Subject: Merging women's file with Household File
Posted by Kwesi.Nkrumah on Wed, 19 Feb 2014 19:38:33 GMT

View Forum Message <> Reply to Message

Dear Users,

I am trying to merge the women's file with the household file. The women's file is my base. I used the following code after reading the merge guidelines from the Measure DHS website. see the code below:

use "/Users/Kwesi/Downloads/GHHR41DT/GHHR41FL.DTA"
keep hhid hv000 hv001 hv002 hv003 hv003 hv213 hv216 hv217 hv221 hv220 hv219

ren hv001 v001
ren hv002 v002

gen unique=v001*1000
replace unique=unique+v002
sort unique

save "/Users/Kwesi/Downloads/GHWI41FL-house.DTA", replace

clear

use "/Users/Kwesi/Documents/Document General/Main Documents/Latest Buckup 29 JAn/Academic Documents/WIP and Project papers/Institutional Projects/AERC/Maternal Health Paper/Data/Ghana-1998.DTA"

keep caseid v001 v002 v003 v005 v012 v024 v025 v101 v106 v107 v113 v115 ///
v116 v130 v131 v133 v136 v137 v138 v149 v150 v151 v152 bord_01 ///
v302 m13_1 m14_1 m15_1 h2_1 h3_1 h4_1 h5_1 h6_1 h7_1 h8_1 h9_1 ///
h0_1 v701 v702 v715 v729 v730

gen unique=v001*1000
replace unique=unique+v002
sort unique

merge 1:m unique using "/Users/Kwesi/Downloads/GHWI41FL-house.DTA"

I got a feedback from Stata that unique does not uniquely identify the variables in the master. I decided to drop the constructed unique variables and its place use v001 and v002, I got the same feedback. Please can you help.

Regards

Subject: Re: Merging women's file with Household File
Posted by Trevor-DHS on Wed, 19 Feb 2014 20:52:30 GMT
View Forum Message <> Reply to Message

You don't need to create your variable unique. You can just sort the household data on v001 and v002:
sort v001 v002

Then for the women's data, you need to sort by these variables plus the woman's line number (v003), e.g.
sort v001 v002 v003

Next (and this is where your problem lies), merge the data using the following:
merge m:1 v001 v002 using "...", keep(match master)

In your version you had 1:m, but you need m:1 as you need to merge multiple women's variable with one household, not one woman with multiple households.

The keep option will keep cases that matched as well as those that were in the master file (the women's data), but did not have matching household data (there shouldn't be any of these, but it is better to do this way to ensure that you are not dropping any records from the women's data by accident).

I hope this will solve the problem for you.