
Subject: Merged HH and Individual Data
Posted by [SR](#) on Thu, 01 Nov 2018 21:11:50 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hello

I am working with NFHS4 to look at women's variables and needed the household information for a full analysis. I have been able to successfully merge the individual and household member data (in Stata). After the merge, the total number of women is 6,99,686 which corresponds to the number of women interviewed in the NFHS4.

My question is:

1. I am assuming that multiple women from some households were interviewed. If so, how do i make sure i only have the one woman who responded to the household level questions in my analysis/ or one woman per household.

I do not want repetitions of household members- that is responses from women belonging to the same household, because i want to look at household level indicators.

My guess is listwise deletion during analysis will solve this problem to an extent, but is there any suggestion on how to choose only one respondent per HH?

Any help will be appreciated. thanks!

Subject: Re: Merged HH and Individual Data
Posted by [Bridgette-DHS](#) on Fri, 02 Nov 2018 14:40:49 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

I'm not sure what your question is. In order to do a merge of the IR and PR files for this survey, you must match hv024, hv001, hv002, hvidx in the PR file with v024, v001, v002, and v003, respectively, in the IR file. When there are multiple women respondents in the same household, this will give you all of them. Are you asking how to randomly select just one of these women, in households that had more than one woman? Why would you want to do that? If you do that, you will lose representativeness of the population of women (although you can restore representativeness by multiplying v005 by the number of women in the household) and you will reduce the effective sample size.

If you want to look at how the women's characteristics are related to household characteristics, you do not need to do any such subsampling. You would keep all the women, and just interpret the household's characteristics as characteristics of the woman. For some purposes, you could think of doing multi-level analysis and treating the household as the level 2 unit (and the cluster as level 3), but there is little gain from doing this because most households contain only one woman.

Subject: Re: Merged HH and Individual Data
Posted by [SR](#) on Fri, 02 Nov 2018 18:30:08 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thanks for your response and clarification. Yes, i was asking if i could select one woman when there were more than one woman in the household. The India report for NFHS4 says 601,509 households and 699,686 women were interviewed. Some households might not have any women, most as you suggest would have only one woman, and some would have more than one. In the latter case, with a merge, the household variables would be replicated for each woman in that household and i wanted to see if there was a way to take care of this replication. Thanks again.

Subject: Re: Merged HH and Individual Data
Posted by [Bridgette-DHS](#) on Mon, 05 Nov 2018 12:27:47 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

You could do something like the following. Before you do the merge, assign the women in the IR file a random number; then sort the women in the household by this number; then keep the one who is first in the sort. Then do the merge with the household file. in the following, I use the HR file rather than the PR file, and just keep the household-level variables.

```
set more off
set maxvar 10000
use ...IAHR74FL.DTA" , clear
drop * _*
sort hv024 hv001 hv002
save ...IAHRtemp.dta, replace

use ...IAIR74FL.DTA" , clear
gen rn=uniform()
sort v024 v001 v002 rn
egen sequence=seq(), by(v024 v001 v002)
tab sequence
keep if sequence==1
gen hv024=v024
gen hv001=v001
gen hv002=v002
sort hv024 hv001 hv002
merge hv024 hv001 hv002 using e:\DHS\DHS_data\scratch\IAHRtemp.dta
keep if _merge==3
drop sequence _merge
```

Subject: Re: Merged HH and Individual Data
Posted by [SR](#) on Tue, 13 Nov 2018 02:28:41 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you Dr Pullum. That is very helpful.
regards

Subject: Re: Merged HH and Individual Data
Posted by [sujata](#) on Thu, 18 Jul 2019 11:19:42 GMT
[View Forum Message](#) <> [Reply to Message](#)

hello,
I am working on NFHS4 dataset and i am having trouble in matching the values with report. I have emerged PR, IR and MR files since some values are not in MR file. But when i am matching the values for variables like Religion, caste, marital status and schooling with the report provided, they are coming to be different. I have merged the files using hv024, hv001, hv002 and hvidx variables from PR file, v024, v001, v002 and v003 from IR file and same variables from MR file.
Any help will be highly appreciated.
