Subject: HH head education level merge
Posted by kate2 on Sun, 02 Nov 2014 22:45:59 GMT

Dear all,

I am working on the Tanzania 2011/12 MIS dataset. I am working in Stata 13.

I currently have a dataset that I have created by merging the observations for all children under five in the household member recode, onto the child recode. This gives me all the malaria test results in the same place as the information on treatment seeking, prevention etc for children under five.
I would also like to add a variable for the head of households educational attainment to this dataset. This information is available in the HH member recode but not the child recode (as far as I can see).

I attempted the following:

use "HH member recode"
keep if hv101 = 1
gen HH_ed = hv106
keep v002 HH_ed
save "HH_member_recode_HHedonly"
clear
use "child merged dataset"
sort v002
merge 1:m v002 using"HH_member_recode_HHedonly"

This gives me the error message:

variable v002 does not uniquely identify observations in the using data

 I am not expecting v002 to uniquely identify observations. I know there will often be several children in each household so I want the HH head ed level variable to be merged and replicated for each child with the same HH number as this HH head. So I assume I do want 1:m; but perhaps not? Could you point me in the right direction here. I am clearly doing something wrong.

Thanks a lot
Kate

Subject: Re: HH head education level merge
Posted by Reduced-For(u)m on Mon, 03 Nov 2014 01:28:18 GMT

I always mess this up to, but I think you need "m:1"

From Stata, with my adaptations in []: http://www.stata.com/manuals13/dmerge.pdf

This is an m:1 merge because [householdID] can correspond to many observations in the master dataset [multiple kids from same household, since you've loaded in child file], but uniquely identifies individual observations in the using dataset ["using HH_member_recode_HHedonly"]

So that should work.

Subject: Re: HH head education level merge
Posted by kate2 on Mon, 03 Nov 2014 08:09:55 GMT

Thanks for the link and the suggestions. Using m:1 gives me the error about not uniquely IDing observations in the 'using' data; and using 1:m gives me the error about not uniquely IDing observations in the master data....

So clearly I should be able to uniquely ID observation in the dataset that I want to merge onto my master data set other wise its not useful. I thought that I had then worked out the problem - that I need to use cluster as well as household id to uniquely identify observations in my HH head education dataset - because HH IDs are not sequential through the whole dataset, rather they (I think) restart for each new cluster. Bingo...thought I had it. So I ran these commands below...but then got the same error result...

use "HH member recode"
keep if hv101 = 1
gen HH_ed = hv106
keep v001 v002 HH_ed
sort v001 v002
save "HH_member_recode_HHedonly"
clear
use "child merged dataset"
sort v001 v002
merge m:1 v001 v002 using"HH_member_recode_HHedonly"

>>> variables v001 v002 do not uniquely identify observations in the using data

So I am clearly missing something as to how I uniquely identify households. Unless some households have more than one HH head recorded -- will look into that possibility.

Any other ideas? Thank for your help!
Kate

Subject: Re: HH head education level merge
Posted by Reduced-For(u)m on Tue, 04 Nov 2014 17:19:00 GMT

"keep if hv101 == 1" seems to do the trick.

Subject: Re: HH head education level merge
Posted by Trevor-DHS on Sat, 08 Nov 2014 02:28:17 GMT

One other thing to try is to check how you created v001. If you just used gen, then sometimes the format of the variable is wrong. Try using
gen long v001 = hv001
See post #3011