Subject: Namibia 2013 - age inconsistencies after merging household and individual datasets
Posted by saf204 on Sat, 25 Feb 2017 21:32:39 GMT
View Forum Message <> Reply to Message

Hello all,

I have recently merged the household (NMHR61FL.dta) and individual (NMIR61FL.dta) datasets using Stata 14.2, following the DHS protocol. However, there are some inconsistencies in the data. We know that blood glucose samples were only taken from individuals older than 35. However, according to the v012 respondent's age variable carried over from the NMIR61FL individual dataset, nearly half of the people with blood glucose measurements have ages <35. This makes me think that something has gone wrong with the merge. Any idea what is going on here? Has anyone else had this problem, or successfully (and consistently) merged the two?

Subject: Re: Namibia 2013 - age inconsistencies after merging household and individual datasets
Posted by Bridgette-DHS on Mon, 27 Feb 2017 16:02:26 GMT
View Forum Message <> Reply to Message

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

You must include the line number. The following lines will work (change the paths):

```
set more off
use e:\DHS\DHS_data\IR_files\NMIR61FL.dta, clear
rename v001 hv001
rename v002 hv002
rename v003 hvidx
sort hv001 hv002 hvidx
save e:\DHS\scratch\temp.dta, replace

use e:\DHS\DHS_data\PR_files\NMPR61FL.dta, clear
sort hv001 hv002 hvidx
merge hv001 hv002 hvidx using e:\DHS\scratch\temp.dta

tab _merge
```

_merge will be 3 for the women in the IR file; _merge will be 1 for the cases in the PR file who are not in the IR file. I never use the version of merge that has 1:1 or 1:m or m:1 or m:m. The older version will do everything you could possibly want to do and is more forgiving.

Subject: Re: Namibia 2013 - age inconsistencies after merging household and individual datasets

Posted by saf204 on Mon, 27 Feb 2017 22:27:46 GMT
View Forum Message <> Reply to Message

Hi all,

Thank you for your reply! That does seem to help. However, I still have roughly 10% of my sample with blood glucose measures with an age <35 (from the merged individual dataset age variable). But according to the household age variable, all of the sample is 35 or older.

I am wondering if I am interpreting the variables incorrectly. In the IR dataset, v012 is "respondent's current age." In the household dataset, sh303 is "age." I understand that v012 was calculated from the respondent's reported day of birth. Do you have any idea why there may be a discrepancy in the age variables between the two datasets?

Thank you for your help on this!

Subject: Re: Namibia 2013 - age inconsistencies after merging household and individual datasets
Posted by Bridgette-DHS on Tue, 28 Feb 2017 17:52:09 GMT
View Forum Message <> Reply to Message

Another response from Tom Pullum:

In the IR file, v012 is age in completed years. (v009 is calendar month of birth and v010 is calendar year of birth; we do not obtain day of birth).

In the PR file, hv105 is age in completed years.

hv105 and v012 do not necessarily agree but if they differ we give priority to v012 because it is always obtained from the woman herself.

I see that sh303, which is always in the range 35 to 64, sometimes differs from both hv105 and v012. However, I can only find 3 cases in which there is a blood sugar measurement (sh336k<.) and hv012 is less than 35. It is 34 for those 3 cases. hv105 is always in the range 35-64 when sh303 is in the range 35-64, even though they do not always agree exactly.

Subject: Re: Namibia 2013 - age inconsistencies after merging household and individual datasets
Posted by saf204 on Tue, 28 Feb 2017 23:30:59 GMT
View Forum Message <> Reply to Message

This is perfect - it works great! Thank you so much. I really appreciate it!