
Subject: Re: Guidance Needed on Weighting for Pooled DHS Data in Logistic Regression

Posted by [Bridgette-DHS](#) on Mon, 04 Nov 2024 16:47:38 GMT

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

Following is a response from Senior DHS staff member, Tom Pullum:

What you have done so far looks fine to me. For the last step, actually applying the weights, please check previous posts on the "svyset" command. This will adjust for the clusters and strata in the surveys, as well as the weights. As you will see in the earlier posts, you need to construct a new variable for the cluster ID, to distinguish between v001 in different surveys. For example, you can enter "egen cluster_ID=group(survey v001)". You will also need a new variable for the statum ID. We recently posted (again) a file that specifies the stratification variable in all the surveys. In recent surveys it is v022=v023 and in most older surveys it is v024 x v025, but there are exceptions, and they are given in that file.

In the analysis I would include a fixed effect for survey with "i.survey". A multi-level model with a random effect for survey is not justified, in my opinion, and it would add complexity.

I expect that your outcome variable has huge variation from one survey to another, as well as variation within most surveys. The whole concept of pooling surveys with this kind of an outcome seems to me to be unnecessary, but of course you can do what you want.
