BIOSTATS 743 (3 credits)
Categorical Data Analysis
Fall 2018 :: T/Th 1:00-2:15 :: LGRT 141

INSTRUCTOR
Nicholas G Reich
425 Arnold House
(413) 545-4534
nick [at] umass.edu
twitter: @reichlab
web: reichlab.io
Office Hours: Tuesday 2:30-3:30pm
Teaching assistant: Zhengfan Wang

MATERIALS
Required Textbooks
  Categorical Data Analysis, Author: Alan Agresti, Publisher: Wiley, 3rd Edition.
Recommended Textbooks
  Andrew Gelman and Jennifer Hill. (2007). *Data Analysis using Regression and Multilevel/Hierarchical Models*. Cambridge University Press.

PREREQUISITES
BIOSTATS 540 (Intro Biostatistics), STAT 515 (Intro Statistics I), STAT 516 (Intro Statistics II), BIOSTATS 650/STAT 525 (both of these are courses on linear regression), or equivalent coursework. Prior programming experience is required. If you have not taken some of the listed pre-requisites but still think the course is the right choice for you, you may petition the instructor directly for permission to enroll in the course.

COURSE DESCRIPTION
This course provides an overview of statistical methods for analyzing data where the outcome variable is categorical or discrete. The course will emphasize the theoretical underpinnings of the methods as well as an applied understanding of the computation and interpretation, both of which are necessary to succeed with real data analysis. We will cover inference for binomial and multinomial variables with contingency tables, generalized linear models, logistic regression for binary responses, logit models for multiple response categories, log-linear models, some machine learning approaches, inference for matched-pairs, and correlated/clustered data. Examples will be taken from public health and biomedical research. Students will be evaluated on homework assignments, a mid-term exam and a final project.

SOFTWARE
The textbook website lists the datasets used for examples, solutions to some exercises, information about using R, SAS, Stata, and SPSS software for conducting the analyses in the text. I will use R throughout the course. Students are allowed to use the language of their choice for all assignments and projects. The instructor will only be able to provide support for assignments done in R and solutions will be given in terms of R only. R is freely available at . R is also installed at every OIT classroom. See for more information.

EXPECTATIONS
Things you should expect from me:
- response to questions via Piazza in < 2 working days (often sooner)
- attention to your questions related to coursework during office hours
- explanations about the statistical methods we cover in class
- instruction in how to write, research, and debug R code

Things you should not expect from me:
- time for frequent non-office hour drop-in questions
- comments on a research project that is unrelated to your coursework
- writing your code for you or *extensive* debugging of your code

TYPES OF ASSIGNMENTS AND ACTIVITIES, WITH GRADE CONTRIBUTIONS

Homework (20%): You will have homework assignments due about every other week throughout the semester. Some parts of the assignments will require you to submit a digital file with reproducible solutions, e.g. a knitr file that reproduces your answers. The homeworks will graded on a discrete scale: Complete (10 points), Submitted but incomplete/inadequate (5 points), Not submitted (0 points). Convincing and thorough attempts at all assigned homework problems will qualify as a "complete" submission. Late homeworks will not be accepted under any circumstances. If a homework is not handed in on time, it will receive a grade of zero.

Midterm exam (25%): The exam will be in-class and closed book. Calculators are not allowed. Make-up exams will only be given for legitimate, documented reasons and when a call has been made to the instructor or the department before the exam.

Scribing (15%): Twice throughout the semester, each student will be required to transcribe the notes from the white-board in class into an RMarkdown file. Additionally, each student will serve twice as a peer-reviewer for the notes scribed by another student. Each set of transcribed notes will be evaluated for accuracy, completeness, formatting, and readability. Students are encouraged to add or embellish the content presented in class with additional examples or explanations that they think others would find helpful in understanding the material. The notes must follow the format used in the notes lectures 1 and 2. (Templates available on the GitHub course repository.) Final draft of notes are due at 5pm one week after the class during which they were transcribed. Notes must be submitted via pull-request to the course GitHub repository.

