SYLLABUS
EDUC/PSYCH 7610

Research Design and Analysis II

Instructor: Kerstin E. E. Schroder, Ph.D.
Office: Utah State University, EDUC-492
Phone: 435-797-1451
Fax: 435-797-1448
Email: kerstin.schroder@usu.edu

Time & Location: M/W 4:30 – 5:45 in EDUC 413C

Office hours: TBD

Required Text: Cohen, Cohen, West, & Aiken: Applied Multiple Regression / Correlation Analysis for the Behavioral Sciences (3rd ed. 2003) (primary reading).

Purposes

This course continues the Research Design and Analysis sequence with an introduction into advanced linear regression techniques, the General Linear Model, ANCOVA, and non-linear regression methods.

Prerequisites and Pretest

To ensure that each student has the necessary background knowledge, EDUC/PSY 6600, or its equivalent, is a course prerequisite. That course and its prerequisites must be taken prior to this course and cannot be taken concurrently. Students who have not taken EDUC/PSY 6600 must provide the instructor with documentation of (an) equivalent course(s) taken at USU or elsewhere. **Further, basic knowledge of SPSS is required.**

To demonstrate mastery of the statistical pre-knowledge necessary for successful participation in this course, each student must complete a pretest administered the first day of class, as well as a homework assignment to be returned in the second session. Both pretest and homework assignment will entail general knowledge regarding hypothesis testing, the standard normal distribution, t-tests and ANOVA designs, and specifically, correlation and regression, as well as alternative correlation techniques. They are based on the course objectives listed in the common syllabi for EDUC/PSY 6570 and 6600. Those syllabi are provided in the office of the Associate Dean for Research (Education 453).

Course Content

Part A. Contents of Pretest and Review

A1 General Rules of Significance Testing

a. Describe the characteristics of the standard normal distribution
b. Explain hypothesis testing based on the standard normal distribution
c. Explain alpha and beta (Type I and Type II) errors.
d. Define and explain sampling distribution, the central limit theorem, standard error of the mean.
e. Calculate confidence intervals around a mean.

A2 t-Test for the comparison of two means

a. Distinguish independent and paired sample t-tests
b. Discuss the assumptions of the independent sample t-test
c. Calculate a t-test for the comparison of two means and test the hypothesis that the two means are equal.
d. Perform t-tests via computer (SPSS) and interpret the results.

A3 One-way and Factorial ANOVA

a. Describe ANOVA as an extension of the t-test
b. Demonstrate how to calculate the sums of squares and the F-tests for one-way and factorial ANOVA.
c. Perform factorial ANOVAs via computer (SPSS) and interpret the results.
d. Interpret significant interactions.

A4 Correlation and Regression

a. Define the concept of a correlation
b. Distinguish between different kinds of correlation coefficients and discuss their applications (Pearson, Spearman, point-biserial ….)
c. Calculate a covariance
d. Explain the relationship between covariance and the Pearson correlation
e. Calculate a Pearson-product moment correlation between two variables
f. Test the significance of a correlation
g. Discuss the assumptions of correlation and regression
h. Calculate and interpret the coefficient of determination
i. Calculate and interpret regression coefficients (intercept, slope)
j. Compare “raw” regression lines and coefficients to standardized ones.
k. Define, calculate, and interpret the standard error of prediction
l. Calculate confidence limits on the dependent variable Y

Part B. Multiple Regression

B1 The Multiple R and R^2 with Two or More Predictors

a. Explain multiple regression as an extension of simple regression
b. Define, calculate, and interpret partial and semipartial (part) correlations
c. Define and explain multiple regression in terms of increments to R^2
d. Perform multiple regression via computer (SPSS) and interpret the output
e. Define multiple regression with multiple independent variables as a series of incremental variance explained
f. Explain the meaning of the "shrunken R^2" in multiple regression
g. Calculate and interpret standard errors and confidence intervals for B and Beta
h. Predict an outcome score based on a multiple regression equation
i. Explain the concept and need for cross-validation.

B2 Data Analytic Strategies in Multiple Regression

a. Distinguish and decide on different strategies in multiple regression analysis
b. Perform hierarchical linear regressions (SPSS) and interpret the results
c. Test the significance of predictor SETS
d. Estimate power for partial coefficients and predictor sets.
e. Test mediator hypotheses via multiple regressions

B3 Diagnosing and Solving Regression Problems I

a. Retrieve graphical display of data for visual inspection and interpret
b. Discuss the assumption in linear multiple regression
c. Detect violations of assumptions
d. Discuss and apply remedies to violations of assumptions.

