ERSH 8320
Applied Correlation and Regression
Methods in Education
Summer, 2005

Instructor: Steve Olejnik
Office: 325M Aderhold Hall
Phone: 542-4263
Office Hours: Monday, 12:30-2:00pm
Wednesday and Friday 11:00-2:00pm
By appointment

e-mail olejnik@uga.edu

Text: **Pedhazur E. J**. (1997). Multiple Regression in Behavioral Research: Explanation and Prediction (3nd Edition) New York: Holt, Rinehart and Winston.

Instructor notes: Applied Multiple Regression available at Bel Jean

THIS COURSE SYLLABUS PROVIDES A GENERAL PLAN FOR THE COURSE; DEVIATIONS MAY BE NECESSARY.

ERSH8320 is a WEBct course. Every student is expected to have an ARCHES account and be familiar with WEBct. This course syllabus, review questions, additional problem sets, useful publications and WEB sites, grades, a calendar of events, and a bulletin board for discussions will be maintained throughout the semester. You are encouraged to check the WEB site frequently for updated materials, discussion postings, and calendar listings. Suggestions for improving the course WEB site are greatly appreciated.

All students are expected to abide by the UGA student honor code:

"I will be academically honest in all of my academic work and will not tolerate academic dishonesty of others." The UGA policy on academic honesty states: Academic honesty means performing all academic work without plagiarism, cheating, lying, tampering, stealing, receiving unauthorized or illegitimate assistance from any other person, or using any source of information that is not common knowledge. The full version of A Culture of Honesty may be found at: http://www.uga.edu/ovpi.

Course Overview

This course is designed to present a general approach to data analysis that is useful in understanding the relationship between a single response (dependent) variable and one or more explanatory (independent) variables. While the explanatory variables may be either qualitative or quantitative the course concentrates primarily on the latter. The general approach that will be considered is referred to as regression analysis. To present this analysis procedure the course is divided into five units: I) Correlation and simple linear regression; II) Regression analysis with a single categorical explanatory variable; III) Regression analysis with two or more quantitative explanatory variables; IV) Regression analysis with one quantitative and one qualitative explanatory variable (Analysis of Covariance); and VI) Regression with two qualitative explanatory variables and (nonorthogonal ANOVA).

Unit I presents the basic ideas on which regression analysis is based. In this unit the relationship between a correlation coefficient and regression analysis is presented as well as the relative simple problem of analyzing the relationship between one quantitative explanatory variable and one quantitative response variable. A linear relationship is assumed. The remaining four units of the course present solutions to more complex problems but are basically generalization of ideas presented in Unit I.

Unit II will apply the regression analysis technique introduced in Unit I to situations involving a single qualitative explanatory variable. In this unit the link between analysis of variance and regression analysis will be established.

Unit III generalizes the procedures discussed in Unit I to situations involving two or more quantitative explanatory variables. Sequential and partial F-tests are discussed as well as the use of multiple, partial, and semi-partial (part) correlation coefficients. This unit also considers the inclusion of interaction terms in the regression model. Finally, Unit III presents several strategies to identify the "best" subset of predictors in the regression model.

Unit IV discusses the problem of analyzing data having one quantitative and one qualitative explanatory variable. This unit is actually an integration of Units I and II. Analysis of Covariance is a special case of this regression model.

Unit V presents the analysis of two qualitative explanatory variables in a factorial design. This section discusses the problem of analyzing data and interpreting the

results of studies when sample sizes are unequal and disproportional. The analysis of these data are sometimes referred to as nonorthogonal analysis of variance.

On successfully completing the course students should be capable of designing studies and analyzing data sets using regression techniques. Furthermore, students should feel comfortable reading journal articles which have used basic regression techniques in analyzing data sets. A word of caution however is needed. Students are not likely to fully understand all research studies using regression techniques. Regression analysis is a very powerful technique which can be appropriately used to analyze very complex problems. This course is designed to introduce students to the major concepts of regression analysis. For more thorough understanding of the analysis procedure further study will be necessary.

