Subject: Convert DHS (SPSS?) missing value codes to Stata codes in Stata dataset
Posted by econ_stata on Fri, 24 Jul 2015 14:42:13 GMT
View Forum Message <> Reply to Message

Greetings from a new member! I am quite new to the DHS survey data, hence please excuse the possibly naive question. Apologies also for a very long post- but I wanted to be quite clear & precise!

For the Indian NFHS-3 survey, I am trying to study the relations between maternal education & child health outcomes. Given the choice of dataset formats available, I chose the Stata data files, as I am quite familiar with Stata. I am looking mainly at the Child data file, KR.dta, but have also needed to merge in the individual & household files. But I find now that the data has many variables (e.g., child/parent height or weight for age percentiles hw4-hw11) which have special missing value codes (e.g., 999, 9999 etc.)   which are not recognized as missing codes by Stata.

E.g., from the NFHS Recode Manual (Description of the Demographic and Health Surveys Individual Recode Data File, 2012) :

Quote:"Special codes are used throughout the data file for certain responses. The general coding scheme is presented below. The codes given apply to 4 digit, 3 digit, 2 digit and 1 digit variables, respectively. If there are other special responses to questions, these are coded in decreasing order from these special codes, i.e., 9996, 996, 96, 6; 9995, 995, 95, 5; etc." 'Coding Standards', p3, Recode Manual) [I am not very familiar with SPSS coding conventions, but from searching around, these appear to be SPSS missing codes.]

Of course, if I know that certain specific variables have missing values coded as a particular set of numbers, it is quite simple to recode them, using, e.g., the mvdecode function in data, or even write a do-file to lop over each wrong code for each variable, & replace them with Stata missing codes, whether standard "." or extended ".a, .b etc. E.g., for some of the anthropometric variables , it is mentioned on p51-52 of the Recode Manual that missing values are coded 9999,9998, 99999, 99998. These I have already replaced with Stata's missing codes, via a do-file.

But is there any way to convert for sure & in one shot ALL the "wrong" (for Stata) codes? So far, the means I have discovered are:
1. Download the SPSS files & from within SPSS, save as Stata datasets.
2. From Stata, convert using a user-written module "-usespss-" but this works only on a windows system
3. Download the ascii file & Import into Stata
4. Use stat/transfer.
[from http://infoguides.gmu.edu/c.php?g=120989&p=1268167]

None of these are very convenient for me. e.g., I am now working with a Macbook, don't have SPSS or Stata/Trnsfer.

Has any other member dealt with this issue, or can anyone suggest what would be the best way for me tackle this? In particular, is there somewhere a program (Stata ado file) to convert all the missing value codes in the Stata files provided by DHS? Some time ago, when working with the SHARE (Survey European Health & Retirement) datasetI, found such a program (sharetom.ado)

provided with it. I hope there is some such program available for DHS data users on Stata. It certainly seems strange that a file which has been distributed as a Stata dataset should have such wrong (for Stata) missing value codes.

Any help much appreciated!