Subject: Re: Merging issue

Posted by Bridgette-DHS on Wed, 17 Jun 2015 18:22:16 GMT

View Forum Message <> Reply to Message

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

As I understand it, you want to merge the AR data with the IR data. Here are Stata lines to do this:

use c:\DHS\DHS_data\AR_files\ETar61FL.dta, clear rename hivclust v001 rename hivnumb v002 rename hivline v003 sort v001 v002 v003 save c:\DHS\DHS_data\scratch\temp.dta, replace use c:\DHS\DHS_data\IR_files\ETIR61FL.dta, clear sort v001 v002 v003 merge v001 v002 v003 using c:\DHS\DHS_data\scratch\temp.dta tab _merge keep if _merge==3 drop _merge

You will need to change the paths, of course. I am using the old version of the merge command, but the version you used would work equally well. Your formula for an id code did not produce unique values. Look at the following results for the IR file:

gen long id=((1000+v001)*10000)+(v002*100)+v003

- . gen n=1
- . collapse (sum) n, by(id)
- . tab n

There are 601 id codes that appear 2 times, 28 that appear 3 times, and 1 that appears 4 times. It is safer to use a hierarchical sort command, such as "sort v001 v002 v003". Also easier. To get a truly unique id you could use this: "egen id=group(v001 v002 v003)". Your formula with powers of 10 will not always work.

File Attachments

1) tab.jpg, downloaded 1070 times