B4 Diagnosing and Solving Regression Problems II

a. Multivariate outliers: Define and interpret leverage and influence statistics as indicators of multivariate outliers
b. Discuss potential sources of outliers and potential remedial actions
c. Define and diagnose multicollinearity
d. Discuss and apply potential remedies for multicollinearity.

1

B5 Interactions Among Continuous Variables in Multiple Regression

a. Define interaction effects relative to additive effects and explain the multiplicative function between predictors
b. Explain the need for centering or standardizing of predictors
c. Perform centering / standardization and create interaction terms
d. Explain the difference between centered and standardized solutions
e. Perform interaction analyses (SPSS) and interpret the results
f. Calculate simple regression equations and simple slopes
g. Plot and interpret the nature of significant interactions
h. Calculate standard errors and confidence intervals for simple slopes
i. Test simple slopes for significance;
j. Discuss various patterns of first-order and interaction effects

Part C. General Linear Model I: Categorical Predictors in Regression

C1 Dummy Coding and Dummy Coding Strategies

a. Discuss the need to dummy code nominal/categorical predictor variables with more than two levels
b. Apply a simple dummy coding scheme
c. Perform a multiple regression with simple dummy codes and interpret the results
d. Predict y-scores based on a multiple regression equation with dummy predictors
e. Calculate an F-test for the set of dummy predictors
f. Calculate confidence intervals and significance tests for the regression coefficients
g. Interpret partial and semipartial correlations for dummy variables.
h. Compare multiple regression with dummy codes to the respective one-way ANOVA
i. Define and discuss weighted and unweighted effect coding strategies
j Create weighted and unweighted effect dummy codes and compare the results
k. Define orthogonal (contrast) coding and create orthogonal codes based on this definition
l. Perform multiple comparisons between groups via beta values
m. Test for differences between betas

C2 Testing Interactions with Categorical Variables

a. Create dummy codes for interactions between two categorical variables
b. Perform multiple regression analyses for two categorical predictors and their interactions
c. Disentangle main and interaction effects in multiple regression
d. Reconstruct the ANOVA summary table from a multiple regression with categorical predictors.

C3 Analysis of Covariance (ANCOVA)

a. Define: Analysis of covariance, covariate, adjusted means
b. Explain typical uses/ research questions that require the use of ANCOVA.
c. Compare and contrast ANOVA and ANCOVA.
d. Discuss the considerations in deciding whether to use a variable as a covariate in an analysis of covariance or as a classification variable in a two-way analysis of variance.
e. Discuss the function of correlation in analysis of covariance.
f. Discuss assumption about slopes of regression lines underlying the use of analysis of covariance.
g. Calculate and discuss the *F*-ratio for homogeneity of regression coefficients (slopes).
h. State the null hypothesis for an analysis of covariance.
i. Interpret the *F*-ratio from an analysis of covariance, including statistical significance and conclusions.
j. Compare and interpret adjusted and unadjusted means for an analysis of covariance.
k. Compute and interpret Eta^2 for an analysis of covariance.

Part D. General Linear Model II: Nonlinear Regressions

D1 Logistic Regression

a. Discuss and develop research questions appropriate for Logistic Regression with a dichotomous outcome
b. Describe the logistic regression curve
c. Discuss the relationship between logistic and linear regression
d. Describe the maximum likelihood procedure used to estimate the coefficients
e. Describe and interpret indices of overall fit of the logistic regression model
f. Define logits, probabilities, odds, log of the odds
g. Interpret logistic regression coefficients
h. Interpret R^2 calculations in Logistic regression (Cox & Snell, Nagelkerke)
i. Define and interpret Odds Ratios
j. Predict group membership based on logistic regression equations
k. Perform hierarchical logistic regressions and interpret the results
l. Apply the AIC and BIC criteria to model comparisons
m. Discuss the limitations of logistic regression

D2 Predicting Counts: Poisson Regression and Negative Binomial Regression

a. Discuss appropriate applications for Poisson regressions
b. Describe the characteristics of a Poisson probability distribution
c. Review formulas for Poisson regression and compare to logistic regression formula
d. Interpret Poisson regression output
e. Define the term “overdispersion” and explain why it is a problem in Poisson regression
f. Discuss negative binomial regression as an extension of Poisson regression for variables with overdispersion
g. Discuss the parallels between Poisson and Logistic Regression in the context of the General Linear Model.

