

---

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 CaselD (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 CaselD provided in the "children" file from a string to a numeric variable, which removed the spaces between the numbers in the CaselD. When I sorted by the woman's CaselD 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 CaselD: example below

CaselD	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!