Syllabus
Sociology 541: Analysis of Sociological Data I
Fall 2021

Class Meetings: Tuesdays 1:00 – 3:40 pm
Location: 128 Davison Hall, Douglass campus

Professor: Lei Lei
Office: 039 Davison Hall, Douglass campus
Email: llei@sociology.rutgers.edu
Office Hours: Tuesdays 10:00 –11:00 am via Zoom
Zoom link:
https://rutgers.zoom.us/j/6743392620?pwd=RklOczVpUWdCNTZWV0txZFZRbmZDQT09

Teaching Assistant: Frederic Traylor
Email: frederic.traylor@rutgers.edu
Office Hours: Thursdays 10:00 – 11:00 am via Zoom
Zoom link:
https://rutgers.zoom.us/j/6292425076?pwd=WUNwOVNORVlKb0JFbCtnQ1NBd3ZnUT09

Course Description
This introductory course is the first part of a two-semester sequence (Soc 541 and 542). It is geared toward graduate students who are training to become professional sociologists. The course introduces students to basic concepts and techniques in quantitative analysis and illustrates how these core statistical ideas can be applied to social science questions and real-world problems. The course covers multiple topics such as descriptive statistics, probability, statistical inference, significance test, and ordinary least square regression. Content covered in Soc 541 serves as the foundation for the second course in the sequence. This course assumes that students have little or no statistical background.

Learning Goals
1. Introduce students to basic concepts, terminology, and process of data analysis.
2. Teach students to intelligently read published quantitative research and creatively design their own quantitative projects
3. Acquire statistical literacy and be able to determine when, why, and how statistical tests are used.
4. Familiarize students with a statistical software package to perform analyses of quantitative data.
5. Foster the ability to critically engage with media or scientific reports of studies based on quantitative data.

Required Texts
1. Agresti, Alan. 2018. *Statistical Methods for the Social Sciences: Fifth edition.* Pearson.
2. De Vries, Andrie and Joris Meys. 2015. *R for Dummies*. John Wiley & Sons, Inc.

3. Wickham, Hadley and Garrett Grolemund. 2017. *R for Data Science: Import, Tidy, Transform, Visualize, and Model Data.* O'Relly Media Inc. **Online version available here: https://r4ds.had.co.nz/**
4. Grolemund, Garrett. 2014. *Hands-on Programming with R.* O'Relly Media Inc. **Online version available here: https://rstudio-education.github.io/hopr/**

Other References
Healey, Joseph F. Statistics: A Tool for Social Research.

Computing
There are several powerful statistical packages. For the purpose of this class, we will use R. There are a tremendous amount of R tutorials and learning resources online. Some helpful links can be found below:
UCLA: https://stats.idre.ucla.edu/r/
Quick-R https://www.statmethods.net/index.html
YaRrr! The Pirate's Guide to R https://bookdown.org/ndphillips/YaRrr/
ListenData https://www.listendata.com/p/r-programming-tutorials.html

Additional learning resources for R will be posted as we go through the course.
Please note that the TA is your go-to person for all R-related questions. The instructor can also help when the TA is unavailable.

COVID-Related Health Policies
In order to protect the health and well-being of all members of the University community, **masks must be worn by all persons on campus** when in the presence of others (within six feet) and in buildings in non-private enclosed settings (e.g., common workspaces, workstations, meeting rooms, classrooms, etc.). Masks must be worn during class meetings; any student not wearing a mask will be asked to leave.

Masks should conform to CDC guidelines and should completely cover the nose and mouth:
https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/about-facecoverings.html

Each day before you arrive on campus or leave your residence hall, you must complete the brief survey on the My Campus Pass symptom checker self-screening app.

Diversity Statement
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, 3 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.

