

STA 4504 Sec 5404
STA 5503 Sec 0043

Categorical Data Analysis

Spring, 2000

Time: Period 6 (12:50 – 1:40 p.m.) MWF

Location: FLO (Griffin-Floyd Hall) 100

Instructor: Dr. Brett Presnell

Office: FLO 220

E-mail: presnell@stat.ufl.edu

Phone: 392-1941 Ext. 236

Web Page: <http://www.stat.ufl.edu/~presnell>

Office Hours: Office hours for this course will be posted on the instructor's web page.

Text: Agresti, *An Introduction to Categorical Data Analysis*, John Wiley & Sons, 1996.

Other Course Information: will be posted on the web at
<http://www.stat.ufl.edu/~presnell/sta4504>

Course Description: Description and inference for binomial and Poisson variables using proportions and odds ratios, multi-way contingency tables, generalized linear models, logistic regression and logit modeling, loglinear models, logit extensions for multiple response categories or matched pairs. Computations will be illustrated using SAS and possibly other statistical software.

Prerequisites: One of the courses STA 3024, STA 3032, STA 4210, STA 4322, STA 6126, STA 6166, or the consent of the instructor (normally a graduate student should also take STA 6127 or STA 6167 before or concurrently with this course). The student should have some exposure to regression analysis and some practice using statistical software, since most of this course deals with extensions of regression to handling categorical response variables.

Exams: *Tentative* dates and times for the exams are as follows:

Exam 1	Wednesday, February 9	Evening	100 pts
Exam 2	Wednesday, March 29	Evening	100 pts
Exam 3	Friday, April 21	In class.	50 pts
Project	Thursday, May 4	9 a.m.	50 pts

The exams are not cumulative. The first two exams are one-hour exams, but will be given in the evening to avoid unnecessary time pressure. No make-up exams will be given except for medical or family emergencies, and must be approved before the time of the exam.

Project: The project will consist of a written report, maximum length 8 pages double-spaced. This report should present a statistical analysis based on modeling a data set containing a categorical response variable. If you have any questions about what to do for your project, please talk to me about it (e.g., if you do not have ready access to a dataset, you can create one from the General Social Survey at the web site <http://www.icpsr.umich.edu/gss/> or from other online sources of data). The report should include parts directed toward: (1) description of the data and its source; (2) statement of questions to be addressed; (3) specification of application of models to data and model-checking; (5) interpretations of

results of model fitting; (6) summary and conclusions. Include, in a separate appendix (not included in the page limit), copies of relevant parts of your computer printouts and (if possible) the data. The project is due on the regularly scheduled date of the final exam, May 4. However, you are welcome to turn it in any time after the third exam on April 21. Also, I am willing to look at a rough draft of your report and provide suggestions at any time up until April 10.

Homework: Required and optional homework problems are listed in the outline below. It is very important that students complete the required problems in a timely manner. Some exam questions will be taken directly from the required homework list. Students should keep a neat, organized file of solutions (including computer printouts) and hand them in with each exam. Each problem should be clearly labeled and in order. Twenty points on the first two exams and 10 points on the third exam will be based on the quality and completeness of this work. A bonus five points for each exam will be given to students who also complete nearly all the optional problems. Students are encouraged to work together in teams to help each other in understanding the course material and completing the homework problems (but each student should write up their own solutions). Partial solutions to some of the problems may be provided by the instructor.

Course Outline

Sec.	Topics	Text Pages	Homework	Optional
1.	Introduction	1–11	1–5, 8	7,9,12,13
2.	Two-Way Contingency Tables			
2.1	Table structure	16–19		30
2.2	Comparing proportions	19–22		
2.3	Odds ratio	22–27	1–4, 7, 8, 9ab	5, 9c, 10
2.4	Chi-squared tests	27–34	13, 15, 16	31–34
2.6	Exact tests for small samples	39–42	23, 25	
3.	Three-Way Contingency Tables			
3.1	Partial Association	53–60	1, 3, 4, 7	5, 6
3.2	Cochran-Mantel-Haenszel methods	60–63	8, 9, 10	
4.	Generalized Linear Models			
4.1	Components of generalized linear models	71–74	1	
4.2	GLMs for binary data	74–78	2(logistic)	
4.3	GLMs for count data	80–87	6, 8	14, 16, 17
4.4	Model inference and model checking	88–93	9, 11	
4.5	Fitting generalized linear models	93–97		
5.	Logistic Regression			
5.1	Interpreting logistic regression	103–108	1, 3, 41	2, 44
5.2	Inference for logistic regression	108–111	4, 8	15
5.3	Model checking	111–118	6, 7	10
5.4	Logit models for qualitative predictors	118–122	11–13, 20	
5.5	Multiple logistic regression	122–129	21, 23ab	24, 30
6.	Loglinear models			
6.1	Loglinear models for two-way tables	145–150		1, 2
6.2	Loglinear models for three-way tables	150–154	3, 4	5, 10, 12
6.3	Inference for loglinear models	155–158	6, 8, 9ab	
6.4	Loglinear models for higher dimensions	158–162	13	20
6.5	The loglinear-logit connection	162–167	14, 15, 17, 19	
8.	Multicategory Logit Models			
8.1	Logit models for nominal responses	205–211	1, 3, 4	
8.2	Cumulative logit model — ordinal responses	211–216	6, 9, 20	7, 8, 10, 21
9.	Models for Matched Pairs			
9.1	Comparing dependent proportions	226–229	1, 2, 4, 6	7, 9a–d