UNIVERSITY OF KENTUCKY
COLLEGE OF ARTS AND SCIENCES

Course Syllabus
STA 665-002 Analysis of Categorical Data
Spring 2021

This is a fully online course which will have asynchronous and synchronous components. Lectures will be delivered asynchronously. You will be given access to lecture notes, plus accompanying video or audio recordings, on a regular basis. You should read the lecture notes and listen to the recordings within one week of their availability. You will be required to have at least seven synchronous meetings with me during the semester, and you may choose to have more. These synchronous meetings can be at mutually agreeable times.

Contact information

Instructor: Dr. Richard Charnigo

Telephone: Because of COVID-19, I am not in the office this semester. E-mail is the best way to reach me.

E-mail: richard.charnigo@uky.edu Please include "STA 665" in the subject line. I try to respond to e-mail (when a response is called for) within one business day.

Office Hours: In addition to required synchronous meetings, I am available for extra Zoom sessions {uky.zoom.us} with individual students, or groups of students, upon reasonable request. Please e-mail me to schedule.

Course description

Multinomial and product-multinomial models; large-sample theory of estimation and testing, Pearson chi-square and modified chi-square statistics, Pearson-Fisher Theorem, Wald Statistics and generalized least squares technique; applications to problems of symmetry, association and hypotheses of no interaction in multi-dimensional contingency tables.

Instructor's remarks about this course

Measuring instruments have finite precision and computers have finite memory. Thus, one may make the provocative assertion that all data are categorical. However, when we speak of categorical data analysis, we usually have in mind that there exists an outcome variable for which fewer than ten – and, sometimes, only two – values are possible. This three credit hour course will help the learner develop practical skills for some of the most frequently employed categorical data analysis techniques, including: logistic regression, Poisson regression and common variants thereof, generalized logit and proportional odds modeling, generalized estimating equations, generalized linear mixed modeling, meta analysis, and others as time permits. Since this course is directed to graduate students in Statistics, the presentation of material will be mathematical. At the same time, students will practice communicating their data analysis findings as if they were writing to an audience which is scientifically literate but not statistically trained.

Course prerequisites

STA 603 and STA 606.

Course Objectives

At the conclusion of this course, you should be able to:

1. fit and interpret the results from a logistic regression model;
2. fit and interpret the results from Poisson, zero-inflated Poisson, and negative binomial regression models;
3. fit and interpret the results from generalized logit and proportional odds models;
4. employ generalized estimating equations to estimate marginal relationships in the presence of clustering;
5. employ generalized linear mixed modeling to estimate conditional relationships in the presence of clustering;
6. perform and understand the computations for a meta analysis; and,
7. express your findings to an audience which is scientifically literate but not statistically trained.

Textbooks and Other Materials

1. The following textbook is required:

Agresti, Alan (2012). *Categorical Data Analysis*, 3rd edition. John Wiley & Sons, Incorporated.

You may access this book online for free, using UKY's institutional subscription. Go to {http://libraries.uky.edu} and use search terms including the edition number.

Please note, however, that the number of pages which can be downloaded in PDF format is limited. I suggest that you give priority to Chapter 5, Chapter 6, and whatever section is relevant to your individual project.

2. Materials for this course will be posted on Canvas. Please check for updates frequently.

Course requirements and learner evaluation

Team Projects: There will be three team projects for you to complete. The team projects will be due at 11:59 p.m. (Lexington KY local time) on the Tuesdays of 09 March, 30 March, and 20 April.

You may select the teams yourselves, but each team is to consist of two or three enrolled students. You may discuss the team projects with members of other teams, but each team must prepare and is responsible for its own submitted work.

Each team member must contribute to the problem-solving process on each item within a project and be able to defend the response which is submitted for grading; a "divide and conquer" approach, whereby the different team members assume sole responsibility for various subsets of items, is not acceptable. (This is not to say that team members need to write redundant R or SAS code; however, please do not simply assign item 1 to person A, item 2 to person B, and so forth.) Please acknowledge any classmates whose insights informed your work, if those classmates are not members of your team.

