
Subject: Re: Convert DHS (SPSS?) missing value codes to Stata codes in Stata dataset

Posted by [Bridgette-DHS](#) on Fri, 07 Aug 2015 13:23:21 GMT

[View Forum Message](#) <> [Reply to Message](#)

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

I can suggest three different ways to deal with these kinds of missing value codes. I use them all the time.

As an example, take hw70, the height-for-age z-score. The anthropometry z-scores have several special codes in the vicinity of 9999. Sometimes you will find values in that vicinity that do not even have a label, but all such values must be excluded.

One approach would be simply to have a line such as "replace hw70=. if hw70>9000". Values with "." are always considered by Stata to be missing and will be ignored from calculations. The problem with this is that you have now lost the original hw70. A second approach would be "gen hw70r=hw70" and "replace hw70r=. if hw70>9000". I add "r" for this kind of simple recode. Then any analysis would use hw70r in place of hw70, and you still have the original hw70. A third approach, when you have several related variables, could be something like the following. "gen hw7x_missing=0", "replace hw7x_missing=1 if hw70>9000 | hw71>9000 | hw72>9000". Then in your analysis, you could limit yourself to the cases that are non-missing on all variables by including "if hw7x_missing==0". I use this third approach if, say, I want to do a series of regression on exactly the same cases.

One more thing --in the DHS data files, the code "." Always means "not applicable". You should not confuse that meaning with what I have implied above, which is "please ignore in any calculations"!