Course Title: Applied Categorical Data Analysis

Course Number: BIST 0615

Course Location: SPH Room 1A&B

Course Date & Time: Wednesday, 6:00-9:00pm

Course Instructor: Pamela Ohman Strickland, PhD, Associate Professor of Biostatistics, Associate Dean for Student Affairs, Rutgers SPH, Pam.Strickland@rutgers.edu, SPH Rm 118, 732-235-9721

Office Hours: By Appointment Only


Course Description: Public health studies, especially those involving questionnaires, contain large amounts of categorical data. This class provides an introduction to descriptive and inferential statistics for univariate and multivariate categorical data with applications to epidemiological and clinical studies. For 2 and 3-way contingency tables, measures of association and tests for homogeneity between populations and independence of variables are presented. Related tests of trend for ordinal data are studied. Loglinear and logistic regression analyses are investigated for data sets with both nominal and ordinal variables.

Computing Language: SAS

Selected Department Competencies Addressed: Each Department identifies competencies for each degree offered. The competencies addressed in this course for the MPH for the Department of Biostatistics include:

- Integrate relevant scientific background to design experimental and observational studies in biomedical, clinical and public health research;
- Use statistical computer packages to organize, analyze and report collected data;
- Apply basic probability theory and standard statistical methods to biomedical, clinical and public health research;
- Review and critique statistical methods and interpretations presented in published research studies, presentations or reports; and
- Communicate the results of statistical studies both in writing and orally to investigators and lay community members.

Please visit the Department webpages on the School of Public Health’s website at http://sph.rutgers.edu/ for additional competencies addressed by this course for other degrees and departments.
Course Objectives: By the completion of this course, students will be able to:

- Formulate appropriate statistical hypotheses for examining cross-classified data from public health and clinical studies
- Justify the basic theoretical models for categorical data
- Create and/or actively participate in the design and analysis plan for a study involving categorical data, whether nominal or ordinal in nature
- Conduct and/or actively participate in the analysis of categorical data
- Interpret results from contingency tables or generalized linear models that evaluate relationships between categorical variables
- Communicate, both verbally and in writing, results with non-statisticians

Writing & Communication

This class includes a heavy emphasis on interpreting and communicating results, verbally and in writing, in English to scientists and lay-people. Particularly if you native language is not English, please realize that you may need to invest some time and effort on writing skills.

Course Requirements and Grading:

Your grade will be determined according to the following assignments:

1. Homework         10%
2. Journal Article Summaries/Discussion    10%
3. Exams         55%
4. Final Project Write-Up       20%
5. Draft of Final Project         5%

Journal Article Summaries

When a journal article is assigned for homework, a two sentence summary of the article will be requested along as responses to one or two additional questions.

Homework policies

1. Homework will be posted to Moodle the day following class. This allows me to tailor the homework so that they only cover material taught in class that week.
2. All homework must be turned in at the beginning of the class period on the day on which they are due.
3. On all homework assignments/problem sets, students are encouraged to discuss with one another, but work should be carried out and written up independently. If any two identical write-ups are found, both homework assignments will be given a grade of zero.
4. The homework assignments will involve computer work using SAS. Relevant portions (only) of the output should be cut and pasted into the homework assignment at the appropriate spot, not attached at the end. Failure to comply will result in a reduction of 50% of points for that homework.
5. Homework solutions will be emailed.
Note that homework and journal article summary assignments will be heavier towards the beginning of the semester and become lighter towards the end so as to encourage focus on final projects and preparation for the final. Keep this in mind as you track your grade in the course over the semester.

**Exam policies**
Exams will be given on October 12, November 9, and December 7. Any unexcused absences for these exams will result in a grade of zero for that exam.

**FERPA issues**
Grades may only be released to the student by the registrar, in person or via a rutgers.edu email address by the instructor. Grades *may not* be released over the phone or posted publicly in any way.

**Course Schedule:**

<table>
<thead>
<tr>
<th>Date</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>September 7</td>
<td>Introduction, Distributions and Sampling</td>
</tr>
<tr>
<td></td>
<td>Definitions, Poisson, Binomial and Multinominal Distributions</td>
</tr>
<tr>
<td></td>
<td>Inference about univariate proportions</td>
</tr>
<tr>
<td>September 14 (on-line)</td>
<td>Two-way Contingency Tables</td>
</tr>
<tr>
<td></td>
<td>Probability Structures for 2-way tables, Chi-square tests of independence</td>
</tr>
<tr>
<td></td>
<td>Comparing proportions, risk ratios and odds ratios</td>
</tr>
<tr>
<td>September 21</td>
<td>Two-way Contingency Tables with Ordinal Data</td>
</tr>
<tr>
<td></td>
<td>Testing independence for ordinal data</td>
</tr>
<tr>
<td></td>
<td>Measures of Association for ordinal data</td>
</tr>
<tr>
<td>September 28</td>
<td>Analyses for Matched Pairs of Categorical Data</td>
</tr>
<tr>
<td>October 5</td>
<td>Three-way Tables</td>
</tr>
<tr>
<td></td>
<td>Justification for studying three-way relationships, partial association,</td>
</tr>
<tr>
<td></td>
<td>Cochran-Mantel-Haenszel Methods, Common odds ratios</td>
</tr>
<tr>
<td>October 12</td>
<td>MID-TERM (Note October 17 is the last day to withdraw with a ‘W’)</td>
</tr>
<tr>
<td>October 19</td>
<td>Generalized Linear Models</td>
</tr>
<tr>
<td></td>
<td>Components of a generalized linear model, linear regression, models for</td>
</tr>
<tr>
<td></td>
<td>binary data and Poisson regression Models for count and rate data,</td>
</tr>
<tr>
<td></td>
<td>model inference and model fitting</td>
</tr>
<tr>
<td>October 26 (on-line)</td>
<td>Logistic Regression</td>
</tr>
<tr>
<td></td>
<td><em>Project Plan due.</em></td>
</tr>
<tr>
<td>November 2</td>
<td>Multi-category Logit Models</td>
</tr>
<tr>
<td></td>
<td>Logit Models for Nominal Responses, cumulative logit models for ordinal</td>
</tr>
<tr>
<td></td>
<td>responses, paired-category logits for ordinal responses</td>
</tr>
</tbody>
</table>
### November 9
(Logistic/Proportional Odds Modeling)
Loglinear Models for Contingency Tables
Loglinear models for two- and three-way tables, inference for loglinear models, loglinear models for higher dimensions, the loglinear-logit connection

