17.871, Political Science Lab
Spring 2006
Take home assignment # 1: Using STATA
Handed out: Feb. 16, 2004
Due: Feb. 23, 2003, *at the beginning of class.*

1. Using a text editor such as EMACS, type the text from Exhibit 1 in the document "How to Use the *STATA* **infile** and **infix** Commands" into Athena and save it in a file named scores.dat on your home directory. Write a "do" file that will create a STATA data set from this raw data and save it as a file called "scores.dta". Turn in a "log" file that documents the STATA commands you issued to read in the data and save it.

2. The purpose of this question is for you to create two data sets and then merge them together using a common identifier.

   Find two tables that interest you in the *Statistical Abstract of the United States* that meet the following criteria: (1) they have between 25 and 52 observations and (2) they have the same units of analysis (e.g., states, years, nations), and (3) the subject matter of the two tables can conceivably be linked some way. (For example, you might find a data table showing the size of the federal deficit per year and then another data table showing the average unemployment rate per year. The link might come because some people think there are bigger deficits when unemployment rates rise.)

   A. Call these two tables Table A and Table B. Create separate STATA data sets for these two tables, naming the saved file tablea.dta and tableb.dta. Merge the two data sets. Save the merged data set, calling the saved file tablesa+b.dta.

   B. Turn in the following:
      i. The "do file" that shows how you create the data sets and merged them.
      ii. A printout of the data.
      iii. A short (one paragraph, 2 or 3 sentences) description of the tables you got your data from and a comment about how the data might be used for research.

3. The Election Assistance Commission conducted an "election day survey," to document a number of statistics concerning the conduct of the 2004 presidential election. The study may be found at the following url: http://www.eac.gov/election_survey_2004/toc.htm. If you follow the link to "State Data Tables" (http://www.eac.gov/election_survey_2004/state_data.htm), you will find individual spreadsheets for each state, with statistics about basic performance measures at the county (or sometimes town) level, such as the number of absentee ballots, the number of ballots, the number of registrants, etc.

   Here's the assignment: For the state I assign you, create (a) an "absentee ballot" data set that contains the number of counted absentee ballots in each county and (b) a "vap" set

that records the voting age population of each county. Save these as separate files. Write a "do file" that will merge the two files. Turn in the following:

i. The "do file" that shows how you did the merging
ii. A printout of the resulting merged data set.

Warning: there are blanks in some of the data fields. What do you do about this?

Daniel Barclay, California
Tabitha Bonilla, Florida
Shannon Greer, Illinois
Adam Groce, Louisiana
Laura Hajj, Minnesota
Jakob Hopping, Nevada
David Nedzel, North Carolina
Zachary Ozer, Pennsylvania
David Tobias, Tennessee