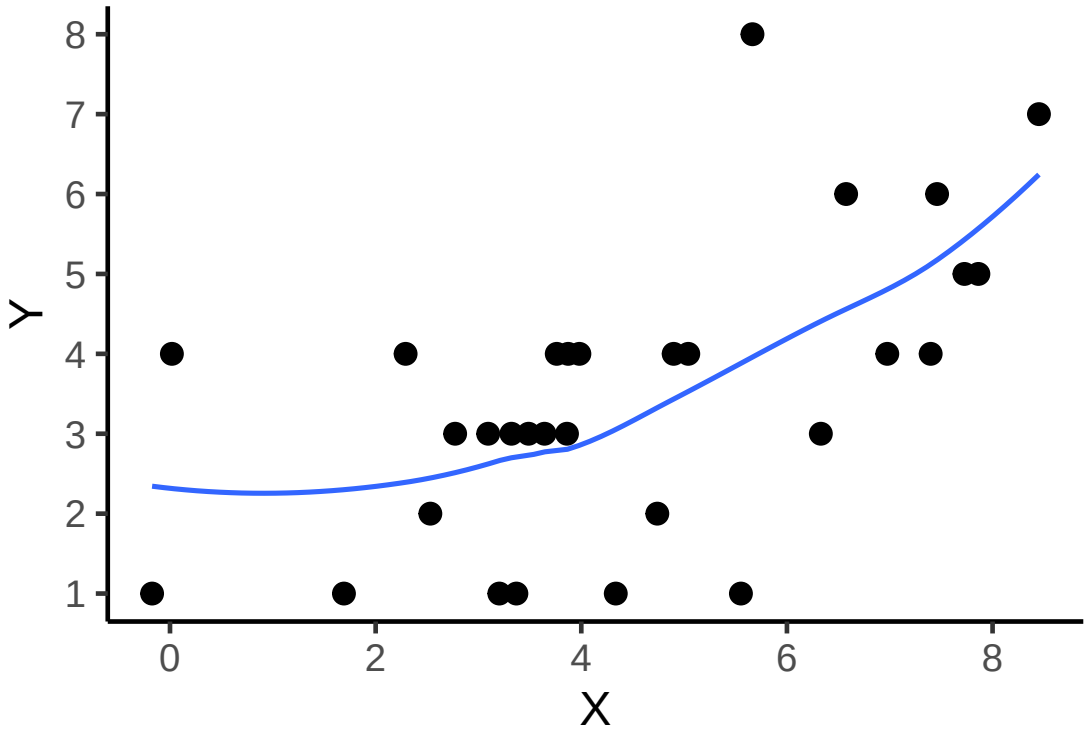
Categorical Data Analysis
Fall 2020 Syllabus

Stefany Coxe



Course Information

- PSY 5939 U04
- Hybrid with synchronous Zoom meetings on Thursdays, 9:30am - 12:15pm

Instructor Information

- Stefany Coxe, Ph.D.
- Email: stefany.coxe@fiu.edu
- Office: DM 275
- Office Hours: By appointment – send me an email to schedule a time

Course description

This course covers topics related to statistical analysis of categorical outcome variables, focusing on methods used in the social sciences. Topics include the generalized linear model (GLiM, including logistic regression and Poisson regression) and repeated measures extensions of GLiM (such as GEE and generalized linear mixed models). You will analyze, interpret, and write up results using these methods.

Learning objectives

- Develop research questions about categorical outcome variables
- Analyze categorical outcomes with regression-based statistical models appropriate to the research question
- Interpret statistical analysis output from common statistical software packages
- Write up the results of these analyses for publication

Prerequisites

Graduate coursework in analysis of variance and linear regression. We will cover a variety of topics in this course, but all of them build on a basic ANOVA and regression (general linear model) framework. A course covering multivariate statistics (such as PSY 5246C) is generally required.

Software

We will use SPSS, SAS, and R in this course. I expect you to be able to use at least one of these software packages to do things like open datasets, transform variables, conduct simple analyses, etc. I will provide information about the specific procedures you will need to know for this course.

Canvas

Lecture notes, lecture recordings, computer code, assignments, and all other materials will be posted on Canvas.

Assignments and grading

Homework (5 assignments - 30% of grade)

Your first homework assignment will be to find a dataset that interests you that includes categorical variables. In particular, you want to think about research questions for which the categorical variable can serve as an **outcome**.

The remaining four homework assignments will cover the major topic areas of the course: (1) models for binary outcomes, (2) models for 3 or more categories , (3) models for count outcomes, and (4) contingency tables. The assignments involve running some analyses, interpreting output, and presenting the results in tables/figures and text.