### November 16
**MINI-EXAM #1 – Building and Applying Logit and Loglinear Models**
Association graphs and collapsibility, modeling ordinal associations, tests of conditional independence, effects of sparse data

### November 23
NO CLASS (Day before Thanksgiving)

### November 30
**Models for Matched Pairs**
Comparing dependent proportions, logistic regression for matched pairs

### December 7
(Building Models/Loglinear Models)
Conditional Logistic Regression
(Peer Reviews Reports Due Electronically December 7, 5pm)
(Final Project Reports Due)

### December 14
**MINI-EXAM #2 – Wrapping Up/Project Reports**
(Peer Review Due Electronically)

---

**School of Public Health Honor Code:** The School of Public Health Honor Code is found in the student bulletin (sph.rutgers.edu/academics/catalog/index.html). Each student bears a fundamental responsibility for maintaining academic integrity and intellectual honesty in his or her graduate work. For example, all students are expected to observe the generally accepted principles of scholarly work, to submit their own rather than another’s work, to refrain from falsifying data, and to refrain from receiving and/or giving aid on examinations or other assigned work requiring independent effort. In submitting written material, the writer takes full responsibility for the work as a whole and implies that, except as properly noted by use of quotation marks, footnotes, etc., both the ideas and the works used are his or her own. In addition to maintaining personal academic integrity, each student is expected to contribute to the academic integrity of the school community by not facilitating inappropriate use of her/his own work by others and by reporting acts of academic dishonesty by others to an appropriate school authority. It should be clearly understood that plagiarism, cheating, or other forms of academic dishonesty will not be tolerated and can lead to sanctions up to and including separation from the Rutgers School of Public Health.

**Policy Concerning Use of Recording Devices and Other Electronic Communications Systems:** When personally owned communication/recording devices are used by students to record lectures and/or classroom lessons, such use must be authorized by the faculty member or instructor who must give either oral or written permission prior to the start of the semester and identify restrictions, if any, on the use of mobile communications or recording devices.
ATTACHMENT: Final Project

There are two options for the final project.

1) Data Analysis

Complete a full analysis of a set of data that contains an ordinal or multinomial response and at least one ordinal covariate. This data set may not be one that you are using for a fieldwork project or dissertation. Present two alternative analyses, one of which must be a generalized linear model. Compare and contrast the two approaches.

2) Report on a Statistical Paper (MUST HAVE SPECIAL APPROVAL FROM INSTRUCTOR)

Report and provide an original simulation study based on methodology presented on a statistical paper addressing the analysis of categorical or ordinal data.
   a. Describe the problem it addresses
   b. Summarize the solution proposed along with advantages and disadvantages of the solution
   c. Perform an original simulation comparing the method to a more standard methodology.
   d. Formulate two or three application or research questions propagated by the research paper. (Try to be as original as possible.)

The emphasis in these final projects should be on the statistical methodology and how it is applied.

Schedule

October 25, 6pm: Project Plan due electronically
December 6, 6pm: Draft of Paper Due
December 13, 6pm: Peer Reviews Due
December 20, 6pm: Final Paper Due

Please pay attention to deadlines and get started on your analyses and paper writing early!

Project Plan

The “Project Plan” for the Data Analysis option, due October 25th, should include a full description of the data set and variables to be used, a set of clearly defined hypotheses and, when possible, an initial outline of what statistical methods might be used. The “Project Plan” for the Statistical Paper option should include a copy of the paper abstract along with the reference as well as a discussion of why this interests you and what type of simulations you envision.

Paper Draft

The Paper Draft, complete with Introduction, Methods, Results, Discussion, and appropriate tables (not just SAS output), is due December 6.
Final Project Peer Review

Final drafts of the Paper will be due on December 6. Each student will be randomly assigned three other students’ papers to review. Using a template, each student will be asked to rate and comment on these three other papers. These reviews will be due on December 13. Ratings will be combined to rank all papers. I will let you know where you stand relative to your peers upon re-distribution of peer reviews. You may use these peer reviews to revise your own paper for a final submission of the project due on December 20.

Note 5% of your final grade is automatically give as a full credit under the following conditions:

1) You hand in a complete draft of your paper for review by December 6. I will read to make sure it contains, Introduction, Methods, Results, Discussion, and appropriate tables (not just SAS output).
2) You hand in completed peer reviews of other students’ papers by December 13.

Otherwise, you will get 0 towards the 5% of your final grade. No exceptions will be made!

Final Paper

The Final Paper is due on December 20. These papers may reflect the peer reviews you received and will be the only document I review for the final grade of the paper. (Note that due to a compressed timeline with the holidays, late submissions cannot be accepted. Please plan accordingly.)