

Family & Consumer Studies 6110/5110

Graduate Multivariate Statistics Spring 2005

Instructor: Dr. Ken R. Smith
Office: 216 Alfred Emery Building
Phone: 581-7847; Main Office 581-6521
Office Hours: By appointment
E-Mail: ken.smith@fcs.utah.edu
Class Hours: Tuesday and Thursday 12:25-1:45 pm, JTB 320 (James Talmadge Building)

Required Textbooks: David Knoke, George W. Bohrnstedt and Alisa Potter Mee. Statistics for Social Data Analysis. Hardcover. 4th Edition, (2002) F.E. Peacock.

Ronald P. Cody and Jeffrey K. Smith. Applied Statistics and the SAS Programming Language. Paperback. 4th edition (April 1997). Prentice Hall.

Course Objectives

- I. Apply principles of empirical research in the social sciences.
- II. Review steps needed to move from generating hypotheses to data analysis.
- III. Become familiar with statistical tools used most commonly in the social sciences. Hope that you get to be quasi-consultants with your peers and real consultants with real clients.
- IV. Learn how to use SAS or Statistical Analysis System, one of the most widely adopted statistical software packages in the galaxy.
- V. Learn how to present their results graphically whenever possible.
- VI. Conduct multivariate analyses on a research question of choosing; all students will analyze data from the same survey (General Social Survey) unless you have your own data.

Course Logistics

LECTURE AND SEMINAR - This is a technical course which covers a fair amount of mathematical material. For this reason, most of the class time will be spent in more formal lectures. However, FCS 5110/6110 is also a graduate seminar where **you are expected (1) to stay on schedule with respect to the readings and (2) to PARTICIPATE ACTIVELY in class discussions.**

HOW MUCH DO YOU NEED TO KNOW ABOUT THE COMPUTER OR SAS - There will be many assignments that require the use of the computer. Everything you need to know about the computer and its operation will be covered in the labs. You will be using the FCS Department's PCs in the lab (Rm 330) in combination with SAS.

QUESTION: HOW WILL MY GRADE BE DETERMINED?

First Midterm	20%
Second Midterm	20%
Final	25%
Homework	“10%”
Poster	25%

CAVEAT EMPTOR

Timetable for topics is subject to change as needed

Poster

You will be exposed to a large national data set that spans a wide variety of topics. You will be asked to raise a research question that can be answered with these data, conduct the necessary (albeit brief) literature and theoretical review, and then conduct an analysis of the data based on your research question. The necessary steps to be completed for this project are:

1. Propose a research question that can be addressed with these data. I would urge you to think about a topic that is as close as possible to your actual thesis topic although no one is likely to find a perfect match between their interests and the data. You will be asked to prepare a 1-page prospectus (**Due THURSDAY January 27th**)
2. You will want to be creative about what topic you use but you need to also be realistic. By this I mean, you should keep in mind the numbers of people who might actually qualify for your study (e.g., single mothers with young children who have less than a high school education) and the variables that were collected. No doubt you will think of variables that were not collected but you will also find interesting variables that you hadn't thought of.
3. Describe and justify (i.e., brief literature review) specific hypotheses that you wish to test. Provide a very brief theoretical basis for your work.
4. Lay out a basic MULTIVARIATE MODEL that can be analyzed using some form of multiple (linear or logistic) regression or ANOVA. I reward creative AND relevant modeling (e.g., interactions, multiple sub-samples, nonlinear effects). Be BRUTALLY CLEAR and PAINFULLY UNAMBIGUOUS about your dependent and independent variables.
5. Using SAS, analyze the data and prepare a poster about your study. This includes your research statement (brief but clear), literature/theoretical review, hypotheses, variable description, model justification, analysis results, and interpretation.
6. We will have a class conference at the end of the quarter where posters will be presented.
7. Parameters
 - a. Clarity and justification of research question
 - b. Specificity and testability of hypotheses
 - c. Motivation for research question (i.e., why is your question important?) - include information about this topic from the published literature
 - d. Model specification - stating independent and dependent variables, how the variables are measured, rationale/type of technique used, variable selection
 - e. Elaborating on the model - interactions, assessing multicollinearity, influential observations, goodness-of-fit, sample stratification.
 - f. Data description - who is in the sample, sample size, how the sample was created, descriptive statistics
 - g. Results - assess whether the hypotheses were supported or not, state in plain English what your results mean, describe problems you faced estimating the models (by the way, the perfect analysis with no problems has yet to be found)
 - h. Graphic presentation
 - i. Conclusions and what you would like to do in the future.
 - j. Overall appearance of the poster - readability, clarity, graphics, compelling question

Week	Dates	Domain	Selected Topics	Readings
1 2	Jan 11,13, 18, 21	Learning SAS on the PC	SAS Reading/Writing SAS Data Sets SAS On-Line Guides	CS Ch 1-2 CS Ch 12-13
3	Jan 25	Introduction Computing and Statistical Overview	Variables & Hypothesis Testing Sampling Role of Theory General Linear Model	KBM Ch. 1
			Data Cleaning and Descriptive Statistics in SAS	CS Ch 2,3
			Central Tendency in SAS Measures of Variation SAS PROSPECTUS IS DUE JAN 27th	KBM Ch 2
3	Jan 27	Statistical Inference	Expected Value Normal Distribution & Central Limit Theorem Point Estimates & Confidence Intervals t distribution Type I and II error	KBM Ch 3
4	Feb 1, 3	Analysis of Variance-ANOVA	1-way ANOVA Eta (η), F test, t test, Z test	KBM Ch 4 CS Ch 7 (pp 150-159)
⇒⇒⇒⇒1st Midterm Exam - Thursday Feb 8 ←←←←				
5 6	Feb 10, 15	Two-Way Tables	χ^2 distribution and test Non-parametric measures of association Odds Ratio	KBM Ch 5 CS Ch 3
6 7	Feb 17, 22, 24	Regression and Correlation	General Linear Model Principle of Least Squares Hypothesis Testing for Intercept & Slope R squared/Goodness-of-fit Standardized coefficients	KBM Ch 6 CS Ch 5
8	March 1	Multivariate Contingency Tables	Making statistical adjustments Confounding, Ceterus Paribus	KBM Ch 7
8 9	March 3, 8	Multiple Regression	Multicollinearity Interaction Dummy Variables Model Selection Regression Diagnostics & Assessing residuals Dependent/Correlated Observations	KBM Ch 8 CS Ch 9 (pp.221-235)
⇒⇒⇒⇒2nd Midterm Exam - March 10th THURSDAY ←←←← ⇒⇒⇒⇒SPRING Break No Class Week of March 14th←←←←				
10 11	March 22, 24	Multiple Regression	continued	
12 13	March 29 April 5, 7 (No class March 31)	Nonlinear Regression	Binary Logistic Regression Problems with "Marginal effects" Odds and Odds ratios Multinomial Logistic Regression Ordinal Logistic Regression Discriminant Analysis	KBM Ch 9 CS Ch 9 (pp. 235-247)
14 15	April 12, 14, 19, 21	Path/Factor Analysis	Direct and Indirect Effects Recursive Models	KBM Ch 11 KBM Ch 12 CS Ch 10
16	April 26	Posters		
⇒⇒⇒⇒Final Exam - TBA←←←←				

