STAT:7510/BIOS:7410 Analysis of Categorical Data, Spring 2013

Home > **SYLLABUS**

Course Web Pages:	Start at http://www.stat.uiowa.edu/~jblang/s220 [username: *xxxx* password: *xxxx]*
Lecture:	MWF 2:30 - 3:20 30 SH
Instructor:	Joseph B. Lang, 207 SH, 335-3129, joseph-lang@uiowa.edu
Office Hours:	MW 11:20-12:20, W 3:20-4:30, or by appointment
Pre-Requisites:	STAT:4101 or STAT:5101, and STAT:5200 or BIOS:5620 (or equivalents)
Department, College:	Statistics and Actuarial Science, Liberal Arts and Sciences
DEO:	Luke Tierney, 241 SH, 335-0712, luke-tierney@uiowa.edu
Main Office:	241 Schaeffer Hall

Required Text	**Supplementary Texts**	**Course Description and Objectives**	**Course Organization**	**Course Outline/Pace**
Course Policies	**Grading**	**Misc**		

Required Text:

Agresti, A. 2002 .*Categorical Data Analysis*, 2nd edn. New York:Wiley. [Supporting website: http://web.stat.ufl.edu/~aa/cda/cda.html]

Supplementary Texts:

Agresti, A. (2007). *An Introduction to Categorical Data Analysis.* 2nd edn. New York: Wiley.

Agresti, A. (2010). *Analysis of Ordinal Categorical Data.* 2nd edn. New York: Wiley.

Aitkin, M., Anderson, D., Francis, B., and Hinde, J.(1989). *Statistical Modelling in GLIM.* Oxford Science Publications. Oxford: Clarendon Press.

Bishop, Y.M.M., Fienberg, S.E., and Holland, P.W. (1975). *Discrete Multivariate Analysis.* Cambridge, MA: MIT Press.

Breslow, N.E. and Day, N.E. (1980). *Statistical Methods in Cancer Research, Volume 1: The Analysis of Case-Control Studies.* Lyon, International Agency for Research on Cancer.

Christensen, R. (1990). *Log-Linear Models.* New York: Springer-Verlag.

Cox, D.R. and Snell, E.J. (1989). *Analysis of Binary Data,* 2nd Edn. London: Chapman and Hall.

Everitt, B.S. (1992) *The Analysis of Contingency Tables,* 2nd Edn. London: Chapman and Hall.

Fahrmeir, L. and Tutz, G. 1994. *Multivariate Statistical Modelling Based on Generalized Linear Models.* New York: Springer-Verlag.

Haberman, S.J. (1974). *The Analysis of Frequency Data.* Chicago: Univ of Chicago Press.

Le, C.T. (1998). *Applied Categorical Data Analysis.* New York: John Wiley and Sons.

Leonard, T. (2000). *A Course in Categorical Data Analysis.* Boca Raton, FL: Chapman and Hall/CRC.

Lloyd, C.J. 1999 *Statistical Analysis of Categorical Data.* New York: Wiley.

McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models,* 2nd edn., London: Chapman and Hall.

Powers, D.A. and Xie, Y. (2000). *Statistical Methods for Categorical Data Analysis.* San Diego, CA: Academic Press.

Stokes,M.E., Davis,C.S. and Koch,G.G. 2000. *Categorical Data Analysis Using the SAS System,* 2nd edn. Cary, NC: SAS Institute Inc.

Zelterman, D. (1999). *Models for Discrete Data.* Oxford: Oxford University Press.

Course Description:

This course surveys theory and methods for the analysis of categorical response and count data. The course begins with an overview of classic results and likelihood-based inference for categorical data analysis. Methods for describing and analyzing contingency tables are surveyed. These include loglinear modeling of association structures, the Cochran-Mantel-Haenszel approach to detecting conditional association, and multinomial-Poisson homogeneous likelihood-based modeling. Dichotomous response models such as the logistic and probit regression models will be described. Poisson regression models will be used to analyze rate data from event history studies. Ordinal and polytomous response models such as the cumulative and multinomial logit models will also be introduced. Depending on class interest, we will cover topics such as exact inference, retrospective and conditional logistic modeling, multinomial choice models, or more on log-linear modeling. [See the course outline below.]

The statistical package SAS (e.g. PROC FREQ, GENMOD, and LOGISTIC) and the freeware package R will be used in this course.

Course Objectives:

The student who successfully completes this course should have a reasonable grasp of the theoretical foundations of categorical data analysis. As examples, the student will be able to work with a variety of univariate and multivariate discrete distributions. The student will understand basic asymptotic techniques (e.g. multivariate CLT, continuous mapping theorems, and delta method) and will understand general maximum likelihood techniques, especially as they apply to categorical data analysis. The student will be familiar with a variety of methods for analyzing contingency tables and categorical response regression data, and understand in what settings they are applicable. The successful student will have a working knowledge of R and the SAS procedures PROC FREQ, GENMOD, LOGISTIC.

