



## BST 466 Categorical Data Analysis

**Term & Year:** Spring 2018  
**Classroom and Time:** SRB 1404, MW 1:45pm - 3:15pm  
**Instructor:** Tongtong Wu, Ph.D.  
**Office Hours:** By appointment  
**Office:** SRB 4141  
**Phone:** (585) 276-6858  
**Email:** tongtong\_wu@urmc.rochester.edu  
**TA:** Jiatong Sui (email: jiatong\_sui@urmc.rochester.edu)

### Course Prerequisites

Required: Fundamental statistical courses

### Required Texts and Other Readings:

Required: Agresti, Alan (2007). An Introduction to Categorical Data Analysis, 2nd Edition. Hoboken, New Jersey: John Wiley & Sons. ISBN: 978-0-471-22618-5

Required: Agresti, Alan (2012). Categorical Data Analysis, 3rd Edition. Hoboken, New Jersey: John Wiley & Sons. ISBN: 978-0-470-46363-5

### Additional Materials Required

SAS (strongly recommended) or other statistical software

### Course Website

Lecture notes, handouts, SAS examples, homework assignments, and homework solutions will be posted on Blackboard. Lecture notes will be uploaded before classes. You can access the website by following the directions

- Direct your URL to <https://learn.rochester.edu/>
- Enter your NetID and Password
- Click "CATEGORICAL DATA ANALYSIS - 2017SPRING"

### Course Description

This course provides an introduction to methods for analyzing categorical data, with emphasis to medical and public health research. It emphasizes the ideas behind the methods and their applications. The course covers descriptive and inferential methods for contingency tables and generalized linear models for binary, ordinal, and count data. It also covers methods for matched-pairs data. Computations are illustrated using SAS.

### Course Aims and Objectives

Upon completing this course, the student will be able to:

1. Understand and describe the major types of categorical data and discrete probability distributions



2. Identify, describe, and apply appropriate descriptive and inferential methods for contingency tables (e.g., odds ratios, independence tests)
3. Identify, describe, and build appropriate models for different types of categorical data (e.g. logistic regression, loglinear models)
4. Interpret results of categorical data analyses in oral and written forms
5. Analyze categorical data using statistical software

### Assignments and Exams

#### Homework:

Four homework assignments will be given. It is important to complete the homework in a timely manner. Twenty percent will be deducted if homework is handed in within three days after the deadline *and* before homework solutions are posted (whichever earlier). No homework will be accepted after three days or homework solutions are posted. Students are encouraged to work together in groups to help each other understand the course materials and complete the homework problems. However, students must finish their homework independently after group discussions. The names of the group members are required to be put on the first page of the homework.

#### Exams:

Midterm and final exams will be in class, closed-book, and closed-note. The content of exams will be cumulative. For the midterm exam, you are allowed to bring one page of letter-size, double-sided formula sheet; for the final exam, you are allowed to bring two pages of letter-size, double-sided formula sheet. You also need to bring a calculator to conduct the calculation. No makeup exams are allowed in general. Exceptions to this rule are evaluated on a case-by-case basis. Students must submit a request before the exam with valid supporting documents. No post-exam request will be considered except the student is hospitalized during the exam period.

### Assignments and Grading Procedures

- 40% Homework
- 30% Mid-term exam
- 30% Final exam

### Additional Literature, Websites and Other Resources:

#### Alan Agresti's Websites:

- Alan Agresti's course website of Intro CDA  
<http://www.stat.ufl.edu/~aa/sta4504/>
- Alan Agresti's SAS appendix for Intro CDA  
<http://www.stat.ufl.edu/~aa/cda/software.html>
- Alan Agresti's website for Categorical Data Analysis (2nd edition, Wiley, 2002)  
<http://www.stat.ufl.edu/~aa/cda/cda.html>

#### Other Useful Websites:

- SAS and Stata code for Intro CDA from UCLA ATS  
<http://www.ats.ucla.edu/stat/examples/icda/default.htm>



- Statistical computing from UCLA ATS (Data Analysis Examples, Textbook Examples, Online Seminars, etc)  
<http://www.ats.ucla.edu/stat/default.htm>
- Laura Thompson's R and S manual for CDA  
<https://home.comcast.net/~lthompson221/#CDA>
- R introduction  
<http://cran.r-project.org/doc/manuals/R-intro.pdf>

