
Subject: Re: Weighting in pooled data

Posted by [Bridgette-DHS](#) on Tue, 17 May 2016 09:59:48 GMT

[View Forum Message](#) <> [Reply to Message](#)

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

My comment in message #6672 may have been incomplete. If you calculate a cluster-level mean, proportion, standard deviation, etc., it will be the same whether or not you use weights. However, for analyses that include the clusters as units, you do need to save the total weight for the cluster.

When pooling multiple surveys, I would first re-scale the weights (e.g. hv005) in each survey by a factor. For example, if you have S surveys, N_i total (weighted=unweighted) cases in survey i , and a total of N cases in all S surveys ($N = \sum N_i$) then you could decide to give equal weight to each survey. You then want the weights in survey i to add to N/S , rather than to N_i . To do that, you multiply the weights in survey i by the ratio $(N/S) / N_i$. (I think of this as the target total divided by the original total.) You can actually do this re-scaling later, not necessarily just at the beginning....

Then when you do the collapsing to get cluster-level means you can ignore the weights for the calculation of cluster level means, etc., as I said, but you must save the weighted total for each cluster. For example, say you are using the IR file and want the mean CEB (which is v201) for each cluster. Part of the within-survey collapse would look like this: "collapse (mean) v201 (sum) v005, by(v001)". Then in your analysis you would treat the collapsed (summed) v005 as the weight. You should also carry along the stratum code and use svyset to adjust for weights and strata, although not for clusters, because the clusters are now your units. Hope this helps.
