BIOS 665: Analysis of Categorical Data

Course Syllabus
Fall 2019

Meeting Times

Lectures:
Tuesdays & Thursdays, 11:00am-12:15pm, TBA

Recitation Session Hours:
Tuesdays 3:30-4:30pm, McGavran-Greenberg 2308
Wednesdays 4:00-5:00pm, McGavran-Greenberg 1304

Office Hours:
Tuesdays 10:00-11:00am, McGavran-Greenberg 3106
Tuesdays 12:30-1:30pm, McGavran-Greenberg 3106

Instructor Information

Instructor: Todd Schwartz, Dr.P.H.
Office: 3106E McGavran-Greenberg Hall
Email: Todd_Schwartz@unc.edu

Instructor: Gary G. Koch, Ph.D.
Office: 3106B McGavran-Greenberg Hall
Phone: 919-966-7282

Coordinating Teaching Assistant:
Reuben Adatorwovor (radat223@email.unc.edu)

Required Text

Stokes ME, Davis CS, Koch GG. *Categorical Data Analysis Using the SAS System*, 3rd ed. SAS Institute Inc., Cary, NC, 2012. (SDK)

Prerequisites

BIOS 550, BIOS 545, and BIOS 662, or equivalent

Website

sakai.unc.edu (assignments, datasets, and exams will be posted here)

Disclaimer

The professors reserve the right to make changes to the syllabus, including dates of coverage for each topic and exam dates. These changes will be announced as early as possible.

Overview

This course presents the conceptual background and computational procedures for statistical methods for categorical data analysis. Attention is given to two complementary strategies:

1) nonparametric (randomization) methods (e.g., Mantel-Haenszel tests) for testing hypotheses of no association under minimal assumptions;

2) regression methods for fitting statistical models to describe multivariate relationships (e.g., logistic regression, Poisson regression, weighted least squares regression, conditional logistic regression, generalized estimating equations).

Consideration of these strategies is motivated through examples from clinical trials, observational studies, and sample surveys. For these examples, the roles for alternative methods and aspects of application are discussed from the perspective of the questions which statistical analyses are to address, the posture of inference, the sampling process, and the data structure.

The specific topics for the course are as follows: logistic regression; Mantel-Haenszel procedures for sets of 2 x 2 contingency tables; proportional odds model extension of logistic regression for ordinal data; extensions of Mantel-Haenszel procedures for stratified ordinal data; weighted least squares methods for ordinal data; Poisson (incidence density) regression methods for categorized times to event; methods for studies with repeated measures and/or matching including generalized estimating equations and conditional logistic regression; computing procedures for implementing methods.

Honor Code at the University of North Carolina

The Honor Code and the Campus Code, embodying the ideals of academic honesty, integrity, and responsible citizenship, have for over 100 years governed the performance of all academic work and student conduct at the University. Acceptance by a student of enrollment in the University presupposes a commitment to the principles embodied in these codes and a respect for this most significant University tradition. Your participation in this course comes with the expectation that your work will be completed in full observance of the Honor Code. Academic dishonesty in any form is unacceptable, because any breach in academic integrity, however small, strikes destructively at the University's life and work.
For additional information about the honor code, please also refer to the University website: https://studentconduct.unc.edu/sites/studentconduct.unc.edu/files/documents/Instrument.pdf

Class participation

There is no formal grading component for class participation, but there are occasions when the instructor will ask for class participation. Additionally, questions are welcomed at any time. Students are encouraged to participate accordingly.

Technology policy

Students are encouraged to use laptop computers/devices for the purposes of taking notes and other class related activities. Use of laptops/devices for non-class related activities is not permitted.

Late Homework

Every effort should be made to submit homework by the due date. With prior permission, late homework will only be accepted until one week past the due date. However, late homework is subject to delays in grading and being returned, as this will be low priority for the graders. Homework submitted later than one week past the due date may not be graded. In summary, to ensure timely grading and return, students should strive to submit their homework problem sets on the assigned due date.

Examination policy

All examinations should be completed independently, with no outside assistance. Any questions should be directed to the instructors only; e.g., not to classmates, teaching assistants, or graders.

The mid-term examination will be assigned on a timed, in-class basis for BIOS majors and those seeking to earn an H for the course. The same mid-term examination may be completed on a take-home basis for all others.

