
Subject: merging HIV results and couples data
Posted by [lberes](#) on Mon, 11 Apr 2016 05:08:20 GMT
[View Forum Message](#) <> [Reply to Message](#)

A 2014 forum post gives extremely helpful advice on how to merge the HIV data and couples data in STATA. The basic code is below. My question is on step 8. I am wondering why drop if `_merge==2` allows us to 'keep only women'. I thought that any combination of v001, v002 and v003 that came from either the couples or the HIV data could only be a woman. How does dropping that combination if it only came from the HIV dataset - and was not in the couples data set - exclude men? In other words, what does drop if `_merge==2` do for us in this code? Thank you!

code from post in Nov 2014 by Trevor-DHS

* Step 1: open AR file

use "xxAR61FL.DTA", clear

* Step 2: rename identifying variables

renvars hivclust hivnumb hivline / v001 v002 v003

* Step 3: sort by identifying variables

sort v001 v002 v003

* Step 4: save results

save "xxAR61FL_mergeprep.DTA", replace

* Step 5: open IR file

use "xxCR61FL.DTA", clear

* Step 6: sort by identifying variables

sort v001 v002 v003

* Step 7: merge!

merge v001 v002 v003 using "xxAR61FL_mergeprep.DTA"

* Step 8: Keep only women

drop if `_merge==2`

Then rename the added hiv variables to something unique for women, e.g.

rename hiv* w_hiv*

and repeat steps 1-8 above using mv003 instead of v003 throughout to merge the men's hiv test result and then finally rename the hiv variables to be for men, e.g.

rename hiv* m_hiv*

Subject: Re: merging HIV results and couples data

Posted by [user-rhs](#) on Tue, 12 Apr 2016 17:00:35 GMT

[View Forum Message](#) <> [Reply to Message](#)

That's because the IR file only contains information on women. If the women's dataset is the one in memory ("master," the dataset that you have opened), as it is in Trevor's code, and the HIV dataset is one you're merging it with ("using," the dataset that comes after "using" in the -merge-command), then, based on the coding for the resultant _merge variable (which, by the way, merge results are shown immediately after a merge):

numeric code	equivalent word (results)	description
1	master	observation appeared in master only
2	using	observation appeared in using only
3	match	observation appeared in both
4	match_update	observation appeared in both, missing values updated
5	match_conflict	observation appeared in both, conflicting nonmissing values

Source: -help merge-

You can see that _merge==2 means that the unique identifier existed in the using/HIV data only, and not the women's/master dataset.

Stata keeps all observations, regardless of matching status, unless you specify to keep just the ones in the master dataset or in the using dataset, i.e., supposing you have IR dataset in memory:

```
merge 1:1 v001 v002 v003 using "HIV.dta", assert(match master) /*Keeps only those who are in the IR dataset, regardless of matching status*/  
merge 1:1 v001 v002 v003 using "HIV.dta",assert(match using) /*Keeps only those in the HIV dataset, regardless of matching status and therefore sex of the respondent*/
```

hth,
rhs
