

Rutgers, The State University of New Jersey  
School of Social Work  
Advanced Statistical Methods I: Applied Regression Analysis  
19:910:639, Spring 2023

**Class**

Instructor: **Lenna Nepomnyaschy**, Associate Professor

Email: [lennan@ssw.rutgers.edu](mailto:lennan@ssw.rutgers.edu)

Bio: <https://socialwork.rutgers.edu/faculty-staff/lenna-nepomnyaschy>

Time: Wednesdays, 1:30 – 4:10 pm

Space: 120 Albany St., Classroom 360

**Required Lab**

Instructor: **Addam Reynolds**, PhD Candidate

Email: [addam.reynolds@rutgers.edu](mailto:addam.reynolds@rutgers.edu)

Bio: <https://socialwork.rutgers.edu/admissions/phd-admissions/student-profiles>

Time: Wednesdays, 4:30 – 6:00 pm

Space: 120 Albany St., Classroom 360

**Class Website**

**Course Overview**

This course, the second of the statistics sequence in the Social Work PhD Program, will focus on generalized linear modeling and related advanced methods. Building and expanding on applied linear regression models in the 1<sup>st</sup> semester course, this course will introduce different functional forms of the dependent variable, including binary, categorical, ordered, censored, and others. This course will also introduce students to more advanced topics. The objective of this course is to help students understand and apply these models to their own research questions as well as to critically analyze quantitative findings from research in their areas of interest.

**Prerequisites**

This course assumes that students have taken a statistics course covering multiple linear regression such as Advanced Statistical Methods I: Applied Regression Analysis and have working knowledge the Stata statistical software package.

**Required Software**

This course requires that students learn and use the **Stata Statistical Software Package** for hands-on data analysis and statistics applications for class assignments.

- Stata is available for all employees (GRAs, staff, faculty) for free download from the OIRT software portal: <https://software.rutgers.edu/>
- Stata is available to all students in any Rutgers computer lab as well as the Doctoral Student Computer lab in the SSW Annex AND in the Virtual Computer Lab through Rutgers Libraries: <https://labgateway.rutgers.edu/>

- Students may also purchase their own version of Stata at discounted rates through the Rutgers Office of Instructional Technology  
<http://www.stata.com/order/new/edu/gradplans/gp-direct.html>

### **Textbooks**

There are no required textbooks to purchase for this course. All required readings will be available online or have PDFs provided.

### ***Recommended Resources and Texts***

Stock, J. & Watson, M. 2020. *Introduction to Econometrics, 4<sup>th</sup> ed.* Pearson Education. **Full text free online w/clickable chapters. Do not purchase:** <https://www.sea-stat.com/wp-content/uploads/2020/08/James-H.-Stock-Mark-W.-Watson-Introduction-to-Econometrics-Global-Edition-Pearson-Education-Limited-2020.pdf>

Pampel, F. C. 2000. *Quantitative applications in the social sciences: Logistic regression.* Thousand Oaks, CA: SAGE Publications, Inc. doi:10.4135/9781412984805. **Available through RU libraries.** <https://dx-doi-org.proxy.libraries.rutgers.edu/10.4135/9781412984805>

Gordon, R. A. (2012). *Applied statistics for the social and health sciences.* Routledge. **Available through RU Libraries.** <https://ebookcentral-proquest-com.proxy.libraries.rutgers.edu/lib/rutgers-ebooks/detail.action?pq-origsite=primo&docID=987910>

Long J. S., & Freese J. 2014. *Regression models for categorical dependent variables using Stata (3<sup>rd</sup> ed.).* College Station, TX: Stata Press. **This is considered a seminal book on non-linear dependent variables. It is somewhat advanced, but very useful.**

Mehmetoglu, M. & Jakobsen, T.G. 2016. *Applied Statistics Using Stata: A Guide for the Social Sciences.* Sage Publishing. **Not available online.**  
*Student Resources (access with Rutgers Netid):* <https://study-sagepub-com.proxy.libraries.rutgers.edu/mehmetogluandjakobsen/student-resources>

