Subject: Re: Merging DHS data in Stata
Posted by Bridgette-DHS on Mon, 30 Jun 2014 17:20:12 GMT
View Forum Message <> Reply to Message

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

You are right to be concerned about whether the merge was correct. It is very easy to make a mistake when merging. The way I would do it is given below. You can change the paths and run it.

During a merge, a variable called "_merge" is constructed. The codes for it are given in "help merge". In your case, you will want to keep the cases with _merge=3. In the lines below I construct "in_PR=1" and "in_IR=1" in the PR and IR files, respectively, and if you run the code you will see that the cases in the merged file with _merge=3 are exactly the same as those with in_PR=1 AND in_IR=1. I sometimes add this extra check to be absolutely sure that I have what I want.

You have to save the sorted PR file. You should save it with another name. Never over-write the basic recode files. I use old syntax for the merge command. I know some other people do too. The current syntax doesn't always mean what you think it means. I dropped the lines to recast v001 and v002. They are not needed. I used rename instead of gen to get v001 and v002 in the PR file. You can do things like that in temporary or scratch files.

Let us know if you have other questions.

```
use c:\DHS\DHS_data\PR_files\PKPR61FL.dta, clear
rename hv001 v001
rename hv002 v002
gen in_PR=1
sort v001 v002

save c:\DHS\DHS_data\scratch\temp.dta, replace

use c:\DHS\DHS_data\IR_files\PKIR61FL.dta, clear
gen in_IR=1
sort v001 v002
merge v001 v002 using c:\DHS\DHS_data\scratch\temp.dta

tab1 _merge in_PR in_IR

keep if in_PR==1 & in_IR==1

tab _merge
drop _merge
```