Columbia University
MAILMAN SCHOOL
OF PUBLIC HEALTH

P8120 - Analysis of Categorical Data

COURSE DESCRIPTION

A comprehensive overview of methods of analysis for binary and other discrete response data, with applications to epidemiological and clinical studies. It is a second level course that presumes some knowledge of applied statistics and epidemiology. Topics discussed include 2 × 2 tables, m × 2 tables, tests of independence, measures of association, power and sample size determination, stratification and matching in design and analysis, and logistic regression analysis.

COURSE LEARNING OBJECTIVES

Students who successfully complete this course will be able to:

- apply statistical tools to make inference about a single binomial proportion or two sample proportions using
  - approximate confidence intervals
  - hypotheses testing using approximate or exact methods
  - calculation of power, sample sizes and detectable effect sizes
- Understand and explain the properties of different measures of association by
  - estimating various forms of measures of association from retrospective, cross-sectional and prospective studies
  - conducting exact and approximate tests of hypotheses
  - constructing approximate confidence intervals
- become familiar with the analysis of many proportions by
  - comparing m proportions that are unordered, quantitatively ordered, or qualitatively ordered
  - testing 2 × m tables for independence
- analyze three-way tables by
  - testing the hypothesis of homogeneous associations across all partial 2x2 tables
  - estimating the common underlying odds ratio
  - testing the hypothesis of conditional independence across all partial 2x2 tables
- understand the fundamental importance of the logistic model, specifically
  - interpret logistic regression coefficients from simple binary logistic regression models, additive and interactive multiple binary logistic regression models and polytomous logistic regression models
  - understand the uses of loglinear model for contingency tables
- appreciate the strengths and limitations of matched sample designs by
  - analyzing data from matched-pair designs
  - testing hypotheses using the McNemar test

PREREQUISITES

P6103 or P6104, P6400

WEBSITES and EMAILS:

- http://courseworks.columbia.edu/
- http://twitter.com/#!/search/realtime/%23P8120S12 (To every tweet, attach hash tag #P8120S12)
- https://twitter.com/#!/P8120S12 (To send me a direct tweet, start with @P8120S12)
- **Official course email:** P8120S12@gmail.com

INSTRUCTOR

Martina Pavlicova: Biostatistics Department, 722 W 168th Street, 6th floor, rm 635
Email: P8120S12@gmail.com **or** mp2370@columbia.edu
Phone: (212) 305-9405 (I prefer the use of email)
Fax: (212) 305-9408
Office hours by appointment

CLASS SESSIONS

Lectures: **Mondays and Wednesdays (both days)**, Rm: Mailman School Auditorium, 8th floor
Session 1: 9:30am - 10:50am
Session 2: 11:30am - 12:50pm

TEACHING ASSISTANTS: TBA

ASSESSMENT OF LEARNING

Quizzes (Best 3 out of 4)	42% (each 14%)
Homeworks (Best 5 out of 6)	30% (each 6%)
Final Exam	28%

If a student obtains 100% on the Final Exam, the student will not receive worse grade than B+.

Late homeworks will not be accepted under any circumstances! There will be no make up quizzes.

The final course grade will be determined using the School's letter grade system. Grades are **A, B, C,** with + and - as applicable. Grades are defined as follows:

A+ Reserved for highly exceptional achievement.
A Excellent. Outstanding achievement.
A- Excellent work, close to outstanding.
B+ Very good. Solid achievement expected of most graduate students.
B Good. Acceptable achievement.
B- Acceptable achievement, but below what is generally expected of graduate students.
C+ Fair achievement, above minimally acceptable level.
C Fair achievement, but only minimally acceptable.
C- Very low performance.
F Failure. Course usually may not be repeated unless it is a required course.

All requirements must be completed by Wednesday, May 9, 2012.
Incomplete grades will not be given except for a serious medical condition documented in advance of the final examination.

COURSE REQUIREMENTS

RECOMMENDED TEXT: *An Introduction to Categorical Data Analysis, 2nd Ed., 2007by Alan Agresti.* John Wiley & Sons, ISBN 9780471226185.

ADDITIONAL TEXTS:

Statistical Methods for Rates and Proportions, 3rd Ed., 2003 by Joseph L. Fleiss, Bruce Levin, and Myunghee Cho Paik. John Wiley & Sons,

Categorical Data Analysis Using The SAS System. Stokes, Davis and Koch. New York: Wiley 2000.