### ***More In-Depth Logistic Regression Texts***

Dunteman, G. H., & Ho, M. R. 2006. *Quantitative applications in the social sciences: An introduction to generalized linear models.* Thousand Oaks, CA: SAGE Publications, Inc. doi:10.4135/9781412983273. **Available through RU libraries**  
<https://dx-doi-org.proxy.libraries.rutgers.edu/10.4135/9781412983273.n1>

Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. 2013. *Applied logistic regression (Vol. 398).* John Wiley & Sons. **Available through RU libraries**  
<https://ebookcentral-proquest-com.proxy.libraries.rutgers.edu/lib/rutgers-ebooks/reader.action?docID=1138225>

James, J. 2001. *Interactions between Qualitative Predictors: Interaction effects in logistic regression.* Thousand Oaks, CA: SAGE Publications, Inc. **Available through RU libraries**  
<https://dx.doi.org/10.4135/9781412984515>

Menard, S. 2002. *Quantitative applications in the social sciences: Applied logistic regression analysis*. Thousand Oaks, CA: SAGE Publications, Inc. doi:10.4135/9781412983433. **Available through RU libraries.** <https://dx-doi-org.proxy.libraries.rutgers.edu/10.4135/9781412983433>

### ***More In-Depth Applied Regression & Introductory Econometrics Texts***

Gelman, A., Hill, J. & Vehtari, A. 2021. *Regression and Other Stories*. Cambridge University Press. (This book is **good**, somewhat advanced, but could be a primary resource as you go forward. It is fully available for free online') <https://users.aalto.fi/~ave/ROS.pdf>

Wooldridge, J. 2006. *Introductory Econometrics: A Modern Approach*, 3<sup>rd</sup> edition. Mason, OH: Thompson. (*The bible of econometrics*).

### ***Writing about Quantitative Analysis***

Miller, Jane E., 2013. *The Chicago Guide to Writing about Multivariate Analysis (2<sup>nd</sup> Edition)*. The Chicago Guides to Writing, Editing, and Publishing. University of Chicago Press.

Study guide: <http://www.press.uchicago.edu/books/miller/multivariate/index.html>

**THIS IS SUPER HELPFUL**

See Jane Miller's website for pdfs, videos and other material:

<http://policy.rutgers.edu/faculty/miller/>

### ***General Stata Books***

- Hamilton, L.C. 2006. *Statistics with Stata*, 6<sup>th</sup> edition. Cengage.
- Long, S.J. 2009. *The Workflow of Data Analysis Using Stata*. College Station, TX: Stata Press.
- Kohler, U. & Kreuter, F. 2009. *Data Analysis Using Stata* (2<sup>nd</sup> ed). College Station, TX: Stata Press.
- Mitchell, M. 2010. *Data Management Using Stata: A Practical Handbook*. College Station, TX

### **Course Requirements**

All course materials, including the course syllabus, readings, and assignment instructions are posted on the course **CANVAS site**. **Students are expected to submit all assignments through the Canvas course site**. In addition, students will receive e-mails and announcements from the instructor through the course Canvas site.

### **Course Assignments & Grading**

Students' work will be evaluated based on the following course requirements (detailed instructions to follow).

#### Attend research seminars or presentations (2 times): 5%

Students will attend (virtually or in person) **TWO** (2) research seminars or presentations **over the course of the semester** that focus on quantitative empirical research in their substantive area of interest. These may be at the SSW, at other schools or departments at RU, at other Universities, or through research think tanks, organizations, conferences, etc. They will write up and submit a short description of what they learned and provide a link to the presentation publicity, email, or

notification. These will be graded on a 3-point scale: 0= no submission; 1= incomplete submission; 2= submitted with required elements.

Submit published empirical quantitative research papers (2): 5%

Students will submit at least **TWO (2)** published quantitative research papers in their substantive area of interest, which may be selected as class readings by the professor. ***One to be submitted by week 5 and one to be submitted by week 10 of the semester.*** The submission will include a very brief summary of what the article found and why this is of interest. These will be graded on a 3-point scale: 0= no submission; 1= late or incomplete submission; 2= submitted on time and with required elements.

