

Applied Categorical Data Analysis
EdPsych/Psych/Soc 589
Fall 2018

Instructor: Carolyn J Anderson
e-mail cja@illinois.edu
office: rm 236C Education Bldg
office hours: Tuesday 1:15 – 3:15 pm

TA: Cory Castaneda
e-mail cjcasta2@illinois.edu
office: rm 230 Education Bldg
office hours: Monday 3:00 – 5:00pm

Lecture: 10:00–11:50 am, Tuesday & Thursday, rm 22 Education Bldg.

Prerequisites: EdPsych 581 or Psych 407. At the *minimum*, you should be familiar with basic concepts of data analysis, hypothesis testing, multiple regression, and ANOVA. You should have a math background through college algebra.

Course Objectives: Introduction to basic concepts and common statistical models and analyses for categorical data; Provides enough theory, examples of applications, and practice using categorical techniques so that students can use these methods in their own research, as well as critically read research papers that use such methods.

Course Web-page: You can find a link to the course web-page¹ at <http://education.illinois.edu/faculty-pages/cja>.

The web-page include copies of lecture notes, homework assignments, answer keys, SAS code, announcements, SAS & R example programs, and some handy programs.

We will use piazza.com for communication. You should receive an email about this. You can post questions, make suggestions, answer each others questions, etc. You can choose to be anonymous or not. I will monitor to ask questions and/or make suggestions. Whereas the course web-site is public, this site is not.

Required Text :

Agresti, A. (2007). *An Introduction to Categorical Data Analysis*, 2nd Edition. NY: Wiley.

- There is a (free) online version of the text at <https://mregresion.files.wordpress.com/2012/08/agresti-introduction-to-categorical-data.pdf>.

¹The college is switching me to a new system, so we'll see how it goes.

- There is a 3rd edition with R that will be coming out in October/November.

Computing: You will have your choice between using SAS or R. SAS is available from [sas.com](https://www.sas.com/en_us/software/university-edition.html) using their SAS Software for Learning program. See https://www.sas.com/en_us/software/university-edition.html You can either get it for download to on demand. There is no cost and have everything that you absolutely need (in some options, PROC Graph is not included). Alternatively, you can get a license and software from web-store; however, this costs money. R is free and can be downloaded from cran-r-project.org.

Which one (or both)? If you have used R and know the basics this will be fine; otherwise, SAS may be a better choice. I have worked thorough all analyses using SAS and am still learning R. SAS requires less programming and R requires more. More will be discussed about choice in the first class.

Evaluation: Homework assignments (60%) and a final exam or project (40%). You have the option of either taking a final exam or doing a project. You are encouraged to do a project rather than the final if you have categorical data from your own research or collaborative research.

Homework: There will be approximately 8–10 homework assignments. Each homework assignment will consist of 2–5 questions and/or problems. They will be due 1 week after assigned and will be due in lecture (hard copy) on the stated due date. No late homework will be accepted without prior approval of the instructor.

Final Exam: The final exams/projects will be take home. The final will be distributed around November 27. The exams/project are due Friday December 14, 4:00 pm. (not late assignments here).

Projects: For those interested in doing a project, you need to either talk to me about your intended project or turn in a proposal describing your intended project by November 13 (or before). The proposal is to ensure that the project is acceptable for this course and it provides an opportunity for preliminary feedback and suggestions. A final paper describing the project is due Friday December 14

The range of possible projects is quite broad. The intent of the project is to provide you an opportunity to apply the methods for categorical data analysis covered in class to your own research and effectively communicate the results. Projects will typically consist of analyses of data from research that you are currently performing (e.g., masters or dissertation research, collaborative research projects, etc.). Possible projects include (but not limited to):

- Use categorical methods to analyze data from your own research or research in which you are involved.

- Critique the use of procedures often used in your field or in a published research paper(s) and present more appropriate alternative analyses. Such a project should include a comparison of results obtained from the different types of analyses (e.g., using log-linear models rather than ANOVA).
- An in depth study of a procedure covered in class or one not covered in class (e.g., latent class analysis, log multiplicative association models, correspondence analysis, random effects models for discrete response data), including an application of it to data.

Illness: If you are sick, do NOT come to lecture, the instructor's office hours or the TAs office hours. If you need to turn in homework, you can either give it to a fellow student (preferred) or it may be sent electronically (to both the TA and instructor).

