Introduction to Quantitative Methodology (G93.2332; E10.2332)

This course provides a basic practical introduction to quantitative sociological methodology. Students will gain an appreciation for both the methodological logic and techniques of statistical data analysis. While the course will cover the purpose, goals and mathematical assumptions behind techniques of statistical analysis, the emphasis will be on actually analyzing data and interpreting results.

The course will meet for two-hours sessions twice a week. Monday’s session will typically be a lecture and discussion of methodological issues and concrete examples of their application. The second session of the week will typically be a lab class, where students will have the opportunity to work on computer data analysis with a teaching assistant present. During weeks 4-6 a slight variation on this scheduling routine occurs to accommodate holiday and work scheduling disruptions.

Course requirements include weekly assignments that will usually involve computer data analysis (10% of semester grade for assignment 8; 10% of semester grade for all other assignments combined), two mid-term exams (30% of semester grade) and a final course paper that makes use of statistical analysis (50% of semester grade). A one or two page memo identifying the topic and direction for your final course paper is due March 10; the final course paper is due May 9.

Materials


SAS (SPSS or STATA) computing manual recommended.
Letters by class dates below indicate: A, Arum lecture; R, Roksa lab; NC, no class.

Week 1 (1/27A, 1/28R)
  Class: Introduction on course structure, research design, quantitative data and methods
  Lab: Assignment 1
  Text: B&K, chapter 1; problems 5,7,9

Week 2 (2/3A, 2/4R)
  Class: Descriptive Statistics
  Lab: Assignment 2
  Text: B&K, chapter 2; problems 6,7,8,10

Week 3 (2/10A, 2/11R)
  Class: Probability Statistics and Statistical Inference
  Lab: Assignment 3
  Text: B&K, Chapter 3; problems 2,3,4,8,10

Week 4 (2/17NC, 2/18A, 2/24R)
  Class: Testing for Differences between Means and Chi-square
  Lab: Assignment 4
  Text: B&K, Chapters 4+5; problems 7,8,9 (chap 4); problem 2 (chapter 5)

Week 5 (2/25R)
  Exam – midterm 1

Week 6 (3/3R, 3/4A)
  Class: Regression with One Variable
  Lab: Assignment 5
  Text: B&K, Chapter 6; problems 3,4,5,8

Week 7 (3/10A, 3/11R)
  Class: Regression with Two Variables I
  Lab: Assignment 6
  Text: B&K, Chapter 8; problems 3,4,5,6

SPRING BREAK

Week 8 (3/24A, 3/25R)
  Class: Regression with Two Variables II
  Lab: Assignment 7
  Text: Chapter 8; problems 7,9,10
Week 9 (3/31A, 4/1R)
   Class: More on Multivariate Regression I
   Lab: Assignment 8
   Text: Berry and Feldman

Week 10 (4/7A, 4/8R)
   Class: More on Multivariate Regression II
   Lab: Assignment 9 (begin)
   Text: Fox

Week 11 (4/14A, 4/15R)
   Exam -- Midterm II
   Lab: Assignment 9 (due)

Week 12 (4/21A, 4/22R)
   Class: Variations on OLS modeling I
   Lab: Assignment 10
   Text: B&K, chapter 9; problems 3,4,5,6,7

Week 13 (4/28A, 4/29R)
   Class: Variations on OLS modeling II
   Lab: Work on papers
   Text: Aldrich and Nelson

Week 14 (5/5NC)

   FINAL PAPERS DUE IN MY SOCIOLOGY MAILBOX -- NOON, MAY 9
COMPUTER ASSIGNMENTS:

All class computer assignments will be based on a subset of questions from the General Social Survey 1972-1996. Please bring assignments to class.

Assignment 1: Data
- Set up computer account
- Examine GSS codebook
- Examine ICPSR data archive website (www.icpsr.umich.edu)
- Examine Berkeley’s data archive website (csa.berkeley.edu:7502/archive.htm) and explore GSS archive (note: the class dataset includes only variables from case identification, personal and family information, workplace and economic concerns, controversial social issues).
- Identify one GSS variable that you think is likely highly reliable and valid; and one GSS variable that is less so. Write a paragraph explaining your choice.