D3 If Data Don’t Fit into a Specific Regression Model: Nonlinear Data Transformation

a. Describe strategies to examine non-linear relationships in multiple regression
b. Define: Power polynomials, quadratic fit, cubic fit
c. Discuss purposes of data transformations.
d. Describe and apply different types of non-linear data transformations.
e. Perform model checks before and after transformation.
f. Discuss limitations of data transformations.
g. Define and discuss non-linear and non-parametric regression as alternatives

Assignments & Tests

Regular class attendance is expected and – based on prior experience – absolutely necessary to prepare for homework assignments and succeed in this class. Written homework assignments will be given each week and will be due one week later. Both the statistical correctness and the quality of interpretations and discussions will be evaluated in the returned homework assignments.

Exams are designed as "learning exams." They are comprehensive and time consuming and typically substitute for homework assignments (i.e., simultaneous assignment of homework during exam weeks is avoided as far as possible). The exams are designed to be challenging and typically require correction. **Students will receive feedback on their work and have the opportunity to make suitable corrections in order to improve their scores.**

Students will have opportunities to analyze data and discuss the outcomes, including SPSS printouts, both orally and in writing. Thus, knowledge of the SPSS software is required.

Grading

Homework Assignments:
Homework assignments will have variable point allocation, dependent on the difficulty of the assignment and the amount of work involved. Point assignment will be disclosed in detail on the assignment at the time of its distribution. All points accrued through homework will be summed into a total score, which will converted into a percentage score (percentage correct).

Altogether, the points accrued though assignments determine **40 % of Total Grade**

Tests & Final	**% of Total Grade**
1. Multiple Regression	15
2. Interactions via regression, GLM - categorical predictors	25
3. ANCOVA, GLM – Nonlinear regression strategies	20

Altogether, the scores obtained through tests determine **60% of Total Grade**

Grading Scale
Final grades will be based on the sum of all points accrued and adhere to the following scale:

A	95 - 100%	C+	78 - 79%
A-	90 - 94%	C	74 - 77%
B+	88 - 89%	C-	70 - 73%
B	84 - 87%	D+	68 - 69%
B-	80 - 83%	D	60 - 67%
		F	Below 60%

Changes in Course Assignments and Schedule:
The instructor reserves the right to adjust the course readings, assignments, tests, and schedule in order to best attain the objectives of the course. Any changes will be announced in class.

Grading Policy

Grading Policy for Assignments

1. Full points can only be earned for homework assignments RETURNED ON TIME. For each day the assignment is returned late, 12.5% of the points will be discounted.

   Example: Due date is Monday, January 23. Feedback is usually provided in the next session after the due date, which would be Wed., January 25.

   Scenario 1: You return your work on Tuesday, January 24; this means that you can only earn 100% – 12.5% = 87.5% of the points possible (which is equivalent to a maximum grade of B+ for this work).

   Scenario 2: You return your work on Wednesday, January 25, the day feedback will be given. This means that you can only earn 100% - 2*12.5% = 75% of the points for your work.
   In addition, the instructor will not be able to give you personal feedback on your work, which you might need to correct it properly.
   Another disadvantage is that the instructor will not be able to address particular problems that you had with these tasks in the feedback session!

   Scenario 3: You return your assignment on Thursday, January 26, after feedback was provided in class and the tasks were worked upon in class. No points will be awarded in this case! (see point 2).

2. Assignments will NOT be accepted once feedback is given and problems were discussed in class.

3. You are eligible for corrections of those tasks that you seriously attempted to solve. In order to prevent that assignments are returned "empty" because the student expects to be able to "correct" everything after feedback has been provided, the following rules will be applied:
   (a) Assignments that were returned with less than 50% correct will be subject of an overall discount of 10% of the points after correction.
   (b) Assignments that were returned with less than 80% of the tasks *even attempted* will be subject to an overall discount of 20% of the points after correction.

4. In order to fully update your points, you need to meet the deadlines set for the corrections of your work. **Typically, you have 1 week (7 days) for corrections after an assignment has been returned to you,** unless announced otherwise in class. If you need more time for corrections, please inform the instructor and agree on a new deadline in order to be eligible for full points.

