

ERSH 8320

Applied Correlation and Regression
Methods in Education
Spring, 2007

Instructor: Steve Olejnik
Office: 325M Aderhold Hall
Phone: 542-4263
Office Hours: 12:30-4:00pm Monday and Wednesday
and by appointment
e-mail olejnik@coe.uga.edu

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

Instructor notes: Applied Correlation and 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 Chapter 1-3,8(195-207) 1-5
Simple Linear Regression
- A. Representing the relationship I 2-8 2
between two variables
a) Scatter plots
b) Correlation coefficient
c) Straight line function
- B. Identifying the best fitting I 9-27 2-3
straight line
a) Least square solution
b) Interpreting the meaning of
the intercept and slope
c) Evaluating the line
- C. Estimation and hypothesis test I 28-35 4
a) Interval estimate for $\beta_{y|x}$
b) Interval estimate for β_0
c) Hypothesis test for $\beta_{y|x}$
d) Using SPSS I 159-170

Computer Problem 1 is due Class 6 January 29

- D. Prediction I 36-42 5
a) Predicting an average
b) Prediction for an individual

Test 1 Unit I A, B, C, D Session 6, January 29

Unit Topic Readings Session

- E. Appropriateness of the model I 43-49 7-8
a) Data assumptions
b) Outliers
c) Using SPSS I 171-198
- F. Summarizing relations with correlations 9-10
a) Pearson Correlation I 50-68
b) Spearman Correlation
c) Point Biserial Correlation
d) Phi Correlation
- II Regression analysis with a p. 340-348 11
qualitative independent variables I 69-76

- A. Developing and interpreting a qualitative model
 - a) Coding qualitative variables
 - b) Interpreting regression coefficients for coded variables
 - c) Hypothesis testing
 - d) Interval estimation

 Test 2 Units I and II Session 13, February 21

Unit	Topic	Readings	Session
III	Regression analysis with two or more quantitative explanatory variables	Chapter 5,7, 8(207-240)	14-19
A.	Overview	I 77-98	14-15
B.	Meaning of the model		
C.	Assumptions		
D.	Multiple Correlation Coefficient		
E.	Testing the overall model		
F.	Partial F-test		
G.	Correlation analysis	I 110-120	16
H.	Using SPSS	I 199-208	17

Computer Problem 2 is due Class 19 March 21

I.	Interaction	I 99-102	18
J.	Multiple Groups (ANOVA)	I 103-109	19

(((((
 Spring Break No Class March 12 or 14
 ((((((

 Test 3 Units III Session 21, March 28

Unit	Topic	Reading	Session
IV	Regression analysis with one continuous and one categorical independent variable	Chapter 15 pp. 592-607 I 128-149	22-26
	A. Overview		
	B. Identifying separate regression equations.		
	C. Testing for the equality of the slopes		
	D. Testing for the relationship between the continuous and dependent variables		
	E. Testing for differences in the intercepts		
	F. Identifying specific differences in treatments		
	G. Using SPSS		
	I. Johnson-Neyman Technique		

Computer Problem 3 is due Class 27 April 18

 AERA Conference No Class April 9 or 11

H. Variable Selection	I 121-127	27-28
-----------------------	-----------	-------

Computer Problem 4 is due Class 30 April 30

V	Regression Analysis with two categorical variables independent variables	pp. 481-512 I 150-158	29
	A. Factorial ANOVA		
	B. Weighted and Unweighted marginal means		
	C. Hierarchical vs Unique solutions		

 Final exam Comprehensive but emphasis given to Units IV and V.

May 4 3:30-6:30pm

Grading

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

	Units	Day
Test 1	I A, B, C, D	January 31
Test 2	I E, F, II	February 21
Test 3	III	March 28
Test 4(Final)	IV, V	May 5 3:30-6:30pm

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

In addition, students will be required to complete four computer assignments 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 resubmit the computer assignment if five points are not awarded when completed within one week of receiving the original feedback. Students must work on the computer programs alone. All interpretation and written work on the assignments **must** be completed by the student alone. If assistance is needed ASK THE INSTRUCTOR.

Final grades will be based on a percentage of the total points earned during the semester. The total points earned will be computed by summing: 1) the two highest test scores from the first three exams; 2) the final exam score; and 3) scores earned on the computer problems. The total points earned will be divided by the total possible points that could be earned and grades will be assigned along the following schedule:

Percent of Points		Earned Grade
100-90	=	A
89-85	=	A-
84-80	=	B+
79-75	=	B
74-70	=	B-
69-65	=	C+
64-60	=	C
59-55	=	C-
54-50	=	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 known 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.