Categorical Data Analysis

BIOS 8110 **SYLLABUS AND COURSE POLICIES** Fall, 2009

Instructor: Woncheol Jang
129B Paul D Coverdell Center
jang@uga.edu
(706) 583-8923
Office Hours: Tuesday/Thursday 3-4PM or by appointment

Lectures: Paul D. Coverdell Center Room 339
Tuesday/Thursday 9:30AM-10:45AM

Course Web Page: `http://jang.myweb.uga.edu/teaching/8110/`

Course description

Social and health science data often consist of discrete categorizations of count of events. In this course, students will learn (1) description and inference for binomial and multinomial variables using proportions and odds ratios, (2) binomial regression models for binary data including logistic regression and probit models, (3) Poisson regression models for count data including log-linear models for contingency tables and (4) inference for matched pairs and correlated clustered data. The course will involve the practical application of the ideas and their implementation through statistical software (SAS and/or R) to make them accessible to social and public health scientists. It is assumed that students have a solid grounding in basic statistical theory and linear regression, as provided, for example, by courses such as STAT 6220, STAT 6320, BIOS 7020 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 Learning Objectives

This course is designed to provide modern statistical methods for analyzing categorical data. Broadly, we will discuss

1. Inference of proportions and odds ratios
2. Perform logistic regression analyses with multiple predictors.
3. Use log-linear models to analyze contingency tables.
4. Use visual and other methods for assessing the adequacy of the fitted model
5. Interpret each coefficient in the model.
6. Describe the methods and results to a non-statistical reader.

Textbooks and Other Required Reading

- Required text:
  - Agresti, A. (2007). *An Introduction to Categorical Data Analysis.* Wiley
- Recommended texts:
  - Stokes, M.E., Davis, C.S., and Koch, G.G. (2000). *Categorical Data Analysis Using the SAS System.* SAS Press.
  - Thomson, L. (2009) *SPLUS (and R) Manual to Accompany Agrestis Categorical Data Analysis (2002) 2nd edition.* `https://home.comcast.net/~lthompson221/Splusdiscrete2.pdf`
- Other useful texts:
  - Agresti , A. (2002). *Categorical Data Analysis*, 2nd edition. Wiley.
  - Christensen, R. (1997). *Log-linear models and logistic regression.* 2nd edition. Springer.
  - Edwards, D. (2000). *Introduction to graphical modeling.* 2nd edition. Springer.
  - Hosmer, D. W. and Lemeshow, S. (2000). *Applied Logistic Regression.* 2nd edition. Wiley.
  - Simonoff, J.(2003). *Analyzing Categorical Data.* Springer.

Course Assignments/Requirements for Grading Purposes

There will be biweekly homework assignments. The assignments are due in class. Homework will be a mix of data analysis part and theoretical materials. Data sets will be provided as needed. You may work with other students on these problems or refer to other sources if you would like. However, the write-up of your homework should be your own. Each assignment will have equal weight on your grade.

Grading Policy

Graded work for the course will consist of homework, class participation, a midterm, and a final exam. Your final grade will be determined as follows:

- Homework: 30 %
- Midterm: 30 %
- Final Exam: 35 %
- Class Participation: 5 %

Computing

We will use SAS and/or R for statistical analysis.

- R is a free download at http://cran.r-project.org/ for many operating systems, and is installed in the CPH computer lab.
- SAS. Available in the CPH computer lab.

Homework done with other statistical package such as SPSS and STATA will be accepted but the instructor will not be available for assistant in using other packages.

Make-Up Policy

Exam dates:

- Midterm: Oct. 6
- Final exam: Dec. 15 8-11AM

You are expected to take the exams at the scheduled times. In case of genuine emergency, illness or hardship, for which you can present written documentation I may agree to arrange for a make-up exam. Make-up exams must always be arranged BEFORE the regular exam is given and always take place AFTER the regular exam. If a student has an excused absence from a midterm exam, then the weight of that material is transferred to the final exam. If a student has an excused absence from the final exam, then special arrangements will be made on a case-by-case basis.

Class Participation/Discussion Policy

You are strongly encouraged to participate in class discussion. Please ask questions when you have them. I will need feedback to understand which elements of the presentation are unclear. Personal experience and interests related with this material are also appreciated. A classlist containing e-mail addresses of all class members will be constructed for e-mail contact. Students are encouraged to direct questions to the instructor. When of general interest, these will be edited. Rendered anonymous as to the sender and forwarded together with the response to the classlist.

University Honor Codes and Academic Honesty Policy

Each student is expected to comply with the University of Georgias ethical code of conduct and the academic work must meet the standards contained on A Culture of Honesty:
`http://www.uga.edu/ovpi/honesty/acadhon.htm`
Please see your student handbook for these policies. Violations of these policies will not be tolerated and may result in the student receiving a failing grade in this course.

Students with Disabilities

Students with disabilities who require reasonable accommodations in order to participate in course activities or meet course requirements should contact the instructor or designate during regular office hours or by appoint.

LECTURE SCHEDULE

The following schedule may be modified as needed. Please read relevant chapters before coming to class.

Date	Topic	Reading	HW due
Aug .18	Categorical data and probability distributions	1.1, 1.2	
Aug. 20	Statistical Inference for a proportion	1.3	
Aug. 25	Table structure and comparing proportions	2.1, 2.2	
Aug. 27	Odds ratio and Chi-square test	2.3, 2.4	
Sep. 1	Cochran-Mantel-Haenszel test	2.5	HW #1
Sep. 3	Exact test	2.6	
Sep. 8	Association in three way tables	2.7	
Sep. 10	GLMs for binary data	3.2	
Sep .15	GLMs for count data	3.3	
Sep. 17	Inference and model checking	3.4	HW #2
Sep. 22	Interpreting logistic regression	4.1	
Sep 24	Inference for logistic regression	4.2	
Sep. 29	Categorical predictors	4.3	
Oct. 1	Multiple logistic regression	4.4, 4.5	HW #3
Oct. 6	**Midterm Exam**		
Oct. 8.	Strategies in model selection	5.1	
Oct. 13	Model checking	5.2	
Oct. 15	Effects of sparse data	5.3	
Oct. 20	Conditional logistic regression and power analysis	5.4, 5.5	
Oct. 22	Logit models for nominal responses	6.1	HW #4
Oct. 27	Cumulative logit model for ordinal responses	6.2	
Oct. 29	Paired-category ordinal logits	6.3	
Nov. 3	Tests of conditional independence	6.4	
Nov. 5	Loglinear models for contingency tables	7.1	
Nov. 10	Inference for loglinear models	7.2, 7.3	HW #5
Nov. 12	Comparing dependent proportions and logistic regression for matched pairs	8.1, 8.2	
Nov. 17	Comparing margins of squared contingency table and symmetry and quasi symmetry models	8.3, 8.4	
Nov. 19	Analyzing rater agreement	8.5	
Nov. 24	**Thanksgiving Break**		
Nov. 26	**Thanksgiving Break**		
Dec. 1	Marginal models	9.1, 9.2	
Dec. 3	GEE for multinomial responses	9.3	HW #6
Dec .15	**Final Exam** 8-11AM		