Throughout the course examples are presented, analyzed and interpreted. Some of the problems may appear artificial. The artificiality is necessary to demonstrate the technique and facilitate instruction. Students are encouraged to identify problems more relevant to their areas of interest and to share those problems with the class. Students are also encouraged to ask questions throughout the semester on the material presented. Because the material is cumulative, it is likely that students will quickly become lost and frustrated if introductory concepts are not understood. Optional exercise problems will be assigned from the end of each section and several computer assignments will also be made. By completing this work the student should attain greater understanding of the application and interpretation of regression techniques.

****

Unit	Topic	Readings	Session
I	Correlation and Simple Linear Regression	P ch 1-3,8(195-207)	1-7
	A. Representing the relationship between two variables a) Scatter plots b) Correlation coefficient c) Straight line function	I pp. 2-11	1
	B. Identifying the best fitting straight line a) Least square solution b) Interpreting the meaning of the intercept and slope c) Evaluating the line	I pp. 11-27	2
	C. Estimation and hypothesis test a) Interval estimate for $ß_{Y\|X}$ b) Interval estimate for $ß_0$ c) Hypothesis test for $ß_{Y\|X}$	I pp. 28-35	3
	D. Prediction a) Predicting an average b) Prediction for an individual	I pp. 36-42	4

Unit	Topic	Readings	Session
	E. Appropriateness of the model a) Data assumptions b) Outliers c) Using SPSS	I pp. 43-49 Computer 1	4-5
	F. Summarizing relations with correlations a) Pearson Correlation b) Spearman Correlation c) Point Biserial Correlation d) Phi Correlation	I pp. 50-68	6
II	Regression analysis with a qualitative independent variables A. Developing and interpreting a qualitative model a) Coding qualitative variables b) Interpreting regression coefficients for coded variables c) Hypothesis testing d) Interval estimation	p. 340-348 I pp. 69-76	7

Test 1 Unit I & II — Session 8, June 29

Unit	Topic	Readings	Session
III	Regression analysis with two or more quantitative explanatory variables	P Ch 5,7, 8(207-240)	9-13
	A. Overview	I pp. 77-85	9
	B. Meaning of the model		
	C. Assumptions		
	D. Multiple Correlation Coefficient		
	E. Testing the overall model	I pp. 86-98	9-10
	F. Partial F-test		
	G. Testing for an interaction	I pp. 99-102	11
	H. Multiple group regression	I pp. 103-109	
	I. R^2 increase (part correlation squared)	I pp 110-120	
	J. Partial Correlation Coefficient	Computer 2	12

Test 2 Units II and III — Session 13 July 11

Unit	Topic	Reading	Session
III	K. Identifying the "best" subset of independent variables	I pp. 121-127 Computer 3	14-15
IV	Regression analysis with one continuous and one categorical independent variable	Chapter 15 pp. 592-607	16-17
	A. Overview	I pp. 128-149, Computer 4	
	B. Identifying separate regression equations.		
	C. Testing for the equality of the slopes		
	D. Johnson-Neyman Technique		
	E. Testing for the relationship between the continuous and dependent variables		
	F. Testing for differences in the intercepts		
	G. Identifying specific differences in treatments		
V	Regression Analysis with two categorical variables independent variables	pp. 481-512 I pp. 150-158	18
	A. Factorial ANOVA		
	B. Weighted and Unweighted marginal means		
	C. Hierarchical vs Unique solutions		
	FINAL REVIEW		19

****

Final exam Comprehensive but emphasis
given to Units III K,IV and V. **July 29 2:15-5:15pm**

****

Grading

Grades in this course will be based on student performance on three tests. The test schedule is tentatively set as follows:

	Units	Day
Test 1	I and II	June 29
Test 2	III A-J	July 11
Test 3(Final)	III K,IV, V	July 29 2:15-5:15pm

All tests will be in-class and will be closed book and notes. Students will be allowed to bring two 8 1/2 by 11 page of summary notes to the first test, four 8 1/2 by 11 pages of summary notes to the second test, and five 8 1/2 by 11 pages of summary notes to the third test (final exam).

Four computer problems are also provided at the Class WEB site. Students can earn “bonus” points by successfully completing these problems. For each problem five questions are asked. To successfully complete a problem, the student must answer 4 out of 5 questions correctly. For each successfully completed computer problem one point will be added to the total points earned by the student. ALL ANSWERS TO THE COMPUTER QUESTIONS MUST BE PROVIDED BY THE STUDENT ALONE WITHOUT CONSULTATION WITH OTHER STUDENTS, COLLEAGUES, OR FACULTY. Students may ask the instructor for clarification if needed. Students may work with other students when running computer programs.

