Subject: Re: Strange Issues w/ Data Formatting from DHS
Posted by Trevor-DHS on Wed, 01 May 2024 17:01:21 GMT
View Forum Message <> Reply to Message

Hi
A few notes:
1) It looks like you are opening the IR file, then keeping just a few variables and sorting the file, and then overwriting the original file. This is generally not considered good practice as you are modifying the original file. Generally, you should start with your original file, but save to an interim file with a different name or in a different folder (or both).
2) The naming of your THESISDATA folder seems to vary - in two cases it has a space between THESIS and DATA and in one case it doesn't (this may be a display issue in the user forum as it occasionally puts extra blanks into the text).
3) In terms of the order of merging, I would start by merging the wealth index to the PR file and saving your output to an intermediate file. They both should have hhid so you should be able to merge those without problem. Then merge the info from the PR/wealth data onto the IR file.
Below is a rough outline of the process (I haven't tested this, so there may be some bugs - this is just to give you the order of operations):
use TZWIxxxx.dta
sort hhid
save TZWIxxxx.dta, replace

use TZPRxxxx.dta, clear
* keep the variables you want from the PR file
keep hhid hv0* ...
sort hhid
merge m:1 hhid using TZWIxxxx.dta
clonevar v001 = hv001
clonevar v002 = hv002
clonevar v003 = hvidx
sort v001 v002 v003
save TZPRxxxx_temp.dta

use TZIRxxxx.dta, clear
sort v001 v002 v003
merge 1:1 v001 v002 v003 using TZPRxxxx_temp.dta
It is also possible to construct hhid from hv001 and hv002 or from v001 and v002, and vie versa, but I don't think you need to.