Subject: Issue merging NFHS 2 women and household dataset
Posted by Mrinal on Tue, 30 Jan 2018 08:45:33 GMT
View Forum Message <> Reply to Message

I am trying to merge the women and the household dataset of NFHS 2. However, after merging the results are not making any sense!! NFHS 2 women dataset has 90,303 observations and household dataset contains 92,486 observations but after merging the merged file shows 104,979 observation, I am trying to understand the result, is it correct or the merging has some issue?? I am also getting an error saying
Quote:variables v001 v002 do not uniquely identify observations in the master data
variables v001 v002 do not uniquely identify observations in
D:\Dropbox\stata\nfhs\IAHR42FL_sort.DTA

Complete result of merging and codes used to merge is given below for reference.

"_merge" result

tab _merge

```
    _merge |      Freq.     Percent        Cum.
------------+-----------------------------------
         2 |      5,722        5.45        5.45
         3 |     99,257       94.55      100.00
------------+-----------------------------------
     Total |    104,979      100.00
```

Merging codes

```
**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
 sort v001 v002
 save "D:\Dropbox\stata\nfhs\IAHR42FL_sort.DTA", replace

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

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

Thanks
Mrinal

Subject: Re: Issue merging NFHS 2 women and household dataset
Posted by Bridgette-DHS on Tue, 30 Jan 2018 13:42:37 GMT
View Forum Message <> Reply to Message

Another response from Tom Pullum:

You should not use the HR file for this merge. The HR file has one record for each household. Instead, use the PR file, which has one record for every person in the household. The following lines will do the merge. In the output file, the cases with _merge=3 are the women in the IR file. The cases with _merge=1 are the other people in the PR file, that is, the household members other than the women in the IR file.

```
set more off
* Prepare IR file for merge
use e:\DHS\DHS_data\IR_files\IAIR42FL.dta, clear
gen hv001=v001
gen hv002=v002
gen hvidx=v003
sort hv001 hv002 hvidx
save e:\DHS\DHS_data\scratch\IAIRtemp.dta, replace


* Prepare PR file for merge
use e:\DHS\DHS_data\PR_files\IAPR42FL.dta, clear
sort hv001 hv002 hvidx

* Merge IR with PR
merge hv001 hv002 hvidx using  e:\DHS\DHS_data\scratch\IAIRtemp.dta
tab _merge
```

Subject: Re: Issue merging NFHS 2 women and household dataset
Posted by Mrinal on Wed, 31 Jan 2018 07:38:32 GMT
View Forum Message <> Reply to Message

Thanks, both of you again!!

But I am really puzzled by the results.

IR file has 90,303 observations
PR file has 517,379 observations

However, in the merged file its showing 143,944 observations for _merge=3 which as you aforementioned are women in IR file.

Whats going on?

One more thing, can I use IR file as the base and merge only the matching observations form PR file?

Regards,
Mrinal

Subject: Re: Issue merging NFHS 2 women and household dataset
Posted by Bridgette-DHS on Thu, 01 Feb 2018 12:13:31 GMT
View Forum Message <> Reply to Message

Another response from Tom Pullum:

The lines I gave yesterday will work for almost every DHS survey. However, I see that for this survey of India the id code includes another number. I saw this by comparing hhid in the PR file with caseid in the IR file. If you check you will find that hhid is a 12-character string and caseid is a 15-character string consisting of hhid in columns 1-12 and v003 in columns 13-15. I answered a related question on the forum yesterday regarding the Mali 2006 survey. In the case of Mali 2006 and some other surveys in West Africa, there is a sub-household code embedded in hhid and caseid. For this India survey, the extra code is the state, not the sub-household, but the strategy is basically the same--that is, to match caseid in the IR file with hhid hvidx in the PR file.

As an alternative, since state is given by v024, you could match v024 v001 v002 v003 in the IR file with hv024 hv001 hv002 hvidx in the PR file.

```
set more off
* Prepare IR file for merge
use e:\DHS\DHS_data\IR_files\IAIR42FL.dta, clear
gen hhid=substr(caseid,1,12)
gen hvidx=v003
sort hhid hvidx
save e:\DHS\DHS_data\scratch\IAIRtemp.dta, replace


* Prepare PR file for merge
use e:\DHS\DHS_data\PR_files\IAPR42FL.dta, clear
sort hhid hvidx

* Merge IR with PR
merge hhid hvidx using  e:\DHS\DHS_data\scratch\IAIRtemp.dta
tab _merge
```

Subject: Re: Issue merging NFHS 2 women and household dataset
Posted by priyoma on Tue, 23 Oct 2018 21:26:01 GMT
View Forum Message <> Reply to Message

Hello!

I followed this particular advice for India (given the discrepancy of state and HHID 12 character long noted earlier on this thread).

Can someone please let me know if this has worked correctly for them?

Mrinal did you notice any errors after using this second stata code?

Best,
Priyoma

Subject: Re: Issue merging NFHS 2 women and household dataset
Posted by Mrinal on Wed, 24 Oct 2018 05:07:32 GMT
View Forum Message <> Reply to Message

Hi Priyoma,

Though, I went by the second advice (To merge IR & PR), I did so in R and everything went fine.

Hence, I suppose it will not give any error in STATA as well.

All the best!

Mrinal