
Subject: Merging Men, Women, Child dataset into Household level unit analysis

Posted by [goomthatha](#) on Mon, 22 Feb 2016 12:55:09 GMT

[View Forum Message](#) <> [Reply to Message](#)

Hi to All,

I am using NFHS India 3 dataset. I would like to merge men, women, and children level data (IR, MR, KR) as a household unit. The individual dataset have caseID as opposed to HHID. When i merge these data, i want to group them as households. Its like adding women, men and children questionnaire in household dataset. I used below sample merge syntax for SPSS, to merge IR and HR dataset. but i am not getting the desired objective. Is there any better way to achieve this objective? I want to integrate variables from IR, MR, KR in household level. Is it possible?

```
GET FILE='C:\DATAUSER\ZMAR51FL.SAV'.  
SORT CASES BY  
HIVCLUST (A) HIVNUMB (A) HIVLINE (A).  
SAVE OUTFILE='C:\DATAUSER\HIV.sav'  
/COMPRESSED.
```

```
GET FILE='C:\DATAUSER\ZMIR51FL.SAV'.  
SORT CASES BY  
V001 (A) V002 (A) V003 (A) .  
SAVE OUTFILE='C:\DATAUSER\WOMEN.sav'  
/RENAME(V001 V002 V003=  
HIVCLUST HIVNUMB HIVLINE)  
/COMPRESSED.
```

```
GET FILE='C:\DATAUSER\WOMEN.sav'.  
MATCH FILES /FILE=*  
/TABLE='C:\DATAUSER\HIV.sav'  
/BY HIVCLUST HIVNUMB HIVLINE.  
EXECUTE.  
SAVE OUTFILE='C:\DATAUSER\ZMAR_IR.SAV'  
/COMPRESSED.
```

Subject: Re: Merging Men, Women, Child dataset into Household level unit analysis

Posted by [Bridgette-DHS](#) on Tue, 23 Feb 2016 21:57:40 GMT

[View Forum Message](#) <> [Reply to Message](#)

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

I do not use SPSS. Perhaps you do not use Stata at all, but I will list below the lines to do this in Stata and you may be able to figure out the logic.

First you combine the IR, MR, and KR files by appending, NOT merging. This is very important. Then you merge the combined IR_MR_KR file with the PR file. If you first merge the IR file with the PR file, and then merge the MR file, and then merge the KR file, you will have a mess.

* prepare the IR, MR, and KR files

```
use IRfile.dta, clear
rename v001 hv001
rename v002 hv002
rename v003 hvidx
save IRtemp.dta, replace
```

```
use MRfile.dta, clear
rename mv* v*
rename v001 hv001
rename v002 hv002
rename v003 hvidx
save MRtemp.dta, replace
```

```
use KRfile.dta, clear
rename v001 hv001
rename v002 hv002
rename b16 hvidx
save KRtemp.dta, replace
```

* append the IR, MR, and KR files

```
use IRtemp.dta, clear
append using MRtemp.dta
append using KRtemp.dta
sort hv001 hv002 hvidx
save IR_MR_KRtemp.dta, replace
```

* prepare the PR file and merge with the IR_MR_KR file

```
use PRfile.dta, clear
sort hv001 hv002 hvidx
save PRtemp.dta, replace
merge hv001 hv002 hvidx using IR_MR_KRtemp.dta
```

Subject: Re: Merging Men, Women, Child dataset into Household level unit analysis
Posted by [lgoyenechec5](#) on Wed, 07 Sep 2016 17:01:11 GMT

[View Forum Message](#) <> [Reply to Message](#)

1. I notice that in some forums you merge all the datasets and for example, in this one you append the IM, MR and KR datasets, and then merge the append file with the PR dataset. I want to understand when I have to append the data sets instead of merge.

2. I used the code that first append and then merge, but as a result I have 18,202 observations not matched from master.

Subject: Re: Merging Men, Women, Child dataset into Household level unit analysis
Posted by [Bridgette-DHS](#) on Wed, 07 Sep 2016 23:17:24 GMT

[View Forum Message](#) <> [Reply to Message](#)

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

For most merges and appends It helps to think of a "case" as a unique individual and a "record" as a line of data. When you merge two files, you are consolidating or combining two records that refer to the same case. That's why you have to identify the cases in both files with id codes such as hv001, hv002, hvidx in the PR file and v001, v002, v003 in the IR file. This is done when you want to attach information about a woman to the information about her household by merging the IR and PR files. Or maybe the case in the merged file will be a couple, a man and a woman, so you merge the IR and MR files to make a CR file, using the stated line numbers of the partners.

You append one file to another if the cases are similar but different (that phrase could be made more precise!). For example, you may have a 2010 survey and a 2015 survey from the same country. The cases are completely different but you can simplify some of the data processing if you append or combine into a single file (keeping an identifier for which survey is which). You would never append an IR file to a PR file, for example.

When I think about manipulating two files, I usually have a physical image in my mind, of two stacks of paper. Do I want to put one stack of paper on top of the other one (append), or do I want to transfer the information in one stack to the other stack, sheet by sheet (merge). Most computer procedures are just a faster way of doing what could be done manually (if we had a LOT of time!).

Sometimes it can be efficient to combine appending and merging, in succession. Let me know if you want to be more specific.
