36-720: Discrete Multivariate Analysis  
Fall 2004

TTh, 12:00--1:20pm, Hamburg Hall 1003  
http://www.stat.cmu.edu/~brian/720

Course Information

Instructor:  
Brian Junker, Statistics  
132E Baker Hall  
268-8874  
brian@stat.cmu.edu

Grader:  
Cari Kaufman  
A60D Baker Hall  
268-6276  
cgk@stat.cmu.edu

Office Hours:  
TBA  
(or by appointment).

Office Hours:  
TBA  
(or by appointment).

Prerequisites

I will assume you have taken 36-705 Intermediate Statistics and 36-707 Applied Linear Regression [or a rough equivalent such as 36-625/626 Probability and Mathematical Statistics and 36-401/402 Undergraduate Advanced Data Analysis].

The computing work for this course is designed to be done in UNIX/LINUX on the Statistics Department workstations, using R, Splus, and SAS. Some homework should be written up in \LaTeX. In-class demo’s will often use the same software in Windows.

Most of what we will do can also be accomplished on Andrew LINUX workstations, or on Windows and Macintosh PC’s, and/or using other software packages. In most of these cases, you are on your own, however.

Required Texts


See also http://www.stat.ufl.edu/~aa/cda/cda.html.

Other Useful Texts


Course Description and Course Objectives

This is a mini-semester course (approx. 7 weeks) on methods for analyzing count data. We will review the theory of maximum likelihood in exponential families as it applies to the analysis of binary and categorical data. Particular models and methods we will consider include:

- Basic contingency table methods;
- Generalized linear models and logistic regression;
- Log-linear models for contingency tables;
- Extensions for displaying and modeling multivariate association and latent structure.

Along the way we will also gain experience in using statistical program packages and interpreting results, primarily in R/Splus, and perhaps also in SAS.

Since the course is only 7 weeks long, we can do nothing in detail. The idea is to give you a taste of each topic, so that when you need to do this kind of analysis in the future, you will have some idea what details are important, and where to look to find out about them.

Homework, Tests and Grading

I intend to give (almost) weekly homeworks. These will be due on Thursdays with some exceptions. Homework will be a mix of developing and exploring theoretical material, and practicing using software to analyze data.

There will also be a final exam, scheduled for December 20 at 1:00pm. The form of the final exam will be discussed later.

Homeworks together count for 75% of your grade. The final exam counts 25%.

You may work with other students on these problems or refer to other sources if you would like. The computations and writeup of your assignment, however, must be your own. Please note that the written interpretation and conclusions from a data analysis are at least as important as generation of data summaries, statistics, tests, etc.

Unless specifically requested, never submit raw computer output pages. Instead, cut out the appropriate parts of the output and neatly tape it onto your homework paper (or better yet, use \LaTeX, with verbatim for text output and psfig for figures, to put appropriate parts of the output files into your writeup). Please label all output, plots, variables, etc., appropriately.

Data Sets, Web Page, Communications

- As far as possible, all materials for this course will be made available on the course web page, http://www.stat.cmu.edu/~brian/720.
- The surest way to contact us anytime outside of class is via email:
  - Instructor: brian@stat.cmu.edu
Also please feel free to drop by our offices or schedule special appointments with either of us.

**Syllabus and Plan of Action**

Following is a **rough** outline of the topics we will cover, together with associated chapters in Agresti. As you can see, we will not cover many topics at Agresti’s level of detail.

**WEEK 1**
1. Tue Oct 26  Ch's 2-3: Contingency Tables: Description and Inference
2. Thu Oct 28  Ch 4: Generalized linear models

**WEEK 2**
3. Tue Nov 2   Ch 5: Logistic regression
4. Thu Nov 4   Ch 6: Building and applying logit models
   [NEED A MAKEUP DATE]

**WEEK 3**
5. Tue Nov 9   Ch 7: Multinomial logit models
6. Thu Nov 11  Ch 8: Loglinear models

**WEEK 4**
7. Tue Nov 16  Ch 9: Building and extending loglinear models
8. Thu Nov 18  More on log-linear models

**WEEK 5**
9. Tue Nov 23  Graphics
10. Thu Nov 25 THANKSGIVING

**WEEK 6**
10. Tue Nov 30 Ch's 12-13: Mixed models and latent variable models
11. Thu Dec 2  Ch's 12-13: Mixed models and latent variable models

**WEEK 7**
12. Tue Dec 7  Ch 10: Methods based on matched pairs
13. Thu Dec 9  Ch 10: Methods based on matched pairs

**FINAL EXAM:** Mon. December 20 1:00p.m.-4:00p.m. Room TBA.