Participation/citizenship (10%) : Being a good class "citizen" plays a large role in your final grade. A few of the characteristics of good class citizens are: attending all course meetings, using office hours, asking questions, offering to answer questions, actively listening when others are talking, correcting typos or errors in course materials on GitHub, and participating on Slack (both asking and answering questions). Citizenship is more a function of quality than quantity. The "default" citizenship score is 5 out of 10. [Acknowledgments to Aaron Swoboda for introducing me to the concept of course citizenship and for some of this text.]

Final Project (30%): A large component of the course will be an independent project which will be presented to your classmates. A separate handout will provide details.

Extra Credit: If you send me an email with "I read the syllabus" as the subject line by the beginning of the second class, you will receive one point of extra credit on your final grade.

COURSE POLICIES

Collaboration on homework is expected and encouraged, although you must write up your own assignment. No copying or cutting and pasting. The exam must be completed without assistance from your classmates. Your independent projects must be your own work. You may discuss your project with others and even solicit ideas and advice, but at the end of the day, you must complete all the analysis and write-up on your own. Any explicitly borrowed ideas must be cited appropriately.

Late assignments: Completing homework assignments on time will be vital to not falling behind in this course. It is expected that you hand in assignments on time. If an assignment is handed in late, you will receive zero credit for that assignment. I will drop your lowest homework grade at the end of the semester.

Make-up exam: Make-up exams will only be given for legitimate, documented reasons and when a call has been made to the instructor or the department before the exam.

Attendance is required. Absences (excused or not) will impact your participation grade.

All mobile devices that can/will be distracting to you or others during class must be turned off at the start of class and may not be used during class time.

Course Schedule

Week	Topic
1	Introduction: Distributions and Inference for Categorical Data
2	Contingency Tables
3	Introduction to Generalized Linear Models
4-5	Logistic Regression
6	Alternative Modeling of Binary Response Data
7	Models for Multinomial Responses
8-9	Loglinear Models for Count Data
10	Models for Matched Pairs
11-12	Clustered Categorical Data: Random Effects Models
13	Other Mixture Models for Discrete Data

Grading Scale

Grade	Percentage
A	93-100
A-	90-92
B+	87-89
B	83-86
B-	80-82
C+	77-79
C	70-76
F	0-69

ACADEMIC HONESTY POLICY STATEMENT Since the integrity of the academic enterprise of any institution of higher education requires honesty in scholarship and research, academic honesty is required of all students at the University of Massachusetts Amherst.

Academic dishonesty is prohibited in all programs of the University. Academic dishonesty includes but is not limited to: cheating, fabrication, plagiarism, and facilitating dishonesty. Appropriate sanctions may be imposed on any student who has committed an act of academic dishonesty. Instructors should take reasonable steps to address academic misconduct. Any person who has reason to believe that a student has committed academic dishonesty should bring such information to the attention of the appropriate course instructor as soon as possible. Instances of academic dishonesty not related to a specific course should be brought to the attention of the appropriate department Head or Chair. The procedures outlined below are intended to provide an efficient and orderly process by which action may be taken if it appears that academic dishonesty has occurred and by which students may appeal such actions.

Since students are expected to be familiar with this policy and the commonly accepted standards of academic integrity, ignorance of such standards is not normally sufficient evidence of lack of intent. For more information about what constitutes academic dishonesty, please see the Dean of Students? website.

DISABILITY STATEMENT The University of Massachusetts Amherst is committed to making reasonable, effective and appropriate accommodations to meet the needs of students with disabilities and help create a barrier-free campus. If you are in need of accommodation for a documented disability, register with Disability Services to have an accommodation letter sent to your faculty. It is your responsibility to initiate these services and to communicate with faculty ahead of time to manage accommodations in a timely manner. For more information, consult the Disability Services website.