Bryn Mawr College
The Graduate School of Social Work and Social Research

Data Analysis II
#541

Spring, 2004

Professor Vartanian

Course Description

This second semester course will give students a more in-depth understanding of the concepts learned in the first semester data analysis class. Many new areas in statistics will also be studied. Analysis and interpretation using analysis of variance, analysis of covariance, and ordinary least squares regression models with dummy and interaction variables as well as log and other transformations will be examined. The Chow test, the likelihood ratio test, and many other tests will be studied to determine, for example, whether separate models for groups is appropriate instead of using all observations within a single model. The class will then examine nominal and ordinal level dependent variables and the types of statistical models that can be used with these dependent variables. Data reduction techniques, such as principal components and factor analysis will be examined near the end of the semester. Each of these statistical techniques will help you to determine the effects of being in particular conditions on outcome variables.

Each student will develop and conduct an original analysis of data from the 1992/1999 Panel Study of Income Dynamics (PSID), the Child/Adolescent sample, the Child Development Supplement, or the National Educational Longitudinal Study (NELS). The data will be available on the Compact disks that I will give you. The first three of these data sets originate from the Survey Research Center at the University of Michigan. I have extensively manipulated the data sets so that they contain variables of interest to doctoral students. Students are encouraged to go to the web page for the data (at http://simba.isr.umich.edu/ALL/) to examine the data further to determine if there are other variables that they would like to use in their analyses. The NELS is data that comes from the National Center for Education Statistics, U.S. Department of Education.

The first two data sets (the ’92/’99 PSID and the Child/Adolescent sample) come from the Panel Study of Income Dynamics (PSID). The PSID is a longitudinal data set that started in 1968 and has continued for 34 years, with the last version available for 2001. The PSID is a representative data set of the U.S. population when weighted (the sample is a disproportionate stratified sample -- we’ll talk more about weighting later). The first data set, the Panel Study of Income Dynamics 1992/1999, includes over 4,000 individuals from the United States. I have included information for a group of individuals that were in the sample in both 1992 and in 1999. The data include information on such characteristics as income, race, education, economic situation while growing up, health status, religious preferences, housing status, whether the person was born with low birth weight, and government income assistance. The variables from 1999 contain information on such health conditions as whether the individual has cancer, heart disease, has ever had a stroke or a heart attack, has mental loss or emotional problems, or has a learning disability, and the number of years the person has had these conditions. You will be able to determine the relationship
between variables for the entire sample or for particular subgroups (such as the elderly, the poor, single mothers, people of different races, husbands, wives, single adults, or those with little education). The PSID is used by a great number of researchers throughout the country because it provides a wealth of information on a nationally representative group of individuals and families.

The Child/Adolescent Sample comes from the 1968 to 1997 panels of the PSID. We examine variables for children aged 10-14 (generally – some variables are from when the child was born) and then examine these same children when they become adults. For example, we examine such childhood/adolescent variables as parental level of education, whether parents are married or not, marital status of the child’s mother when the child was born, birth weight, income, wealth, and government assistance. Also included in the data are adult variables such as age when first married, number of marriages, number of children, health status, religious preference, level of education, income, occupation, wealth and government assistance received (such as welfare). From this data, you can determine which childhood/adolescent factors help predict adult outcomes. There are nearly 4,000 individuals in this child/adolescent sample.

The third data set is called the Child Development Supplement (CDS). It is a supplement to the Panel Study of Income Dynamics. These data contain information on children, aged 0 to 12, and their primary caregivers. From these data, you can help determine what factors affect children’s health, behavior, cognitive ability, and a great variety of other outcomes. For example, you can determine how family circumstances while growing up, such as the mother’s emotional state, parental expectations of the child, parental disciplinary practices, whether the child comes from an immigrant family, the level of cognitive stimulation for the child, child’s perception of self, family income, perceived neighborhood safety, race, wealth, family composition, welfare receipt, religious practices, and many other variables, affect how children fare. Some of the principle outcome variables for these data include reading and math abilities, both internal and external behavioral problems as well as overall behavioral problems, health issues, including a great variety of problems, and many other possibilities.

The fourth data set, the NELS, was first assembled in 1988. The sample includes over 26,000 eight graders in 1988 and follows many of these students until 2000, the latest year of the NELS. Information was collected by interviewing students, parents, teachers, and school administrators. The data includes a wide variety of questions relating to grades (including grades in specific subjects), parental involvement in the students lives, violence at school, drug use, self image, highest level of schooling, parental expectations of the student, sexual activity, and many demographic and other variables.

Each of you is advised to take a close look at the codebooks for each of the data sets to get a more complete picture of what each set of data contains. In the case of the NELS data, the codebook is only available in electronic form, which you will be working with in one of the computer assignments. Each student will select variables to use in hypothesis testing and use SAS to execute appropriate analyses of the relationships between these variables. Late submission of papers must be approved by the instructor.
Note: All data used in this class are the property of Professor Vartanian. Under no conditions can any student claim title to any of these data or publish without the written consent of Professor Vartanian.

Grades for the class will be determined on the basis of five scores: a mid-term exam, a final exam, a research paper, a series of computer projects, and class participation. The mid-term and final exams each count for 35% of your grade, the research paper counts for 20% of the grade, and the computer projects count for 10% of the grade. You are allowed one computer project to be up to one week late without any penalty. After this first late submission, all subsequent computer projects turned in late will be reduced by half the allowable points up to the first week late and by another point for each additional day late. Students must pass both exams and have at least a 70% average grade level to receive a passing grade. Students with an average grade between 70% and 73% will receive a grade of marginal satisfactory. While some students were given the opportunity to retake the first exam in the first semester of this class, retakes of exams will not be available in this second semester class.

The students and the instructor will decide when in the final exam period the final exam will be given. Those who are unable to take the final exam during the scheduled exam period determined in class should talk to the instructor about the possibility of rescheduling the exam for another time.

All computer assignments and other materials are located on my web site:

http://www.brynmawr.edu/Acads/GSSW/Vartanian/


Sage Series of Quantitative Analysis:

#13 Jae-On Kim and Charles W. Mueller, Introduction To Factor Analysis: What It Is And How To Do It

#46 Paul D. Allison, Event History Analysis

#69 George H. Dunteman, Principle Components Analysis

#92 William D. Berry, Understanding Regression Assumptions
Grading:

Mid-Term Exam: 35%
Final Exam: 35%
Research Paper 20%
Computer Assignments: 10%

Session

1-3  Review; Dummy Variables; Chow Test; Review of Log and other transformations.

Readings: Agresti and Finaly, Chapter 14, 543-561; Cody and Smith, 234-235. On Blackboard.

3  Two-Way Analysis of Variance, Analysis of Covariance.


4. Understanding Regression Assumptions.


5. Logistic Regression Analysis.


7. Mid-term Exam.


11. Event History and Hazard Rate Analysis.


12 + 13. Principal Components Analysis.

