Course: BIOS 765: Models and Methodology in Categorical Data
Instructor: John Preisser
Term: Fall 2014

Time: Tuesdays and Thursdays, 9:30am-10:45am
Place: 2303 McGavran-Greenberg Hall

Instructor John Preisser, 3105F McGavran-Greenberg, email: john_preisser@unc.edu

Office hours: Immediately after class or by appointment

Textbook: Stokes ME, Davis CS, Koch GG (2012). Categorical data analysis using the SAS system, 3rd edition, Cary, NC: SAS Institute Inc. serves as the primary reference source. The course relies heavily on course notes provided in the course pack and various selected readings from the literature.

Course pack: Students should purchase the course pack from student stores containing the lecture notes.

References: A reading list is provided at the end of this syllabus. The following reference is optional and may be useful. It will not be used for required reading:


Course website: www.bios.unc.edu/~jpreisse/bios765/

Please note faculty course notes are protected under University copyright policies: www.unc.edu/faculty/faccoun/news/special/2010FacultyCopyrightAnswers.pdf

Prerequisites: Biostatistics 661, 663, and 665 or equivalents

Course Description: Theory and application of methods for models with categorical response data. The course will 1) present the theory of statistical methods for analyzing categorical data with maximum likelihood, estimating equation, and chi-square methods for large samples, and exact inference for small samples; 2) provide many illustrations of the various methods; and 3) provide ample opportunity for students to demonstrate methodological understanding and to apply the various methods to data from public health, dental, medical and biopharmaceutical settings.
Problem Sets

There will be five take-home problem sets over the course of the semester. The student will have at least two weeks (usually three weeks) to complete each one.

Grading

Problem sets are weighted according to length. Individual problems are labeled as either “Applied” or “Theory”. General expectations for grades are as follows:

H – The student will be expected to perform exceedingly well on both theoretical and applied homework problems.

P – The student will be expected to perform satisfactorily on applied homework problems.

Note: An overall 90% score or greater based on all the problems (applied and theory) will earn an “H” grade. At the instructor’s discretion, a score slightly lower than 90% may be awarded an “H” grade depending upon the overall distribution of student scores. If an “H” grade is not awarded, an overall 70% score or greater on all “Applied” problems will earn a “P” grade. At the instructor’s discretion, a score lower than 70% on “Applied” problems may be awarded a “P” if the student works some “Theory” problems, or parts thereof, correctly. Regular class attendance is expected for all enrolled students.

Late Homework

Late homework will be deducted 1 point for each day it is tardy for a maximum possible deduction of 3 points per assignment (maximum score for a problem set is typically 30 to 40 points). Homework assignments should be turned in before 4pm of the due date. Late homework should be submitted via email to the instructor if possible, otherwise placed in the instructor’s mailbox and email instructor stating when the problem set was turned in.