Assignment 2: Dummy variables and descriptive statistics
- Restrict the sample to individuals aged 25-60.
- Create a set of dummy variables for gender and race as follows:
  - MALE
  - BLACK, WHITE, OTHER
- Identify the mean, median and modal responses for years of education; conduct the analysis separately for white men, white women, African-American men and African-American women.

Assignment 3: Recoding categorical and continuous variables; t-tests of means
- Restrict the sample to individuals aged 25-60
- Create a new categorical variable called NEWEDUC and code it based on years of education: 1(<12), 2(12), 3(13-15), 4(16+)
- Create a new variable called PAREDUC and code it as a categorical variable similar to NEWEDUC. Base the variable on the highest education reported for either father or mother.
- Create a new variable called INHEIRED assigning individuals to categories 1(<12), 2(12), 3(13-15), 4(16+) if both respondents and the highest reports of parental education are identical; assign all other individuals missing values
- Create a new variable called ABSCALE based on the sum of positive responses to ABNOMORE, ABDEFECT, ABRAPE, ABPOOR, ABSINGLE, ABHLTH (assign as missing if missing data on any prompt)
- Run a cross tabulation of NEWEDUC by PAREDUC
- Use a two-tailed t-test at the .05 level to test the hypothesis that the mean value on the abortion scale in the general population is not 4.2.
- Run descriptive statistics (n, means, sd, min, max) on all recoded variables.
Assignment 4: Testing mean differences and chi-square
   Use a two-tailed z-test at the .05 level of significance to test the hypothesis that respondents in INHEIRED categories 1 and 4 have different mean values on the abortion scale.
   Calculate chi-square statistics for the table NEWEDUC by PAREDUC.

Assignment 5: Univariate analysis
   Restrict the sample to individuals aged 25-55.
   Create a new variable PAREDUC2 for highest years of education either parent
   Regress RINCOME on EDUC
   Regress EDUC on PAREDUC2
   Regress RINCOME on PAREDUC2
   Regress ABSCALE on EDUC

Assignment 6: Multivariate analysis
   Restrict the sample to individuals aged 25-55.
   Regress EDUC on PAREDUC2, MALE, BLACK, OTHER
   Regress RINCOME on EDUC, PAREDUC2, MALE, BLACK, OTHER

Assignment 7: Interaction terms
   Restrict the sample to individuals aged 25-55.
   Create an interaction term MALEEDUC (MALE*EDUC) and MALEPAED (MALE*PAREDUC2)
   Regress RINCOME on EDUC, PAREDUC2, BLACK, and OTHER separately for men and women. Conduct t-test for significant differences in coefficients
   Regress RINCOME on EDUC, PAREDUC2, MALE, BLACK, OTHER, MALEEDUC
   Regress EDUC on PAREDUC2, BLACK, and OTHER separately for men and women. Conduct t-test for significant differences in coefficients
   Regress EDUC on PAREDUC2, BLACK, OTHER, MALE, and MALEPAED

Assignment 8: Generation and analysis of original data
   Survey the students in our class for information on the following: gender, highest years of education of either mother or father, and their age when they first entered a post-graduate program (not necessarily this one).
   Create a SAS dataset that has the following four variables ID, GENDER, PARED, AGEGRAD (where ID is a variable coded 1,2,3… for the different observations)
   Regress AGEGRAD on GENDER, PARED.
   Regress AGEGRAD on GENDER, PARED and an interaction of GENDER*PARED.

Assignment 9: Secondary analysis of existing data
   Download, extract and create an individualized (smaller, workable) dataset to be used for your own research papers
   Recode a set of variables necessary for analysis.
   Generate descriptive statistics on all (recoded) variables to be used in the analysis (n, means, sd, min, max).
Assignment 10: Dichotomous dependent variables
   Restrict the sample to those individuals aged 25-55.
   Create a dummy variable HIGHGRAD for high school graduates.
   Create a dummy variable COLGRAD for college graduates.
   Regress HIGHGRAD on PAREDUC2, MALE, BLACK, OTHER (using both OLS and logistic regression).
   Restrict the sample to those who finished high school and regress COLGRAD on
   PAREDUC2, MALE, BLACK, OTHER (using both OLS and logistic regression).