Course Organization:

Lecture. We will cover many of the topics introduced in Agresti (2002) and include several topics from Stokes, Davis, and Koch (2000). Although the majority of the meetings will be in the lecture/discussion format, some meetings will be set aside for student presentations and/or discussion.

Homework. Homework problems and short independent projects (applied, computational, and theoretical) will be assigned on a regular basis. Some will be handed in and graded. The textbook (Agresti 2002) includes a supporting website: http://web.stat.ufl.edu/~aa/cda/cda.html, which includes solutions to odd numbered problems.

Portfolios. You will include all of your worked exercises (graded and un-graded) in a portfolio. You will be asked to hand in your portfolio at least once during the semester. The quantity, quality, and neatness of the work in this portfolio will be assessed and will count toward your participation score.

Point-Earning Opportunities (PEO). On occasion, there will be in-class PEO's, some of which will be pre-announced. PEOs will be in the form of in-class exercises, minute papers,

and attendance checks. Missed PEOs cannot be made up unless you have a legitimate excuse. See Section VIII. Grading--Attendance and Examinations in the CLAS Student Academic Handbook.

Projects. By March 15th at the latest you will choose a journal article (and any supporting articles) to report on. The article(s) must be approved by the instructor. (A list of candidate papers will be provided; you should choose early because each student must report on a distinct paper.) Your report will consist of three parts: a 5-10 page summary, a computational component (e.g. small-scale simulation study), and a 15-20 minute oral presentation. Oral presentations will be given during the last two weeks of the semester. You must come prepared to field questions. You will hand in your written summary and computational component when you present. More details will be given during the semester.

Course Outline/Pace [approx number of lectures in brackets]:

I. Intro to Categorical Data Analysis (AA Chapter 1) [7]

II. Basic Asymptotic and Maximum Likelihood Tools (AA Sections 1.3, 3.1.5-3.1.7, 14.1) [4]

III. Analysis of Contingency Tables (MPH modeling, AA Ch. 2, 3, 10) [9]

IV Regression Models for Dichotomous and Count Responses (AA Chapters 4-6) [12]

- A. GLM for categorical data
- B. Dichotomous Response Models
- C. Poisson Regression

V. Polytomous Response Regression Models (AA 7) [6]

- A. Ordinal Response Models
- B. Nominal Response Models

VI. Loglinear Models for Contingency Tables (AA Ch. 8-9) [4]

VII. Special Topics (one or more of Exact Inference, Logistic for Retrospective Data, Exact Conditional Logistic, Multinomial Choice Modeling, more on Log-Linear Models) [3]

Course Guidelines and Policies:

Course Web Page. Reading assignments, announcements, homework, exam descriptions, and supplementary materials will be made available on the password-protected course web page; start at http://www.stat.uiowa.edu/~jblang/s194 [not on ICON!]. You should check the course pages daily.

Reading Ahead. It is vitally important that you read ahead. If the material in a lecture is completely new to you, you will find it very difficult to get much out of lecture.

Participation and Attendance. Students are expected to attend, and participate in, class. You will be asked many questions, and you will be strongly encouraged to ask lots of questions. Your portfolio will also count toward your participation score.

Working Together. Unless instructed otherwise (e.g. for the two class projects), you may work together on the homework problems. However, you must write up your own solutions

in your own words. If you are personally asked to please write up your own solutions and subsequently turn in material that is obviously in the same words as a fellow student, the work will be considered to be plagiarized. Plagiarism will be dealt with according to the policies of the University.

Late Homework. Unless otherwise instructed, homework is due at 2:30 PM. Late homework has a half-life of 24 hours; that is you get 50% credit if it is handed in late, but within 24 hours of the due time; you get 25% credit for the next 24 hours; etc. Homework not handed in directly to me must be handed in to a department secretary (located in 241 SH). The homework must include a hand-in time and date, and must be signed by the department secretary. (It follows that you cannot hand in homework after the main office is closed.)

Grading Questions. Questions about grading must be asked within one week of the graded work's return.

Electronic Etiquette. While in the classroom, you will not be allowed to send or check text messages, send or check email, browse the web, or use a cell phone. Social networking of any kind is not allowed. Please keep cell phones in your bag/backpack. If your cell phone is visible, it will be taken from you and placed in the front of class until the period has ended.

Grading and Components for Evaluation

Your final score $\boldsymbol{S}$ will be computed as $\boldsymbol{S} = 0.55\boldsymbol{H} + 0.25\boldsymbol{P_1} + 0.20\boldsymbol{P_2}$, where $\boldsymbol{H}$ = percent credit on homework, $\boldsymbol{P_1}$ is your score on the project (written + computational + oral) and $\boldsymbol{P_2}$= participation score (which includes credit on portfolio and PEOs).

