
Subject: Re: Pooled datasets - weighting data problem
Posted by [Reduced-For\(u\)m](#) on Sun, 01 Jun 2014 19:36:35 GMT
[View Forum Message](#) <> [Reply to Message](#)

I think one issue we haven't dealt with (yours) is re-normalizing weights when using a sub-population. For a simple solution, I'd suggest going with the recommendation from the DHS staff (buried somewhere in this thread) for when you have multiple survey rounds from the same country - just use the regular weights*.

The idea is that, if the sample sizes are similar across survey rounds, the re-normalizing shouldn't matter much (you are implicitly weighting each survey by sample size when using the regular weights). And with the sub-pop command, I'm not sure how you would want to re-normalize anyway (that bit of survey design inference I'm not real teched up on and the Stata documentation isn't super helpful to me - I think the right re-normalizing might somehow relate to the ratio of the prevalence of the sub-population to the full population, probably across strata or something really difficult and nuanced).

*Note - you want to use the "subpop" command and you want to create new identifiers for "cluster" and "strata" that are survey-round specific (say, replacing cluster "10" with cluster "svy2011_10" or something like that).

**Additional option: just take out all the sub-pop observations from both rounds, sum (within round) all the weights, and divide the old weights by the sum of the new weights. Then you have sub-populations only in each survey round, and each round has the same total weight, but nothing is "population representative", it's just corrected for selection probability across the sub-population.

Any thoughts from the DHS staff on either of these options?