Weekly reading responses (12 weeks): 20%

Students will submit brief responses to the required readings each week via the Canvas course site. Responses can be very brief (couple of sentences) describing any questions, thoughts, or ideas that the readings brought up for you. Responses must informally cite (author) the reading that they are referencing. Reading responses will be due by 8am, the morning prior to class: ***Wednesdays by 8am.*** These will be graded on a 3-point scale: 0= no submission (or not submitted prior to start of class); 1= late (but prior to class) or incomplete submission; 2= submitted on time and with required elements.

Weekly Class Activities: 30%

On specified weeks (approx. 7) students will bring statistical output (tables, figures, etc.) from their own work to class for group work. They will present their output to peers and work together to interpret, understand, and critique each others' work. Submission of output will be due by noon on Wednesday so that it can be printed and distributed to classmates: ***Wednesdays by noon.***

On other weeks, students will work together to understand, interpret, and critique published (or in progress) papers provided by the professor, submitted by other students, or presented by guest speakers.

Final Assignment: 40%

- Presentation of Final Assignment (5%) - Students will present a working draft of their final assignment to the class in a PowerPoint presentation in order to receive feedback and provide feedback for peers (10-15-minute presentations). ***Week 14 and/or 15***
- Final Written Assignment: (35%) - The final assignment will consist of a complete data analysis project building on linear models from the 1<sup>st</sup> semester and adding the modeling techniques from this semester. Analyses will include descriptive results and estimation of various multiple regression models, including models with continuous, binary, and perhaps other forms of the dependent variable, and models with interaction effects. The written assignment will take the form of the *Methods, Results, and Discussion* sections of a journal-style quantitative empirical paper. Students will describe their data, sample, measures, and analytic strategy, describe the sample characteristics, interpret results from all of the regression models estimated, and provide a summary of their findings, a discussion of the rationale for their different modeling strategies and the trade-offs of

different techniques, and identify the limitations of their analyses related to violations of regression assumptions and other sources of bias. **Monday, May 8.**

***It goes without saying that students are expected to attend every class, come to class on time, remain in class for the entire session, and to be prepared for class having read the required readings and submitted the required materials (if something is due). While there is not specific course credit associated with attendance, absences, being late to class, and lack of participation and preparation will substantially impact students' overall grades.***

### **Grading**

Grade cut-offs are as follows (scores of 0.5 and above will be rounded up):

A	92-100
B+	87-91
B	82-86
C+	77-81
C	70-76
F	0-69

### **Academic Integrity**

As per Rutgers University Academic Integrity Policy, “Students are responsible for understanding the principles of academic integrity and abiding by them in all aspects of their work at the University. Students are also encouraged to help educate fellow students about academic integrity and to bring all alleged violations of academic integrity they encounter to the attention of the appropriate authorities.” All SSW students are expected to review and familiarize themselves with the [RU Academic Integrity Policy](#) in its' entirety.

All work submitted by students may be run through the Turnitin plagiarism detection software at the professors' discretion.

## **Detailed Course Outline**

**Please note:** Readings & topics may change as we move forward. Please note announcements and emails for changes. **ALSO:** Some of the more technical readings include formulas, equations, and mathematical proofs. I encourage you to take a look at these, **BUT**, do not worry about fully understanding the details. **ALWAYS focus on the substantive content in these articles.**

### **Week 1: January 18, 2023**

**Topic: Intro to course, topics, assignments & debrief about prior semester**

*Class Activity: Discuss possible topics for papers going forward*

### **Required Readings**

- Read over your final assignments, make sure to note any issues/mistakes/questions raised.
- Be prepared to discuss next steps as you go forward

### **Week 2: January 25, 2023**

**Topic: Missing data, why it's a problem and what can be done**

*Class Activity: Students bring output showing missing data in their work (no submission), discussion of missing data in students' own work and how to address*

### **Required Reading:**

- Newman, D. A. (2014). Missing data: Five practical guidelines. *Organizational Research Methods*, 17(4), 372-411.  
<https://journals.sagepub.com/doi/full/10.1177/1094428114548590>
- Gorard, S. (2020). Handling missing data in numeric analyses. *International Journal of Social Research Methodology*, 23(6), 651-660. <https://web-p-ebshost-com.proxy.libraries.rutgers.edu/ehost/pdfviewer/pdfviewer?vid=0&sid=fe82b5d2-07b5-48d1-8a89-df98b4807ea4%40redis>

