BIOS:7410/STAT:7510 Analysis of Categorical Data

Spring Semester, 2018
MW, 1:30 p.m. – 2:50 p.m., C410 CPHB

Instructor: Joseph (Joe) Cavanaugh
Departmental Executive Officer, Biostatistics
Professor of Biostatistics
Professor of Statistics and Actuarial Science
Professor of Applied Mathematical and Computational Sciences

Office: N332A CPHB
Office Hours: Wednesdays: 3:30 p.m. – 5:00 p.m.; Fridays: 3:00 p.m. – 4:00 p.m.
Phone: 384–1602
e–mail: joe-cavanaugh@uiowa.edu
web: http://myweb.uiowa.edu/cavaugh/


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 $2 \times 2$ 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
- Multiple Hypergeometric Distribution / Exact Tests of Independence for $I \times J$ Tables

Generalized Linear Models (GLMs) (Chapter 4)
- Components of the GLM
- GLMs for Binary and Count Data
- Moments, Likelihood, and Likelihood Equations for GLMs
- Inference for GLMs
Deviance / Model Fit / Estimation of Dispersion Parameter
Pearson and Deviance Residuals
Maximum Likelihood / Newton-Raphson / Fisher Scoring
Quasi-Likelihood Estimation
Overdispersed GLMs and Quasi-Likelihood Estimation
Generalized Additive Models (GAMs)

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 / Bayesian Information Criterion
Measures of Predictive Power / Classification Tables
ROC Curves / Concordance Index

Logit Models for Multicategory Responses (Chapter 8)
Nominal Responses and Baseline Category Models
Ordinal Responses / Cumulative Logit Models / Proportional Odds Models

Loglinear Models (Chapter 9)
Loglinear Models for Two-Way and Three-Way Tables
Conditional and Marginal Associations / Independence Relations
Generalized Linear Models for Longitudinal Data Analysis
Quasi-Likelihood Estimation for Longitudinal Data
Generalized Estimating Equations (GEEs)

Generalized Linear Mixed Models and Longitudinal Data Analysis
Longitudinal Data Structure
Traditional Linear Mixed Models for Longitudinal Data
Generalized Linear Mixed Models (GLMMs) for Longitudinal Data

Coverage:

In general, the topics covered in the lectures and notes will be organized according to the format of the corresponding chapters in Agresti. Some of the preceding topics will be explored in greater depth than others. Certain sections and subsections in the text will be assigned as independent readings, yet not covered in class.

Required readings will be listed by chapter and section along with each assignment. Those subsections that can be omitted will be clearly specified.

Agresti maintains a website for the text, which features a number of valuable resources, including datasets, software guides, errata, and short solutions to selected exercises. The site can be accessed at http://www.stat.ufl.edu/~aa/cda/cda.html.

Exams:

This course will feature one midterm examination and a comprehensive final examination. Both exams will be closed notes and closed book.
The midterm examination will cover material from Chapters 1, 2, and 3, and the first part of Chapter 4. This exam is scheduled for March 21.

The date and time of the final examination will be announced later in the semester.

**Homework:**

There will be seven assignments during the semester. Most homework sets will include theoretical, conceptual, and applied work. For problems involving data analysis, the SAS source code needed to obtain the required results will be covered in class. The topics for each assignment along with corresponding chapters from the text are listed in the following table.

<table>
<thead>
<tr>
<th>Nbr</th>
<th>Chapter</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>1</td>
<td>Distributions and Inference for Categorical Data</td>
</tr>
<tr>
<td>II</td>
<td>2</td>
<td>Descriptive Methods for Contingency Tables</td>
</tr>
<tr>
<td>III</td>
<td>3</td>
<td>Inferential Methods for Contingency Tables</td>
</tr>
<tr>
<td>IV</td>
<td>4</td>
<td>Generalized Linear Models (Part 1)</td>
</tr>
<tr>
<td>V</td>
<td>4</td>
<td>Generalized Linear Models (Part 2)</td>
</tr>
<tr>
<td>VI</td>
<td>5,6</td>
<td>Logistic Regression: Logit Models for Binary Responses</td>
</tr>
<tr>
<td>VII</td>
<td>8</td>
<td>Logit Models for Multicategory Responses</td>
</tr>
<tr>
<td></td>
<td>9</td>
<td>Loglinear Models</td>
</tr>
</tbody>
</table>

**Data Analyses:**

Data analyses will be primarily conducted using SAS. The fundamental procedures in SAS for the analysis of categorical data are FREQ (FREQuency tables), LOGISTIC (LOGISTIC regression), GENMOD (GENeralized linear MODels), and GLIMMIX (Generalized LInear MIXed Models). It will be assumed that all students have had prior exposure to SAS.