Disability Resources Statement: To obtain disability-related academic adjustments and/or auxiliary aids, students with disabilities must contact the course instructor and the Disability Resources and Educational Services (DRES) as soon as possible. To contact DRES you may visit 1207 S. Oak St., Champaign, call 333-4603 (V/TDD), or e-mail a message to disabilityillinois.edu . To insure that disability-related concerns are properly addressed from the beginning, students with disabilities who require assistance to participate in this class are asked to see the instructor as soon as possible. If you need accommodations for any sort of disability, please speak to me after class, or make an

Academic Integrity Statement: The Illinois Student Code should also be considered as a part of this syllabus. Students should pay particular attention to Article 1, Part 4: Academic Integrity. Read the Code at
URL: <http://www.admin.illinois.edu/policy/code/> .

Academic dishonesty may result in a failing grade. Every student is expected to review and abide by the Academic Integrity Policy: <http://www.admin.uiuc.edu/policy/code/>. Please note that you are responsible for reading this policy. Ignorance is not an excuse for any academic

Emergency Planning Statement: Plan for emergency situations in the classroom by reviewing the important material found at <http://police.illinois.edu/emergency-preparedness/> . The more prepared you are, the safer you will be. In an emergency in this building, we have three choices: RUN (get out), HIDE (find a safe place to stay inside), or FIGHT (with anything available to increase our odds for survival).

If you sign up for emergency text messages at emergency.illinois.edu, you'll receive information from the police and administration during these types of situations.

If you have any questions, go to police.illinois.edu, or call 217-333-1216

Course Plan/Outline 2017

Week	Topic	Section in Agresti
1	Overview & Introduction: history, data, computing. Sampling models, Inference: a proportion	1.1 – 1.2 1.2–1.3
2	2–way tables: Structure & Proportions 2–way tables: Odds Ratios	2.1–2.2 2.3, 11.1–11.2
3	Inference: Chi-squared tests of independence Inference: ordinal data, exact tests	2.4 2.5–2.6
4	3–way tables: partial association Inferential methods for conditional independence and homogeneous association	2.7 2.7
5	Generalized linear models (GLM) Optional reading available online: Anderson, Verkuilen & Johnson, ch 2	3.1
6	GLMs for binary data Poisson regression (possibly negbin & ZIP)	3.2 3.3
7	Inference and model checking Logistic Regression (numerical predictors)	3.4 4.1

Week	Topic	Section in Agresti
8	Logistic regression: model checking Logit models (qualitative predictors)	4.2, 5.1–5.2, 5.5 4.3
9	Multiple logistic/logit regression Loglinear models: 2–way tables	4.5–4.5 7.1
	***November 17–25 Fall Break ***	
10	Loglinear models: 3–way tables & inference Higher–way tables and the logit/loglinear model connections	7.2 7.3
11	Model building: association graphs Modeling ordinal association	7.4 7.5
12	Tests of conditional association Effects of sparse data	7.5
13	Multicategory logit models: nominal responses Optional supplemental reading: Anderson & Rutkowski (2007) and Anderson (2009) Ordinal responses & paired responses	6.1 6.2–6.4
14	Matched pairs ***Final exam will be available on or before November 27 ***	8.1–8.3
15	Square tables: quasi-independence, symmetry, quasi-symmetry, marginal homogeneity Wrap-up and/or catch-up	8.4

*** Projects & Final Exams Due Friday December 14 4:00pm. ***

- Agresti, A. (1984). *Analysis of Ordinal Categorical Data*. NY: Wiley.
- Agresti, A. (2013). *Categorical Data Analysis*, 3rd Edition. NY: Wiley.
- Agresti, A. (2007). *An Introduction to Categorical Data Analysis*, 2nd edition. NY: Wiley.
- Andersen, E.B. (1994). *The Statistical Analysis of Categorical Data*, 3rd Edition. Berlin: Springer-Verlag.
- Andersen, E.B. (1997). *Introduction to the Statistical Analysis of Categorical Data*. Springer-Verlag.
- Bishop, Y.M.M, Fienberg, S.E., & Holland, P.W. (1975). *Discrete Multivariate Analysis*. Cambridge, MA: MIT Press.
- Blasuius, J & Greenacre, M. (Editors) (1998). *Visualization of Categorical Data*. San Diego: Academic Press.
- Christensen, R. (1990). *Log-Linear Models*. NY: Springer-Verlag.
- Clogg, C.C., & Shihadeh, E.S. (1994). *Statistical Models for Ordinal Variables*. Thousand Oaks, CA: Sage.
- Cox, D.R., & Snell, E.J. (1989). *Analysis of Binary Data*, 2nd Edition. London: Chapman and Hall.
- Dobson, A.J. (1990). *An Introduction to Generalized Linear Models*. London: Chapman and Hall.
- Edwards, D. (1995). Graphical modeling. In W.J. Krzanowski (ed) *Recent Advances in Descriptive Multivariate Analysis*, pp 135–156. NY: Oxford.
- Edwards, D. (2000). *Introduction to Graphical Modelling*, 2nd Edition. NY: Springer-Verlag.
- Fahrmeir, L, & Tutz, G. (2001). *Multivariate Statistical Modelling Based on Generalized Linear Models*, 2nd edition. NY: Springer.
- Fienberg, S.E. (1980). *The Analysis of Cross-Classified Categorical Data*, 2nd Edition. Cambridge, MA: MIT Press.

