Subject: Re: Confirming the correctiness of mereging two datasets
Posted by Bridgette-DHS on Mon, 05 Dec 2022 18:13:36 GMT
View Forum Message <> Reply to Message

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

Your merge is ok. The WI file has one record for every household in the survey, and there were 7095 households that had no children under 5. Below I will give the Stata code for this merge, because it shows how to unpack whhid in the WI file. I use an older version of the merge command, which I prefer because it does not require spedifying 1:m, etc.

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

* Prepare the WI file
use "C:\Users\26216\ICF\Analysis - Shared Resources\Data\DHSdata\ETWI41FL.DTA"
describe hhid

* whhid is str12

forvalues li=1/12 {
gen col`li'=substr(whhid,`li',1)
}

list col* if _n<=20, table clean
tab1 col*

* It appears that hv001 is cols 7-9 and hv002 is cols 10-12
gen hv001=substr(whhid,7,3)
gen hv002=substr(whhid,10,3)

destring(hv001), generate(cluster)
destring(hv002), generate(hh)

sort cluster hh
save ETWItemp.dta, replace

* Prepare the KR file
use "C:\Users\26216\ICF\Analysis - Shared Resources\Data\DHSdata\ETKR41FL.DTA"
summarize v001 v002
* v001 and v002 have 1-3 columns
gen cluster=v001
gen hh=v002

list cluster hh if _n<=20, table clean
sort cluster hh

```
* Do the merge
merge cluster hh using ETWItemp.dta
tab _merge

* _merge=2 for 7095 cases; these are households that have no children under 5; drop them

drop if _merge==2
drop _merge
```