Quizzes (6 quizzes - 15% of grade)

Short multiple choice quizzes to assess your understanding of the material in the recorded lectures. (So please watch the lecture *first*!) You can attempt each question as many times as you need to in order to get it correct – please use this as a learning tool.

Discussions (6 discussions - 20% of grade)

There will be regular Canvas-based discussions of readings and other class material. These are designed to get you to think about the topics in more detail. In particular, they will focus on the application side of things – how these analysis methods translate to *your* research.

Final project

The final project is a self-directed project to develop research questions involving categorical outcome variables, conduct the appropriate analyses to answer those questions, write up the results in a **brief report**, present the results to your peers in a **short presentation**, and carry out a **short discussion** of your presentation with your peers.

Presentation (10% of grade)

A short presentation about your final project. The main purpose of this presentation is to give you **practice presenting your analysis findings**. Approximately 15 to 20 minutes. You will record your presentation (on Zoom or similar) and upload to Canvas by 8pm on Sunday, October 4, 2020.

Presentation discussion (5% of grade)

Each student should ask a question of at least 2 other students about their presentations. The original student should attempt to answer the questions. (Feel free to have further discussion as well!) Discussions should be complete by 8pm on Wednesday, October 7, 2020.

Final paper (20% of grade)

The final paper will be a brief report of your research question(s), your analysis plan, your actual analyses, and your conclusions. A major focus is **developing research questions involving categorical outcomes** and **mapping them onto the appropriate analyses**. The final paper is due on Canvas on Friday, October 9, 2020 by end of day (midnight).

Grade distribution

A	A-	B+	B	B-	C+	C	F
>= 93	90 - 92.99	87 - 89.99	83 - 86.99	80 - 82.99	77 - 79.99	70 - 76.99	<= 69.99

Tentative Course Outline

Week of...	Topic
24 August	Introduction and generalized linear model (GLiM)
31 August	Binary outcomes: logistic regression
07 September	3+ categories: ordinal and multinomial logistic regression
14 September	Count outcomes: Poisson regression family
21 September	Contingency tables
28 September	Mixed models for categorical outcomes
05 October	Presentation, discussion, final paper

Due dates subject to change due to hurricane, emergency, scheduling changes, etc.

Each week....

- Monday: Lecture available by end of day
- Tuesday: NA
- Wednesday: Quiz is due by 8pm
- Thursday: Zoom meeting for code and interpretation from 9:30am to 12:15pm
- Friday: Discussion is due by 8pm
- Saturday: NA
- Sunday: Homework assignment is due by 8pm

Final project...

- Presentation is due by 8pm on Sunday, October 4, 2020
- Presentation discussion is due by 8pm on Wednesday, October 7, 2020
- Final paper is due on Friday, October 9, 2020 by end of day (midnight)

Extended Reading list

Linear regression and GLiM

Agresti, A. (2003). Categorical data analysis (Vol. 482). John Wiley & Sons.

Agresti, A. (2018). An introduction to categorical data analysis. John Wiley & Sons.

Cohen, J., Cohen, P., West, S.G. & Aiken, L.S. (2003). Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences. L. Erlbaum Associates, Mahwah, N.J.

Fox, J. (2015). Applied regression analysis and generalized linear models. Sage Publications.

Gelman, A., & Hill, J. (2006). Data analysis using regression and multilevel/hierarchical models. Cambridge University Press.

Long, J. S. (1997). Regression models for categorical and limited dependent variables (Vol. 7). Advanced quantitative techniques in the social sciences, 219.

Proportions, binary outcomes, and ordered categories

Allison, P. D. (2012). Logistic regression using SAS: Theory and application. SAS Institute.

Bürkner, P. C., & Vuorre, M. (2019). Ordinal regression models in psychology: A tutorial. Advances in Methods and Practices in Psychological Science, 2(1), 77-101.

Chen, K., Cheng, Y., Berkout, O., & Lindhiem, O. (2016). Analyzing Proportion Scores as Outcomes for Prevention Trials: A Statistical Primer. Prevention Science, 1-10.

DeMaris, A. (2002). Explained variance in logistic regression: A Monte Carlo study of proposed measures. Sociological Methods & Research, 31(1), 27-74.

Hayes, A. F., & Matthes, J. (2009). Computational procedures for probing interactions in OLS and logistic regression: SPSS and SAS implementations. Behavior research methods, 41(3), 924-936.

Menard, S. (2002). Applied logistic regression analysis (No. 106). Sage.

Counts and frequencies

Atkins, D. C., & Gallop, R. J. (2007). Rethinking how family researchers model infrequent outcomes: a tutorial on count regression and zero-inflated models. Journal of Family Psychology, 21(4), 726.

