Subject: Re: Sample weights and stratification - Nigeria 2008 and 2018
Posted by Goethe2014 on Wed, 20 May 2020 15:57:12 GMT
View Forum Message <> Reply to Message

Dear Tom,
Thank you for your reply (I assume the second reply was not meant to answer my question?).

If I understood correctly: First, I do not have to reweigh the sampling weights when combining the Nigerian DHS 2008 and 2018 IR data so I will stick to v005 as sampling weights.

Second, I have to create new stratum and cluster identifiers having combined/appended the surveys. I would code
gen survey=.
replace survey=1 if v007==2008
replace survey=2 if v007==2018

Regarding stratification: Both DHS surveys (according to the final reports) have been stratified by state and urban/rural. Now the issue is that in the DHS2008 v023 accounts only for the state. In the DHS 2018 v023 accounts for region, state and rural/urban. How could I deal with this difference in v023 definition? In general the surveys make use of the same stratification process but v023 only indicates this stratification process completely in the DHS 2018. (i.e. "label list V023" for DHS 2018 shows that there exist 74 unique values of v023 (37 states either rural or urban) whereas "label list v023" for DHS 2018 shows that there exist 37 unique values (37 states only).

Regarding clusters: The DHS 2008 data lists 888 unique cluster, the DHS 2018 data lists 1400 clusters. These cluster are created based on the same National Census from 2006 (DHS cluster/PSU= 2006 Census Enumeration Area (EA)). If i now make use of the command "egen cluster_ID=group (survey v001) Stata now lists 2275 unique clusters (1-2275). This seems odd to me as the clusters should be the same.

I have seen that in published papers using the Nigerian DHS 2008 and 2013 data the authors do NOT generate any new variables and stick to v005 as sampling weight, v001 as clusters and do not account for the stratification process at all. Would that also be an option?

Thank you very much in advance!