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: TBC.

LOGISTICS

Class meetings: Mondays 5:00-8:00 p.m. We will meet on Zoom in January and return to in-person classes in Davison Hall from 2/7 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 multivariate statistics to sociological research. The course begins with an overview of the quantitative approach to sociological research and a review of basic statistics and bivariate regression. We then cover multivariate linear 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 conduct and interpret various different forms of multivariate statistical 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**


**Recommended**


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

Recommended


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

Readings

- GHV 6-7

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

Readings

- McElreath 2-3
- GHV 8-9

Recommended


**Week 4 - February 14 - Multivariate 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*


**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

Recommended

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)
• TBD

Week 13 - April 25 - Causal inference using observational data

Readings
• McElreath 6, 14 (selections)
• TBD

Week 14 - May 2 - Student presentations