Subject: Re: Clustered Standard Errors
Posted by Bridgette-DHS on Wed, 24 Feb 2021 14:01:54 GMT
View Forum Message <> Reply to Message

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

The generic form of svyset is this:

svyset cluster_ID [pweight=v005], strata(stratum_ID) singleunit(centered)

In this syntax, the first variable after "svyset" (here "cluster_ID") is the PSU. "singleunit(centered)" is related to the "strata(stratum_ID)" term. It keeps the program from crashing if it encounters only one PSU within a stratum. There are a couple of alternatives to "centered" but I have done comparisons and the results are indistinguishable for the different options. You are usually ok without the singleunit option but I usually include it because I hate crashes.

Below I will paste an example from one time when I was pooling two surveys from 2008 and 2018. The two surveys had different specifications of strata. You can see how "egen group" was used. Note that "egen group" does NOT combine or pool. It does just the opposite. For example, say that in the PR file you wanted to construct a variable for all combinations of urban/rural (hv025=1 or 2) and male/female (hv104=1 or 2). You would use "egen place_sex=egen(hv025 hv104)" to get a four-category variable for the combinations of hv025 and hv104. This can be handy for making tables or interpreting interaction terms. In the example below, it basically distinguishes the designs of the 2008 and 2018 samples. Hope this helps.

```
* In the 2008 survey
egen stratum_ID_2008=group(shstate v025)
gen tempvar=stratum_ID_2008

* In the 2018 survey
gen stratum_ID_2018=v023
gen tempvar=stratum_ID_2018

* Append, and construct "survey" using v007...

* In the combined file
egen stratum_ID=group(tempvar survey)
drop tempvar

egen cluster_ID=group(v001 survey)

svyset cluster_ID [pweight=v005], strata(stratum_ID) singleunit(centered)
```