EPIB 652 Categorical Data Analysis

Semester: Fall 2014
Classroom and Time: SPH 1302, Wednesday 4:00pm - 6:45pm
Instructor: Shuo Chen, Ph.D.
Office: 2234B SPH Building
Office Hours: Wednesday 2:00pm - 4:00pm or by appointment
Phone: (301) 405-3085
Email: shuochen@umd.edu

Course Pre- and Co-requisites:
Required: EPIB 650 Biostatistics I & EPIB 651 Biostatistics II

Required Texts and Other Readings:

Additional Materials Required:
SAS (strongly recommended) or other statistical software

Course Description:
This course provides an introduction to methods for analyzing categorical data, with emphasis to public health research. It emphasizes the ideas behind the methods and their applications. The course covers descriptive and inferential methods for contingency tables and generalized linear models for binary, ordinal, and count data. It also covers methods for matched-pairs data. Computations are illustrated using SAS.

Course Learning Objectives:
Upon completing this course, the student will be able to:
1. Understand and describe the major types of categorical data and discrete probability distributions
2. Identify, describe, and apply appropriate descriptive and inferential methods for contingency tables (e.g., odds ratios, independence tests)
3. Identify, describe, and build appropriate models for different types of categorical data (e.g. logistic regression, loglinear models)
4. Interpret results of categorical data analyses in oral and written forms
5. Analyze categorical data using statistical software

Program Competencies Addressed in this Course:
The following competencies for the Biostatistics MPH program are addressed in this course:
1. Describe basic concepts of probability, random variables, and commonly used statistical probability distributions
2. Apply appropriate descriptive statistical methods for summarizing public health data
3. Conduct descriptive and inferential statistical methods that are appropriate to the different study designs used in public health research
4. Critically review and summarize statistical analyses presented in public health literature
5. Draw appropriate inferences based on statistical analyses used in public health research
6. Demonstrate statistical methods as applied to public health data, research study designs, and tests of inference
7. Select appropriate inferential statistical methods to answer research questions relevant to public health research
8. Investigate statistical approaches to address threats to validity in epidemiologic studies
9. Communicate results of statistical analyses to lay and professional audiences

Course Requirements:

Course Website:
Lecture notes, handouts, SAS examples, homework assignments, and homework solutions will be posted on the ELMS (Enterprise Learning Management System). Lecture notes will be uploaded before classes. You can access the website by following the directions
• Direct your URL to https://elms.umd.edu/
• Enter your Directory ID and Password
• Click “Course Sites” on the ELMS home tab
• Click : EPIB652 Sec - 0101 Fall 2013:Categorical Data Analysis”

Homework:
Four homework assignments will be given. It is important to complete the homework in a timely manner. Twenty percent will be deducted if homework is handed in within three days after the deadline and before homework solutions are posted (whichever earlier). No homework will be accepted after three days or homework solutions are posted. Students are encouraged to work together in groups to help each other understand the course materials and complete the homework problems. However, students must finish their homework independently after group discussions. The names of the group members are required to be put on the first page of the homework.

Exams:
Exams will be in class, closed-book, and closed-note. The content of exams will be cumulative. For the midterm exam, you are allowed to bring one page of letter-size, double-sided formula sheet; for the final exam, you are allowed to bring two pages of letter-size, double-sided formula sheet. You also need to bring a calculator to conduct the calculation. No makeup exams are allowed in general. Exceptions to this rule are evaluated on a case-by-case basis. Students must submit a request before the exam with valid supporting documents. No post-exam request will be considered except the student is hospitalized during the exam period.

Grading Procedures:
40% Homework
30% Mid-term exam
30% Final exam

Additional Literature, Websites and Other Resources:

Alan Agresti’s Websites:
• Alan Agresti’s course website of Intro CDA
  http://www.stat.ufl.edu/~aa/sta4504/
• Alan Agresti’s SAS appendix for Intro CDA
  http://www.stat.ufl.edu/~aa/cda/software.html
• Alan Agresti’s website for Categorical Data Analysis (2nd edition, Wiley, 2002)
  http://www.stat.ufl.edu/~aa/cda/cda.html

Other Useful Websites:
• PennState online course: STAT 504 - Analysis of Discrete Data
  http://www.stat.psu.edu/online/development/stat504/
• SAS and Stata code for Intro CDA from UCLA ATS
  http://www.ats.ucla.edu/stat/examples/icda/default.htm
• Statistical computing from UCLA ATS (Data Analysis Examples, Textbook Examples, Online
  Seminars, etc)
  http://www.ats.ucla.edu/stat/default.htm
• Laura Thompson’s R and S manual for CDA
  https://home.comcast.net/~lthompson221/#CDA
• R introduction http://cran.r-project.org/doc/manuals/R-intro.pdf

Course Policies:

Email – The Official University Correspondence:
Verify your email address by going to www.my.umd.edu.