5. If you miss the **deadline for *corrections***, 10% of the points earned through your corrections will be discounted for each day you submit late.
   Example: In your first attempt, you achieved 70% of the points. Your subsequent corrections add 25% to your points (lifting your score to 95% correct). However, because your corrections were submitted 2 days late, only 80% of the additional 25 percentage points will be granted, reducing your overall score to 90% correct).

Grading Policy for Exams

Exams typically require applying the exercises trained in assignments to new problems with the aim to demonstrate mastery of the topics covered in class. At the time of the exam, you are expected to have obtained the knowledge to complete these tasks independently and correctly.

Accordingly, stricter rules are applied to exams: Only 50% of the points missed on exams can be regained through corrections.

Further, the following rules apply to remain eligible for point updates:

1. 50% of the missed points can be regained for exams **RETURNED ON TIME**. No grace period for late exam submission!!!

2. Minimum requirements to be eligible for reworks:
   (a) You need to have at least 60% of the points earned in your first attempt.
   (b) You can only update exam tasks that you attempted to solve.

3. You need to meet the deadlines set for the corrections of your work. Except for the final exam, you **typically, you have 1 week (7 days) for corrections after receiving feedback,** unless announced otherwise in class. If you need more time for corrections, please inform the instructor and agree on a new deadline.

4. If you miss the **deadline for *corrections***, 10% of the points earned through your corrections will be discounted for each day you submit late.
   Example: In your first attempt, you achieved 80% of the points. In your subsequent corrections, you show a perfect performance, but you can only add 10 points (50% of the 20 points) to your exam, (lifting your score to 90% correct). . However, because your corrections were submitted 2 days late, only 80% of the additional 10 percentage points will be granted (8 out of 10), reducing your overall score to 88% correct.

Course Schedule

Week	Date	Topics	Readings & Assignments (A)
1	Aug. 25/27	Introduction and Pretest Review of 6600 statistics	Handout; **A1** distributed
2	***Sep. 01***	***Holiday, no class***	
2	Sep. 03	Correlation and Regression, Intro to Multiple Regression	Handouts; Cohen-Chapter 2, or Howell-Chapter 9, **A2** distributed
3	Sep. 08/10	Multiple Regression	Handouts; Cohen-3, or Howell-15, **A3**
4	Sep. 15/17	Regression Diagnostics	Cohen-4 (Howell-15), **A4**
5	**Sep. 22/24**	Regression diagnostics; Review and **Exam I**	Cohen-4, Cohen-10, ***Exam I due: Oct. 01***
6	Sep. 29/Oct. 1	Interactions with continuous predictors; testing and plotting interactions	Handout, Cohen-7, **A5**
7	Oct. 06/08	Interaction with continuous predictors continued: testing simple regression equations	Handout, Cohen-7, **A6**
8	Oct. 13/15	Interaction with continuous predictors continued: standardized solution	Handout, Cohen-7 **A7**
9	Oct. 20/22	Intro to the GLM: Categorical predictors in regression, dummy coding strategies	Handout, Cohen-8, **A8**
10	Oct. 27/29	GLM continued: Factorial ANOVA via regression	Handout, Cohen-8 (Howell-13), **A9**
11	Nov. 03/05	Review of Interaction Analyses, Categorical Predictors, Dummy Coding, **Exam II**	***Exam II due date: March 31***
12	Nov. 10/12	ANCOVA Review Exam II	Handout; Howell-16, Pedhazur-13; **A10**
14	Nov. 17/19	Non-linear Regression Techniques I: Logistic Regression	Handout; Cohen-13, **A11**
15	Nov. 24	Nonlinear Regression II: Poisson & Negative Binomial Regression; Linear regression following non-linear data transformations	Handout; Cohen-13, Cohen-6; **A12**
15	**Nov. 26**	***Thanksgiving weekend, no class***	
16	**Dec. 01/03**	Review of non-linear regression techniques; **Exam III**	***Exam III due date: April 28***
17	Dec 08-12	Final's week	***Feedback Exam III: Dec. 10***

Plagiarism

Plagiarism is considered academic dishonesty or academic fraud and is concerned with the issue of false attribution. The UNC Honor Court defines plagiarism as "the deliberate or reckless representation of another's words, thoughts, ideas as one's own without attribution in connection with submission of academic work, whether graded or otherwise." (*Instrument of Student Judicial Governance*, Section II.B.1.). Plagiarism has serious consequences, which may include grade adjustment, academic failure, probation, expulsion, and denial or revocation of academic degrees.