Subject: Re: Merging and appending Kenya DHS
Posted by Bridgette-DHS on Tue, 07 Nov 2017 12:56:14 GMT
View Forum Message <> Reply to Message

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

I think there are still a couple of questions hanging from your August posts about using the PR, IR, and MR files from the Kenya 2014 DHS. Sorry for the delay. I will try not to repeat what I said earlier. First, on "when to renormalize the weights", this is mainly an issue when pooling surveys from different countries or several surveys from the same country. This amounts to finding some survey-specific number for survey i, call in ki, to re-scale each survey up or down. ki could be the population size (at the time of the survey) Ni divided by the sample size, ni, in which case the weighted number of cases can be interpreted as population estimates. That is, ki=NI/ni. Then, in the pooling, the relative weight of each survey will be proportional to the population size. This sounds good but there is a down side--the pooled results are hardly affected at all by the smaller countries. The alternative is to weight each survey equally. For example if n is the total sample size in a pooling of 20 surveys, and ni is the sample size for survey i, then you ki will be ki=(n/20)/ni = n/(20*ni). This would be my preference. (To be very specific, I am saying that you have a command such as "gen hv005_rev=hv005*ki".)

In Stata, pweights are always rescaled so that they have a mean of 1. Thus [pweight=mv005] will give you exactly the same result as [pweight=weight] where weight=mv005/1000000. Try it both ways and you will see.

Second, I said, "I see that this survey only had a subsample of men". This would be a problem if, say, you merged the IR and MR files with the PR file and then wanted to analyze, say, men and women age 15-49. The PR file includes 32,172 women who are age 15-49 and de facto residents (hv103=1); all of them were eligible for the interview of women (hv117=1). The PR file includes 29,514 men who are age 15-49 and de facto residents. Of them, 13,337 lived in households selected for the male interview, i.e. were eligible for the male interview. If you want to pool the men and women, using variables that are in both the IR and MR file, to get an estimate for men and women combined, you will have to weight up the men, basically with a factor 29514/13337, but actually the factor should be the ratio of the sums of the weights for the 29,514 and the 13,337 cases.

You only need to make these adjustments to the weights if you want to produce pooled estimates. If you just want to compare surveys or compare men and women, it is better to leave the weights alone.