### **Suggested Readings:**

- Stuart, E. et al. 2009. Multiple Imputation with Large Data Sets: A Case Study of the Children's Mental Health Initiative. *American Journal of Epidemiology* 169(9):1133-1139.
- Allison, Paul. Missing Data. Handbook. [https://statisticalhorizons.com/wp-content/uploads/Allison\\_MissingData\\_Handbook.pdf](https://statisticalhorizons.com/wp-content/uploads/Allison_MissingData_Handbook.pdf)
- Johnson, D. & Young, R. 2011. Toward Best Practices in Analyzing Datasets with Missing Data: Comparisons and Recommendations. *Journal of Marriage and Family* 73: 926-945.
- Young, Rebekah, and David Johnson. 2013. Methods for handling missing secondary respondent data. *Journal of Marriage and Family* 75(1): 221-234.

### **Week 3: February 1, 2023**

**Topic: Power analysis, what is it & how you do it**

Guest speaker: Addam Reynolds

*Please download this free software package prior to class:*

<https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower>

*Class Activity: Discuss Reynolds et al. paper; Run power analysis exercise*

#### **Required Readings:**

- Introduction to Power Analysis. UCLA Statistical Methods and Data Analytics. <https://stats.oarc.ucla.edu/other/mult-pkg/seminars/intro-power/>
- Reynolds, A., Greenfield, E. A., Moorman, S., & Reyes, L. (2022). Race, Childhood Socioeconomic Status, and Region of Childhood Residence as Intersectional Life-Course Predictors of Cognitive Aging in the United States. *Innovation in Aging*, 6(3). <https://academic-oup-com.proxy.libraries.rutgers.edu/innovateage/article/6/3/igac020/6563354>

### **Week 4: February 8, 2023**

**Topic: Binary/dichotomous Outcomes, cross-tabs, probability, chi-square**

*Class Activity: Cross-tabs & probability group work w/examples provided*

#### **Required Readings - TBD**

- [The SAGE Encyclopedia of Communication Research Methods](#). Motel, Laura. 2018: [Odds Ratios \(and probability\) PDF](#)
- [Cross-Tabulation, Boston University School of Public Health](#) - online tutorial, very clear and brief.

#### **Suggested Readings:**

### **Week 5: February 15, 2023**

**Topics: Linear probability models, interpretations, pros & cons & Introduction to Logistic Regression**

*Class Activity: Students submit cross-tab output for group work*

#### **Required Readings: TBD**

Read **ONE** of these:

- Stock & Watson (2020). Chapter 11.1. Binary Dependent Variables and the Linear Probability Model (p. 392-397 ONLY)

- Wooldridge Chapter 7.5 A Binary Dependent Variable: The Linear Probability Model (p. 248-253 ONLY)

**AND this one:**

- Pampel, Fred. 2022 (2nd edition). Logistic Regression: A Primer, Ch. 1: The Logic of Logistic Regression (p. 2-24).

### **Week 6: February 22, 2023**

**Topic: Logistic regression**

*Class Activities:*

- *Guest Speaker Sarah Gold, PhD (alum of our PhD program), to present & discuss her paper with class and talk about her career trajectory*
- *Students submit linear probability models for group work; Cross-tabs and odds ratios group work*

### **Required Readings: TBD**

- Gold, S., & Nepomnyaschy, L. (2018). Neighborhood physical disorder and early delinquency among urban children. *Journal of Marriage and Family*, 80(4), 919-933.
- IF you did not read this chapter last week; MAKE SURE TO READ THIS WEEK: Pampel, Fred. 2022 (2nd edition). Logistic Regression: A Primer, Ch. 1: The Logic of Logistic Regression (p. 2-24).

