

---

Subject: Re: Pooled data analysis from 25 countries  
Posted by [Bridgette-DHS](#) on Wed, 06 Feb 2019 15:50:16 GMT  
[View Forum Message](#) <> [Reply to Message](#)

---

Following is a response from Senior DHS Stata Specialist, Tom Pullum and Senior DHS Sampling Specialist, Mahmoud Elkasabi:

1. For certain countries cluster and PSU variables are not the same. In such cases which variable shall I use to create the otherwise PSU level variables.

In most of the countries, variables for cluster and PSU are the same. In cases where the two are different, you can use the cluster.

2. Do I need to specify the strata variable while using svyset command in Stata. If yes, how do I deal with missing strata information.

Yes, you can use HV022. In cases where HV022 is not consistent with the stratification described in sampling design in Appendix A of the final report, you can recode a new stratification variable. If you are using HV022, you should not expect to find missing strata information.

3. I have already de-normalized the weights as suggested in earlier posts on this forum. Do I need to re-normalize the weights before using them? If yes, how shall I do it?

I do not see a reason to re-normalize the de-normalized weights. This should not affect your regression.

4. In one post on this forum I read that in multi country analysis data must be clustered at country level. Do I need to do that for this analysis? If yes, how do I cluster data at two different levels i.e., country level and then individual PSU level?

I would recommend fixed effects for country or survey, rather than random effects. Just assign a survey code such as survey=1, 2, 3, etc. to each survey and include a term such as "i.survey" in the model specification.

Someone else could prefer random effects for surveys, especially if there are MANY surveys in the analysis. That would require a 3-level hierarchical model (for respondents / clusters / surveys).

When combining surveys, the strata and cluster codes must be unique. For example, you do not want cluster 1 in survey 1 to be confused with cluster 1 in survey 2. For example, you could have "egen cluster\_id=group(survey hv021)" and "egen stratum\_id=group(survey hv022)."

---