For each of the team projects, each team is required to have two Zoom meetings with me: once before the project is formally submitted for grading (to ask questions about items and/or request my comments on draft solutions) and once afterward (to receive my comments on submitted solutions and a grade). The first meeting should be within 14 days of the due date, and the second meeting should be within 7 days following the due date. Plan on 40-60 minutes for the first meeting and 20-30 minutes for the second meeting.

You are not limited to the required Zoom meetings; I am available for other Zoom meetings upon reasonable request, to discuss team projects, answer questions about lecture notes or recordings, or attend to other course-related matters.

Team projects must be typed and submitted by e-mail as PDF files. Please include "STA 665" in the subject line of the e-mail.

Individual Project: This individual project will take the place of a final examination. You will select a topic not covered by me (options appear below) and prepare a lecture of length 40 – 60 minutes. You will record your lecture using Zoom and provide me with the hyperlink by 11:59 p.m. on Tuesday 11 May. (If a later deadline is required by the registrar's final examination schedule, then I will make an adjustment.) The hyperlink will be shared with other members of the class on Canvas. This means that other students may view your lecture, and you may view theirs. I will not impose any formal requirement for you to comment on or even listen to other students' lectures. However, you are encouraged to view at least some of them.

You are required to have a Zoom meeting with me, sometime between Thursday 08 April and Thursday 29 April, at which you can ask me questions about your topic and/or ask me to review a draft version of your lecture notes or slides. Plan on 40 – 60 minutes for this meeting.

Some options are as follows; others may be proposed but are subject to my prior approval. Topics will be assigned on a first come, first served basis. Students taking my CPH 636 course are requested not to choose one of the topics marked with an asterisk, since those topics will be addressed in CPH 636.

Section 7.1 Probit and Complementary Log-log Models
Section 7.2 Bayesian Inference for Binary Regression
*Section 7.4 Smoothing: Kernels, Penalized Likelihood, Generalized Additive Models
Section 8.5 Discrete Choice Models
Section 11.5 Measuring Agreement Between Observers
Section 14.1 Latent Class Models
*Section 15.1 Linear Discriminant Analysis
*Section 15.2 Tree-Structured Prediction
*Section 15.3 Cluster Analysis for Categorical Data

Grading Components: The three team projects will count for 100 points (90 points for content, 10 for attending the two required synchronous meetings), and the individual project will count for 100 points (90 points for content, 10 for attending the one required synchronous meeting). Some extra credit may be available. The thresholds will be 90%, 75%, and 60% for "A", "B", and "C" letter grades respectively, although an outstanding individual project may result in the higher letter grade if you just barely miss the threshold. There will not be a "D" letter grade because this is a graduate course. Actually, no one should earn less than a "B". (I do not promise that outcome, but I expect it.) This is an elective course and is meant to strengthen your data analysis and communication skills, not stress you out.

Instructor expectations

1. I expect you to listen to all of my video/audio commentaries (and review the corresponding lecture notes) within one week of their being posted on Canvas.

2. Please check the e-mail address under which you registered for the course regularly. As a courtesy, I will add alternate e-mail addresses to my mailing list upon request. You are responsible for all material and announcements conveyed by e-mail; a full mailbox or bouncing of messages by your e-mail provider does not remove this responsibility.

3. You are encouraged to ask questions by e-mail. Besides the required Zoom meetings, you may request appointments with me on Zoom. Prior permission from me (and from any other attendees, if applicable) is required for a student to initiate recording of a Zoom meeting.

4. Grading of written work will be based primarily on appropriateness of concept or methodology, technical accuracy or logic, and soundness of conclusions. I may also consider clarity, succinctness, and adherence to appropriate conventions of English language.

5. If you wish to appeal my grading, you may present an appeal in writing (by e-mail). However, this must be done within one week of the time my grading is conveyed to you.

6. You are not expected to read, word-for-word, every section of the textbook mentioned in the video/audio commentaries. Read what you need to read for better understanding. If you are not sure what to read, then ask me for advice.

Other Helpful Information

Please view material on excused absences, accommodations due to disability, and plagiarism/cheating at {https://www.uky.edu/universitysenate/helpful-components-course-syllabus}. If religious observances affect your ability to complete a project by the appointed time, please let me know at least two weeks in advance of the project deadline.

