Subject: Re: ARI data and commands in STATA
Posted by Janet-DHS on Thu, 21 Nov 2024 14:34:46 GMT
View Forum Message <> Reply to Message

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

When you combine KR files from different surveys into one file, you "append", not "merge". Appending and merging are very different, and since you have already combined the files, I assume that you did use the append command.

If you are looking at changes over time, with a sequence of cross-sectional surveys, then you do not need to change the weights at all. You would only need to change the weights if you wanted to construct a single estimate for the full sequence of surveys. In earlier responses I have said that I think pooled estimates are meaningless, but there have been several posts on how to re-weight for pooled estimates.

You may need to add "singleunit(centered)" to the end of the svyset command to avoid an error message that often comes up.

Construct an identifier for each survey that you call "survey". It can be 1, 2, 3, for the successive surveys or the year of the survey or just about anything, such as the year of the survey (but not v000, because it can be the same in successive surveys).

When selecting a specific survey in the combined file for an estimation command, use the "subpop" option in svy rather than, say, "if survey==1". The difference in results is very small but (as you will find from Stata documentation) "subpop" is preferred.

The stratum variable is NOT v024--that variable is "region". Stratum is given in recent surveys by v022 or v023 (they are now duplicates) but for older surveys you need to check the stratum specification file. A link has been posted on the forum several times, including twice recently. For most surveys the strata are combinations of v024 and v025, but you should check. Construct or identify the stratum variable for each survey and give it the same name in every survey, such as "stratumID". Then "egen stratumID_all = group(survey stratumID" will produce a stratum identifiers that is unique across all surveys. Within svyset, the stratum component will be "strata(stratumID_all)".

Similarly, you can get unique cluster identifiers with "egen clusterID_all = group(survey v001)" and the cluster component will be "clusterID_all". As I said, I would not alter the weight variable v005. The steps I described for strata and clusters are only needed if, say, you want to fit a line through several surveys or compare surveys. For example, if you wanted to estimate the increase between survey 3 and survey 4 in the percentage of children with ARI symptoms who are taken for medical treatment, and test whether the increase was statistically significant, you would need unique identifiers for strata and cluster and you would not need to change the weights.

Hope this is helpful and not too confusing!