Subject: Re: De-normalizing weights and svyset command in Stata
Posted by Bridgette-DHS on Tue, 17 Nov 2015 15:20:32 GMT

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

No problem--don't hesitate to ask questions. If you put several IR files (for example) into a single file, perhaps by appending them, you need to construct a variable called "survey" (for example) that is 1 for the first survey, 2 for the second survey, etc. Ideally you could just us v000 for this purpose, but there have been some instances of two successive surveys in the same country that have the same code for v000, so I always construct a new variable such as "survey". The command to construct the new cluster codes could be "egen cluster_all=group(survey v001)". Here v001 is the cluster variable (v021 is always exactly the same as v001) and there is no need to name it "cluster" but you can if you want. Both "survey" (after you construct it) and v001 are in all the surveys, and they do not need to have separate names within each survey--in fact that would defeat the purpose. You are right that hiv05r (or whatever you want to call it) would also be the same variable across the combined file. You do not need separate names within the individual surveys.

I can't tell what you mean by "should be using all of the women's variables for this weighting process" in your CR+AR merged file. The recommended weight in this file would be hiv05 for the men in the AR file. It would be applied to all runs using this file, including a run that, say, only happened to use variables from the women in the couples, because their probability of inclusion in this file is affected by the participation of the matched men in the HIV testing, and their participation (usually, at least) is more problematic than the participation of the women. No variables other than *v005 or hiv05 are ever involved in the calculation or selection of weights, so that phrase (in quotes) does not seem relevant.