Blevins, D. P., Tsang, E. W., & Spain, S. M. (2015). Count-Based Research in Management Suggestions for Improvement. Organizational Research Methods, 18(1), 47-69.

Coxe, S., West, S. G., & Aiken, L. S. (2009). The analysis of count data: A gentle introduction to Poisson regression and its alternatives. Journal of personality assessment, 91(2), 121-136.

Gardner, W., Mulvey, E. P., & Shaw, E. C. (1995). Regression analyses of counts and rates: Poisson, overdispersed Poisson, and negative binomial models. Psychological bulletin, 118(3), 392.

Green, J. (2020). A tutorial on modelling health behaviour as count data with Poisson and negative binomial regression.

Land, K. C., McCall, P. L., & Nagin, D. S. (1996). A comparison of Poisson, negative binomial, and semiparametric mixed Poisson regression models with empirical applications to criminal careers data. Sociological Methods & Research, 24(4), 387-442.

Interactions in GLiMs

Ai, C. & Norton, E. C. (2003). Interaction terms in logit and probit models. Economics Letters, 80 (1), 123–129. doi:10.1016/S0165-1765(03)00032-6

McCabe, C., Halvorson, M. A., King, K. M., Cao, X., & Kim, D. S. (2020, April 8). Estimating and interpreting interaction effects in generalized linear models of binary and count data. https://doi.org/10.31234/osf.io/th94c

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.

Non parametric statistics

Daniel, W. W. (1990). Applied nonparametric statistics. Duxbury.

Gibbons, J. D. (1993). Nonparametric statistics: An introduction (No. 90). Sage.

Siegel, S., & Castellan, N. J. (1988). Non Parametric Statistics for the Behavioral Sciences. McGraw-Hill.

Contingency tables

Bradley, D. R., Bradley, T. D., McGrath, S. G., & Cutcomb, S. D. (1979). Type I error rate of the chi-square test in independence in R x C tables that have small expected frequencies. Psychological Bulletin, 86(6), 1290.

Camilli, G., & Hopkins, K. D. (1978). Applicability of chi-square to 2 x 2 contingency tables with small expected cell frequencies. Psychological Bulletin, 85(1), 163.

Clusters and repeated measures (GEE and GLMM)

Aiken, L. S., Mistler, S. A., Coxe, S., & West, S. G. (2015). Analyzing count variables in individuals and groups: Single level and multilevel models. Group Processes & Intergroup Relations, 18(3), 290-314.

Hedeker, D. (2005). Generalized linear mixed models. Encyclopedia of statistics in behavioral science.

Hu, F. B., Goldberg, J., Hedeker, D., Flay, B. R., & Pentz, M. A. (1998). Comparison of population-averaged and subject-specific approaches for analyzing repeated binary outcomes. American Journal of Epidemiology, 147(7), 694-703.

Hubbard, A. E., Ahern, J., Fleischer, N. L., Van der Laan, M., Lippman, S. A., Jewell, N., Bruckner, T., & Satariano, W. A. (2010). To GEE or not to GEE: comparing population average and mixed models for estimating the associations between neighborhood risk factors and health. Epidemiology, 21(4), 467-474.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. Journal of memory and language, 59(4), 434-446.

Leckie, G., Browne, W. J., Goldstein, H., Merlo, J., & Austin, P. C. (2020). Partitioning Variation in Multilevel Models for Count Data. Psychological Methods.

Stroup, W. W. (2012). Generalized linear mixed models: modern concepts, methods and applications. CRC press.

Course and University Policies

Attendance

If you need to miss class for a good reason (such as illness, religious event, professional activity, or university-sanctioned event), please contact me as soon as possible to make any necessary arrangements.

Special Needs

Any student with a disability or other special need that may require special accommodations for this course should make this known to the instructor during the first week of class.

Information	Disability Resource Center
Website:	http://drc.fiu.edu
Email:	drcupgl@fiu.edu
Office:	Graham Center 190
Phone:	305-348-3532

Academic Misconduct

Students at Florida International University are expected to adhere to the highest standards of integrity in every aspect of their lives. Honesty in academic matters is part of this obligation. Academic integrity is the adherence to those special values regarding life and work in an academic community. Any act or omission by a student which violates this concept of academic integrity shall be defined as academic misconduct and shall be subject to the procedures and penalties set forth herein. All students are expected to adhere to a standard of academic conduct, which demonstrates respect for themselves, their fellow students, and the educational mission of the University. All students are deemed by the University to understand that if they are found responsible for academic misconduct, they will be subject to the Academic Misconduct procedures and sanctions, as outlined in the Student Handbook.

Academic Dishonesty

Please refer to your student handbook for a description of what constitutes academic dishonesty. *I expect all students to complete and turn in their own work.*