

STA4504/5503 CATEGORICAL DATA ANALYSIS SPRING 2008

Period 2-3 Tuesday, 3 Thursday, CSE E 222

Instructor: Alan Agresti

Office: Griffin-Floyd 204

Phone: 392-1941, ext. 234

E-mail: aa@stat.ufl.edu

Office Hours: Tuesday and Thursday 1:45-3:45, and by appointment.

Teaching assistant: Tezcan Ozrazgat

office hours Monday and Wednesday 1-3 pm, Friday 9:30-11:30 am, Griffin-Floyd 117a
e-mail torazgat@stat.ufl.edu, phone 392-1941, ext. 205.

Tezcan will handle questions about the homework exercises, including software questions, and in my office hours I will handle questions about the methods themselves.

Course Homepage: www.stat.ufl.edu/~aa/sta4504/index.html

Course description: Description and inference for binomial and multinomial variables using proportions and odds ratios, multi-way contingency tables, generalized linear models for discrete data, logistic regression for binary responses, models for multiple response categories, loglinear models, inference for matched-pairs and correlated clustered data.

Prerequisites: familiarity with statistical methods as covered in courses such as STA 3024, STA 3032, STA 4210, STA 4322, STA 6127, STA 6167, or the consent of the instructor. Since much of this course deals with extensions of regression modeling to handle categorical response variables, students should be comfortable with multiple regression (including dummy variables for incorporating categorical predictors in a model) and should have had practice using statistical software.

Course text: *An Introduction to Categorical Data Analysis, 2nd edition*, by A. Agresti (2007), published by John Wiley & Sons. A copy is on 2-hour and overnight reserve at the Science library.

Software: My lectures will illustrate computations using SAS statistical software. For the homework exercises that require software, you are welcome to use whatever software you prefer. There is some information about software at the course homepage listed above. This includes a pdf file about using SPSS to analyze categorical data, taken from my book *Statistical Methods for the Social Sciences* (4th edition 2008). The

website for my more advanced text *Categorical Data Analysis* (2nd edition 2002)

<http://www.stat.ufl.edu/~aa/cda/cda.html>

has a section with more detailed information about software for categorical data methods including the link

<https://home.comcast.net/~lthompson221/Splusdiscrete2.pdf>

to a comprehensive manual prepared by Dr. Laura Thompson showing how to use R and S-Plus to conduct all the types of analyses presented in this course. The TA is available to help you with the software that you decide to use. SAS programs and data sets from the text are available at the website for the text for the course,

<http://www.stat.ufl.edu/~aa/intro-cda/appendix.html>

Exam dates:

Exam 1 Thursday, February 14 (100 pts.)

Exam 2 Thursday, March 27 (100 pts.)

Exam 3 Tuesday, April 29, 7:30-9:30 am (100 pts.)

The exams are not cumulative. Make-up exams will not be given except for medical or family emergencies, and must be approved before the time of the exam. The final exam will not be prepared until shortly before the date of that exam and cannot be taken early.

Homework: Required and optional homework problems are listed in the outline of course topics on the next page. (You are not responsible for the “optional” problems in the list, but those students who want to extend their knowledge of the methods further and have practice with more difficult exercises are encouraged to work them.) To provide you with feedback about your solutions, outlines of the solutions to the homework problems are available in a pdf file of the brief solutions at

<http://www.stat.ufl.edu/~aa/restricted/index.html>

Short answers for odd-numbered exercises are also available at the end of the textbook. You are encouraged to get help from the TA (Tezcan) with homework problems that you are unable to do and/or to work together in teams to help each other in understanding the course material and completing the homework. Some exam questions will be similar if not identical to those on the required homework list. Keep a neat, organized file of solutions (including computer printouts) and hand them in with each exam. Twenty points on each exam will be based on the quality of this work.

Topics	Text Pages	Homework	Optional
1. Introduction			
1.1-1.3 Statistical inference for a proportion	1-10	1-4, 8, 12	15, 16
2. Contingency Tables			
2.1 Table structure	21-25	2	
2.2 Comparing proportions	25-28	3	
2.3 Odds ratio	28-34	5-8, 12	
2.4 Chi-squared tests	34-40	17-19	21, 24-26
2.6 Exact tests for small samples	45-48	29	
2.7 Association in three-way tables	49-54	33-36, 39	37, 38
3. Generalized Linear Models			
3.1 Components of generalized linear model	65-68	1, 22ab	
3.2 GLMs for binary data	68-73	2, 5	6
3.3 GLMs for count data	74-84	11-12, 16	17-18, 20-21
3.4 Inference and model checking	84-87	9, 13	14
3.5 Fitting generalized linear models	88-90		
4. Logistic Regression			
4.1 Interpreting logistic regression	99-106	1, 4	35, 36
4.2 Inference for logistic regression	106-110	2, 8	
4.3 Categorical predictors	110-115	11, 16-17, 37	
4.4 Multiple logistic regression	115-120	19, 21, 23, 24	
4.5 Summarizing effects	120-121	28	27
5. Building and Applying Logistic Regression Models			
5.1 Strategies in model selection	137-144		
5.2 Model checking	144-150	4, 15, 19, 30	20
5.3 Effects of sparse data	152-156	22	
6. Multicategory Logit Models			
6.1 Logit models for nominal responses	173-179	1, 6	
6.2 Cumulative logit model for ordinal responses	179-189	5, 7, 12, 22	
8. Models for Matched Pairs			
8.1 Comparing dependent proportions	244-247	2, 4, 7	8, 10
8.5.5 Measuring agreement	264	20ac	
9. Modeling Clustered Responses (Repeated Measures)			
9.1 Marginal models vs. conditional models	276-279		
9.2 Marginal modeling: The GEE approach	279-284	2-4, 7, 18	
9.3 GEE for multinomial responses	285-287		
7. Loglinear Models			
7.1 Loglinear models for 2-way and 3-way tables	204-212	27	
7.2 Inference for loglinear models	212-223	5-7	8