Subject: Re: NFHS3 - Weights and Survey command
Posted by Bridgette-DHS on Sat, 15 Aug 2015 14:44:14 GMT
View Forum Message <> Reply to Message

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

For summary statistics, you can usually manage with iweights. You should virtually never use aweights. The definition "analytical" for the "a" is very misleading. Read "help aweight"--they can be a big help if you are combining means that are based on different numbers of cases but not otherwise. Use, for example, "summarize x [iweight=v005/1000000]"

For virtually all estimation models, you can use svy, with pweights. I am not aware of the warning you mention. Some people, usually economists, are opposed to using any weights in estimation models. Perhaps that's where the warning came from, but it would not just be for ML estimation.

You can get the syntax for svyset with "help svyset". It has changed slightly with version 14. In DHS surveys, the cluster variable is v001 or v021, they should be identical. You use "pweight=v005"; no need to divide by 1000000. Stata automatically normalizes pweights to have a mean of 1. See other posts for the stratum variable; it is not always the same. After specifying svyset, to apply it, you put "svy:" at the beginning of an estimation command, e.g. "svy: regress y x".

Summary statistics with weights can be difficult. For example, "tab A B, summarize(x)" doesn't like iweights, let alone pweights. The "collapse" command can be problematic. However, if all else fails, for such commands, you can use "[fweight=v005]" and just ignore the huge frequencies; means will be correctly weighted.