Applied logistic regression, David W. Hosmer, Stanley Lemeshow, New York :Wiley 2000.

A scientific calculator is necessary. It should have exponential, log and square root capabilities.

We will be using SAS® for our statistical computing needs. SAS® is available in the Hammer Library or may be purchased by students. A free online version of SAS® (SAS® Enterprise Guide) will be available to students as well. To access this free version of SAS®, you will need to register for SAS® OnDemand for Academics and then access the Enterprise Guide.

- Access the following website: http://support.sas.com/ondemand/index.html
- Review the information and follow the steps at this site (register under Columbia University)
- If you have additional questions, see http://support.sas.com/ondemand

The course will be given in consecutive lectures. Students are encouraged to ask questions and to utilize the extensive teaching assistant's office hours. Attendance at lectures is absolutely essential. Lectures are meant to enrich the student's reading and understanding of the textbook, often by examples not contained in the reading or homework problems.

MAILMAN SCHOOL POLICIES AND EXPECTATIONS

Students and faculty have a shared commitment to the School's mission, values and oath.
http://mailman.columbia.edu/about-us/school-mission/

Academic Integrity

Students are required to adhere to the Mailman School Honor Code, available online at http://mailman.columbia.edu/honorcode.

Disability Access

In order to receive disability-related academic accommodations, students must first be registered with the Office of Disability Services (ODS). Students who have, or think they may have a disability are invited to contact ODS for a confidential discussion at 212.854.2388 (V) 212.854.2378 (TTY), or by email at disability@columbia.edu. If you have already registered with ODS, please speak to your instructor to ensure that s/he has been notified of your recommended accommodations by Lillian Morales (lm31@columbia.edu), the School's liaison to the Office of Disability Services.

COURSE SCHEDULE:

WEEK 1	
1/18/2012	**L1: Review of Probability and Statistical Inference for a Single Proportion 1** Homework #1 assigned at the end of the week.
WEEK 2	
1/23/2012	**L2: Review of Probability and Statistical Inference for a Single Proportion 2**
1/25/2012	**L3: Sampling Plans and Measures of Associations 1**
WEEK 3	
1/30/2012	**L4: Sampling Plans and Measures of Associations 2, SAS** Homework #1 due 1pm.
2/1/2012	**Quiz I**
WEEK 4	
2/6/2012	**L5: Assessing significance in fourfold table 1**
2/8/2012	**L6: Assessing significance in fourfold table 2** Homework #2 assigned.
WEEK 5	
2/13/2012	**L7: Determining sample size and effect size 1**
2/15/2012	**L8: Determining sample size and effect size 2** Homework #2 is due 1pm.
WEEK 6	
2/20/2012	**President’s Day - No classes**
2/22/2012	**Quiz II**
WEEK 7	
2/27/2012	**L9: Mantel-Haenzel’s methods**
2/29/2012	**L10: Assessing Confounding and Effect Modification** Homework #3 assigned.
WEEK 8	
3/5/2012	**L11: Contingency tables 1**
3/7/2012	**L12: Contingency tables 2** Homework #3 is due 1pm.
WEEK 9	
3/12/2012	**Spring Break - No classes**
3/14/2012	**Spring Break - No classes**

WEEK 10	
3/19/2012	**Quiz III**
3/21/2012	**L13: Logistic Regression 1 (Intro)**
WEEK 11	
3/26/2012	**L14: Logistic Regression 2 (Interpretation)** Homework #4 assigned.
3/28/2012	**L15: Logistic Regression 3 (Inference)**
WEEK 12	
4/2/2012	**L16: Logistic Regression 4 (Evaluating Bias and Confounding)** Homework #4 is due 1pm, Homework #5 assigned.
4/4/2012	**L17: Logistic Regression 5 (Goodness of Fit)**
WEEK 13	
4/9/2012	**L18: Logistic Regression (Model Selection)** Homework #5 is due 1pm.
4/11/2012	**Quiz IV**
WEEK 14	
4/16/2012	**L19: Logistic Regression (Matching and Conditional)**
4/18/2012	**L20: Logistic Regression (Polytomous)** Homework #6 assigned.
WEEK 15	
4/23/2012	**L21: Logistic Regression (Additional material)**
4/25/2012	**L22: Logistic Regression (Additional material)** Homework #6 is due 1pm.
WEEK 16	
4/30/2012	**L23: Review**
WEEK 17	
5/9/2012	**Cumulative Final Exam** Session 1: 9-10:50am; Session 2: 11-12:50pm.