Subject: Merging IR, KR and HR files
Posted by Aarthi1410 on Mon, 05 Aug 2024 16:46:10 GMT
View Forum Message <> Reply to Message

I am trying to merge datasets for DHS India 2015-16 and 2019-21. I need variables from all three files since there are mother's education levels employment status and ownership of land/house included as controls which I am planning to include in my analysis. I use the following codes to merge the three for both rounds. For 2015-16, the code works fine, however for 2019-21, when I merge the IR and KR files _merge == 3 is only 333 and when I finally merge the HR file the total observation for mother's education and employment is approximately 2000. Is there some other way to merge in a correct manner?

```
use "${cleaned}\cleaned_household_2015",clear

gen cluster=hv001
gen hh=hv002
sort cluster hh
save HH_2015temp.dta, replace

use "${cleaned}\cleaned_women_2015", clear
gen cluster=v001
gen hh=v002
gen mo_line=v003
sort cluster hh mo_line
save women2015_temp.dta, replace

use "${cleaned}\cleaned_children_2015", clear
gen cluster=v001
gen hh=v002
gen mo_line=v003

sort cluster hh mo_line
quietly merge cluster hh mo_line using women2015_temp.dta
tab _merge
keep if _merge==3
drop _merge

* Merge the child and mother with the household data
sort cluster hh
quietly merge cluster hh using HH_2015temp.dta
tab _merge
keep if _merge==3
drop _merge
```

Subject: Re: Merging IR, KR and HR files
Posted by Bridgette-DHS on Tue, 06 Aug 2024 20:00:10 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS staff member, Tom Pullum:

You would only do this kind of merge if you need IR and HR variables that are not already in the KR file. That seems to be your situation. The Stata program below will do this merge. Let us know if you have questions about it.

```
* Merge needed variables from the HR and IR files onto the KR file
* This illustrative example just adds one HR variable and some v75* variables
*  from the IR file.  You would add other variables.

* Illustrated with NFHS-5

* Specify workspace
cd e:\DHS\DHS_data\scratch

* Prepare HR file
use "...IAHR7EFL.DTA", clear
rename hv001 cluster
rename hv002 hh

* keep the merging variables and other needed HR variables
keep cluster hh hv201
save HRtemp.dta, replace

* Prepare IR file
use "...IAIR7EFL.DTA", clear
rename v001 cluster
rename v002 hh
rename v003 line
*keep the merging variables and other needed IR variables
keep cluster hh line v75*
save IRtemp.dta, replace

* Prepare KR file
use "...IAKR7EFL.DTA", clear
rename v001 cluster
rename v002 hh
rename v003 line

* keep just the merging variables
keep cluster hh line
save KRtemp.dta, replace

* Do the merge, starting with KR, then adding IR and HR
```

```
use KRtemp.dta, clear

* m:1 because many children to one woman
merge m:1 cluster hh line using IRtemp.dta
tab _merge

* _merge=2 for women in the IR file who do not have a child in the KR file
keep if _merge==3
drop _merge

* m:1 because many children to one household
merge m:1 cluster hh using HRtemp.dta
tab _merge

* _merge=2 for households in the HR file that do not have a child in the KR file
keep if _merge==3
drop _merge
```