Subject: Re: Setting up pooled DHS data as panel
Posted by Reduced-For(u)m on Thu, 03 Jul 2014 01:56:09 GMT
View Forum Message <> Reply to Message

Hi Dan,

  I wish I could be super helpful, but probably only a little helpful.

  A good place to start is with the document I'm attaching, which comes from this thread (the "weighting" thread in general has a lot of discussion on this topic):

  http://userforum.dhsprogram.com/index.php?t=tree&goto=82
&S=e3a92c3e8a765f0217181d74f8127581#msg_82

  The major problem is that the DHS weighting design is intended for national (and sometimes sub-national) representativeness, but not for combining survey rounds - either within country over time, or across countries. The sample weights sum to the total sample size, and thus the meaning of some weight within some survey is lost when it is compared to an observation in another survey. The DHS way of handling that is to "de-normalize" the weights as described above. Then, in theory, the re-computed weights should work across survey rounds, and have implicit weighting for population size as well.

  Unfortunately, if you read that document above, the "de-normalizing" process is complicated and requires you to bring in outside data that may or may not be good/useful. I've proposed other methods, such as forcing weights within a survey to sum to 1, and then multiplying them by some population size estimate to overlay population weights on the (DHS given) probability weights. I haven't done the math to see if this reduces to the formula provided by DHS, but I hope to get around to it at some point (I'm not a statistician/econometrician, so I'm worried I'm missing something with this method, even if the algebra seems reasonable).

  You also have to re-define strata and cluster variables when using multiple survey rounds, so that cluster "101" in one survey is differentiated from cluster "101" in some other survey (and same with strata).

  I think the "do all the surveys separately" thing might be easiest for you, since you are effectively looking at one survey as one observation - so all the weighting and p-value/standard-error problems reduce to the regular single-round DHS method. Then again, if the sample sizes are fairly constant within-country, you should be able to just use the DHS-provided weights and get a very similar answer. You could also ignore weighting altogether, cluster your standard errors using ", cluster(clustervar)", skip the "svy" part, and just estimate an effect within the sample population that is not scaled to be population representative (only a problem if there are heterogeneous effects of X on Y across the population - otherwise, in a lot ways, one observation is as good as any another one).

Its up to you. That's most of what I know. I'm happy to follow up on anything, but don't know if I'll have much to offer.