UM Clinical Scholars Program
2005 Summer Data Analysis Assignments

Assignment #2: **Oral Presentation of Completed Data Analysis**

The *Introduction to Data Mgt & Analysis* sections in weeks 6-8 have three main components: 1) didactic sessions on biostatistics and data analysis, 2) structured reading and homework, and 3) hands-on labs in which you will analyze a dataset using Stata. This third task culminates in an oral presentation of the results of your results. Below, I detail the data analysis assignment for this section week. Biostatistics and data analysis cannot be learned in the classroom - you learn them by sitting at the keyboard, thinking through problems for yourself, and preferably, working on research questions of your own construction. There is no substitute for learning by doing, asking your own study questions, making your own errors, and developing experience and confidence in task-related problem solving.

Therefore, much of weeks 6-8 will be devoted to you conducting your own data analysis. You will be given the following materials:

1. National Health and Nutrition Examination Surveys (NHANES) 1999-2000 data.
   The NHANES survey is a nationally representative, multistage, stratified probability sample of the US population aged ≥2 months living in households. NHANES 1999-2000 includes over-sampling of low-income persons, adolescents 12-19 years, persons 60+ years of age, African Americans and Mexican Americans. The major objectives of NHANES are: to estimate the number and percent of persons in the U.S. population and designated subgroups with selected diseases and risk factors; to monitor trends in the prevalence, awareness, treatment and control of selected diseases; to monitor trends in risk behaviors and environmental exposures; to analyze risk factors for selected diseases; to study the relationship between diet, nutrition and health; to explore emerging public health issues and new technologies.

2. Information on the questions in the survey. See codebook.

3. Readings on using Stata. Much of the readings for week 8 are introductory. Browse these readings and then refer back to them once you need to use this information. You will also probably need to refer back to the manual, or use the "lookup" and "help" commands for electronic support. I strongly recommend making your own "crib" sheet that lists commands or code that you think that you will want to use again in the future.

Goals:

By the end of the week 8 you will have a general idea of:
1. The 4 steps in "A Basic Analysis Plan for Examining Bivariate and Multivariate Associations: Four Easy Steps to Greatness."
2. Understand the basics of dataset structure.
3. How to use Stata do-files and log-files to do basic data management and transformations.
4. How to use Stata do-files and log-files to do basic data analysis.
5. Know how to keep excellent documentation of your data analyses.

Materials Needed for this Assignment:

1. A computer equipped with Stata 8.0 or 9.0
2. A copy of "NHANES_transformed.dta"
3. Information on the variables in the dataset
4. On the course materials website a) "A Basic Analysis Plan for Examining Bivariate and Multivariate Associations", b) "Measures of Effect-Size & Associations" c) "What Statistical Test Should I Use?" d) "A Comparison of Linear and Logistic Regression"

Although NOT essential reading, your handouts for this assignment include two papers which ultize the NHANES data:

1. Muntner, P.; He, J.; Cutler, J. A.; Wildman, R. P., and Whelton, P. K. **Trends in blood pressure among children and adolescents**. JAMA. 2004 May 5; 291(17):2107-13.

2. Saydah, S. H.; Fradkin, J., and Cowie, C. C. **Poor control of risk factors for vascular disease among adults with previously diagnosed diabetes**. JAMA. 2004 Jan 21; 291(3):335-42.

Assignment:

Each scholar will conduct an independent analysis in which s/he will:

1) Ask a study question about bivariate and multivariate associations between selected independent (predictor) variables and (a) one continuous dependent variable and (b) one dichotomous dependent variable. The selection of the study question, which variables to include in the analysis and how the analyses should be done are left to your curiosity and discretion.

2) Using NHANES_transformed.dta, conduct the analyses that answer (to the greatest extent possible given the information in the dataset), the study questions that you have proposed. We will work through an example analysis plan together in class which is designed to answer the study question, "What are the bivariate and independent associations between several patient attributes (such as HbA1c and hypoglycemic meds) and 1) self-reported hyperglycemic symptoms, and 2) self-perceived general health status.

3) Give a 5-minute presentation of your findings at the end of the week.

Timeline:

August 9: Lottery for picking your 2 dependent variables (one continuous, one dichotomous). Realize that you can create a dichotomized variable from an ordinal or continuous variable (e.g. use the "a1c" variable to make a dichotomous variable in which 0= a value < 9 and 1= a value >=9). No two people can select the same dependent variables, therefore be sure to have a second, third and fourth choices.

By Aug 14: Have an initial plan of the independent variables that you want included in your analyses and why.

By Aug 19: Complete analyses on bivariate analyses.

By Aug 23: Complete regression analyses.

Sept Aug 24: Informal presentation of your results to the group.

See codebook handout for potential dependent variables.