Final grades will be based on the following criteria: The total points earned will be divided by the total possible points that could be earned. It is anticipated that the first two tests will consist of 30 points and the final exam will consist of 50 points. Grades will be assigned along the following schedule

85 - 100%	A
70 - 84%	B
60 - 69%	C
50 - 59%	D
Below 50%	F

Objectives for Unit I

1. Given summary statistics compute the least squares regression equation.

2. Given summary statistics determine whether a statistically significant relationship exists between an explanatory and response variable.

3. Given a regression equation:
   a) Interpret the meaning of the slope and intercept
   b) Test hypotheses concerning the slope and intercept
   c) Provide interval estimates for the slope and intercept
   d) Use the regression equation for prediction of individual performance at a given level of X

3. Given an ANOVA summary table for a regression problem:
   a) Compute and interpret R^2
   b) Test the hypothesis that there is no relationship between X and Y
   c) Interpret residual plots

4. Given residuals and/or diagnostic information determine
   a) if assumptions are violated
   b) if influential outliers are present in the data

5. Given a computed correlation and/or summary statistics:
   a) Test the hypothesis that there is no relationship between the measures
   b) Test the hypothesis that the relationship between two measures equals a know constant
   c) Provide a confidence interval for the population parameter

6. Given a short description of a data set identify the most appropriate correlation coefficient for the scale of measurement.

7. Given computed correlation between two measures
   a) From two independent samples
   b) From related samples
   test the hypothesis that the relationships are equal

Objectives for Unit II

1. Given a research problem involving a categorical independent variable:
   a) Interpret the meaning of the coefficient for the regression solution
   b) Test the hypothesis for the difference between group means

c) Provide interval estimates for the difference between group means

Objectives for Unit III

1. Given the results of a research study involving two or more quantitative explanatory variables:
   a) Interpret the overall results of the analysis
   b) Calculate R^2 and interpret its meaning
   c) Test for statistical significance each independent variable using the R^2 for the full and reduced models and interpret the results
   d) Determine R^2 increase for each independent variable
   f) Develop interval estimates for each regression coefficient and interpret its meaning

2. Given summary statistics calculate the partial and semi-partial correlation coefficient and interpret its meaning.

3. Using Venn diagrams explain the meaning of the squared multiple correlation, squared partial correlation, and squared semi-partial correlation.

4. Given a description of a research problem present appropriate regression models used to answer specific questions and provide the appropriate test statistic.

5. Given a research context correctly interpret the meaning of an interaction.

Objectives for Unit IV

1. Given the results of a research study involving a continuous and a categorical explanatory variable:
   a) Identify and interpret the separate regression lines
   b) Test for statistical significance the equality of the separate regression slopes
   c) Test for significance differences between intercepts (or adjusted means)
   d) Provide interval estimates for differences between intercepts (or adjusted means)

Objectives for Unit V

1. Explain the difference in the calculation of weighted and unweighted marginal means.

2. Determine under what conditions weighted or unweighted marginal means would be of interest.

3. Interpret the results of a study based on a weighted or unweighted marginal means as a solution to nonorthogonal analysis of variance.

4. Write the regression models that reflect the hierarchical and unique solutions.

Final objectives

1. Given a research context identify the appropriate regression model and the specific statistic of interest.

2. Given a regression model, interpret the meaning of each regression coefficient.

In addition, students are encouraged to complete four computer problems. Assignments may be submitted to the instructor for feedback. Adequate performance and submit them by a designated date. Late papers will not be accepted without sufficient justification. Each assignment will require the student to analyze/interpret the results of a research study and answer several questions regarding the results. **THE COMPUTER OUTPUT MUST BE SUBMITTED ALONG WITH THE ANSWERS TO THE QUESTIONS.** Each assignment will be graded on a five point scale. Students may work on the computer programs in teams but all interpretation and written work on the assignments must be completed by the student alone. Data for the computer assignments will be made available on the class WEB site. Students may analyze their own data sets if they choose but must answer the same type of questions asked by the instructor.