Subject: Assigning characteristics of head of household to the entire household
Posted by ssysahan on Mon, 27 Jan 2014 23:04:02 GMT
View Forum Message <> Reply to Message

I am trying to merge individual men and women files to the household files so that I can assign the ethnicity of the head of the household to each household (the ethnicity variable is only available in the individual files). I am using Ghana DHS4 (GHIR41DT, GHMR41DT, GHHR41DT). There are 4,843 women; 1,549 men; and 6,003 households.

To do this, I first merge men and women files to household files independently in Stata, following the instructions on the website:

    merge m:1 v001 v002 using [household level file]
    merge m:1 mv001 mv002 using [household level file]

I then append both files to create a dataset with individual men and individual women, each referenced to a household. However, there are still about 1,700 of the 6,003 households that are not matched to any individual respondents. Therefore I can't assign ethnicities to these households, which was the main goal of merging in the first place.

On top of this, I should be able to identify the head of the household because the respondent's line number should be equal to 1. But when I use the commands "tab v003" and "tab mv003", the number of respondents with line number equal to 1 does not equal the total number of households. Shouldn't this be the case if the merge occurred properly (i.e., for every household there should be a respondent with line number equal to 1 for the household head).

My questions are:
1) Am I correctly merging the files?
2) If my end goal is to assign the ethnicity of the household head to each household, is this the right way to go about it?

Subject: Re: Assigning characteristics of head of household to the entire household
Posted by Sarah B on Mon, 03 Feb 2014 00:47:04 GMT
View Forum Message <> Reply to Message

The good news is that your code looks fine, assuming that you've renamed the household variables hv001 and hv002 to v001 and v002 when you use
merge m:1 v001 v002 using [household level file]

And that you rename the same variables to mv001 and mv002 when you use
merge m:1 mv001 mv002 using [household level file].

The bad news is that you can't do what you're trying to do. As you note, ethnicity wasn't collected in the household questionnaire -- it was only collected for the men and women who were interviewed using the women's and men's questionnaires. Here's the problem: not every household head was an interviewed woman or man. The easiest way to see this is to look in the household (HR) file at the households in which no one was eligible for interview:

ta hv010 hv011

There are 1,635 households in which 0 women and 0 men were eligible for interview. I looked at this briefly and found that
1. In 110 households, the household head did not stay in the household last night and so was ineligible for interview (hv103_01==0)
2. In 786 households, the head of household was not in age range (15-59 for men, 15-49 for women)
3. In 739 households, the head of household was male and in age range but the house was not selected for the male survey (hv027==0)
These are the households that are not matched to any individual respondents.

Further, there are about 2,000 households in which at lease one person listed in the household roster was eligible for interview but the household head was ineligible for one of the 3 reasons listed above. In these households, there will (almost always) be household members listed in the MR and/or IR file, but none of them will be line number 1.

I hope that helps -- please re-post if you have further questions.