Subject: Re: Question re-weighting combined survey data
Posted by Reduced-For(u)m on Thu, 12 Sep 2013 00:27:19 GMT
View Forum Message <> Reply to Message

bsayer,

This was a great post that summarizes really well a lot what has been talked about on these weighting threads. In particular, I think the whole "modify the stratum variable to be survey specific" is something we should put in a sticky thread at the top of the weighting thread, along with the code for svyset (and its SAS/SPSS equivalent code) and a list of which weights to use for which dataset/recode.

Now just between us - since we've gone a little back-forth on this in the past - let me give one instance in which I think you might want to re-scale the weights in some way.

Suppose you want to know the correlation between sanitation access and child mortality in all of sub-Saharan Africa. So you pool together the last DHS from each of a bunch of countries into one dataset, and regress mortality on sanitation access. Now, you could do this separately by each country, but maybe you just want to know the average impact across the region.

If you use the original DHS weights, which, within country sum to the sample size, are you not implicitly weighting the pooled regression by the sample sizes in each country (so each country's total regression weight comes out to (Nsurvey/Nallcountries)- the fraction of the total observations that came from that country)? Wouldn't a better weighting scheme be to have each country's weights sum up to that countries population? Then, when STATA re-normalizes all the weights (summed for all observations in all countries) to 1, people in small countries will have (by design) less weight than people from large countries. And isn't that what we'd want? Wouldn't we want the larger countries to get more weight in this case - assuming we are looking for some "population average" in our regression coefficient?

I know we've been over this a bit before, and I have conceded that I might be missing something, but I still think that in some cases you would want to re-scale the DHS weights so that they could simultaneously act as population weights. Or, if you want each country to have equal weight, then re-normalize so that each country's weights sum to 1/Ncountries.

If anyone thinks it is worth it, I'll try to do some regressions of this sort weighted in various ways and post the results here, but I can't do it at the moment and wanted to respond before I forgot to.