

Categorical Data Analysis
QNME 0614 (CRN#17036)
UMDNJ-School of Public Health, Newark Campus
Fall 2006 (Mondays, 5:30 – 8:30)

Instructor: Michael Brimacombe, PhD, Associate Professor, Dept. of Preventive Medicine & Community Health, NJMS & SPH – UMDNJ Office: MSB F-647.
Phone: 973-972-5229. **E-mail:** brimacmb@umdnj.edu. **Office hours:** by appointment.

Overview:

This course examines the analysis of categorical data in clinical and epidemiological settings. It focuses on the ways in which exposure, risk and association are applied in the investigation of disease and the causes of disease. At the end of this course the student should be able to analyze the various aspects of categorical data. The computing package to be used is SAS.

Textbook: An Introduction to Categorical Data Analysis, by Alan Agresti. Published by John Wiley & Sons, 1996. On reserve in the GF Smith Library, NJMS.

Suggested supplemental text: Categorical Data Analysis Using the SAS System, by Stokes, Davis and Koch, 2nd edition, SAS Institute, 2000.
(To order: www.sas.com/apps/pubscat/welcome.jsp)

Additional book: Statistical Methods for Rates and Proportions, by Joseph Fleiss, John Wiley & Sons, 2nd edition, 1981. On reserve in the GF Smith Library, NJMS.

Assignments:

Homework assignments (3 - 4) will account for approximately 50% of the grade. You will usually have at least two weeks to complete an assignment. You may collaborate with your fellow students on computing and discussion, but must write up your assignment independently. Be selective in submitting SAS output. Less is more!! If you begin working on your assignment early enough, you should be able to identify SAS coding problems. I will be available to assist you with coding problems. No late assignments.

Grading:	Homework:	50%
	Midterm Exam:	20%
	Final Exam:	30%

Final Exam: The final exam will be a take-home exam. You may NOT collaborate on the final exam, although you may consult directly with me regarding any SAS coding or other problems.

COURSE OUTLINE

Lecture 1: Binomial, multinomial and Poisson distributions. Inference for a single proportion. Maximum likelihood. Introduction to two-by-two tables (two-way contingency tables). Comparison of two proportions.

Reading: Agresti, pp. 1-21.

Lecture 2: Two-by-two tables continued. Relative risk, odds ratios, tests of independence. Cohort and case-control studies. Introduction to SAS for statistical computing (“proc freq”).

Reading: Agresti, pp. 21-34 (skip section 2.4.6 “Partitioning Chi-squared”)

Lecture 3: Testing independence for ordinal data. Small sample inference. Inter-rater agreement. McNemar’s test for matched data. More SAS.

Reading: Agresti, pp. 34-45, Fleiss (handout).

Lecture 4: Introduction to three-way contingency tables. Partial association. Cochran-Mantel-Haenszel methods. Marginal vs. conditional independence. More SAS.

Reading: Agresti, pp. 53-64

Mid-term Exam.

Lecture 5: Introduction to Logistic Regression. Interpretation of model coefficients. SAS “proc logistic.” Model checking, diagnostics, influence.

Reading: Agresti, pp. 103, 107- top of 110, 111 – 118, 118-120, 122-129

Lecture 6: Logistic Regression continued.

Reading: Agresti, pp. 130-132, 205-211.

Lecture 7: Logistic Regression continued. Count data and Poisson regression

Reading: Agresti, pp. 71-80, pp. 80-93.

Lecture 8: Generalized linear models. SAS “proc genmod.” General sample size considerations. Reading: Agresti, pp. 229-233

Lecture 9: Categorical data analysis with dependent data. Logistic regression for matched pairs. Logistic regression with more than two response categories. (polytomous regression).

Lecture 10: Correspondence analysis. Review of course.

Take home Final exam.