Subject: A little question about pooling data
Posted by belladaisy2018 on Fri, 06 Apr 2018 02:44:44 GMT
View Forum Message <> Reply to Message

Hello!

I have seen several studies (Maheu-Giroux et al, 2015; Helleringer et al, 2014) pool congruent DHS data from different countries. I wanted to know if anyone has more information regarding how to do this exactly. Specifically, do you measure the weighted rates of the outcome for each country and then backtrack that rate/percentage to the actual study population (so if study pop is 18000 and weighted perc is 18% you would take 18% of the study population) and then use those numbers for each country's study?

I'm just a little confused how they can pool different countries and take into account the weighting as well as adjustment for key factors, all at the same time. I looked into the methods sections for these two studies and about 4-5 others that pool DHS data from different countries and the methods to pool data were not clear.

Thanks for any help in advance!!

Subject: Re: A little question about pooling data
Posted by Reduced-For(u)m on Fri, 06 Apr 2018 22:47:49 GMT
View Forum Message <> Reply to Message

There are several methods discussed in detail on this forum, but the answer is basically "it depends on what you want to do".

Several options include:

1. Not doing any weighting at all, and then sacrificing the idea that your estmate corresponds to any particular "population"

2. Using the DHS weights given, in which case total influence of one survey is equal to its sample size (as above) but within each country the estimate is population-representative...also not great.

3. Re-normalizing the weights within each country (essentially dividing each weightby the sum of weights within that country, so that all the weights for each country/survey add up to 1) giving each survey both a nationally representative estimate and the overall estiamte the interpretation of weighting each survey/country as equally important (so unit of observation is sort of like the country).

4. Taking (3) but then multiplying those weights by some reference population (say, number of households in a country) so that the sum of weights in one survey adds up to the number of households in that country - this is probably the closest you can get to a "population

representative" weighting, if the population you want is the people in all of the countries in your sample.

Choice of these optinos depends on what you want to do. Most published work prior to very recently that merges multiple DHS rounds probably did not pay enough attention to the problem, so I can't promise following the previous literature is the way to go (I don't know those papers you mentioned). I tend to lean towards (1) or (3) these days (partly because 4 means Nigeria is basically everything if you do Africa...or India is everything if you do all DHS countries). But there is no clearly 100% right way...at least until you state what you are trying to estimate in terms of the population you want your numbers to represent.

This is all about getting population-level parameters right. For causal effects estimation, there are a whole set of other arguments that apply, but they all basically relate to 1-4 in terms of what can be done, they just differ on why you might prefer one to the other (for instance, if you a prior believe the causal effect is constant across everyone, you actually don't need to weight at all).

Hope this helps.

Subject: Re: A little question about pooling data
Posted by Reduced-For(u)m on Sun, 15 Apr 2018 20:14:40 GMT
View Forum Message <> Reply to Message

There are a number of ways to do that. One would be what you describe - take a weighted average of the (weighted) country specific averages. Your second wieghting (the average of the country-level estimates) could be weighted in any way you want. One difficulty with that is getting good confidence-interval estimates... which is why people sometimes pool the data together to do the estimation in one step (and because that simple method won't work for more complicated estimates).

The second option is to simply append all the data together and re-scale the DHS weights in such a way that you effectively get back out the weighting that you do want. One way would be to normalize each individual survey/country to have their weights sum to 1* and then apply those weights in a regression context (or using the svy: prefix in front of a Stata command). This would be effectively weighting each country equally.

Or, you could take those "sum-to-1" weights from the previous step, and multiply them by some population of interest to get a "representative population weighted average", which is more like the thing you describe (the thing in my first paragraph). The difficulty here is getting the appropriate population size for the appropriate population... that has to come from outside the DHS.

If you need more details on these, many of these problems have discussed on the boards under the name of "de-normalizing" weights... if you don't find what you need there, feel free to ask for more specifics about one of the methods.

*To get within-country weights to sum up to 1, you just get the total sum of weights for each survey round, and divide the DHS-given weights with the sum-of-weights (the Stata "egen, by"

command is good for this).