Categorical Data Analysis

PC 419

Wednesday 9am – 11:45am

Instructor Stefany Coxe, Ph.D.

Office: DM 275
Office hours: by appointment
Email: stefany.coxe@fiu.edu
Website: http://faculty.fiu.edu/~scoxe

NOTE: Anything on this syllabus is subject to change at the Instructor's discretion.

Course Description

This course covers topics related to statistical analysis of categorical outcome variables, focusing on methods used in the social sciences. Topics include chi-square and other non-parametric methods for categorical outcomes, the generalized linear model (GLiM, including logistic regression, Poisson regression, and survival analysis), 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.

Prerequisite

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 highly recommended but it is not required.

Teaching Assistant

Our teaching assistant, Kelly Cromer, will be available as an additional resource. You can contact her via email: kcromer@fiu.edu

Software

We will use SPSS and SAS in this course. I expect you to be able to use one of these software packages (e.g., open datasets, transform variables, conduct simple analyses, etc.). I will teach specific procedures for this course.

Blackboard

Lecture notes, computer code, assignments, and other materials will be posted on Blackboard. You should have them available during class.

Textbook

Generalized Linear Models for Categorical and Continuous Limited Dependent Variables
Michael Smithson and Edgar C. Merkle
ISBN: 978-1466551732
Other readings: Additional helpful readings are listed at the end of the syllabus.

Assignments

Homework: 5 assignments, 12% each

Five homework assignments covering the five major topic areas of the course: (1) non-parametric tests, (2) two- and three-way contingency tables; (3) generalized linear models for binary, ordered, and unordered categories; (4) generalized linear models for rates and counts; and (5) models for clustered and repeated-measures categorical outcomes.

The assignments involve running several analyses, making some decisions based on the analyses, interpreting output, and presenting the results in tables/figures and text.

Assignment grading. Homework assignments will be graded as if they are journal submissions; that is, you will receive written feedback and grades of ACCEPT (100), MINOR REVISIONS (95), MAJOR REVISIONS (85), REJECT (75). You may re-submit up to two assignments for which you receive a REJECT grade. For the re-submission, you should provide the original submission, your revised submission, and a "letter to the editor" explaining where you erred and how you revised the errors. Please contact me if you want more information about the revision procedure.

Final Project: 5% for proposal, 10% for presentation, 25% for final paper

I will post several datasets with categorical variables to Blackboard, along with brief descriptions of each. You will propose a project using one of these datasets or your own data. This will culminate in a short paper. I want you to focus on developing research questions and mapping them on to appropriate analyses. More details to follow throughout the semester.

Proposal. 2 to 3 page proposal due March 22. Similar to a thesis or dissertation proposal, but shorter.

Presentation. Presentations on April 12 and 19. (If you are unable to present on one of these dates, please contact me ASAP to make sure you get scheduled appropriately.) I expect that your analyses should be complete (or nearly so) at this point; preparing the presentation should help you organize your thoughts to write the paper. The main purpose of this presentation is to give you practice presenting your analysis findings in a group setting. Approximately 10 minutes.

Paper. The final paper is due April 26.

Letter Grades

>= 93.00	A
90.00 - 92.99	A-
87.00 - 89.99	B+
83.00 - 86.99	B
80.00 - 82.99	B-
77.00 - 79.99	C+
73.00 - 76.99	C

Course and University Policies

Attendance and Late Policy

- I shouldn't have to tell you to attend every class. This is graduate school.
- Assignments are late if they are turned in after the due date and time. A 5 point late penalty will be deducted for each 24 hour period late — maximum score of 95/100 if 1 day late, maximum score of 90/100 if 2 days late, etc.
- Legitimate, verifiable cases of illness and emergencies, religious holy days, and conference travel can be accommodated. You need to contact me as soon as possible to make arrangements.

Drop Dates

Tuesday, January 17: Last day to drop courses or withdraw from the University without incurring financial liability for tuition and fees
Monday, March 20: Deadline to drop a course with a DR grade

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.

Disability Resource Center
Graham Center (GC) 190
(305) 348-3532
drcupgl@fiu.edu
drc.fiu.edu

Academic Misconduct

Florida International University is a community dedicated to generating and imparting knowledge through excellent teaching and research, the rigorous and respectful exchange of ideas, and community service. All students should respect the right of others to have an equitable opportunity to learn and to honestly demonstrate the quality of their learning. Therefore, 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.

NOTE: Anything on this syllabus is subject to change at the Instructor's discretion.

Tentative Course Outline

Date	Topics	Readings	Assignments
Date	Topics	Readings	Assignment due
Jan 11	Introduction		
Jan 18	Non parametric		
Jan 25	2 way cont		HW1
Feb 01	3 way cont		
Feb 08	GLIM	Chapter 1	HW2
Feb 15	GLIM	Chapter 2	
Feb 22	GLIM	Chapters 3 and 4	HW3
Mar 01	GLIM	Chapter 5	
Mar 08	Intro to GEE and GLMM	Chapter 8	HW4
Mar 15	SPRING BREAK		
Mar 22	GEE	Chapter 8	Proposal
Mar 29	GLMM	Chapter 8	
Apr 05	Censoring?	Chapter 7	HW5
Apr 12	Presentations		
Apr 19	Presentations		
Apr 26	FINALS WEEK		Final paper

Extended Reading list

These are additional resources if you want to learn more about a specific topic. I used many of these resources when developing the course.

Linear regression

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.

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 C tables that have small expected frequencies. Psychological Bulletin, 86(6), 1290.

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

Proportions, binary outcomes, and ordered categories

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

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.

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.

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.

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