
Subject: How to merge DHS databases of different countries

Posted by [AmsP](#) on Wed, 17 May 2017 21:31:27 GMT

[View Forum Message](#) <> [Reply to Message](#)

Hello,

I expect to create a single database that combines databases of various countries. I have already de-normalised different survey rounds of a particular country and created single database for each country. Then what should I take care if I append various countries' databases together? My purpose is to find out if living in different countries have any impact on v133 of a particular ethnic group that live in all these countries.

Because different countries have different weighting systems, I think the "svyset" command may not work for such a combined database.

Thank you very much!

Subject: Re: How to merge DHS databases of different countries

Posted by [Bridgette-DHS](#) on Thu, 18 May 2017 14:55:44 GMT

[View Forum Message](#) <> [Reply to Message](#)

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

There have been several posts on this topic. You need to construct new variables that distinguish ALL the strata in all the surveys, and ALL the clusters in all the surveys. You can assign a numeric code (1, 2, ..., for example) called "surveyid" to each survey. For each survey, you can use v021 as the cluster id. The stratification variable is not completely consistent across surveys, so you can give a name such as "stratumid" to v023 or whatever is the correct identifier for each survey. You then use lines such as "egen allclusterid=group(survey v021)" and "egen allstratumid=group(survey stratumid)". Even if the strata are the same in two successive surveys in the same country, say, you need to distinguish them this way. You can leave the weight variable alone or you can re-weight as described in earlier postings. Then you put "allclusterid" and "allstratumid" and the weight variable into svyset for the pooled file.

Subject: Re: How to merge DHS databases of different countries

Posted by [AmsP](#) on Thu, 18 May 2017 21:30:57 GMT

[View Forum Message](#) <> [Reply to Message](#)

Thank you very much! So shall I do the manipulation on the original cluster and stratum variables (e.g. v021 and v022) in each survey or on the recoded cluster and stratum (e.g. v021+2000 and v022+2000)? Because STATA cannot distinguish cluster and stratum across different surveys, Dr. Ruilin Ren suggested, in previous posts, the method to add 1000 or 2000 to cluster and stratum in different survey rounds.

Thank you again!

Subject: Re: How to merge DHS databases of different countries
Posted by [Bridgette-DHS](#) on Fri, 19 May 2017 16:05:56 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

If you are using Stata, then the "egen group" command will probably be easier and safer than adding 1000, or 2000, etc. If you are not using Stata, then adding those numbers may be easier. You do not need to do any manipulation before appending the files, so long as you have a unique identifier for each survey, as I suggested. So long as you have a "survey" variable or other unique identifier, you can do all your survey-specific edits, etc., AFTER the appending. Note: do not rely on v000 or hv000 as survey identifiers. There are many cases of two different surveys, conducted close together in time, having the same values of v000 or hv000.

Subject: Re: How to merge DHS databases of different countries
Posted by [AmsP](#) on Fri, 19 May 2017 17:24:16 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you very much! Just one more issue. Shall I still de-normalise the weights (v005) before appending surveys of different countries? Just like the de-normalisation before appending different surveys of the same countries.

Thank you again.

Subject: Re: How to merge DHS databases of different countries
Posted by [Bridgette-DHS](#) on Fri, 19 May 2017 17:55:35 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

Yes, whatever re-normalization you would do for multiple surveys from the same country, you would do the same thing for surveys from different countries. There's not just one way to do this, as discussed in several earlier postings.

Subject: Re: How to merge DHS databases of different countries
Posted by [AmsP](#) on Fri, 19 May 2017 18:30:55 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you very much again!

Subject: Re: How to merge DHS databases of different countries
Posted by [AmsP](#) on Mon, 22 May 2017 12:23:35 GMT
[View Forum Message](#) <> [Reply to Message](#)

Sorry for a perhaps naive question. But after appending surveys of different countries and running `allclusterid=group(survey cluster)` and `allstratum=group(survey stratum)`, in the following `-svyset-` I think I should use `psu(allclusterid)` and `strata(allstratumid)`, instead of `psu(v021)` and `strata(v022)`. Is this right?

Thank you!

Subject: Re: How to merge DHS databases of different countries
Posted by [Bridgette-DHS](#) on Mon, 22 May 2017 12:58:37 GMT
[View Forum Message](#) <> [Reply to Message](#)

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

Yes, that's right. The `svyset` statement for the combined surveys should use the variables produced by the "egen group" commands.
