
Subject: Re: Sample weights and stratification - Nigeria 2008 and 2018

Posted by [Goethe2014](#) on Thu, 21 May 2020 09:13:22 GMT

[View Forum Message](#) <> [Reply to Message](#)

Dear Tom,

Thanks a lot for the advice.

I applied the command first to the DHS 2008 data [egen stratum_ID_2008=group(ssstate v025)]. If I tabulate stratum_ID_2008 it now shows me 74 distinct values. If I do the same in the DHS 2018 data [egen stratum_ID_2018=group(ssstate v025)] this also gives me 74 distinct values.

I then appended the DHS 2018 data to the DHS 2008 data [append using "MY_DHS_2018_FILE"]. If I now order the total appended dataset [order stratum_ID_2008 stratum_ID_2018] and browse I can see that of course there is no entry (.) for the stratum_ID_2018 for observations from DHS 2008 which makes sense. If I now make use of the command you suggested [egen stratum_ID=group(stratum_ID_2008 stratum_ID_2018)] this creates missing values only as either stratum_ID_2018 is missing for DHS 2008 data and vice versa stratum_ID_2008 is missing for DHS 2018 data. Therefore the newly generated stratum_ID variable in the appended/combined dataset has 0 entries.

Did I miss something or is there any solution for this issue?

Having solved this I would now have the weights (v005) and the stratification indicator (stratum_ID). If I want to define svyset I therefore know 2 of 3 needed values [svyset [pweight==v005], psu (???) strata (stratum_ID)]. You wrote that I could not assume the clusters (which are equal to the Primary Sampling Unit/psu v021) are the same in DHS 2008 and 2018. How would therefore the complete command for svyset look like?

Thanks in advance.

Greetings
