
Subject: Weighting in pooled data
Posted by [owraza](#) on Mon, 16 May 2016 11:55:36 GMT
[View Forum Message](#) <> [Reply to Message](#)

I have tried to read extensively about weighting posted on this forum (also, re-watched the DHS webinar on weighting) and while doing so I found one comment by Tom Pullum (DHS) in message # 6672, quoting:

"If you construct a cluster-level variable using the collapse command, it is not necessary to use weights at all, because everyone in the same cluster has the same weight. To confirm this, you could collapse WITH weights and then collapse WITHOUT weights, and compare the two sets of numbers. They should be exactly the same.

However, if you want to collapse for a larger aggregate, such as a district or region, which includes more than one cluster, you definitely should use weights as part of the collapse."

My concern in this regard is, does this information still valid for a scenario where I have to pool various countries together (after collapsing at cluster level) and run regression analyses? Any suggestion where do I apply weighting?

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.

Subject: Re: Weighting in pooled data
Posted by [owraza](#) on Wed, 18 May 2016 15:16:03 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thanks for your post. Just one more question, will I be able to re-scale once I have merged and collapsed data (at cluster level)?

Subject: Re: Weighting in pooled data
Posted by [Bridgette-DHS](#) on Thu, 19 May 2016 14:13:30 GMT
[View Forum Message](#) <> [Reply to Message](#)

Another response from Tom Pullum:

Yes, you should be able to do the rescaling later. I would go back to the original files and calculate the survey-specific ratios or factors that I described. That's one number for each survey. Some factors will be greater than one, some will be less than one. Then go to the final (merged and collapsed) files and multiply the weights (or weighted frequencies) by those factors, survey by survey. To check, you can confirm that in the final file each survey has the same total weight. There are other ways to get the same result but this should work.

Subject: Re: Weighting in pooled data
Posted by [owraza](#) on Thu, 19 May 2016 20:52:58 GMT
[View Forum Message](#) <> [Reply to Message](#)

I will follow this process and let you know the result.
Thanks