**Final Project:**

Near the end of the term, you will be required to perform a categorical data analysis on a data set of your choosing and to prepare a written report summarizing your results. You may complete this project either on your own or in collaboration with another student. Details concerning the project will be provided later in the semester.

**Grading:**

Your overall grade will be based on a weighted average of your midterm examination score (26%), homework average (18%), final project score (20%), and final examination score (36%).

**Course Website:**

The course notes, assignments, solution sets, and other relevant materials will be made available on the ICON (Iowa Courses Online) website for the course. The ICON login page is at [http://icon.uiowa.edu](http://icon.uiowa.edu).

The course notes will be made available prior to the lectures in which the notes are covered.
Prerequisites:

The prerequisites for this course include a sequence in mathematical statistics (STAT:4100/4101, Mathematical Statistics I and II; or STAT:5100/5101, Statistical Inference I and II), and a statistical methods course that covers regression analysis (STAT:5200, Applied Statistics I; or BIOS:5720, Biostatistical Methods II).

Competencies:

The successful completion of this course will facilitate the following doctoral-level competencies, pertaining to biostatistics and its application to public health.

- The ability to demonstrate an increased level of knowledge and understanding of current statistical theory, methods, and practices in the health sciences.
- The ability to manage, analyze, and interpret data from a variety of experimental and observational studies.
- The ability to effectively communicate research findings, including new statistical methods, to various audiences in writing and through oral presentation.

Administrative Home:

This course is offered by the College of Public Health. Thus, class policies on matters such as requirements, grading, and sanctions for academic misconduct are governed by the College of Public Health. Students wishing to add or drop this course after the official deadline must receive the approval of the Associate Dean for Academic Affairs in the College of Public Health. For more information, see https://www.provost.uiowa.edu/sites/provost.uiowa.edu/files/crossenroll.pdf.

Electronic Communication:

University policy specifies that students are responsible for all official correspondence sent to their standard University of Iowa e-mail address (@uiowa.edu). This account should be frequently checked.

Academic Misconduct:

Plagiarism and any other activities where students present the work of others as their own constitute academic misconduct. Academic misconduct is a serious matter and is reported to the Departmental Executive Officer (DEO) and to the Associate Dean for Academic Affairs in the College of Public Health. Instructors and DEOs decide on appropriate consequences at the departmental level while the Associate Dean may enforce additional consequences at the collegiate level. Details concerning the consequences associated with acts of plagiarism, including a student appeals process, are provided in the Graduate College Manual of Rules and Regulations, section IV.F. The Graduate College Manual is available online at http://www.grad.uiowa.edu/graduate-college-manual.

Additional information on academic misconduct and its possible consequences may be found in the University of Iowa Code of Student Conduct. The Code of Student Conduct is available online at https://dos.uiowa.edu/policies/. Academic misconduct is discussed in Section II.C; see https://dos.uiowa.edu/policies/academic-misconduct/.
Disabilities:

The University of Iowa is committed to providing an educational experience that is accessible to all students. A student may request academic accommodations for a disability (which includes but is not limited to mental health, attention, learning, vision, and physical or health-related conditions). A student seeking academic accommodations should first register with Student Disability Services (SDS), and then meet with the course instructor privately in his/her office to make particular arrangements. Reasonable accommodations are established through an interactive process between the student, instructor, and SDS. For more information, see https://sds.studentlife.uiowa.edu/.

Concerns about Faculty Actions:

Students who have a concern about a faculty action should first address the issue with the instructor, and then the Departmental Executive Officer. Students may also contact the Associate Dean for Academic Affairs in the College of Public Health. Another resource for students is the Office of the University Ombudsperson. If a complaint cannot be resolved at the departmental and/or collegiate level, students may file a formal complaint utilizing the procedure specified in the UI Operations Manual (II-29.7). The UI Operations Manual is available online at http://opsmanual.uiowa.edu.

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 to the UI Office of Sexual Misconduct Response Coordinator (OSMRC), or to the Office of Equal Opportunity and Diversity (EOD). Students impacted by a Title IX issue (sexual misconduct, dating/domestic violence, or stalking) may be eligible to request an academic accommodation. Information regarding the OSMRC is available online at https://osmrc.uiowa.edu/. Specific information regarding confidential resources for students subject to sexual harassment and/or misconduct is available at https://osmrc.uiowa.edu/victim-resources/confidential-support.

The UI policy on sexual harassment in its entirety may be found in the UI Operations Manual (II-4): https://opsmanual.uiowa.edu/community-policies/sexual-harassment.

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 https://hawkalert.uiowa.edu/.