STAT 665 – Categorical Data Analysis

Administrative

Classes: Wednesdays, 4:30 p.m. - 7:10 p.m., January 27 – May 5 (Note: No class March 17.)
Instructor: Dr. Linda Davis
Office: Room 1703, Nguyen Engineering Building (ENGR)
Office Phone: (703) 993-4835
Address: Dr. Linda Davis
George Mason University
Statistics Dept., Mail Stop 4A7
4400 University Drive
Fairfax, VA 22030
E-Mail: ldavisc@gmu.edu
Statistics Dept: (703) 993-3645, Fax: (703) 993-1700

Prerequisites: STAT 554 or equivalent. (Familiarity with basic statistical terminology concerning hypothesis testing and estimation.)
STAT 656 (Familiarity with regression modeling for continuous data.)
STAT 501 (Familiarity with basic SAS programming such as how to use the DATA step and some of the elementary procedures such as PROC FREQ and PROC MEANS.)


- Available (free of charge to registered students) via the Mason Virtual Lab (VCL). The Virtual Lab allows access to SAS over the Internet, so that SAS does not have to be available on a local machine. (For details, see http://doit.gmu.edu/studentSection.asp?page=vcl)
- Available on campus on some of the computers in the ITE Lab, Engineering Building, Room 1506.
- A copy of SAS that you can load on your personal computer can be licensed for a fee through Patriot Computers in the Johnson Center. (For details, see http://compstore.gmu.edu/products/software/#SAS). There is NO installation support available to the end user. Note: Prior to leasing SAS, make sure your computer meets the minimum system requirements. List of system requirements can be found at: http://support.sas.com/resources/sysreq/92/index.html

Resources: Copies of lecture notes and examples discussed in class will be made available via Blackboard; hardcopies will not be distributed in class. So, if you want to follow along during class, you will need to bring copies with you. Since SAS is an integral part of this course (specifically we will be using SAS procedures FREQ, LOGISTIC, CATMOD, and GENMOD), you may want to purchase either: The Little SAS Book by Delwiche and Slaughter (great reference on basic SAS programming and data manipulation) or Applied Statistics and the SAS Programming Language, 5th Edition by Ronald P. Cody and Jeffrey K. Smith (provides a more tutorial review of SAS).

Web Sites: http://www.stat.ufl.edu/~aa/cda/cda.html – Contains links to SAS code discussed in Appendix A of Agresti’s book as well as errata for the text (which you need to review).

Communication: E-mails concerning this class will be sent to your GMU e-mail address. So, if you choose to utilize another e-mail account, please make sure your GMU e-mail account is forwarded to your preferred e-mail address. When you send an e-mail to me, please place STAT 665 at the beginning of the subject line. Also, please restrict use of e-mails to administrative matters. If you have questions about course material or assignments, please ask after class, stop by during office hours, or call to make an appointment.
Course Description

A categorical variable is one measured on a qualitative or discrete quantitative scale. This course covers the basic techniques for analyzing categorical response variables in two and higher dimensions. Since categorical variables naturally group observations in a data set, techniques discussed in this course involve analyzing and modeling frequencies (i.e., counts or proportions), not measured characteristics of individuals.

Much of the material in Chapters 1–8 in the required text by Agresti will be discussed. The main topics covered are:

1. Classical (asymptotic) and exact analysis techniques applied to two-way and three-way contingency tables – Chi-squared Tests, Measures of Association.
2. Poisson Regression – Methods analogous to regression, but for a discrete count response (e.g., number of traffic accidents in a week).
3. Logit Models – Methods analogous to regression, but for a binary response (e.g., response to yes/no question). Comparison to Cochran-Mantel-Haenszel techniques for $2 \times 2 \times K$ tables.
4. Loglinear Models – Methods analogous to ANOVA, but for a categorical response variable with categorical explanatory variables.

Goals

The goal of this course is to provide students with an overview of available techniques and models for analyzing categorical data. Both theoretical derivations and applications using SAS will be discussed. And, emphasis will be placed on the correct interpretation and reporting of the results from a categorical data analysis.

Grading

50% (Weekly) Homework
20% In-Class Midterm (Open Book, Calculator Required), March 10, 5:45 p.m.–7:10 p.m.
30% Final Exam (In-Class, Cumulative, Open Book, Calculator Required), May 12, 4:30 p.m.–7:15 p.m.

