

STAT 645 – Categorical Data Analysis – Spring 2016

Instructor

	Office	Email	Office Hours
Darren Homrighausen	Stat 204	darrenho@stat.colostate.edu ¹	After class by request

Lecture TR 2:30 PM - 3:45 PM Weber 223J

Teaching assistant *Yewon Kim*

Required Text *Categorical Data Analysis* (Second Edition)
Alan Agresti

Web Site <http://www.stat.colostate.edu/~darrenho/CDA.html>

Prerequisites Some mathematical statistics/probability and a course on multiple regression/ANOVA.

Expectations As this is a 3 credit class, there is a CSU expectation that you spend 6 hours outside class on homeworks and review each week.

Categorical Data Analysis is an extension of methods for continuous valued data to discrete valued. By the end of this course, you should be able to analyze any type of data where a regression analysis is appropriate. This includes any combination of qualitative/quantitative predictors as well as qualitative/quantitative responses.

¹Please put “STAT645” at the beginning of the subject line.

Administrative Remarks

Lectures

Attendance. Attendance is mandatory. You can do it.

Software

SAS and R. In this class you will be provided the opportunity to work with two statistical packages. The first, R, is a free and widely used statistical computing platform. It is free, moderately well documented, and has a vast user community (download it from www.r-project.org). This is the primary software for this class. Alternately, SAS is commonly used by practicing statisticians (and is actually mandatory in some professions). I will be providing some code for doing various analyses in SAS.

Some types of analysis are more suitable to R. But, SAS is the industry workhorse for linear modeling and hence its interface is usually better suited to the tasks of this course. I have set up a free online version of SAS that you can use (please contact me if you have not gotten a link or are having problems).

Homework and Tests

Homework. There will be six assignments during the semester. Most homework sets will include theoretical, conceptual, and applied work. The topics for each assignment along with corresponding chapters from the text are listed in the following table.

Assignments

#	Chapter(s)	Topic
I	1	Distributions and Inference for Categorical Data
II	2	Descriptive Methods for Contingency Tables
III	3	Inferential Methods for Contingency Tables
IV	4	Generalized Linear Models
V	5, 6	Logistic Regression: Logit Models for Binary Responses
VI	7	Logit Models for Multicategory Responses
	8	Loglinear Models

Feel free to discuss homework assignments with others, but realize that the work you hand in must be your own. Homeworks will be equally weighted and comprise 42% of the course grade. *No late homeworks will be accepted, regardless of the reason they are late*

Quizzes. There will be three class-long quizzes (Feb. 11, Mar. 10, Apr. 28) comprising the entirety of lecture that day. Each quiz is worth 10% of the course grade.

Final Policy. There will be no final exam nor class during the final exam period. Near the end of the term, you will be required to perform a categorical data analysis on a data set and to prepare a written report summarizing your results. You may complete this project either on your own or in collaboration with another student. You can use a data set that you already have or I can provide one for you. Details concerning the project will be provided later in the semester.

Miscellaneous

Disability Resources. If you require a special accommodation, such as needing more time to finish exams, contact me as soon as possible.

Topics:

Distributions and Inference for Categorical Data (Chapter 1)

- Categorical Response Data
- Common Study Designs
- Binomial, Poisson, and Multinomial Distributions
- Overdispersion / Negative Binomial Distribution
- Likelihood Functions / Maximum Likelihood Estimation
- Wald, Score, Likelihood Ratio (LR) Tests
- Confidence Intervals Based on Test Inversion
- Inference for Binomial Parameters
- Inference for Multinomial Parameters / Pearson's and LR Chi-Squared Tests

Analysis of Contingency Tables (Chapters 2 and 3)

- Sampling and Probability Distribution Models
- Relative Risk, Odds Ratio, and Measures of Association for 2×2 Tables
- Conditional and Marginal Associations in Three-Way Tables
- Odds Ratios and Measures of Association for $I \times J$ Tables
- Confidence Intervals for Association Measures
- Testing Independence in Two-Way Tables
- Pearson Residuals
- Partitioning Chi-Squared Test Statistics
- Two-Way Tables Based on Ordinal Variables
- Fisher's Exact Test / Exact Tests of Independence for $I \times J$ Tables

Generalized Linear Models (GLM's) (Chapter 4)

- Components of the GLM
- GLM's for Binary and Count Data
- Moments, Likelihood, and Likelihood Equations for GLM's
- Inference for GLM's
- Deviance / Model Fit / Estimation of Dispersion Parameter
- Pearson and Deviance Residuals
- Maximum Likelihood / Newton-Raphson / Fisher Scoring
- Quasi-Likelihood Estimation
- Overdispersed GLM's and Quasi-Likelihood Estimation

Logistic Regression: Logit Models for Binary Responses (Chapters 5 and 6)

- Parameter Interpretation and Model Structure
- Logistic Regression with Case-Control Studies
- Inference and Model Fit
- Categorical Explanatory Variables
- Logit Models for $I \times 2 \times K$ and $2 \times 2 \times K$ Tables
- Model Selection / Akaike Information Criterion
- Measures of Predictive Power / Classification Tables and ROC Curves

Logit Models for Multicategory Responses (Chapter 7)

- Nominal Responses and Baseline Category Models
- Ordinal Responses / Cumulative Logit Models / Proportional Odds Models

Loglinear Models (Chapter 8)

- Loglinear Models for Two-Way and Three-Way Tables
- Conditional and Marginal Associations / Independence Relations

Further topics..