

---

Subject: Re: Merging data files in Stata

Posted by [Bridgette-DHS](#) on Thu, 04 Apr 2013 18:28:54 GMT

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

---

Here is a response from one of our DHS Stata experts Tom Pullum, that should answer your question.

In Stata, this would be done with an append command, rather than merge. I suggest that you "use" (open) the 2008 file, then "append using" the 2003 file, "append using" the 1998 file, and then save with a different file name.

I have encountered situations in which starting with the earlier file and then appending later files will not work; it's safer to start with a later file and then append an earlier file. I think Stata likes to start with the file that is larger, in terms of the number of variables, and with DHS that would usually mean starting with a later file.

You are correct that it is necessary to re-number the clusters. There are two ways to do this. One is to add some larger number to the codes. For example, you could have the original id numbers in the first survey, but in the second survey, add 1000 to the id numbers and in the third survey add 2000 to the id numbers. An alternative would be to use the "egen group" command. For example, if you had a line "egen v001r=group(v000 v001)", it would completely renumber the clusters, consecutively, from 1 to the total number of clusters in the three surveys. This is elegant but will make it just a little harder to figure out what was the original number of the cluster if you ever needed to do that.

The strata may be the same across surveys, and in that case you would want to just make sure they have the same id numbers. For example, Metro Manila should have the same number in all three surveys.

The weights should be ok. Sometimes surveys from several countries are pooled, and then the weights may need to be changed by a different multiplier for each survey.

We will often combine multiple surveys into a single file, but you have to be careful when you do this. For example, I would advise against treating them as a single survey and calculating the mean of some variable in all three surveys. Looking at differences or changes between surveys is fine.

Bridgette-DHS

---