Subject: Bangladesh 2011 DHS: Merging "children" and "male" files
Posted by skatumuluwa on Mon, 24 Mar 2014 19:09:21 GMT
View Forum Message <> Reply to Message

Hi,

I am working with the Bangladesh 2011 DHS dataset and am trying to merge the "children" and "male" files and have run into some difficulties.

Basically, I cannot figure out how to create a unique variable in both files so that I can merge based on that variable. Details follow.

Since the "male" file contains all the variables needed to re-create the women's CaseID (cluster + household + woman's line number) that is found in the "children" file, I went ahead and created a numeric variable in the "male" file combining these 3 variables. Since SPSS cannot merge files based on a string variable, I also converted the CaseID provided in the "children" file from a string to a numeric variable, which removed the spaces between the numbers in the CaseID. When I sorted by the woman's CaseID in both of my files and tried to merge the datasets, I got an error message saying that there were duplicate cases. When I went back to the data I realized that there are 2 ways to get each CaseID: example below

CaseID	V001(cluster)	V002 (household)	V003 (respondent's line number)
1882	1	88	2
1882	18	8	2

I also tried merging the data sets based on the 3 individual variables (V001, V002, and V003) but you can only sort the data by one variable. When I sorted by 1 variable and then tried to merge the data based on the 3 variables, Stata gave me an error message saying that the variables had not been sorted.

Any suggestions of how to get around this dilemma?

Many thanks!

Subject: Re: Bangladesh 2011 DHS: Merging "children" and "male" files
Posted by user-rhs on Mon, 24 Mar 2014 20:13:40 GMT
View Forum Message <> Reply to Message

Hi Skatumuluwa,
You mentioned both SPSS and Stata in your post. The instructions provided below works with Stata.

Sort and merge on v001 and v002. For the male dataset, create variables v001 and v002 from mv001 and mv002 so that both datasets now contain the variables v001 and v002. In Stata, you

can sort on multiple variables, e.g.
sort v001 v002

Sort both datasets and save. Now you can merge on v001 and v002. Stata will also ask you to specify if it is a one-to-many (1:m), one-to-one (1:1), many-to-one (m:1), or many-to-many (m:m) match, so be sure to do that in your merge command.

If you are using SPSS, are you using SPSS syntax? You may have greater flexibility in sorting etc. if you are syntax instead of point and click.

Finally: re: caseid: they are created from concatenating v001, v002, and v003, but they are "right-justified," meaning that if any of these variables contain fewer than 3 characters, the blank would go to the left of the number where it starts. For example, if mv001 = 543, mv002=24, and mv003=1, the mcaseid is 543.24..1. If mv001=51, mv002=115, and mv003=1, then mcaseid is .51115..1. Where I have used red-colored periods to denote what would be indicated as a blank space (" ") in the dataset.

hth,
RHS