
Subject: Use of svyset when standard errors are already clustered

Posted by [ale_sovera](#) on Wed, 28 Jun 2017 13:08:53 GMT

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

Good afternoon,

if in my regression analysis I need to cluster the standard errors by month-year of birth, should I use also svyset in Stata in order to take into account survey strata?

I have read some posts about that and I am still not completely sure about how to use svyset when running analysis with DHS data.

Subject: Re: Use of svyset when standard errors are already clustered

Posted by [Bridgette-DHS](#) on Wed, 28 Jun 2017 13:56:10 GMT

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

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

The purpose of svyset is to adjust the statistical estimates (mostly the standard errors) for the survey design. You do not need to make any adjustments for the variables in the analysis, such as month and year of birth. The only clustering in the survey design is for the PSUs, given by v021 (v001 is almost always the same as v021). The stratification variable is usually v023 or v022; if in doubt, use the combinations of v024 and v025. The weight variable is v005. There have been many postings on the syntax of svyset--hope you can find them.

Subject: Re: Use of svyset when standard errors are already clustered

Posted by [ale_sovera](#) on Wed, 28 Jun 2017 14:02:17 GMT

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

Got it. But I need to cluster standard errors by month-year of birth because of my design, a regression discontinuity. It is routine.

Does my clustering compete in some way with the one suggested by DHS?

Subject: Re: Use of svyset when standard errors are already clustered

Posted by [Bridgette-DHS](#) on Wed, 28 Jun 2017 14:13:28 GMT

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

Another response from Senior DHS Stata Specialist, Tom Pullum:

DHS cannot advise you on your methods of analysis or Stata syntax. If what you are doing is "routine" then you should have no trouble finding out how to do it from the Stata forum or other websites.

Subject: Re: Use of svyset when standard errors are already clustered

Posted by [Reduced-For\(u\)m](#) on Sat, 12 Aug 2017 22:32:10 GMT

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

I think I see what both of you are saying:

For Tom (DHS) the nature of the data collection/sampling is such that the only "built in" issues of correlated standard errors (from a sampling standpoint) is in the PSU, and hence the DHS suggests (tightly) "clustering" in some way on PSU.

For Ale, the question isn't about data-collection design but about econometric/statistical design, where in month-year of birth probably determines some "treatment assignment" and then the whole "correlated treatment assignment" issue arises... if "treatment" is assigned by the cohort, any within-cohort correlation in Eps gets blown up by the serial-correlation within-group in T (whatever the treatment assignment variable of interest is). This is a version of the problem in Bertrand et al "how much should we trust diff in diff?" paper.

Clustering on one won't fix the other, because they are not nested (there is not a "higher" group to cluster on that nests the other groups. An obvious way to deal with both concerns would be to use the Cameron, Gelbach and Miller approach to multi-way clustering. Then you could deal with the spatial auto-correlation in the error term the survey design generates, and the cohort level auto-correlation in X that the identification strategy generates (and risks blowing up any within-cohort correlations in the error term).

I'd also advise you to google about birth date measurement in the DHS (it isn't as good as we'd like) and consider that this may (I have no idea) pose a threat to your identification strategy.