

Applied Categorical Data Analysis
EdPsych/Psych/Soc 589
Spring 2010

Instructor: Carolyn J Anderson
e-mail cja@uiuc.edu
office: rm 236C Education Bldg 244-3537
office hours: Tuesday 1-3pm

Lecture: 10:00–11:50 am, Mon & Wed, 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: To introduce basic concepts and common statistical models and analyses for categorical data; to provide 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: <http://faculty.ed.uiuc.edu/cja/homepage/teaching.html>
The web-page include copies of lecture notes, homework assignments (and afterward the due date, answer keys and SAS program(s)), announcements, SAS example programs and output, and some handy programs.

Required Text:

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

Recommended (if you're new or novice using SAS):

Delwiche, L.D., & Slaughter, S.J. (2008). *The Little SAS Book*. Cary, NC: SAS Institute Inc.

Computing: We will be using SAS version 9.2¹. Access to SAS will be available via remote desktop connection to the College of Education server. Instruction on both connecting to the software and use will be given in class. If you prefer to have a copy of SAS on your own computer you may purchase a one year license from webstore (55.00). The media can be downloaded or purchased for an addition 75.00

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, especially if you have categorical data from your own research or collaborative research.

¹You may use R, STATA or MATLAB is you wish; however, I will not be providing instruction on the use of these programs.

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

Final Exam: The final exams will be take home. It will be available by Wednesday April 28 and due Friday May 7 before 5:00pm.

Projects: For those interested in doing a project, you need to turn in a proposal describing the intended project must be turned in by March 17. 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 May 7 by 5 pm.

The range of possible projects is quite broad. The intent of the project is to provide 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.

Applied Categorical Data Analysis
EdPsych/Psych/Soc 589

Course Plan/Outline 2010

Week	Date	Topic	Section in Agresti
1	Jan 20	Overview & Introduction: history, data, computing.	1.1 – 1.2
2	Jan 25	Sampling models, Inference: a proportion	1.2–1.3
	27	2–way tables: Structure & Proportions	2.1–2.2
3	Feb 1	2–way tables: Odds Ratios	2.3, 11.1–11.2
	3	Inference: Chi-squared tests of independence	2.4
4	Feb 8	Inference: ordinal data, exact tests	2.5–2.6
	10	3–way tables: partial association	2.7
5	Feb 15	Inferential methods for conditional independence and homogeneous association	2.7
	17	Generalized linear models (GLM) Optional reading available online: Anderson, Verkuilen & Johnson, ch 2	3.1
6	Feb 22	GLMs for binary data	3.2
	24	Poisson regression (possibly negbin & ZIP)	3.3
7	Mar 1	Inference and model checking	3.4
	3	Logistic Regression (numerical predictors)	4.1
8	Mar 8	Logistic regression: model checking	4.2, 5.1–5.2, 5.5
	10	Logit models (qualitative predictors)	4.3
9	Mar 15	Multiple logistic/logit regression	4.5–4.5
	17	Loglinear models: 2–way tables ** Project proposal due **	7.1
	***	March 20 – March 28 Spring Break ***	
10	Mar 29	Loglinear models: 3–way tables & inference	7.2
	31	Higher–way tables and the logit/loglinear model connections	7.3

Week	Date	Topic	Section in Agresti
11	Apr	5 Model building: association graphs	7.4
		7 Modeling ordinal association	7.5
12	Apr	12 Tests of conditional association	7.5
		14 Effects of sparse data	
13	Apr	19 Multicategory logit models: nominal responses Optional supplemental reading: Anderson & Rutkowski (2007) and Anderson (2009)	6.1
		21 Ordinal responses & paired responses	6.2–6.4
14	Apr	26	8.1–8.3
		28 Matched pairs Final exam will be available	
15	May	3 Square tables: quasi-independence, symmetry, quasi-symmetry, marginal homogeneity	8.4

*** Projects & Final Exams Due Friday May 7, by 5pm. ***

Applied Categorical Data Analysis
EdPsych 590AT/ Psych 593CA

- Agresti, A. (1984). *Analysis of Ordinal Categorical Data*. NY: Wiley.
- Agresti, A. (2002). *Categorical Data Analysis*, 2nd Edition. NY: Wiley.
- Agresti, A. (1996). *An Introduction to Categorical Data Analysis*. 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*, 2^{small nd} edition. NY: Springer.
- Fienberg, S.E. (1980). *The Analysis of Cross-Classified Categorical Data*, 2nd Edition. Cambridge, MA: MIT Press.

- 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.
- 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). *Multiway 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.