Subject: STATA codes to merge women and household datasets
Posted by Mrinal on Fri, 26 Jan 2018 16:53:03 GMT
View Forum Message <> Reply to Message

I am working with NFHS datasets and need to merge women and household datasets using STATA, which I hardly use. May I please have the Stata codes to merge aforementioned datasets of all four rounds of NFHS. I will certainly appreciate it.

Thanks and regards,
Mrinal

Subject: Re: STATA codes to merge women and household datasets
Posted by Bridgette-DHS on Tue, 30 Jan 2018 01:05:16 GMT
View Forum Message <> Reply to Message

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

If you want to combine, say, the household (PR) files from the successive surveys, you should use the "append" command. This is distinct from a "merge", in which, say, the KR records and PR records from a single survey could be combined child by child.

Some variables will have different codes and categories in different surveys. For example, v023 may not be defined the same way in every survey. That must be taken into account. When you use the append command, the variable names and labels from the last survey in the append command will over-ride any previous names and labels.

The main reason for appending files is for convenience of file manipulation. In the case of the India surveys, all the files are very large and after appending they will be enormous--and slow to work with. I would recommend trimming the files and just carrying along the variables you need for analysis.

Issues related to appending have been discussed in other forum postings.

Subject: Re: STATA codes to merge women and household datasets
Posted by Mrinal on Tue, 30 Jan 2018 08:28:30 GMT
View Forum Message <> Reply to Message

Thank you, Bridgette and Pullum. Actually, I was more interested in merging codes for stata package than appending. However, I managed to construct the merging code for NFHS-2 and is given below.

Merging household on women dataset
 Round 2
 use "D:\Desktop\dhs\data\nfhs\2\IAHR42FL.DTA", clear

```
gen int v001 = hv001
gen int v002 = hv002
gen byte v003 = hv003
sort v001 v002 v003
save "D:\Dropbox\stata\nfhs\IAHR42FL_sort.DTA", replace

use "D:\Desktop\dhs\data\nfhs\2\IAIR42FL.DTA", clear
sort v001 v002 v003
merge v001 v002 v003 using "D:\Dropbox\stata\nfhs\IAHR42FL_sort.DTA"

save "D:\Desktop\dhs\data\nfhs\2\IA_HR_IR_42FL.DTA", replace
```

Thanks and regards,
Mrinal

Subject: Re: STATA codes to merge women and household datasets
Posted by boyle014 on Thu, 01 Feb 2018 16:14:20 GMT
View Forum Message <> Reply to Message

Mrnal,

You should consider using IPUMS-DHS, which has already harmonized all of the variables across
the surveys. You select the samples you want--all of the Indian ones it sounds like--and then the
variables you want. You don't have to download separate files and merge or append them. You
can download a single file with multiple surveys. The latest Indian sample is being uploaded into
the system now. It will be available in March. The other three Indian samples are already there.

Liz Boyle

Subject: Re: STATA codes to merge women and household datasets
Posted by Mrinal on Mon, 05 Feb 2018 06:07:32 GMT
View Forum Message <> Reply to Message

Dear Elizabeth,

First of all, I would like to thank and congratulate you and your team for the much needed
PUMS-DHS initiative. It is such a relief to have a resource like this when working with huge
databases such as NFHS. Interestingly, the first thing I did after getting access to DHS was to
browse the IPUMS-DHS project, however, I realized that I should take it as an opportunity to learn
a thing or two while preparing the database for analysis. But again I heartily appreciate your
contribution in smoothening the path to analyze DHS database.

Mrinal

Subject: Re: STATA codes to merge women and household datasets
Posted by boyle014 on Mon, 05 Feb 2018 18:08:59 GMT
View Forum Message <> Reply to Message

Thanks, Mrinal!

Subject: Re: STATA codes to merge women and household datasets
Posted by Gowokani on Wed, 18 Apr 2018 22:52:28 GMT
View Forum Message <> Reply to Message

This IPUMS-DHS , is just unavailable. See the message I am getting (attached)

Gowo


```
File Attachments
1) Capture.PNG, downloaded 772 times
```

Subject: Re: STATA codes to merge women and household datasets
Posted by boyle014 on Fri, 20 Apr 2018 18:00:21 GMT
View Forum Message <> Reply to Message

Dear Gowo,

Sorry to hear that you got a Temporarily Unavailable page. We uploaded lots of new data this
week. When we do this, the IT people sometimes have to take the website offline for a minute or
two to fix bugs. It's working again now!

Liz Boyle

Subject: Re: STATA codes to merge women and household datasets
Posted by priyoma on Tue, 23 Oct 2018 21:18:22 GMT
View Forum Message <> Reply to Message

Hi!

I've a similar question. I wish to merge PR file with IR file for NFHS 2(India). I had a look at
IPUMS. It is indeed very helpful. However when I try to select variables from two different types of

---

files (IR and PR), the webpage says my selection (data cart) will not be retained.

Any way around this?

I primarily need all of IR variables and education variables of PR (ie education information of each and every member of the household)

Best,
Priyoma

Subject: Re: STATA codes to merge women and household datasets
Posted by boyle014 on Thu, 01 Nov 2018 21:20:37 GMT
View Forum Message <> Reply to Message

Hi priyoma,

Thanks for the query. You've uncovered a temporary weakness with the system when using household members as the unit of analysis--the Household Number (HHID) variable is not available for selection.

To put all the household members' education on each woman's record in IPUMS DHS, you would first create a woman's data file (extract) with all the other variables you need. Then you would go back into the system, select household members as the unit of analysis and create a second extract with the additional variables. You would then merge the files on the HHID variable.

