

STA 6505 Analysis of Categorical Data – Spring 2013

Section 05F7: M W F 6 (12:50 - 13:40) in FLO 230

Instructor:

Demetris Athienitis (dathien@stat.ufl.edu)

Office: Griffin Floyd 116B,

Phone: 352 – 273 – 2972

Office Hours: Office Hours subject to change and will be posted on the course website

Teaching Assistant:

TBA

Office:

Phone:

Office Hours: Office Hours subject to change and will be posted on the course website

Course Web Address: <http://www.stat.ufl.edu/~dathien/sta6505.html>

Course Description:

This course surveys methods for the analysis of categorical response variables, from the maximum likelihood (frequentist) perspective. The main subject areas covered are descriptive and inferential statistics for two-way and three-way contingency tables, generalized linear models for discrete responses, binary regression models (emphasizing logistic regression), multi-category logit models for nominal and ordinal responses, loglinear models for contingency tables, and matched pairs.

Course Materials Required:

Categorical Data Analysis, third edition, by A. Agresti (Wiley, 2012).

The website for the text is

<http://www.stat.ufl.edu/~aa/cda/cda.html>

Software:

Examples will given and analyzed in R. Some complementary handout will given for SAS and the website for the text contains links for R, SAS and some additional software packages.

That site has a link

<https://home.comcast.net/~lthompson221/Splusdiscrete2.pdf>

to a detailed manual prepared by Dr. Laura Thompson showing how to use R and S-Plus to conduct all the analyses in the text. I highly recommend this resource if you would like to use R for statistical analyses of categorical data. There is also a link there to a website of Dr. Chris Bilder, whose link to R has examples of the use of R for many methods for categorical data (organized in terms of my lower-level text, *An Introduction to Categorical Data Analysis*).

Homework:

Complete homework solutions will be handed in to the TA to be graded on the exam dates. Homework problems are listed in the outline and it is acceptable for students to work together as long as the work handed in is uniquely yours. You are encouraged to address the TA if you have any difficulties.

Brief solutions are available at <http://www.stat.ufl.edu/~aa/cda/solutions-part.pdf>

Some homework require the use of statistical software and your solutions must show the use of software for exercises that require it by attaching relevant computer output.

Exams: There will be three exams, given on the dates that will be agreed to in class. The first exam should cover the first 4 topics and should be around **February 13th**. Exams will NOT be given in class but in the afternoon/evening.

If you are unable to take an exam at the scheduled time, you must notify the instructor as early as possible. If an emergency situation precludes an advance arrangement, you should let the instructor know within 24 hours of the missed exam. Each case will be reviewed individually. You will be required to provide official documentation to be eligible for make-up examination.

Grades:

Exam 1 and 2 account for 25% of the grade each, Exam 3 for 30% and the homework for 20%.

Grading Scale:

The grading scale will be as follows:

A:	90-100%,	A-:	87- <90%,		
B+:	84- <87%,	B:	80- <84%,	B-:	77- <80%,
C+:	74- <77%,	C:	70- <74%,	C-:	67- <70%,
D+:	64- <67%,	D:	56- <64%,		
E:	< 55%.				

To see the effect of the + and – grades on your GPA, look at the following link:

<http://www.registrar.ufl.edu/catalog/policies/regulationgrades.html>.

Course Policies:

Academic Dishonesty: I adhere to the University of Florida rules and guidelines for handling instances of academic dishonesty. Please refer to the Office of Students Services for detailed information about the current policies.

Instructor's Honor Code: We the members of the University of Florida community pledge to hold ourselves and our peers to the highest standards of honesty and integrity.

Grading: Grades will be changed only when an error has been made by the instructor.

Incomplete: Incompletes are only assigned when extraordinary circumstances, arising after the date for dropping off the course, prevent the student from completing the course requirements. The student must be currently passing the course and discuss the circumstances with the instructor before the final exam takes place. Having a failing grade in the course is not a valid reason for requesting an incomplete.

Getting Help: Students may ask questions during the lectures (preferred) or the office hours. The TAs will answer questions during office hours. A list of private tutors (if needed) may be obtained from the Statistics Department Secretary in Griffin Floyd 103.

Students with disabilities: Students requesting classroom accommodation must first register with the Dean of Students office. The Dean of Students will provide documentation to the students who must then provide this documentation to the Instructor when requesting information.

Privacy Policies: Student records are confidential. Only information designated “UF directory information” may be released without your written consent. UF views each student as the primary contact for all communication. If your parents contact me about your grade, attendance or other information that is not “UF directory information”. I will ask them to contact you.

The instructor reserves the right to update any parts of this syllabus as necessary. Students will be notified of any changes.

Note: There will be no classes on the following days

January 21 st	M.L. King Day
March 2 nd -9 th	Spring Break

Topics

1. Introduction: Distributions and Inference

Discrete distributions

Inference for categorical data

2. Describing Contingency Tables

Probability structure

Comparing proportions

Stratified tables

3. Inference for Contingency Tables

Deriving large-sample normal distributions

Chi-squared tests of independence

Exact tests for small samples

4. Introduction to Generalized Linear Models

Generalized linear models

GLMs for binary data

Inference and fitting GLMs

5. Logistic Regression

Interpreting parameters

Inference for logistic regression

Categorical and multiple predictors

Fitting logistic regression models

6. Building and Applying Logistic Regression Models

Model selection

Diagnostics

Inference in stratified tables

Power

Probit and complementary log-log link

7. Models for Multinomial Responses

Baseline-category logit models

Cumulative logit models

8. Loglinear Models

Loglinear models for two-way tables

Loglinear models for three-way tables

Inference for loglinear models

Loglinear - logit connection

10. Models for Matched Pairs

Comparing dependent proportions

Time permitting:

Inter and Intra rater reliability and classification models