Subject: Merging data sets for DHS 2011 for Nepal
Posted by pjoshi on Mon, 06 Oct 2014 13:55:48 GMT
View Forum Message <> Reply to Message

I would like to look into the differences in women's status in Nepal based on whether or not their husbands have migrated outside Nepal for work. I have the information on the migrants in the household data set and the women's information (indices for health, education etc.) in the IR data set. I want to merge these and match each individual woman with whether or not the household has a member who migrated abroad for work. I carried out the following steps to do this.

First, I converted the household data set from wide to long so that each migrant member is an observation using the following syntax.
reshape long SH22_ SH24_ SH26_ SH27_ SH28C_ SH28D_, i(HHID) j(line)

I then created a new dummy variable to separate migrant households (households with a member who has been living outside Nepal and working) from non-migrant households using the following syntax.

gen var1=2
replace var1=1 if SH21==1 & SH27_==1 & SH28C_>=2

So, var1=1 for migrant households and var1=2 for non-migrant households

I now want to merge this household level data with individual data for women. So, I need to match each observation in IR data with var1 (whether the woman is from a migrant household or not). So, I sorted both the HR and IR data set by the identifying variables v001, v002 and v003. I then opened the HR dataset as the master data set and the used the following command to merge it to the individual dataset.

merge m:m v001 v002 v003 using "the file location for IR data set"

I did the many to many merge because for the same household there is multiple migrant members on the HR data set and multiple women in the IR data set. Does this make sense?

Thanks in advance for any suggestions you can make to help me out with this.