Subject: Re: Pooling 3 rounds of DHS Nepal -- weights?
Posted by Bridgette-DHS on Mon, 21 Oct 2019 10:59:05 GMT
View Forum Message <> Reply to Message

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

Say that you have combined the three surveys (by appending them) and you use v021 as the psu id and v023 and the stratum id. Then run these two lines:

egen clusterid=group(survey v021)
egen stratumid=group(survey v023)

This is the easiest way to get unique ids in the combined file for clusters and strata. There is no need to do any other kind of spatial reconciliation.

What you have done with the weights seems ok to me. Yes, the means will be 1 within each survey and overall. However, there is a problem of scale. You divided v005 by 1000000, but when you divide again by 2770.8256, etc. you will have a number which is much less than 1. You should avoid having decimal points or anything to the right of a decimal point in a weight. There are some kinds of weights (such as fweight in Stata) that require a weight to be an integer. I have encountered weighting procedures that will truncate a weight and ignore anything to the right of the decimal point, which means that a weight between 0 and 1 will be treated like 0--that is, the case will be ignored entirely! You might want to rescale your weights somehow with an arbitrary multiplier such as 1000000.