Grading Scale

<table>
<thead>
<tr>
<th>Grade</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>90 – 100</td>
</tr>
<tr>
<td>A-</td>
<td>85 – 89.9</td>
</tr>
<tr>
<td>B+</td>
<td>80 – 84.9</td>
</tr>
<tr>
<td>B</td>
<td>75 – 79.9</td>
</tr>
<tr>
<td>B-</td>
<td>70 – 74.9</td>
</tr>
<tr>
<td>C</td>
<td>60 – 69.9</td>
</tr>
<tr>
<td>F</td>
<td>Below 60</td>
</tr>
</tbody>
</table>

Homework Assignments

1. Homework assignments will include both theoretical and data analysis problems. Most problems will be taken from the required text by Agresti.
2. Please make your assignments as readable and neat as possible. Please organize the problems sequentially. Also, clearly indicate the demarcation between problems and number pages across entire assignment.
3. Write in complete sentences as much as possible.
4. You will be generating a great deal of computer output. Applicable figures/tables must be incorporated into your solutions; and, only those parts of the computer output specifically cited in your solution should be included. If you don’t discuss the computer output, don’t turn it in as part of your assignment.
5. Unless otherwise indicated, you do not need to turn in your actual SAS program or Log file.
6. Check and double check the accuracy of your data entry.
7. If you do not understand a problem or are having difficulty with your SAS program, please ask for help before the assignment is due.
8. Unless otherwise indicated, you may work together on the homework.
9. Solutions to homework problems will not be discussed in detail in class. I encourage you to review your graded assignments. If the correct solution to any problem is unclear, please come to see me during office hours or make an appointment.
Class Policies

1. **Contact Information**: If you need to contact me between classes, please call my office: (703) 993-4835. If you leave a voice mail message or send an e-mail and I have not returned your call/e-mail within two business days, please call/e-mail again. I do not check voice mails/e-mails over the weekend.

2. **Attendance**: Regular attendance for the full period of each class is expected (but not required).

3. **Class Etiquette**:
   a. Class will start on time at 4:30 p.m. and end on time at 7:10 p.m., with a 10-minute break around 5:45 p.m. Although situations arise making it impossible for you to arrive on time and/or require you to leave early, please remember that late arrivals and early departures can be quite disruptive to your peers as well as me. So, please make arriving to class late or leaving early an exception, not a habit.
   b. Please turn off cell phones during class.

4. **Homework Assignments**: Late homework assignments, with 10% penalty, will be accepted only until 5 p.m. on the day after the assignment is due. Please do not call/e-mail to ask for an extension; just turn in what you have finished. Due dates may be changed depending upon class situations; any changes will be posted in Blackboard. No extra-credit assignments will be given.

5. **Missed Midterm/Final**: If observance of major religious holidays or non-emergency personal issues will make it impossible for you to attend either the midterm or final, you must notify me in advance and provide appropriate verification in writing. In case of an emergency, you should obtain proper written documentation as well as notify me as soon as possible. Except in case of prior arrangement or a genuine emergency (which must be documented in writing), you will receive a zero on the missed midterm and an F in the course (University Policy) for a missed final.

6. **Special Accommodations**: If you have a disability that might require accommodations in this course, you must provide the instructor with a letter from the GMU Office of Disability Services (ODS) early in the course so that arrangements can be made to address your needs. See ods.gmu.edu, or call (703) 993-2474.

7. **Incomplete Policy**: An incomplete grade can only be given if you are passing the course and have a documented non-academic reason (e.g., health and/or family problems) why you cannot complete the course in the usual time frame.

8. **Honor Code**: You are responsible for becoming familiar with your rights and responsibilities as defined by the George Mason Honor Code; and you are responsible for knowing the requirements for this particular course. All violations of these rules will be referred to the Honor Committee; I take the Honor Code seriously and so should you. In this course, you are required to work alone on the midterm and final. Unless otherwise indicated, you may work together on the homework assignments. However, working together does not mean:
   a. handing in a photocopy/verbatim copy of someone else’s solutions.
   b. using someone else’s computer program or output or a copy of someone else’s computer session.
   c. handing in a single “group” copy of the assignment with multiple names.

Each individual must submit his/her own work to receive credit. To avoid complications with the Honor Code on homework assignments, please list on the copy of the homework you turn in the names of ALL persons with whom you worked (no matter how minimal the interaction). Please see me if you have any questions concerning the Honor Code as it relates to this particular course.