

Understanding Society User Support - Support #2215

PSU, Strata

02/26/2025 06:34 PM - Ioannis Rotous

Status:	In Progress	Start date:	02/26/2025
Priority:	Normal	% Done:	10%
Assignee:	Olena Kaminska		
Category:	Survey design		
Description			
I read through the documentation provided in Understanding Society web page for the data design and data collections, however, I'll need some clarifications if possible. :)			
The variables <code>_psu</code> and <code>_strata</code> bring some kind of nested hierarchy across individuals, however, I'm not fully sure the order of the hierarchy. An individual in <code>_psu</code> equal to <code>i</code> will be in <code>_strata</code> equal to <code>j</code> , or the opposite, an individual on <code>_strata</code> equal to <code>j</code> will be in <code>_psu</code> equal to <code>i</code> ? In plain words, do we have school, class, student with respect to the <code>_psu</code> , <code>_strata</code> and <code>pidp</code> .			
Also, I've read that the <code>_psu</code> corresponds to the postal code sector, however it takes values from 1 up to 50000+, hence, how can those values be interpreted as postal codes?			
Best, Ioannis			

History

#1 - 03/02/2025 07:26 PM - Understanding Society User Support Team

- Category set to Survey design
- Status changed from New to In Progress
- Assignee changed from Understanding Society User Support Team to Olena Kaminska
- % Done changed from 0 to 10
- Private changed from Yes to No

#2 - 03/03/2025 01:52 PM - Olena Kaminska

Ioannis,

Thank you for your question. The data structure within one wave is PSU, household, person, and in a longitudinal context it is `psu`, person, time-point observation.

We use implicit stratification. Depending on statistical model, sometimes they can't account for stratification. Stratification gains statistical efficiency, hence making your CIs smaller. But if the model can't take them into account your results will still be valid, but conservative.

Hope this helps,
Olena

#3 - 03/03/2025 02:03 PM - Ioannis Rotous

Hi again,

I use a Bayesian model so I can include them as random effects, so I can include the PSU as a random effect, and also the STRATA as another random effect. But the STRATA is a finer or coarser representation of the data compared to PSU? From my understanding it is a coarser.

Best,
Ioannis

#4 - 03/04/2025 02:23 PM - Olena Kaminska

Strata is not nested within PSUs. It is also not used as clustering in the design. Of course you could define your clusters based on strata information, if you wish, if your research design aligns with the definition of our stratification, but more likely you would use other variables for this. In such situation you may need a cross-classified modelling.

Hope this helps,
Olena