All enrolled students are provided access to the University’s email system and an email account. All
official University email communication will be sent to this email address (or an alternate address if
provided by the student). Email has been adopted as the primary means for sending official
communications to students, so email must be checked on a regular basis. Academic advisors,
faculty, and campus administrative offices use email to communicate important and time-sensitive
notices.

Students are responsible for keeping their email address up to date or for redirecting or
forwarding email to another address. Failure to check email, errors in forwarding email, and
returned email (from “full mailbox” or “unknown user” errors for example), will not
excuse a student from missing University announcement, messages, deadlines, etc. Email
addresses can be quickly and easily updated at www.my.umd.edu or in-person at the Student
Service Counter on the first floor of the Mitchell Building.

For technical support for University email: www.helpdesk.umd.edu or call 301-405-1400.

Absence Policy:
In accordance with University policy if you are absent for a single (1) lecture due to illness or some
form of personal or family emergency, this absence will be considered “excused” and the instructor
will accept a note from you attesting to the date of the illness/incident, along with an
acknowledgement that the information is true. Whenever feasible, you should try to contact the
instructor in advance.
Multiple or prolonged absences, and absences that prevent attendance at a major scheduled grading event (like an exam or test) will require written documentation from an appropriate health care provider/organization.

A link to pull information on the new policy covering absences from class can be found at [http://www.president.umd.edu/policies/v100g.html](http://www.president.umd.edu/policies/v100g.html)

**Late work and Missed Exams / Assignments:**
Twenty percent will be deducted if homework is handed in within three days after the deadline and before homework solutions are posted (whichever earlier). No homework will be accepted after three days or homework solutions are posted. No makeup exams are allowed in general. Exceptions to this rule are evaluated on a case-by-case basis. Students must submit a request before the exam with valid supporting documents. No post-exam request will be considered except the student is hospitalized during the exam period.

**Religious Observances:**
The University System of Maryland policy provides that students should not be penalized because of observances of their religious beliefs; students shall be given an opportunity, whenever feasible, to make up within a reasonable time any academic assignment that is missed due to individual participation in religious observances. It is the student’s responsibility to inform the instructor in advance of any intended absences for religious observance.

**Special Accommodations / Disability Support Services:**
If you have a documented disability and wish to discuss academic accommodations for test taking or other needs, you will need documentation from Disability Support Service (301-314-7682). If you are ill or encountering personal difficulties, please let the instructor know as soon as possible. You can also contact Learning Assistance Services (301-314-7693) and/or the Counseling Center (301-314-7651) for assistance.

**Academic Integrity:**
The University's code of academic integrity is designed to ensure that the principle of academic honesty is upheld. Any of the following acts, when committed by a student, constitutes academic dishonesty:

- **CHEATING:** intentionally using or attempting to use unauthorized materials, information, or study aids in an academic exercise.
- **FABRICATION:** intentional and unauthorized falsification or invention of any information or citation in an academic exercise.
- **FACILITATING ACADEMIC DISHONESTY:** intentionally or knowingly helping or attempting to help another to violate any provision of this code.
- **PLAGIARISM:** intentionally or knowingly representing the words or ideas of another as one's own in any academic exercise.

For more information see: [http://www.shc.umd.edu/code.html](http://www.shc.umd.edu/code.html).

The Honor Pledge is a statement undergraduate and graduate students should be asked to write by hand and sign on examinations, papers, or other academic assignments. The Pledge reads:
I pledge on my honor that I have not given or received any unauthorized assistance on this assignment/examination.

The University of Maryland, College Park has a nationally recognized Code of Academic Integrity, administered by the Student Honor Council. This Code sets standards for academic integrity at Maryland for all undergraduate and graduate students. As a student you are responsible for upholding these standards for this course. It is very important for you to be aware of the consequences of cheating, fabrication, facilitation, and plagiarism. For more information on the Code of Academic Integrity or the Student Honor Council, please visit http://www.shc.umd.edu.

Inclement Weather / University Closings:
In the event that the University is closed for an emergency or extended period of time, the instructor will communicate to students regarding schedule adjustments, including rescheduling of examinations and assignments due to inclement weather and campus emergencies. Official closures and delays are announced on the campus website (http://www.umd.edu) and snow phone line (301-405-SNOW), as well as local radio and TV stations.

