

There are two required texts. You are encouraged to purchase copies, but both are available online for free on the links below.

- (**ROS**) Gelman, Andrew, Jennifer Hill, and Aki Vehtari. 2020. *Regression and Other Stories*. Cambridge University Press.
- (**BR!**) Johnson, Alicia A., Miles Q. Ott, Mine Dogucu. 2021. *Bayes Rules! An Introduction to Applied Bayesian Modeling*. CRC Press.

Recommended

The following four texts are all useful companions for the course. There will be a few readings drawn from the McElreath and Cunningham books. The Wickham and Grolemund and Healy books are both indispensible references for data manipulation and visualization in R.

- (**SR**) McElreath, Richard. 2020. *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. 2nd ed. Chapman and Hall/CRC.
- (**CIM**) Cunningham, Scott. 2021. *Causal Inference: The Mixtape*. Yale University Press.
- Wickham, Hadley, and Garrett Grolemund. 2016. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data*. (*R4DS*). O'Reilly Media, Inc.
- Healy, Kieran. 2018. *Data Visualization: A Practical Introduction*. Princeton University Press.

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.

To maintain accreditation, all graduate programs have overall program learning goals for their graduate students on the department website: https://sociology.rutgers.edu/academics/graduate/learning-goals

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.

OUTLINE

Week 1 (1/23) Statistics review and course overview

Readings

- Gelman, Hill, and Vehtari (**ROS**) Chapters 1-5

Recommended

- McElreath (**SR**) Chapter 1
- Cunningham (**CIM**) p. 16-36 / Sections 2.1-2.4, 2.7-2.12
- 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 (1/30) Linear regression with a single predictor

Readings

- **ROS** 6-7
- 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

- **CIM** p. 37-76 / 2.13-2.25

HOMEWORK 1 RELEASED, DUE 2/3

Week 3 (2/6) Frequentist and Bayesian estimation

Readings

- Johnson, Ott, and Dogucu (**BR!**) 1-2
- **SR** 1-2
- **ROS** 8-9

Recommended

- **BR!** 3-4, 6-8 (skim)
- 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.

Week 4 (2/13) Multiple regression

Readings

- **ROS** 10.1-10.2, 10.7-11.6
- **CIM** 3
- 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.

Recommended

- **BR!** 9, 11.2
- **SR** 4-4.4, 5-5.2

Week 5 (2/20) Dummy, categorical, and non-linear variables

Readings

- **ROS** 10.3-10.6, 12-12.5, 12.7-12.8
- **BR!** 11.1, 10.2
- Johfre, Sasha Shen, and Jeremy Freese. 2021. "Reconsidering the Reference Category." *Sociological Methodology* 51 (2): 253–69. https://doi.org/10.1177/0081175020982632.

Recommended

- **SR** 4.5.1, 5.3-5.4

HOMEWORK 2 RELEASED, DUE 2/27

Week 6 (2/27) Interactions

Readings

- **ROS** 10.3, 12.2
- **BR!** 11.3-11.4

Recommended

- **SR** 8

PAPER PROPOSAL DUE 3/7

Week 7 (3/6) Model checking, comparison, and missing data

Readings

- **ROS** 11.7-11.9, 17.3-17.8
- **BR!** 10.3-11.5

- 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.
- Western, Bruce. 2018. "Comment: Bayes, Model Uncertainty, and Learning From Data." *Sociological Methodology* https://doi.org/10.1177/0081175018799095.

Recommended

- **SR** 7, 15.2
- 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.
- Steegen, Sara, Francis Tuerlinckx, Andrew Gelman, and Wolf Vanpaemel. 2016. "Increasing Transparency Through a Multiverse Analysis." *Perspectives on Psychological Science* 11(5):702–12. doi: 10.1177/1745691616658637.

SPRING BREAK

Week 8 (3/20) GLMs I: Binary outcomes and logistic regression

Readings

- **ROS** 13, 15.1, 15.4
- **BR!** 13
- Gomila, Robin. 2021. "Logistic or Linear? Estimating Causal Effects of Experimental Treatments on Binary Outcomes Using Regression Analysis." *Journal of Experimental Psychology: General* 150(4):700–709. https://doi.org/10.1037/xge0000920.

Recommended

- **SR** 10.1-10.4, 11.1

HOMEWORK 3 RELEASED, DUE 3/31

Week 9 (3/27) GLMs II: Logistic regression and marginal effects

Readings

- **ROS** 14

Recommended

- 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.

Week 10 (4/3) GLMs III: Count outcomes and overdispersion

Readings

- **ROS** 15.2-15.3, 15.8
- **BR!** 12

Recommended

- **SR** 11.2, 12.1-12.2

Week 11 (4/10) GLMs IV: Categorical and ordered outcomes

Readings

- **ROS** 15.5

Recommended

- **SR** 11.3, 12.3-12.5

HOMEWORK 4 RELEASED, DUE 4/14

Week 12 (4/17) Fixed effects, hierarchical models, and clustered standard errors

Readings

- **ROS** 22
- **BR!** 15-17
- 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.

Recommended

- **BR!** 18-19 (skim)
- **CIM** 2.25

PRELIMINARY RESULTS DUE 4/21

Week 13 (4/24) Regression and causal inference

Readings

- **ROS** 18-21, Appendix B
- **CIM** p. 96-198 / 3-5

Recommended

- **CIM** p. 241-509 / 6-9 (skim)
- Morgan, Stephen L., and Jennifer J. Todd. 2008. "A Diagnostic Routine for the Detection of Consequential Heterogeneity of Causal Effects." *Sociological Methodology* 38(1):231–81.
- King, Gary, and Richard Nielsen. 2019. "Why Propensity Scores Should Not Be Used for Matching." *Political Analysis* 27 (4): 435–54. https://doi.org/10.1017/pan.2019.11.

Week 14 (5/1) Final resentations

PRESENTATION ON 5/1

FINAL PAPERS DUE 5/5