Course Requirements
1. **Readings:** Readings from the text will be assigned each week, and it is expected that you will have completed the reading before the class. I recommend that you read the assigned material once before class and then again afterward. It is also strongly encouraged that you complete the selected review exercises at the end of each chapter to ensure that you've understood the concepts. Learning statistics requires lots of practice.
2. **Class participation:** Students are expected to attend every class and participate fully in class discussions and exercises. Your participation in these activities constitutes **10%** of your final grade.
3. **Assignments:** You will complete 5 assignments independently. The problems will include applying statistical theory, conducting analysis using R, and interpreting the results. You are encouraged to ask other students or the TA for help, but copying other people's work is not acceptable. Assignments are due on Canvas and they constitute **25%** of your final grade.
4. **Exams**: There will be two <u>in-class</u> exams addressing the concepts covered in the course. Each exam is worth 15% of your final grade (**30%** in total).
5. **Final paper and presentation:** You will turn in a one-page proposal of your topic in November (see instructions at the end of the syllabus) and a final research paper of approximately10 pages (double spaced with Times New Roman, 12-point font size, with 1-inch margins, excluding references, figures, and tables) by the end of the semester**.** The last class will be devoted to oral reports of the findings from your projects. Each class member will be allotted approximately 15 minutes toward this end. The final paper and presentation account for **35%** of your grade.

The primary goal of the paper is for you to learn the statistical methods and if you can couple that with advancing your interest in a relevant sociological subject, even better! You will need to analyze a secondary data set for the final project. I would recommend that you look into data sets such as the General Social Survey (GSS), the World Value Survey (WVS), the National Health Interview Survey (NHIS), and the Panel Study of Income Dynamics (PSID). <u>I will provide the GSS data and 2013 Global Dataset on our Canvas site.</u> Feel free to come to speak with me or the TA if you need help with finding an appropriate data set.

The final paper is anticipated to have at least the following sections:
1) Introduction: a concise description of the topic, why it is worth pursuing, and what the contribution is.
2) Background and hypotheses: a clear presentation of research hypotheses to be tested and the theoretical rationale for each hypothesis.
3) Research methods: a description of your data source(s), the operationalization of variables, and a clear explanation of the analytic strategy.
4) Analytic results: a systematic presentation of your analysis, including, but not limited to, descriptive statistics, bivariate analysis, and multivariate analysis results. Careful construction of tables and figures, with a discussion and interpretation of each, is important here.
5) Conclusions and discussion: a discussion that reviews key findings vis-à-vis the research hypotheses, considers scholarly and policy implications and takes note of the limitations of your study.

I strongly suggest that you select a single, continuous dependent variable, several (perhaps 3 or 4) continuous independent variables, and at least one categorical (or discrete) independent variable. You should have some theoretical reasons for the relationships between the independent and dependent variables. However, you are not required to review the available literature extensively on the topic. Specifying clearly the hypotheses to be tested will be fine. You will want to begin the analytical section of the paper with relatively simple techniques, e.g., descriptive statistics (including frequency distributions, means, standard deviations, etc.), bivariate correlations, the difference of means tests, etc.; then work the way up to more sophisticated procedures, e.g., multiple regression with dummy explanatory variables. Please feel free to consult with me as needed.

Course Schedule

Week 1   **September 7**
Topic: Introduction to statistics & Syllabus Walkthrough
          R lab: Review R basics, data reading and writing
Reading: Agresti, Chapter 1
          R for Dummies, Chapters 1–4
          Hands on Programming with R, Chapters 4–7

Week 2   **September 14**
Topic: Basic concepts and exploring data
          R lab: Cleaning and preparing data
               Frequency distributions, bar chart, pie chart
Reading: Agresti, Chapter 2
          R for Dummies, Chapters 5 and 7
          "In Calculation of Military Rates, the Numbers are not all
          Straightforward", *New York Times*, May 16, 2013

Week 3   **September 21**
Topic: Measures of central tendency and meausres of dispersion
          R lab: Use appropriate statistics to describe variables
               Graphics, histogram, density plot and box plot
Reading: Agresti, Chapter 3
          R for Dummies, Chapters 14 and 8
          "When Numbers Mislead", *New York Times*, May 25, 2013
          <u>***Assignment 1 due***</u>

