Subject: using design elements correctly
Posted by gebretsh@gmail.com on Fri, 16 Dec 2022 08:42:40 GMT
View Forum Message <> Reply to Message

Dear Dr Tom and other DHS experts,

I'd like to thank you for your usual invaluable firsthand assistance with DHS data analysis.
I would like now to ask questions on the specifications of the three design elements of DHS data:
cluster, weight and strata

1) I have a habit of using these three elements whenever I do analysis using DHS data via the
svyset function in Stata.
svyset psu [pw=weight], strata(strata var) singleunit(centered)
Now, I read today the Stata's survey data reference manual and recommends the specification of
a secondary sampling unit (ssu), which is the household ID in DHS, as follows:

svyset psu [pw=weight], strata(strata var) || household Id (v002)

I have already analyzed my data using the first command and sent it to a journal for publication.
Should I re-analyze the data using the second code?

2) I want to use a Stata command that does not support "svy". The Stata command that I want to
use is "mvdcmp", a tool used to do decomposition analysis between two groups. Now, In place of
the svy command, I just opt to use another way of supplying the design elements into my syntax,
as follows:
 mvdcmp place of residence: logit skilled_onc_2days wealth_early1 wealth_early2 [pw=w1],
robust cluster(id)
To add to the problem, this "mvdcmp" command does not support/accept strata, and supports
only weight and cluster, as indicated above. Is there a severe problem If ignore the strata variable
from being taken into account in my analysis.

Thanks so much for your advice.

Regards,

Subject: Re: using design elements correctly
Posted by Bridgette-DHS on Fri, 16 Dec 2022 16:04:55 GMT
View Forum Message <> Reply to Message

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

We recommend the version of svyset that you are currently using. I just ran the lines below on the
Philippines 2017 DHS, for an example. #1 includes only the weights, v005. #2 adds the usual
adjustments for clustering and strata with svyset. #3 is your proposed modification of svyset, with
subsampling of households.

All three models give exactly the same estimates of coefficients. #2 and #3 give estimates of standard errors, test statistics, and confidence intervals that are different from #1. However, the estimates of standard errors, etc. are exactly the same in #2 and #3. That is, you can use #3 if you want but it appears from this simple check that the results will be the same as with #2.

Note: I am not proposing that you would analyze CEB with linear regression! This is just an example of a statistical model.

* Estimation #1
regress v201 i.v013 i.v190 [pw=v005]

* Estimation #1
svyset v001 [pw=v005], strata(v022) singleunit(centered)
svy: regress v201 i.v013 i.v190

* Estimation #1
svyset v001 [pw=v005], strata(v022) || v002
svy: regress v201 i.v013 i.v190

Subject: Re: using design elements correctly
Posted by Bridgette-DHS on Fri, 16 Dec 2022 18:09:23 GMT
View Forum Message <> Reply to Message

Here's a response to your second question, from Tom Pullum:

mvdcmp is just one example of an estimation command that does not allow svyset. There are also some commands that allow svyset now, but not in earlier versions of Stata. When this happens, your only option is to make as many of the adjustments for weights, clustering, and stratification as possible. Sometimes, if svyset is not accepted, you can still include [pweight=v005] before the comma and cluster(v001) as an option after the comma. The adjustment for stratification is the only one of the three adjustments that can only be done with svyset and svy. If the estimation command does not accept these adjustments, all you can do is to put a comment in your paper or report saying which adjustments were not possible.