Classroom Recording and Copyright Statement

The University of Kentucky Code of Student Conduct defines Invasion of Privacy as using electronic or other devices to make a photographic, audio, or video record of any person without their prior knowledge or consent when such a recording is likely to cause injury or distress.

Prior permission from me (and from any other attendees, if applicable) is required for a student to initiate recording of a Zoom meeting. I will not ordinarily initiate recording of Zoom meetings myself.

All video and audio recordings provided or initiated by me are not to be shared with those not enrolled in the class nor uploaded to other online environments. However, you may download them to a location accessible only to you.

Students with specific recording accommodations approved by the Disability Resource Center should present their official documentation to me.

Course Copyright

All original instructor-provided content for this course, which may include handouts, assignments, and lectures, is the intellectual property of the instructor. Students enrolled in the course this academic term may use the original instructor-provided content for their learning and completion of course requirements this term, but such content must not be reproduced nor sold. Students enrolled in the course this academic term are hereby granted permission to use original instructor-provided content for reasonable educational and professional purposes extending beyond this course and term, such as studying for a comprehensive or qualifying examination in a degree program, preparing for a professional or certification examination, or to assist in fulfilling responsibilities at a job or internship; other uses of original instructor-provided content require written permission from the instructor in advance.

Any use of another student's recorded lecture (from the individual project) besides your own viewing requires that student's prior permission.

Attendance Policy

Attendance at required synchronous meetings is incorporated into the grade as described above.

Late work policy

Cases involving the following will be handled individually: excused absences (including religious observances), University-prescribed academic accommodations, and recommendations for special consideration from the office of an appropriate Dean or the Ombud.

Otherwise, the team projects and individual project will be accepted up to 24 hours past their respective deadlines without question or penalty. Submissions after 24 hours but within 48 hours will be accepted subject to a 25% penalty. Submissions after 48 hours will not be accepted.

Note: I understand that life is different with COVID-19. If you have difficulty keeping up with course requirements for a reason related to COVID-19 which does not fall under the University policy on excused absences, please let me know. I do not promise to agree to every request (for extension, accommodation, etc.), but I will try to be reasonable.

Other course-related information

If an unforeseen contingency arises that requires a new course policy, or if some clarification is warranted, then I will make an appropriate announcement.

Course schedule and topics

This schedule is very tentative and subject to change, because this is my first time teaching STA 665. If our pace is faster than anticipated, then additional material may be drawn from Chapter 11 (Sections 1 – 2). If our pace is slower than anticipated, then material from Chapter 9 (Sections 1 – 3) may be omitted.

Week of 25 January: Chapter 5 (Sections 1 – 2)
    Introduction to logistic regression
Week of 01 February: Chapter 5 (Sections 3 – 5)
    Some fine points of logistic regression
Week of 08 February: Chapter 6 (Sections 1 – 2)
    Model selection and diagnostics
Week of 15 February: Chapter 6 (Sections 3, 5 – 6)
    Predictive ability, complete separation, and sample size
Week of 22 February: Chapter 4 (Sections 1 – 2, 4 – 6)
    Generalized linear modeling
Week of 01 March: Chapter 4 (Section 3) and Chapter 14 (Section 4)
    Poisson, zero-inflated Poisson, and negative binomial regression
Week of 08 March: Chapter 8 (Section 1)
    Generalized logit regression
Week of 15 March: Chapter 8 (Section 2)
    Proportional odds regression
Week of 22 March: Chapter 12 (Section 1)
    Marginal modeling and maximum likelihood
Week of 29 March: Chapter 12 (Section 2)
    Generalized estimating equations
Week of 05 April: Chapter 13 (Section 1)
    Random effects
Week of 12 April: Chapter 13 (Sections 2 – 3)
    Generalized linear mixed modeling
Week of 19 April: Chapter 6 (Section 4)
    Mantel-Haenszel and meta-analysis
Week of 26 April: Chapter 9 (Section 1)
    Log-linear modeling for two-way tables
Week of 03 May (Dead Days and Reading Days): Chapter 9 (Sections 2 – 3)
    Log-linear modeling for three-way tables
Week of 10 May (Finals Week): No new topics
    Submit individual project