

<b>STA 6238</b> <b>LOGISTIC REGRESSION MODEL</b> <b>SPRING 2009</b>
---

**MEETING TIMES:** T & R 6:00-7:15 pm,  
 Instruction Period: 01/07/2009 - 4/27/2009

**Holidays:**

Martin Luther King Jr. Day Monday, January 19  
 Spring Break March 9-14

**PLACE:** CL1- Room 218  
**INSTRUCTOR:** Xiaogang Su  
 Room 102, Computer Center II  
 (407) 823-2940 [H]  
[xiaosu@mail.ucf.edu](mailto:xiaosu@mail.ucf.edu)

**OFFICE HOURS:** Tuesday 3:00-4:00pm and Thursday 2:00-3:00pm

**TEXT:** *Applied Logistic Regression Ed. 2* by David W. Hosmer and  
 Stanley Lemeshow 2000, John Wiley & Sons, Inc

**Reference Books:**

Generalized Linear Models, 2<sup>nd</sup> Ed. By McCullagh, P. and Nelder, J. A. (1989). Chapman & Hall / CRC.

**WEBSITE:** Class notes and homework assignments will be made available at  
 the class website. <http://pegasus.cc.ucf.edu/~xsu/STA6238>

**PREREQUISITE:** STA6236 or Basic Knowledge on Statistical Modeling and SAS

**COURSE OBJECTIVE:** The logistic regression model is one of most popular models in statistical data analysis for binary outcome. In this course, I will cover from the fundamental of the logistic regression for categorical data to the cutting-edge statistical analyses associated with special study designs. SAS programming language will be the primary statistical analytical tool for the course. Throughout the semester, I will emphasize the application of logistic regression on biomedical studies in two major angles: modeling and interpreting data. I hope this course will provide students an adequate training in using their statistical knowledge and SAS programming skill to solve real problem in practice and to be able to communicate with others based on statistical results. The following topics are tentatively included for the lectures:

- Simple logistic regression model
- Multiple logistic regression model
- Model building strategies
- Model Assessment and Diagnostics
- Logistic regression for Multi-Level and Ordinal Responses
- Logistic Regression for case-control studies

- Generalized Estimating Equation (GEE) for data with repeated measures
- Introduction to Survival Analysis

**ASSIGNMENTS:** There will be about 5 midterm projects plus a final project during the semester. For the midterm projects, students are given the data sets arisen from some biomedical studies to analyze. A technical report (essay) should be submitted after the analysis for each project. In each report, students are required to present the background of the study, perform some explanatory studies for some key variables collected in the study and fit statistical models to answer the scientific questions concerned in the study.

For the final project, students are given the freedom to select data from whatever field they are interested in (and at the same time, a topic that is adequate for the course). Students should make their own plans to collect data, raise some interesting questions to study, and consult me for the adequacy of the project. Also, each student will have the opportunity to present their work in class.

The grade of this course is based on the performance of projects: each midterm project is worth 15 points and the final 20 points. There will be also several in-class quizzes, which make up 15 more points.

The grade is given according to the following scale. Note that the total is 110.

<b>Range</b>	94+	93-90	89-87	86-83	82-80	79-77	76-73	72-70	69-67	66-63	62-60	59-
<b>Grade</b>	A	A-	B+	B	B-	C+	C	C-	D+	D	D-	F

**NOTE:**

1. Students are allowed to have discussions for programming in their midterms or projects but no group work is allowed for the technical writing. No credit will be assigned to either for two nearly identical reports.
2. In any situation, late submissions of projects within 3 days after the deadline will be counted only *half* towards your final scores. Submissions after 3 more days will not be counted for any reason.