### **Suggested Readings:**

### **Week 7: March 1, 2023**

**Topic: Logistic Regression Continued & Marginal effects**

*Class Activity: Students submit logistic regression models for group work*

### **Required Readings:**

- Pampel, F. 2022. Logistic Regression: A Primer
  - Chapter 2: Interpreting Logistic Regression Coefficients. This is an excellent and very accessible presentation of logistic regression with the most updated issues. Make sure to read about logged odds, predicted probabilities, and marginal effects. These are key!
  - Chapter 3: Estimation & Model Fit. this is complex, just read for substance and skim rest.

### **Week 8: March 8, 2023**

**Topic: Maximum likelihood, probit, marginal effects (continued)**



*Class Activities:*

***Students submit output comparing results from LPM, logit, and probit models (w/marginal effects) for group work***

**Required Readings:**

- Mood, C. (2010). Logistic regression: Why we cannot do what we think we can do, and what we can do about it. *European sociological review*, 26(1), 67-82.
- Pampel, Fred. 2022. Logistic Regression: A Primer. Chapter 4: Probit Analysis (read for substance, skim the rest).

**SPRING BREAK: MARCH 15, 2023 – NO CLASS**

**Week 9: March 22, 2023**

**Topic: Interactions w/ binary outcomes**

*Class Activity: Guest speaker, Alex Haralampoudis, PhD Candidate, RUSSW, to present & discuss her dissertation paper*

***Students submit output comparing results from LPM, logit, and probit models (w/marginal effects) for group work***

**Required Readings: TBD**

- Haralampoudis, A. 2022. State Income Tax Burden and Household Food Insecurity. In progress, dissertation paper #2.
- Mize, T. D. (2019). Best practices for estimating, interpreting, and presenting nonlinear interaction effects. *Sociological Science*, 6, 81-117.
- Norton, E. C., Wang, H., & Ai, C. (2004). Computing interaction effects and standard errors in logit and probit models. *The Stata Journal*, 4(2), 154-167.

**Week 10: March 29, 2023**

**Topic: Continue interactions with binary outcomes**

*Class Activity: Students submit output of binary interaction models & figures for group work*

***Students submit published paper in areas of interest #2***

**Required Readings: TBD**

- Sample papers
- Make sure to have read last week's papers:

- Mize, T. D. (2019). Best practices for estimating, interpreting, and presenting nonlinear interaction effects. *Sociological Science*, 6, 81-117.
- Norton, E. C., Wang, H., & Ai, C. (2004). Computing interaction effects and standard errors in logit and probit models. *The Stata Journal*, 4(2), 154-167.

**Week 11: April 5, 2023**

**Topic: Multinomial (polytomous or categorical) outcomes**

*Class activity: Discussion of published papers*

**Required Readings: TBD**

- Pampel, Fred. 2022. Logistic Regression: A Primer. Chapter 5: Ordinal & Multinomial Logistic Regression (**p. 20-37 ONLY**)
- Viruell-Fuentes, Edna, Morenoff, Jeffrey, Williams, David, & James House. 2011. Language of Interview, Self-Rated Health, and the Other Latino Health Puzzle. *American Journal of Public Health*, 101(7): 1306-1313. (multinomial logit, predicted probabilities)

**Week 12: April 12, 2023**

**Topic: Other discrete outcomes: count, ordered, truncated, censored, etc.**

*Class Activity: Discuss published papers*

**Required Readings: TBD**

- Pampel, Fred. 2022. Logistic Regression: A Primer. Chapter 5: Ordinal & Multinomial Logistic Regression (**p. 1-19 ONLY**)
- Bell, Janice & Frederick Zimmerman. 2010. Shortened Nighttime Sleep Duration in Early Life and Subsequent Childhood Obesity. *Archives of Pediatric and Adolescent Medicine*, 164:9(840-845). (ordered logit)

**Week 13: April 19, 2023**

**Topic: Comparing across models, trade-offs, pros & cons**

*Class activity: Students bring output comparing different analytic techniques for group work*

**Required Readings: TBD**

**Week 14: April 26, 2023**

**STUDENT PRESENTATIONS OF FINAL ASSIGNMENT**

**Week 15: May 3, 2023**

**Topic: Wrap up & Debrief & Intro to Advanced Topics**

**Required/Suggested Readings: TBD**

**FINAL ASSIGNMENT DUE, Monday, May 8, 2023**