Subject: Link the DHS Individuals (IR) with DHS HIV (AR)
Posted by DHS user on Tue, 02 Sep 2014 12:39:13 GMT
View Forum Message <> Reply to Message

I am a Demography and Population Studies student at the University of Witwatersrand, South Africa. As part of our curriculum we are required to conduct some research using the DHS data. Our current topic is on age at first sex and HIV infection among women in Swaziland. We would like to know of the possible ways in which we can link the DHS Individuals Data with the DHS HIV data since we would like to check the profiles of these people. We know on the Individuals Data there is a variable serves as an indicator of whether a blood sample was taken and also on the HIV Data there is a unique barcode identifier for each person tested. We will be using STATA as our analysis tool.

Subject: Re: Link the DHS Individuals (IR) with DHS HIV (AR)
Posted by Bridgette-DHS on Tue, 02 Sep 2014 12:40:58 GMT
View Forum Message <> Reply to Message

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

Inserted below are the lines to do this. You will have to change the paths.

use c:\DHS\DHS_data\AR_files\SZar51fl.dta, clear
ren hivclust v001
ren hivnumb v002
ren hivline v003

sort v001 v002 v003

save c:\DHS\DHS_data\scratch\temp.dta, replace

use c:\DHS\DHS_data\IR_files\SZIR51fl.dta, clear
sort v001 v002 v003
merge v001 v002 v003 using c:\DHS\DHS_data\scratch\temp.dta
tab _merge

keep if _merge==3

* hiv03 is the result of the test
* all analysis of the hiv data should use hiv05 for weights, not v005

Subject: Re: Link the DHS Individuals (IR) with DHS HIV (AR)
Posted by bwbennett09 on Tue, 17 Mar 2015 13:57:42 GMT
View Forum Message <> Reply to Message

I realize the above answer was a few years ago, but I do have a question regarding the code. Why only keep if _merge==3? Also, is this specific to one country or all countries?

I am following a similar procedure for 7 different nations looking at HIV status and measures of women's empowerment. Thanks!

Subject: Re: Link the DHS Individuals (IR) with DHS HIV (AR)
Posted by Bridgette-DHS on Wed, 18 Mar 2015 11:56:06 GMT
View Forum Message <> Reply to Message

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

The answer you refer to is about six months old, not several years.... Like many Stata users, I have stayed with old syntax for the merge command, rather than the one that was introduced with Stata 11 or 12. Both the old and the new versions produce a diagnostic variable called "_merge". The variable has three values of most interest, 1, 2, and 3. "1" means the case was only in the "master" file, i.e. the first one in the sequence. "2" means the case was only in the "using" file, i.e. the second one in the sequence. "3" means the case was in both files.

Usually, when I do a merge, I want just the cases with _merge==3, but that is definitely not always the case. Sometimes I also want the cases with 1 or 2 (almost never 1, 2, and 3, but even that could happen). You have to think about what makes sense. Say, for example, that you wanted to merge the children under 5 in the household survey (the PR file) with the children in the KR file. Say you started with the KR file and then merged the PR data (again for children under 5) with the KR file. The PR and KR files would be the master and using files, respectively. The PR file includes all children in the household, including children whose mother is not a resident of the household. Those children will get _merge=1. The KR file includes children who do not live in the same household as the mother, and they will get _merge=2. If the mother and child both live in the household, the child will get _merge=3. They are the only ones for whom you will get both PR and KR variables, so you probably would want to keep just them--but that's not necessarily the case.