



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, $m \times 2$ 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 $2 \times m$ tables for independence
- analyze three-way tables by
 - testing the hypothesis of homogeneous associations across all partial 2×2 tables
 - estimating the common underlying odds ratio
 - testing the hypothesis of conditional independence across all partial 2×2 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/>
- <https://twitter.com/search/realtime?q=P8120S13> (To every tweet, attach hash tag #P8120S13)
- <https://twitter.com/P8120S13> (To send me a direct tweet, start with @P8120S13)
- **Official course email:** p8120s13@gmail.com

INSTRUCTOR

Martina Pavlicova: Biostatistics Department, 722 W 168th Street, 6th floor, rm 635
Email: p8120s13@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)
Room: HSC 401 (except 1/28: A-7)
Time: 5:30pm - 6:50pm

TEACHING ASSISTANTS: TBA

ASSESSMENT OF LEARNING

Quizzes (Best 3 out of 4).....	39% (each 13%)
Homeworks (Best 6 out of 7).....	36% (each 6%)
Final Exam.....	25%

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 Monday, May 13, 2013.

Incomplete grades will not be given except for a serious medical condition documented in advance of the final examination.

RECOMMENDED TEXT: *An Introduction to Categorical Data Analysis, 2nd Ed., 2007* by 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://www.mailman.columbia.edu/about-us/mission-and-history>

Academic Integrity

All students are required to adhere to the Mailman School Honor Code, available online at

<http://mailman-handbook.com/2009/node/165>

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	
W 1/23/2013	L1: Review of Probability and Types of Distributions
WEEK 2	
M 1/28/2013	L2: Statistical Inference for a Single Proportion (different location: A-7)
W 1/30/2013	L3: Contingency Tables and Measures of Associations
WEEK 3	
M 2/4/2013	L4: Sampling plans and Introduction to SAS Homework #1 due 5:30pm.
W 2/6/2013	Quiz I (L1 - L4)
WEEK 4	
M 2/11/2013	L5: Confidence Intervals for Different Measures of Associations
W 2/13/2013	L6: Assessing Significance in 2x2 Table 1
WEEK 5	
M 2/18/2013	<i>President's Day - No classes</i>
W 2/20/2013	L7: Assessing Significance in 2x2 Table 2
WEEK 6	
M 2/25/2013	L8: Determining Sample Size, Power and/or Effect Size Homework #2 is due 5:30pm.
W 2/27/2013	Quiz II (L5 - L8)
WEEK 7	
M 3/4/2013	L9: Logistic Regression : Intro + 1 Binary Predictor
W 3/6/2013	L10: Logistic Regression : Different Predictors
WEEK 8	
M 3/11/2013	L11: Logistic Regression : Confidence Intervals and Assessing Significance Homework #3 is due 5:30pm.
W 3/13/2013	Quiz III (L6 - L11)
WEEK 9	
M 3/18/2013	<i>Spring Break - No classes</i>
W 3/20/2013	<i>Spring Break - No classes</i>

WEEK 10	
M 3/25/2013	L12: Assessing Effect Measure Modification in 2x2 Tables
W 3/27/2013	L13: Assessing Confounding in 2x2 Tables
WEEK 11	
M 4/1/2013	L14: Logistic Regression : Assessing Interaction and Confounding Homework #4 is due 5:30pm.
W 4/3/2013	L15: Logistic Regression : Goodness of Fit and Saturated Model
WEEK 12	
M 4/8/2013	L16: Logistic Regression : Akaike Information Criterion, Model Building Homework #5 is due 5:30pm.
W 4/10/2013	Quiz IV (L11 - L16)
WEEK 13	
M 4/15/2013	L17: Matched Pairs, McNemar's Test, SAS
W 4/17/2013	L18: Conditional Logistic Regression : Matching and Likelihood Function
WEEK 14	
M 4/22/2013	L19: 2xK Tables (Unordered and Ordered), Cochran- Armitage Test Homework #6 is due 5:30pm.
W 4/24/2013	L20: Polytomous Logistic Regression
WEEK 15	
M 4/29/2013	L21: Ordinal Logistic Regression
W 5/1/2013	L22: Logistic Regression (Additional material)
WEEK 16	
M 5/6/2013	L23: Review Homework #7 is due 5:30pm.
WEEK 17	
M 5/13/2013	Cumulative Final Exam (L1-L23)