### **Academic Integrity**

Academic integrity is a core value of the University of Rochester. Students who violate the University of Rochester University Policy on Academic Honesty are subject to disciplinary penalties, including the possibility of failure in the course and/or dismissal from the University. Since academic dishonesty harms the individual, other students, and the integrity of the University, policies on academic dishonesty are strictly enforced. For further information on the University of Rochester Policy on Academic Honesty, please visit the following website:

[http://www.rochester.edu/college/honesty/docs/Academic\\_Honesty.pdf](http://www.rochester.edu/college/honesty/docs/Academic_Honesty.pdf)

### **Accommodations for Students with Disabilities**

Students needing academic adjustments or accommodations because of a documented disability must contact the Disability Resource Coordinator for the school in which they are enrolled:

<http://www.rochester.edu/eoc/DisabilityCoordinators.html>



## Course Schedule

Course Schedule Summary			
Week	Date	Topic	Reading / Assignment
0	1/17/2018	No class	
1	1/22/2018 1/24/2018	Introduction and review Categorical response data Distributions for categorical data (Bernoulli, binomial, multinomial, Poisson)	Sections 1.1-1.2
2	1/29/2018 1/31/2018	Introduction and review (cont'd) Likelihood function and MLE Three likelihood-based inference methods (Wald, LRT, score tests)	Sections 1.3-1.4
3	2/5/2018 2/7/2018	Two-way Contingency tables Table structures 3 ways of comparing proportions (difference, relative risk, odds ratio) and their inferences	Sections 2.1-2.3 Homework #1 due
4	2/12/2018 2/14/2018	Two-way Contingency tables (cont'd) Chi-square tests ( $X^2$ , $G^2$ ) Following-up chi-square tests: (Pearson residuals, partitioning of $G^2$ ) Extension for I-J tables (local odds ratio, uncertainty coefficient U) Testing independence for ordinal data (linear or monotone trend) Fisher's exact test for small samples	Sections 2.4-2.6
5	2/19/2018 2/21/2018	Three-way Contingency tables Conditional vs. marginal tables Simpson's paradox Conditional vs. marginal odds ratios Conditional vs. marginal independence CMH test Homogeneous association for 2-2-K tables	Section 2.7
6	2/26/2018 2/28/2018	Generalized linear models Components of generalized linear models GLMs for binary data GLMs for count data Statistical inference and deviance Fitting generalized linear models	Sections 3.1-3.5 Homework #2 due
7	3/5/2018 3/7/2018	Logistic regression Interpreting logistic regression Inference for logistic regression Categorical predictors Multiple logistic regression	Sections 4.1-4.5



		Probability-based interpretations	
8	3/12/2018 3/14/2018	No class during spring break	Homework #3
9	3/19/2018 3/21/2018	Mid-term exam	
10	3/26/2018 3/28/2018	Building and applying logistic regression models Strategies in model selection Model checking Model diagnostics Effects of sparse data SAS Examples (AIDS, clinical trial with sparse data, grouped vs. ungrouped crab data)	Sections 5.1-5.3
11	4/2/2018 4/4/2018	Multicategory logit models Baseline-category logit models for nominal responses Cumulative logit model for ordinal responses Paired-category ordinal logits Tests of conditional independence	Sections 6.1-6.4
12	4/9/2018 4/11/2018	Loglinear models Loglinear models for count data Loglinear models for rate data Loglinear models for 2-way tables SAS Examples (belief in afterlife)	Sections 7.1-7.2
13	4/16/2018 4/18/2018	Loglinear models (cont'd) Loglinear models for 3-way tables Loglinear-logit model connection Modeling ordinal association (L-by-L association model) SAS Examples (drug use, auto injury, sex opinions)	Sections 7.3-7.5
14	4/23/2018 4/25/2-18	Models for matched pairs Comparing dependent proportions (McNemar's test) Kappa measure of agreement Bradley-Terry model for paired preference Symmetry and quasi-symmetry models for square tables	Sections 8.1, 8.4, 8.5.5, 8.6 Homework #4 due
15	4/30/2018	Review	
16	TBA	Final exam	