Honor Code

Students are expected to abide by the University Honor Code which is described fully at: http://honor.unc.edu. Write-ups of problem sets should be the student’s own work. Violations may result in severe consequences.
<table>
<thead>
<tr>
<th>Target Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aug 19</td>
<td>Notes 1. Sampling distributions</td>
</tr>
<tr>
<td>Aug 21</td>
<td>Notes 2. Generalized linear models</td>
</tr>
<tr>
<td>Aug 26</td>
<td>Notes 3. Poisson regression</td>
</tr>
<tr>
<td>Aug 28</td>
<td>Notes 4. Piecewise exponential regression</td>
</tr>
<tr>
<td>Sep 2, 4</td>
<td>Notes 5. Negative binomial regression and zero-inflated models</td>
</tr>
<tr>
<td>Sep 9</td>
<td>Notes 6. Single multinomial</td>
</tr>
<tr>
<td>Sep 11</td>
<td>Notes 7. Application to multiple recapture census</td>
</tr>
<tr>
<td>Sep 16</td>
<td>Notes 7 cont’d. Multivariate Delta Method, Loglinear models</td>
</tr>
<tr>
<td>Sep 18, 23</td>
<td>Notes 8. Loglinear models for the analysis of association</td>
</tr>
<tr>
<td>Sep 25</td>
<td>Notes 9. Logistic regression</td>
</tr>
<tr>
<td>Sep 30, Oct 2</td>
<td>Notes 10. Multi-category logistic regression</td>
</tr>
<tr>
<td>Oct 7</td>
<td>Notes 11. Quasi-likelihood and overdispersion</td>
</tr>
<tr>
<td>Oct 9</td>
<td>Notes 12. Randomization model methods &amp; nonparametrics (2x2)</td>
</tr>
<tr>
<td>Oct 14</td>
<td>Notes 13. Randomization model methods for s x r tables</td>
</tr>
<tr>
<td>Oct 16</td>
<td>NO CLASS – FALL BREAK</td>
</tr>
<tr>
<td>Oct 21</td>
<td>Notes 13 cont’d.</td>
</tr>
<tr>
<td>Oct 23</td>
<td>Notes 14. Exact Contingency Table Inference</td>
</tr>
<tr>
<td>Oct 28</td>
<td>Notes 15. Conditional logistic regression/Exact logistic regression</td>
</tr>
<tr>
<td>Oct 30</td>
<td>Notes 16. Weighted least squares</td>
</tr>
<tr>
<td>Nov 4</td>
<td>Notes 17. Repeated measures analysis with weighted least squares</td>
</tr>
<tr>
<td>Nov 6</td>
<td>Notes 18. Linear models for rank measures of association</td>
</tr>
<tr>
<td>Nov 11</td>
<td>Notes 19. Covariance analysis using weighted least squares</td>
</tr>
<tr>
<td>Nov 13, 18</td>
<td>Notes 20. Generalized estimating equations (GEE)</td>
</tr>
<tr>
<td>Nov 20</td>
<td>Notes 21. Weighted GEE for longitudinal data with dropouts</td>
</tr>
<tr>
<td>Nov 25</td>
<td>Notes 22. Modeling within-cluster association/estimating equations</td>
</tr>
<tr>
<td>Nov 27</td>
<td>NO CLASS – THANKSGIVING RECESS</td>
</tr>
<tr>
<td>Dec 2</td>
<td>Notes 22 cont’d. Alternating logistic regressions</td>
</tr>
</tbody>
</table>
Lecture Topic

Sampling distributions

Optional reading for sampling distributions

Agresti (2002), sections 1.1, 1.2

Generalized linear models


Optional reading for generalized linear models


Poisson regression


Piecewise exponential regression


Negative Binomial, Zero-inflated Poisson, and Zero-inflated Negative Binomial Regression


Optional reading

Single Multinomial and application to multiple recapture census


*Optional reading for generalized linear models*


Loglinear Models for Contingency Tables (Analysis of association)


*Optional reading*

Agresti (2002), Chapter 8.

Logistic regression and Multi-category logistic regression


*Optional reading for logistic regression*


Quasi-likelihood and overdispersion


Optional reading on quasi-likelihood and overdispersion


Randomization methods and nonparametric methods


Optional reading for randomization methods

Stokes, Davis and Koch (2012), Chapter 6 and 7.


Exact inference (for contingency tables and logistic regression)


Optional reading for exact inference


Conditional logistic regression


*Optional reading for conditional logistic regression*


Weighted least squares (WLS)


*Optional Reading on WLS*


Linear models for rank measures of association


*Optional Reading on linear models for rank measures of association*


Nonparametric ANACOVA (Covariance analysis using WLS)


Optional Reading on Nonparametric ANACOVA


Generalized estimating equations (GEE)


Optional reading on GEE


Weighted Generalized Estimating Equations


Optional reading on weighted GEE


Alternating logistic regressions (ALR)

Stokes, Davis, Koch (2012). Section 15.12

Optional reading on ALR