Week 4   **September 28**
Topic: Sampling, probability and the normal distribution
          R lab: Area under the normal curve, more on data preparation
Reading: Agresti, Chapter 4
          R for Data Science, Chapter 3
          <u>***Assignment 2 due***</u>

Week 5   **October 5**
Topic: Basis of statistical inference (point estimates and confidence intervals)
          R lab: Confidence intervals
Reading: Agresti, Chapter 5
          R for Dummies, Chapters 10 and 11

Week 6   **October 12**
Topic: Hpothesis testing and significance test
          R lab: one sample test of means and proportions
Reading: Agresti, Chapter 6
          Baker, M. 2016. "Statisticians issue warning on p-values." *Nature* 531:
          151

R for Dummies, Chapter 12
Assignment 3 due

Week 7   **October 19**
                    Exam 1

Week 8   **October 26**
Topic: Comparing two groups using t-test
                    R lab: Statistics by groups, density graph by groups, t-tests
Reading: Agresti, Chapter 7
                    R for Dummies, Chapter 15 (pages 299-310)
                    R for Dummies, Chapter 13 (pages 239-251)

Week 9   **November 2**
Topic: Analysis of variance (comparing several means)
                    R lab: Statistics by three or more groups,
                           Density graph and boxplot by groups, ANOVA
Reading: Agresti, Chapter 12.3 and 12.4
                    R for Dummies, Chapter 15, pages 313-318
                    One-page proposal due

Week 10   **November 9**
Topic: Cross-tabulation and Chi-squared test
                    R lab: Crosstabulation, add margins and proportions to tables,
                           Chi-squared test
Reading: Agresti Chapter 8
                    R for Dummies, Chapter 15, pages 310-312
                    Assignment 4 due

Week 11   **November 16**
                    Exam 2

Week 12   **November 23**
Topic: Correlation and regression
                    R lab: Scatterplot, correlation coefficient, and OLS regression model
Reading: Agresti, Chapter 9
                    R for Dummies, Chapter 15, pages 318-324

Week 13   **November 30**
Topic: Multiple Regression and Causality
                    R lab: multiple regression, post-regression predictions,
                           access modeling results and make publishable tables
Reading: Agresti, Chapter 10 and 11
                    Assignment 5 due

Week 14   **December 7**

Topic: Final presentation

Final paper is due on December 15 11:59 pm.

Advice for Succeeding in the course
Advice given to students (Inspired by the A. Rossman's Instructor Notes to *Workshop Statistics*):

1. Ask Questions!
2. Use all your resources (the book, the labs, the lectures, class discussion, office hours).
3. Continually review and integrate the material throughout the semester - don't fall behind or get overconfident. Don't leave early topics behind, but build on them.
4. Work together, ask questions of me and other students.
5. Practice the material in new situations.
6. Thoughtfully read the assigned sections before coming to class.
7. Always provide clearly written explanations of your answers.
8. Have fun.
9. Be responsible for your own learning.

Reading is fundamental, re-reading is inevitable, and even then ultimate comprehension is not guaranteed. Other readings that supplement content covered in the textbooks or that clarify a particularly difficult or important concept will be distributed during class or make available via Canvas. Further, I will put several introductory statistics textbooks for social scientists on reserve in my office, just stop by and borrow one if you feel the need to see material addressed in another way. If you are struggling with the readings, talk to other students in the class, and then set up a time to talk with the teaching assistant and/or the professor.

One-page proposal
A one-page summary of your topic is due on November 2. The summary should appropriately address the following questions:

1. What is your research question? What is the central relationship you are interested in? Identify your outcome variable and the explanatory or independent variables. You want to identify 4-5 variables with which you'll work. You can select one outcome variable and 4-5 independent variables or you can select multiple outcome variables and 2-3 independent variables.

2. What hypotheses can you draw about the relationship? Be sure to link these hypotheses to sociological theory. What guides your thinking and leads to these hypotheses?

3. How will you operationalize your hypotheses? What data set will you use? Which variables will you use from the data set to test your hypotheses?

Please feel free to come to see me with any questions about this project. I'd like to see you early on in the semester, so that I can make sure you are on the right track with the paper.