Subject: Re: When/How to use (de-normalized) weights for pooled data analysis (four waves of BD child anthro data
Posted by Reduced-For(u)m on Mon, 17 Mar 2014 21:00:00 GMT
View Forum Message <> Reply to Message

Hi Anna,

  Not a whole lot I can offer, but as for your options 1/2 on weighting: what I suggest is a kind of de-normalizing, just one where you only want to preserve relative within-survey probability of selection. The other version you mentioned earlier tries to re-adjust the weights from the household weights to get a "fraction of total households in the area that were surveyed" and would, I think, allow you to use a fully pooled sample and make that "population representative" by (sort of) accounting for changing population over time (supposing you had different Nhouseholds estimates for each survey year). But since you want separate estimates for each region-X-survey (yes?), you shouldn't need to do that.

Like I said, I tend to think of each survey as population-representative, and usually conceive the population as being static (I don't want to weight my 2000 results less than my 2007 just because the population was larger in 2007). But that is a preference, not a rule.

Question - do you really need to pool these surveys at all to do what you want? Why not just collapse each survey down into regional cells, weighting by the given weights? Oh, except that 2004 claims (in the coding, but not the documentation) it is not regionally representative, which is odd.

Also, a last comment on stratification (which obviously I'm still not great on): accounting for strata shouldn't affect point estimates, only standard errors. In fact, accounting for strata should decrease the size of your se's a little bit. So you could be "conservative" with your p-values/CIs and not account for strata at all. You could even, to be more conservative, skip the "svyset" part, specify the weight in the regression itself, and use the cluster robust se option clustered on PSU. Those SEs will be, maybe, slightly too big (conservative), but not by much, and if they are significant you should know that regardless of how you stratify they will remain significant (I know, this is a somewhat non-compelling argument for many reasons, but I find stratification specification doesn't matter too much here).

Can I ask what you want region-X-survey-round estimates for? You'll end up with like N=10regions-X-3-surveys = 30 mean anthropometric measurements. Not a lot to do further analysis on, right? I ask because I have special interest in Bangladesh, but would totally understand if you had some really cool idea you didn't want to share yet.