

STAT 5514
Topics in Regression
Fall Semester, 2009

INSTRUCTOR: Dr. Jeffrey B. Birch
Professor
Department of Statistics

OFFICE: 210-C Hutcheson Hall, (540) 231-7934

OFFICE HOURS: 1:00 – 2:00 M, T, W, Th or by appointment

TEXT: Schabenberger, Oliver and F. J. Pierce. 2002. *Contemporary Statistical Models for the Plant and Soil Sciences*. CRC Press. Boca Raton.

HANDOUT PACKET: A handout packet has been prepared containing most course handouts, sample midterms with solutions and sample finals with solutions. This packet is required and is available at the bookstore on the Virginia Tech campus. Students are held responsible for material contained in the handout packet. This includes all questions from assignments and exams and their solutions.

OBJECTIVES AND PREQUISITES: This course is designed to continue the student's education in the area of regression analysis. It is expected that each student has already been exposed to the first course in regression analysis (both simple and multiple regression), the theory of matrices and the general linear model. This course will cover both the theory and applications of modern regression techniques. Therefore, a high level of mathematical ability is expected of each student. Prerequisite material is that covered in the courses STAT 5044, STAT 5124, and MATH 5524. In addition, all students are expected to have a strong working knowledge of SAS.

SYLLABUS:

<u>TOPIC</u>	<u>BOOK REFERENCE</u>
I. Regression Overview	Chapter 1
II. Multiple Linear Regression Review	Chapter 4
(a) Model Description	
(b) Method of Least Squares with Matrix Notation	
(c) Maximum Likelihood Estimators	
(d) Tests on Coefficients	
(e) Confidence Intervals and Prediction Intervals	
(f) A Quick Look of Multicollinearity	
(g) Properties of the HAT Matrix	
(h) Indicator Variables	
(i) Properties of Residuals	
(j) Partial, Sequential SS	
(k) Other Estimators: LAV, M, LMS	
(l) WLS, EWLS	
(m) Heterogeneity of variance: the dual model	
(n) Example	

III. Model Selection

Chapter 4 and Lecture notes

- (a) R^2 properties
- (b) Cross Validation (PRESS Statistics)
- (c) Impact of Overfitting and Underfitting
- (d) The C_p Statistic
- (e) Brief Review of Sequential Procedures
 - Forward Selection
 - Backward Elimination
 - Stepwise Regression
- (f) Partial R^2 , partial correlation, partial H matrix
- (g) All Possible Regressions
- (h) Example

IV. Analysis of Residuals

Chapter 4 and Lecture notes

- (a) Plotting of Residuals
- (b) Studentized Residuals
- (c) Outlier Analysis
- (d) Further Diagnostic Plots

V. Influence Diagnostics

Chapter 4 (4.4)

- (a) What is Influence and Leverage?
- (b) Diagnosis of Influence
- (c) Diagnostics Measures
 - DFFITs
 - DFBETAS
 - Cooks-D
- (d) Example

VI. Collinearity Revisited

Chapter 4 (4.4)

- (a) Collinearity Diagnostics
 - Eigenvector-eigenvalue-based diagnostics
 - Spectral decomposition theorem
- (b) Ridge Regression: theory and application
- (c) Principal Component Regression: theory and application
- (d) Example

VII. Natural Nonnormal Theory Regression

Chapter 6

- (a) Regression with a Binary Response: Quantal (Binary) Regression
 - Logistic Regression
 - Model and assumptions
 - Estimation and inference
 - numerical methods
 - Example
- (b) The Generalized Linear Model
 - Model and assumptions
 - Estimation and inference
- (c) Poisson regression
 - Theory and inference
 - Example

VII. Non Linear Regression

Chapter 5

- (a) Model and assumptions
- (b) Estimation and Inference
- (c) Diagnostics
- (d) Numerical Methods
 - Gauss-Newton Procedures
- (c) Non Linear Regression or Transform to Linearize?
- (d) Example

IX. Mixed Linear Models

Chapter 7

- (a) Laird-Ware Model
- (b) Theory
 - Estimation of parameters
 - Prediction of random terms
- (c) Application

GRADING:	Homework (approximately 6 assignments	
	including project)	25%
	Two Midterm Exams	50%
	Final Exam	25%

Make-up examinations will be given only under unusual circumstances and only if prior arrangements have been made. The Virginia Tech Honor System will be strictly enforced on exams. Students may share ideas on homework problem solutions, but all work presented on solutions must be the student's own. The two midterms will not be given during our regularly scheduled lecture period but at times agreed upon by the class (tentative dates below).

ADDITIONAL COMMENTS:

1. The use of a cell phone during an in-class exam will be viewed as an honor code violation. The use of a calculator for retrieving stored formulas or theorems or for performing algebra during an exam is viewed as an honor code violation.
2. Tentative dates for the midterms are: midterm 1: Wednesday, October 7, 6:30 P.M.
midterm 2: Wednesday, November 18, 6:30 P.M.
3. The SAS statistical package will be used a great deal in this course. Handouts in the course packet illustrate all procedures required for the course.
4. One of the assignments will involve a "project" consisting of a thorough analysis of a regression data set, written in "report format" using your best writing skills. The data set may be data from your research, from a journal article, or from the Web. The data set cannot come from your textbook or from another textbook.

More details concerning the project will be given during the course.

4. The final will be given during our scheduled time (Wednesday, December 16, 4:25 PM-6:25 PM).