Letter grades (including +'s and -'s) will be awarded according to a 90-80-70-60 schedule (e.g. if $\boldsymbol{S} \geq 90$ then a grade of A- or better will be awarded). These are guaranteed cutoffs, so it is possible (but unlikely) that everyone receives an 'A.' I do, however, reserve the right to lower (but not raise) the cutoffs. Note that with this grading scheme you are not "graded on a curve," and so you are not competing with fellow students. Therefore, you are not penalized for working together to better understand concepts.

Miscellaneous (Help and General Policies)

Textbook (Agresti 2002) support:

Start at http://web.stat.ufl.edu/~aa/cda/cda.html. (Includes datasets and solutions to odd numbered problems.)

Help outside of class:

I have regular office hours. Sometimes it is effective to ask specific questions via email.
Course web pages; start at http://www.stat.uiowa.edu/~jblang/s220.

Help with SAS:

- Creating (temporary and permanent) SAS data sets

- Some helpful links...
    - SAS in a nutshell, SAS basics for Windows, SAS INSIGHT (pdf file), SAS INSIGHT intro, SAS basics (inc. INSIGHT)
    - or go to search engine Google, enter the key words "SAS basics".
- Use SAS Options statement
- Using SAS ODS (Output Delivery System)

Help with R software:

- SimpleR. Go to http://www.math.csi.cuny.edu/Statistics/R/simpleR/index.html
- An Introduction to R, by Elizabeth Slate and Elizabeth Hill.
- R for CDA. For example... http://www.stat.ufl.edu/~presnell/Courses/sta4504-2000sp/R/R-CDA.pdf or http://statistics.unl.edu/faculty/bilder/stat875/

College of Liberal Arts and Sciences: Policies and Procedures

Administrative Home
The College of Liberal Arts and Sciences is the administrative home of this course and governs matters such as the add/drop deadlines, the second-grade-only option, and other related issues. Different colleges may have different policies. Questions may be addressed to 120 Schaeffer Hall, or see the CLAS *Student Academic Handbook.*

Electronic Communication
University policy specifies that students are responsible for all official correspondences sent to their University of Iowa e-mail address (@uiowa.edu). Faculty and students should use this account for correspondences. (*Operations Manua*l, III.15.2 Scroll down to k.11.)

Accommodations for Disabilities
A student seeking academic accommodations should first register with Student Disability Services and then meet privately with the course instructor to make particular arrangements. See www.uiowa.edu/~sds/ for more information.

Academic Fraud
Academic fraud, including plagiarism and other forms of cheating, is a serious matter and is reported by the instructor to the departmental DEO and to the Associate Dean for Undergraduate Programs and Curriculum. All students in the College of Liberal Arts and Sciences should review and understand the CLAS Code of Academic Honesty.

CLAS Final Examination Policies
Final exams may be offered only during finals week. No exams of any kind are allowed during the last week of classes. Students should not ask their instructor

to reschedule a final exam since the College does not permit rescheduling of a final exam once the semester has begun. Questions should be addressed to the Associate Dean for Undergraduate Programs and Curriculum.

Making a Suggestion or a Complaint
Students with a suggestion or complaint should first visit the instructor, then the course supervisor, and then the departmental DEO. Complaints must be made within six months of the incident. See the CLAS Student Academic Handbook.

Understanding Sexual Harassment
Sexual harassment subverts the mission of the University and threatens the well-being of students, faculty, and staff. All members of the UI community have a responsibility to uphold this mission and to contribute to a safe environment that enhances learning. Incidents of sexual harassment should be reported immediately. See the UI Comprehensive Guide on Sexual Harassment for assistance, definitions, and the full University policy.

Reacting Safely to Severe Weather
In severe weather, class members should seek appropriate shelter immediately, leaving the classroom if necessary. The class will continue if possible when the event is over. For more information on Hawk Alert and the siren warning system, visit the Public Safety web site.

*These CLAS policy and procedural statements have been summarized from the web pages of the College of Liberal Arts and Sciences and The University of Iowa *Operations Manual*.

University Examination Policy

Final Examinations. An undergraduate student who has two final examinations scheduled for the same period or more than three examinations scheduled for the same day may file a request for a change of schedule before the published deadline at the Registrar's Service Center, 17 Calvin Hall, 8-4:30 M-F, (384- 4300).

Missed exam policy. University policy requires that students be permitted to make up examinations missed because of illness, mandatory religious obligations, certain University activities, or unavoidable circumstances. Excused absence forms are required and are available at the Registrar web site:
http://www.registrar.uiowa.edu/forms/absence.pdf

I hope you all have an enjoyable and successful semester. Good luck in all of your courses.

This page was last updated: 1/23/13 (Joseph B. Lang)