- Friendly, M., & Meyer, D. (2016). *Discrete Data Analysis with R*. Boca Raton, FL: Taylor & Francis Group/CRC Press.
Or see <http://datavis.ca/courses/VCD/>
- Goodman, L.A., & Kruskal, W.H. (1979). *Measures of Association for Cross Classifications*. NY: Springer-Verlag. (reprint of articles appearing in the *Journal of the American Statistical Association* in 1954, 1959, 1963, and 1972.)
- Haberman, S.J. (1975). *The Analysis of Frequency Data*. Chicago, IL: University of Chicago Press.
- Hilbe, J.M. (2011). *Negative Binomial Regression*, 2nd Edition. Cambridge University Press: United Kingdom.
- Hosmer, D.W., & Lemeshow, S. (1989). *Applied Logistic Regression*. NY: Wiley.
- Le, C.T. (1998). *Applied Categorical Data Analysis*. NY: Wiley.
- Liao, T.F. (1994). *Interpreting Probability Models: Logit, Probit, and Other Generalized Linear Models*. Thousand Oaks, CA: Sage.
- Lindsey, J.K. (1995). *Modelling Frequency and Count Data*. NY: Oxford.
- Lloyd, C.J. & Lloyd, C.J. (1999). *Statistical Analysis of Categorical Data*. NY: Wiley.
- Lauritzen, S. (1996). *Graphical Models*. NY: Oxford.
- Lindsey, J.K. (1997). *Applying Generalized Linear Models*. NY: Springer-Verlag.
- Long, J.S. (1997). *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage.
- McCullagh, C.E., & Searle, S.R. (2001). *Generalized, Linear, and Mixed Models*. NY: Wiley.
- McCullagh, P., & Nelder, J.A. (1983). *Generalized Linear Models*, 2nd Edition. London: Chapman and Hall.
- Moldenberghs, G., & Verbeke, G. (2005). *Models for Discrete Longitudinal Data*. Springer.
- Powers, D.A. & Xie, Y (1999). *Statistical Methods for Categorical Data Analysis*. Academic Press.
- Read, T.R.C., & Cressie, N.A.C. (1988). *Goodness-of-fit Statistics for Discrete Multivariate Data*. NY: Springer-Verlag.

- Sobel, M.E. (1995). The analysis of contingency tables. In G. Arminger, C.C. Clogg, & M.E. Sobel (eds) *Handbook of Statistical Modeling for the Social and Behavioral Sciences*, pp 251–310. NY: Plenum Press.
- van der Ark, L.A., Croon, M.A., & Sijtsma, K. (editors) (2005). *New Developments in Categorical Data Analysis for the Social and Behavioral Sciences*. Mahwah, NJ: Lawrence Erlbaum.
- Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*. NY: Wiley.
- Wickens, T.D. (1989). *Multway Contingency Tables Analysis for the Social Sciences*. Hillsdale, NJ: Lawrence Erlbaum.
- Zelterman, D. (1999). *Models for Discrete Data*. Clarendon Press.

Generalized Linear Models:

- Dobson, A.J. (1990). *An Introduction to Generalized Linear Models*. London: Chapman and Hall.
- Lindsey, J.K. (1997). *Applying Generalized Linear Models*. NY: Springer-Verlag.
- McCullagh, P., & Nelder, J.A. (1983). *Generalized Linear Models*, 2nd Edition. London: Chapman and Hall.

Graphical Models

- Edwards, D. (2000). *Introduction to Graphical Modelling*, 2nd Edition. NY: Springer-Verlag. 2nd Edition.
- Lauritzen, S. (1996). *Graphical Models*. NY: Oxford.
- Whittaker, J. (1990). *Graphical Models in Applied Multivariate Statistics*. NY: Wiley.