Understanding Society User Support - Support #1080

Weights for pooled cross-sectional analysis - accounting for clustering

10/21/2018 06:39 PM - Lewis Anderson

Status:	Resolved	**Start date:**	10/21/2018
Priority:	Normal	**% Done:**	100%
Assignee:	Lewis Anderson		
Category:	Weights		

Description

Dear Support Team,

This can be seen as a follow-up to #758, which presents a similar problem.

I am trying to explore the cross-sectional association between two time-varying variables (a value of interest of one of the variables is relatively rare). To do this I would like to pool data from the various waves of Understanding Society.

In comment #7 on #758 Nico Ochmann writes: "I run logrealhourlywage on x1 x2 [pw=newwgt], cluster(pidp) / Is this reasonable or am I still completely off?", to which Peter Lynn replies "Looks fine!".

However I would also like to account for the survey design by using ~~svyset~~ in Stata. ~~svyset~~ does not allow the cluster option. Is there a straightforward way around this? Or is it not possible to cluster on pidp because I am effectively already clustering on psu by specifying: svyset psu [pweight=weight_indsc_xw], strata(strata) singleunit(scaled) -- where weight_indsc_xw is a_indscus_xw from wave 1, b_indscub_xw from wave 2, etc.? Is it in fact satisfactory to cluster on the higher level (PSU) and ignore clustering within individuals at the lower level?

Or - would it be better to run this as a multilevel model, with observations clustered in individuals, individuals (in households, and households) in PSUs? According to the Stata help file for ~~mixed~~, and the parts of the Stata Reference Manual to which it refers, this raises a few difficulties with regard to sampling weights:

"...it is not sufficient to use the single sampling weight wij , because weights
enter into the log likelihood at both the group level and the individual level. Instead, what is required
for a two-level model under this sampling design is wj , the inverse of the probability that group j
is selected in the first stage, and wijj , the inverse of the probability that individual i from group j is
selected at the second stage conditional on group j already being selected."

Any help much appreciated.
Regards,
Lewis

History

#1 - 10/22/2018 11:33 AM - Stephanie Auty

- Category changed from Weights to Data analysis

- Assignee changed from Peter Lynn to Stephanie Auty

- % Done changed from 0 to 10

- Private changed from Yes to No

Many thanks for your enquiry. The Understanding Society team is looking into it and we will get back to you as soon as we can.

Best wishes,
Stephanie Auty - Understanding Society User Support Officer

#2 - 10/22/2018 04:32 PM - Peter Lynn

- Status changed from New to Feedback

- Assignee changed from Stephanie Auty to Lewis Anderson

- % Done changed from 10 to 50

Lewis,

With "svyset psu ..." you have indeed already specified PSUs to be the clusters. This will give you unbiased standard error estimates even if there are

additional levels of clustering (e.g. individuals within households, and observations within individuals (as you are pooling)), provided that those additional levels are hierarchical to PSUs (which they are, in this case). It will not however apportion the variance between the levels. For that, you would need to specify the levels explicitly, which you can do in Stata. An example would look something like this:

svyset psu [pweight=weight_indsc_xw]|| pidp, strata(strata) singleunit(scaled)

For a multilevel model you should indeed specify weights at each level, as described in Pfeffermann et al (1998).

Regards,

Peter

Reference: Pfeffermann, D., Skinner, C. J., Holmes, D. J., Goldstein, H., & Rasbash, J. (1998). Weighting for unequal selection probabilities in multilevel models. Journal of the Royal Statistical Society: series B (statistical methodology), 60(1), 23–40.

#3 - 10/23/2018 02:41 PM - Lewis Anderson

Great, that answers my question. Thank you.

Regards,
Lewis

#4 - 11/08/2018 04:46 PM - Stephanie Auty

- Status changed from Feedback to Resolved

- % Done changed from 50 to 100

#5 - 08/22/2023 01:07 PM - Understanding Society User Support Team

- Category changed from Data analysis to Weights