Subject: Merging IR and KR with PR - no unique identifiers (for particular countries only)
Posted by nibiti on Wed, 23 May 2018 11:21:05 GMT
View Forum Message <> Reply to Message

Hi guys,

I've posted this in another thread but I guess it makes more sense to have it as a stand-alone topic.

I am working with Sub-Saharan African countries and have already merged PR with the GE files (for GPS coordinates). Now, I am merging these combined PR/GE files with MR, IR and KR datasets. This works fine except for a few countries/waves for IR and KR. Following the guiudelines from other forums, I am using as identifiers hv001 hv002 and hvidx (PR), v001, v002 and v003 (IR) and v001, v002 and b16 (KR). I have renamed the variables in IR and KR to match the PR variable names.

As said, I am using Sub-Saharan African countries in waves 3 to 7. I have excluded KR from wave 3 because the b16 identifier does not exist there. Also, I am also using the Stata commands "drop if b16==0" and "drop if b16==." and have also removed duplicates in the Using dataset with the command "duplicates drop v001 v002 v003, force" (for IR) and "duplicates drop v001 v002 b16, force" (for KR).

There are several countries for which the three variables are still not unqiue identifiers when trying to merge. Here's a list of those countries and waves:

Wave 3:
IR: ML, NI, TG --> all not uniquely identified in the Using data

Wave 4:
IR: ML, SN --> all not uniquely identified in the Using data

Wave 6:
KR: BJ, LS, ML --> all not uniquely identified in the Using data

Separately for each wave, I am using a loop over all countries. For all other countries it works fine, which means that the code is not the problem and all the merge commands are correct. It should be something about the raw data, is what I suspect. Maybe someone has an idea?

Many thanks in advance
Best
Timo

Subject: Re: Merging IR and KR with PR - no unique identifiers (for particular

countries only)
Posted by Bridgette-DHS on Wed, 23 May 2018 11:53:17 GMT
View Forum Message <> Reply to Message

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

Regarding what's in what file--you may have found some exceptions to the general pattern. Also there have been some changes over time; what's standard now has not always been standard.

Yes, there are some surveys with files that are difficult to merge. If you don't have b16, then you can't merge the KR and PR files with complete reliability, but apart from that some merges are difficult.

Try the following example of a PR / KR merge, which uses caseid and hhid. (You will want to include more variables in the "keep" lines and change the paths.) It should work on some of those difficult merges but probably not all. Let me know which ones remain.

I want DHS to prepare a library of merge programs for these difficult cases. The basic problem is that in some surveys there is a sub-household id and it's not well documented.

```
use "C:\Users\2626I\ICF\Analysis - Shared Resources\Data\DHSdata\PEKR6IFL.DTA" , clear
gen hhid=substr(caseid,1,12)
rename b16 hvidx
gen in_KR=1
keep hhid hvidx in*
sort hhid hvidx
save e:\DHS\DHS_data\scratch\PEKR6Itemp.dta, replace

use "C:\Users\2626I\ICF\Analysis - Shared Resources\Data\DHSdata\PEPR6IFL.DTA" , clear
gen in_PR=1
keep hhid hvidx in*
sort hhid hvidx
merge hhid hvidx using e:\DHS\DHS_data\scratch\PEKR6Itemp.dta
tab _merge
```

Subject: Re: Merging IR and KR with PR - no unique identifiers (for particular countries only)
Posted by nibiti on Wed, 23 May 2018 13:25:25 GMT
View Forum Message <> Reply to Message

Thanks for this and the explanations. Does this also work with the IR datasets? I guess, I would have to replace b16 by v003, but what do I take instead of hhid?

So far I applied the code for the countries in wave 6, because these had problems with the KR file.

The code resolved the issues for LS, but did not solve it for BJ and ML unfortunately. Any other

ideas?

Many thanks and best
Timo