Course Evaluations
Your participation in the evaluation of courses through CourseEvalUM is a responsibility you hold as a student member of our academic community. Your feedback is confidential and important to the improvement of teaching and learning at the University as well as to the tenure and promotion process. CourseEvalUM will be open for you to complete your evaluations starting about two weeks prior to the last day of the term before exams begin. Please go directly to the website (www.courseevalum.umd.edu) to complete your evaluations. By completing all of your evaluations each semester, you will have the privilege of accessing online evaluation reports for the thousands of courses for which 70% or more students submitted their evaluations. You can access results at www.CourseEvalUM.umd.edu, the same link you use to submit your evaluations. Click View Past Results instead.

Use the CourseEvalUM URL and choose Take Evaluations to discover upcoming evaluation dates:
### Course Outline / Course Calendar:

#### Course Schedule Summary

<table>
<thead>
<tr>
<th>Session</th>
<th>Date</th>
<th>Topic</th>
<th>Reading / Assignment</th>
</tr>
</thead>
</table>
| 1       | 9/3/2014  | Introduction and review  
Categorical response data  
Distributions for categorical data (Bernoulli, binomial, multinomial, Poisson) | Sections 1.1-1.2               |
| 2       | 9/10/2014 | Introduction and review (cont’d)  
Distributions for categorical data (cont’d)  
Likelihood function and MLE  
More on inference (Wald, LRT, score tests) | Sections 1.3-1.4               |
| 3       | 9/17/2014 | Two-way Contingency tables  
Table structure  
3 ways of comparing proportions (difference, relative risk, odds ratio) and their inferences | Sections 2.1-2.3               |
| 4       | 9/24/2014 | Two-way Contingency tables (cont’d)  
Chi-square tests ($X^2$, $G^2$)  
Following-up chi-square tests: (Pearson residuals, partitioning of $G^2$)  
Extension for I-J tables (local odds ratio, uncertainty coefficient U)  
Testing independence for ordinal data (linear or monotone trend) | Sections 2.1-2.4               |
| 5       | 10/1/2014 | Two-way Contingency tables (cont’d)  
Fisher’s exact test for small samples  
Three-way Contingency tables  
Conditional vs. marginal tables  
Simpson’s paradox | Section 2.5  
Homework #1 due |
| 6       | 10/8/2014 | Three-way Contingency tables (cont’d)  
Conditional vs. marginal odds ratios  
Conditional vs. marginal independence  
CMH test  
Homogeneous association for 2-2-K tables | Sections 2.6-2.7               |
| 7       | 10/15/2014| Generalized linear models  
Components of generalized linear models  
GLMs for binary data  
GLMs for count data  
Statistical inference and deviance  
Fitting generalized linear models | Sections 3.1-3.5  
Homework #2 due |
| 8       | 10/22/2014| Mid-term exam                                                        |                               |
| 9       | 10/29/2014| Logistic regression  
Interpreting logistic regression  
Inference for logistic regression  
Categorical predictors  
Multiple logistic regression | Sections 4.1-4.5               |
<table>
<thead>
<tr>
<th></th>
<th>Date</th>
<th>Session Content</th>
<th>Sections/Assignments</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
<td>11/05/2014</td>
<td>Building and applying logistic regression models</td>
<td>Sections 5.1-5.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Strategies in model selection</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Model checking</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Model diagnostics</td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>11/12/2014</td>
<td>Building and applying logistic regression models (cont'd)</td>
<td>Sections 5.1-5.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Effects of sparse data</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SAS Examples: AIDS, clinical trial with sparse data, grouped vs. ungrouped crab data</td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>11/19/2014</td>
<td>Multicategory logit models</td>
<td>Sections 6.1-6.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Baseline-category logit models for nominal responses</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Cumulative logit model for ordinal responses</td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>11/26/2014</td>
<td>Loglinear models</td>
<td>Sections 7.1-7.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Loglinear models for count data</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Loglinear models for rate data</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Loglinear models for 2-way tables</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SAS Examples: belief in afterlife</td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>12/03/2014</td>
<td>Loglinear models (cont'd)</td>
<td>Sections 7.3-7.5</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Loglinear models for 3-way tables</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Loglinear-logit model connection</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Modeling ordinal association (L-by-L association model)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>SAS Examples: drug use, auto injury, sex opinions</td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>12/10/2014</td>
<td>Models for matched pairs</td>
<td>Sections 8.1, 8.5.5, 8.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Comparing dependent proportions (McNemar's test)</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kappa measure of agreement</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Bradley-Terry model for paired preference</td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>12/17/2014</td>
<td>Final exam</td>
<td></td>
</tr>
</tbody>
</table>