
Subject: Re: sampling stratification of the Cameroon DHS 2011, 2004, 1998 and 1991 data

Posted by [Bridgette-DHS](#) on Mon, 11 Jul 2016 14:46:33 GMT

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

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

Just some minor modifications. In the syntax for svyset, both the psu and the weight go before the comma, and the psu does not go in parentheses.

It is not necessary to divide v005 by 1,000,000; Stata will automatically normalize the weights to have a mean of one. You can do it if you want but it will have no effect.

I have not looked at the four surveys but for many of the older DHS surveys the stratum variable is incorrect. I would use the combinations of regions and place of residence.

It's a good idea to include the singleunit option. Otherwise you will probably crash.

So this is what I would recommend:

```
egen strata = group(v024 v025), label  
svyset v021 [pweight=weight], strata(strata) singleunit(centered)
```

If you put all of the surveys into a single file, then the only change I would recommend is that the surveys be numbered (survey=1, 2, 3, 4, for example) and you do the following:

```
egen clusters=group(survey v021), label  
egen strata = group(survey v024 v025), label  
svyset clusters [pweight=weight], strata(strata) singleunit(centered)
```