The final examination is given on a take-home basis for all students. It is open book, and you are expected to use SAS or comparable software to complete the exam. BIOS majors and those seeking an H may be asked to complete a greater number of problems than all others. This will be clearly specified when the examination is assigned.

Due to university policy, our course is required to meet during the university-assigned final exam period. Please plan to attend the beginning of that time period in order to submit the final exam. If this policy causes you considerable hardship, we would ask that you contact both instructors as soon as possible. Exemptions will need to be approved by the Chair of the Department of Biostatistics or by the Dean of the Gillings School of Global Public Health; instructors cannot exempt students from this requirement without such approval.

Grading policy

Student grades are computed based on their performance on the mid-term and final examinations. Those seeking an H grade for the course will need to perform sufficiently well on both exams (e.g., averaging 85% or above) to justify this grade. All others will be expected to perform satisfactorily on both exams, as well as having satisfactory homework (e.g., averaging 70% or above) to earn a course grade of P. Any unsatisfactory work will be subject to grades lower than a P (i.e., L or F), or instructors may ask students to rework and resubmit such unsatisfactory exercises.

Course Evaluation

The course evaluation is scheduled to be available online during the final weeks of the semester. We encourage and expect students to complete the evaluation and provide comments to assist us in structuring the course in a helpful manner. We appreciate your participation in this process and your thoughtful feedback.

Date	SDK Chapter	Topic Description
08/20	Chapter 1	Introduction to analysis of categorical data: scale of measurement, sampling frameworks, analysis strategies, contingency tables
	Chapter 2	The 2×2 table: hypothesis testing, exact methods, difference in proportions, measures of association, sensitivity and specificity, McNemar's test for matched pair data
08/22	Chapter 2	
08/27	Chapter 2	
08/29	Chapter 2	
	Sample Size	Based on Normal distribution, based on Wald statistic, based on counterpart to Pearson's chi-square, two-sided tests, unequal sample sizes
09/03	Chapter 8	Logistic regression I: dichotomous response; dichotomous, nominal, continuous, and ordinal explanatory variables; model fitting; goodness of fit; testing hypotheses; maximum likelihood estimation; using LOGISTIC and GENMOD
09/05	Chapter 8	
09/10	Chapter 8	
09/12	Chapter 8	
09/17	Chapter 10	Conditional logistic regression: paired observations from cohort study, paired observations in retrospective matched study, 1:1 matching, 1:*m* matching
09/19	Chapter 10	
09/24	Chapter 3	Sets of 2×2 tables: Mantel-Haenszel test, measures of association
09/26	Chapter 3	
10/01	Chapter 4	Sets of 2×*r* and *s*×2 tables: comparing two groups with ordered columns, comparing ordered groups with dichotomous response
10/03	Chapter 4	
10/08	Chapter 9	Logistic regression II: polytomous response, proportional odds model, generalized logits model, model fitting

Date	SDK Chapter	Topic Description
10/10	Chapter 9	
10/15	**Mid-Term**	**In-Class / Take-Home Mid-Term Exam** (Covers Chapters 1, 2, 3, 4, 8, 9, 10, and Sample Size)
10/17	**Fall Break**	**No Class**
10/22	Chapter 11	Quantal bioassay analysis: estimating tolerance distributions, comparing two drugs, analyzing example data
10/24	Chapter 11	**Take-Home Mid-Term Exam Due in Class at 11:00AM**
10/29	Chapter 12	Poisson regression
10/31	Chapter 12	
11/05	Chapter 13	Categorized time-to-event data: life table methods, Mantel-Cox test, Poisson regression, piecewise exponential model
11/07	Chapter 15	Generalized Estimating Equations (GEE)
11/12	Chapter 15	
11/14	Chapter 15	
11/19	Chapter 5	The $s \times r$ table: tests for association, measures of association, exact test, observer agreement
11/21	Chapter 6	Sets of $s \times r$ tables: Mantel-Haenszel statistic in matrix terminology, applications, repeated measures analysis
11/26	Chapter 6 Chapter 7	Nonparametric methods: Kruskal-Wallis test, Friedman's chi-square test, rank analysis of covariance **Final Exam Assigned by This Date**
11/28	**Thanksgiving Break**	**No Class**
12/03	Chapter 14	Analysis using Weighted Least Squares methods **Question & Answer Period for Final Exam**
12/12 (Thursday)	**Final Exam**	**Due at 12:00 NOON (Room TBA)** **Attendance is <u>required</u> at the session to submit exams**