We are in the process of fixing this now. HHID and a few other technical variables will become available for household members extracts next week. We will be adding them when we release a new set of samples from Afghanistan, Angola, Burundi, Lesotho, Myanmar, Namibia, and Senegal. I will post again when that process is complete.

Subject: Re: STATA codes to merge women and household datasets
Posted by Isabelle on Sun, 11 Aug 2019 12:48:46 GMT
View Forum Message <> Reply to Message

Hello Mrinal and the community,

I am working with the NFHS-2 and I am trying to merge the datasets "household" and "individual" exactly like you described earlier. I used the same code, however, I am not able to merge the datasets as
"variables v001 v002 v003 do not uniquely identify observations in the master data".

I added the commands I used. I cannot identify what I am doing wrong, so I am thankful for any advice.

_____

use "C:\Users\Isabelle\Desktop\DHS Data India\1998-99\Household\IAHR42FL.dta" // household dataset
gen int v001 = hv001
gen int v002 = hv002
gen byte v003 = hv003
sort v001 v002 v003
save "C:\Users\Isabelle\Desktop\DHS Data India\Single Datasets Recode\1998-99Recode\1998-99HouseholdSort.dta", replace

use "C:\Users\Isabelle\Desktop\DHS Data India\1998-99\Individual\IAIR42FL" //women dataset
sort v001 v002 v003
merge 1:m v001 v002 v003 using "C:\Users\Isabelle\Desktop\DHS Data India\Single Datasets Recode\1998-99Recode\1998-99HouseholdSort.dta"

save "C:\Users\Isabelle\Desktop\DHS Data India\Single Datasets Recode\1998-99Recode\1998-99RecMerge.dta"
_____

Kind regards,
Isabelle

Subject: Re: STATA codes to merge women and household datasets
Posted by Bridgette-DHS on Fri, 23 Aug 2019 17:40:10 GMT
View Forum Message <> Reply to Message

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

For all of the India surveys (and ONLY the India surveys), the household ID and case ID include state (generically "region"), given by hv024 in the PR file and v024 in the IR file. Just add that into the merge, along with v001, v002, v003 (in any order) and you should be ok. Let us know if you still have a problem.

Subject: Re: STATA codes to merge women and household datasets
Posted by Isabelle on Sat, 24 Aug 2019 13:56:26 GMT
View Forum Message <> Reply to Message

Thank you Bridgette for your response and help.

However, I have a follow up question after merging the datasets including forth identifier v024/hv024.

When merging the datasets IR and PR using

household member:
gen int v001 = hv001
gen int v002 = hv002
gen byte v003 = hv003
gen int v024 = hv024
sort v001 v002 v003 v024

individual:
sort v001 v002 v003 v024
merge 1:m v001 v002 v003 v024 using "C:\Users\Isabelle\Desktop\DHS Data India\Single
Datasets Recode\1998-99Recode\V21998-99HHMSort.dta"

I get the following result:
not matched: 337,481
from master 49,203
from using: 288,278

matched: 229,101

The household dataset has 517,379 observations
The individual dataset has 90,582 observation

Does this mean that from my master (individual) file, only (90,582-49,203) 41,379 individuals are
not matched to a household; meaning I cannot use them in my further analysis?

Thank you in advance for your help!
Best
Isabelle

Subject: Re: STATA codes to merge women and household datasets
Posted by Bridgette-DHS on Tue, 03 Sep 2019 19:37:55 GMT
View Forum Message <> Reply to Message

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

Hi Isabelle--Here is how I would do the merge. I use the old version of the merge command but
you would get the same thing if you used 1:1. I also introduce a variable called "in_IR", which is
coded 1 for every case in the IR file. It just clarifies the "_merge" code, which is described in your
results with different terms.

ALL of the 90,303 cases in the IR file are also in the PR file. 427,076 people in the PR file are
NOT also in the IR file. Hope this makes sense. Tom

```
cd e:\DHS\DHS_data\scratch

use "C:\Users\26216\ICF\Analysis - Shared Resources\Data\DHSdata\IAIR42FL.DTA" , clear
gen hv024=v024
gen hv001=v001
gen hv002=v002
gen hvidx=v003
gen in_IR=1
sort hv024 hv001 hv002 hvidx
save IAIR42_temp.dta, replace

use "C:\Users\26216\ICF\Analysis - Shared Resources\Data\DHSdata\IAPR42FL.DTA" , clear
sort hv024 hv001 hv002 hvidx
merge hv024 hv001 hv002 hvidx using IAIR42_temp.dta
replace in_IR=0 if in_IR==.
tab _merge
tab _merge in_IR
```

The following table is produced:

Subject: Re: STATA codes to merge women and household datasets
Posted by vpatil on Sun, 26 Apr 2020 04:54:07 GMT
View Forum Message <> Reply to Message

Hi

I am trying to merge NFHS4 household and Individual datasets and I have tried all the options discussed in the previous thread on this topic but no success. This is what I get:

merge 1:m hv001 hv002 hv003 hv024 using "/Users/drvaishalipatil/Desktop/DHS datasets/RO1 India/RO12015sort.dta"

```
    Result                # of obs.
    -----------------------------------------
    not matched            1,280,793
       from master           591,308  (_merge==1)
       from using            689,485  (_merge==2)

    matched                   10,201  (_merge==3)
```

The final numbers are wrong. I have tried 1:1, 1:1, m:1, and all different possibilities. Please tell me what am I doing wrong.