

Understanding Society User Support - Support #1273

Weights_considering survey complex structure ?

11/13/2019 11:52 AM - Lydia Palumbo

Status:	Resolved	Start date:	11/13/2019
Priority:	High	% Done:	100%
Assignee:			
Category:	Weights		
Description			
Dear Support,			
I have a question on how to analyse weights. In the handouts of the Moodle, I found that I should consider the survey complex design and set data this way with STATA.			
svyset psu [pw=weight], strata(strata) singleunit(centered).			
Now, I am interested in clustering the standard errors by individual (pid) or in running a random effect model to account for individual heterogeneity, but STATA does not allow to svy data in this way and run this kind of model (which should be reasonable).			
Would you think that it is possible to avoid controlling for the complex structure of the survey? If not, what would I actually miss?			
Thank you. Best, Lydia			

History

#1 - 11/13/2019 01:19 PM - Stephanie Auty

- Category set to Weights

- Assignee set to Olena Kaminska

- Private changed from Yes to No

#2 - 11/13/2019 02:16 PM - Olena Kaminska

Lydia,

Indeed some statistical analyses do not work with svy command. One of such is multilevel models, random effect model is one of these. Two options are available to you: check with an expert on multilevel modelling and whether you may need to use a specialist software to run the model that would enable you to take into account full sample design.

Second option, use Stata and random effects model with weights - you should be able to use weights with this. Also, use person ID and PSU as two nested clusters (depending on your data setup - this may be different). This way you will be ignoring only stratification which does not influence your point estimates and makes your confidence intervals slightly wider. Talk to an expert though about potential effect on within and between variance estimates if you are interested in them.

Hope this helps,
Olena

#3 - 11/13/2019 05:12 PM - Lydia Palumbo

Thank you. It does.

I think that in this way it is possible to use svyset because I can create a variable that considers both PSU and ID within a strata. So the command should be:

```
svyset psupid [pw=weight], strata(strata) ...
```

I will check if this makes sense with an expert in multilevel.

Thank you again.
Lydia

#4 - 11/14/2019 10:18 AM - Olena Kaminska

Lydia,

No, you shouldn't combine PSU and ID - as the results will be wrong. If you have to choose you should use the higher level clustering: PSU. You could also explore an option of using SSU (as ID) with PSU, as in this example:

<https://www.stata.com/support/faqs/statistics/stratified-multiple-stage-designs/>

Best,
Olena

#5 - 11/14/2019 11:35 AM - Lydia Palumbo

Thank you. Yes, I realized that the combination was nonsense. Thank you for the link. I was not sure how to set the individual level in the design. This should help.

Best,
Lydia

#6 - 11/20/2019 02:56 PM - Stephanie Auty

- Status changed from New to Feedback

- Assignee changed from Olena Kaminska to Lydia Palumbo

- % Done changed from 0 to 80

#7 - 11/22/2019 03:51 PM - Lydia Palumbo

Hi,

I do have another question about multistage sampling.

My sampling strategy is to consider only those who formed their union within the panel. Should I apply a factor of correction for this? I am using longitudinal weights.

Thank you and best,
Lydia

#8 - 11/25/2019 12:57 PM - Olena Kaminska

Lydia,

Yes, you can take couples into account. If you want you can specify three levels of clustering: PSUs, couples and individuals. Then I assume you will have observations within individuals.

Hope this helps,
Olena

#9 - 12/18/2019 02:38 PM - Lydia Palumbo

Hi,

I am using longitudinal weights to analyse the event between t and $t+1$.

Now I am questioning whether I am using the correct weights. I noticed that once the boost for Scotland/NI is done, those who attrited before wave 10/12 and had a weight of 0, are then given a positive weight.

Which weight should I use for these units? Should I consider them as part of the sample (like as they were truncated for some time)?

I would say yes because otherwise there would not be the representativeness of GB, but I am not very sure.

I would appreciate your input.

Thank you and best,
Lydia

#10 - 12/18/2019 02:50 PM - Stephanie Auty

- Assignee changed from Lydia Palumbo to Olena Kaminska

#11 - 12/20/2019 07:33 AM - Olena Kaminska

Lydia,

Thank you for your question. Could you clarify? Are you creating your own weights or are you using ours? If you are using ours - they are correct and

you don't need to worry about zeros etc. In our weights u_i weight may be positive while u_b or u_s weights are zero. This is due to how they are calculated and this is correct. More importantly in a pooled analysis using u_s , u_b and u_i weights together throughout time will give you correct results.

If this doesn't answer your question could you provide more details on which weights you are using in your analysis?

Olena

#12 - 12/20/2019 08:01 AM - Lydia Palumbo

Hi Olena,

Sorry. I was not clear. I will rephrase the issue.

I am doing a pooled cross-sectional analysis (with the events in $t+1$) with all the waves from BHPS and UKHLS, including Scottish, Northern Irish Sample, IEMB and EMB. I am using longitudinal weights, as you said.

For BHPS

Waves from 1 to 9 I am using $b^*w^*_{lrwght}$.

Waves 10 to 12 $b^*w^*_{lrwtsw1}$;

Waves 13 to 18, $b^*w^*_{lrwtuk1}$.

For UKHLS

Wave 1 $w^*_{indinus_xw}$,

Waves 2 to 6 $w^*_{indinub_lw}$

Waves from 7 on $w^*_{indinui_lw}$.

Are they correct?

I noticed that those individuals that were having $b^*w^*_{lrwght} = 0$ between wave 1 and 9 (because they missed one wave) are given a positive weight from wave 10 (by using $b^*w^*_{lrwtsw1}$) or from wave 13 (by using $b^*w^*_{lrwtuk1}$). They would have a weight of 0 if I used $b^*w^*_{lrwght}$ for all the waves (I do not do that because I want also to have boosts).

So my question is whether I have to include those who had a weight of 0 between wave 1 and 9 and then a positive one from wave 10 or 13 on.

Hope this is clear. Please tell me if I could be more explicit.

Thank you and best regards,

Lydia

#13 - 12/20/2019 08:06 AM - Lydia Palumbo

I forgot one part. If possible, I would like to do a robustness check by using cross-sectional (or design) weights in t and apply my own correction for individual and partners' non-response, as we spoke on the phone.

Would it be possible to be advised on how to perform it?

Thank you again.

Best,

Lydia

#14 - 12/23/2019 02:30 PM - Olena Kaminska

Lydia,

My suggestion would be use longitudinal enumeration weight at time t , and conditional on it being positive predict personal response to wave t and wave $t+1$ at the same time. This will be valid for OSMs only.

Hope this helps,

Olena

#15 - 12/23/2019 02:38 PM - Olena Kaminska

And to respond to your earlier question, the weights that you suggested are correct. The reason that some zero weights become non-zero at wave 10 and 13 is also correct and indeed is related to new modelling for boosts. I suggest that you include everyone in the model and rely on weights to exclude people (people are excluded if a weight is zero). If you want to change anything like have people with zero weights in your model you have to create your own tailored weights. Otherwise I always suggest that your choice to include or exclude people should be only substantive (categories of social groups etc.) and never related to their response pattern or samples - as long as you use weights your results are representative.

Thanks,

Olena

#16 - 08/15/2022 03:30 PM - Understanding Society User Support Team

- Status changed from Feedback to Resolved
- Assignee deleted (Olena Kaminska)
- % Done changed from 80 to 100