Subject: Merging datasets from multiple countries
Posted by bwbennett09 on Tue, 17 Mar 2015 14:32:11 GMT
View Forum Message <> Reply to Message

I am working on an ecological study across 7 countries regarding the relationship of measures of Women's Empowerment to HIV status. What would the best procedure be for merging the country level datasets in stata? I have already merged the HIV datasets to the Individual Recode datasets for each individual country.

Once I merge these country level datasets into one large dataset for analysis, what would be the best way to weight the dataset? Thanks!

Subject: Re: Merging datasets from multiple countries
Posted by Reduced-For(u)m on Wed, 18 Mar 2015 01:15:46 GMT
View Forum Message <> Reply to Message

1) You don't want to merge them, you want to append them. It is easy: append using "filename"

2) Yes. You have to consider weights carefully and de-normalize them before using them. There is a lot of discussion on the boards about that - look in the "weighting data" threads.

Subject: Re: Merging datasets from multiple countries
Posted by Bridgette-DHS on Wed, 18 Mar 2015 12:25:03 GMT
View Forum Message <> Reply to Message

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

You can combine several countries/surveys in the way you describe with "append". It's an easy but powerful command. You can get the syntax with "help append". Here are a few suggestions. First, be sure to have some variable (I call it "survey") to identify the different surveys. You cannot always rely on hv000 or v000 to do this. Second, understand that some variables, such as region, are country-specific and the codes will mean different things in different surveys. Third, the variable and value labels for the final survey in the append will be the only ones that are saved. Fourth, do not keep more variable than you need because the file can get very large. Fifth, if you use svyset, you need to re-define the cluster and stratum variables, for example with "egen cluster=group(survey v001)", and probably re-normalize the weights. I prefer to weight each survey equally, for example by forcing the total weight in each survey to be one billion (hv005 or v005 is constructed to have mean value of 1 million). The steps to do this sort of thing are described elsewhere on the forum.