

Understanding Society User Support - Support #1329

Enumeration of strata variable and

04/01/2020 10:22 AM - Andreas Wejs Andersen

Status:	Resolved	Start date:	04/01/2020
Priority:	Normal	% Done:	100%
Assignee:			
Category:	Weights		
Description			
Dear Support			
I have two questions regarding strata formation and clustering in analysis:			
1. After reading Lynn (2009) - Sample Design for Understanding Society (and consulting both your online class and the manual), I am left with a question regarding the enumeration of strata seen in the variable 'strata'. This surely comes down to my a lapse in understanding of the survey design. For the GPS Lynn details the stratification of postal code into 103 distinct strata (12 regions X 3 SEG-bands X 3 pop. density bands), however when tabulating Strata for GPS members in wave 1 of UKHLS I see 1200 strata. Where is the disconnect? I also find a discrepancy between strata_bh and the characterization given in Lynn (2006) "Quality Profile: British Household Panel Survey". Lynn details 82 minor strata while the strata_bh variable takes 75 values at wave 1 of BHPS.			
2. Should I specify two levels of clustering if studying individuals (say in stata svy enviroment)? If my interest is in individuals (adult respondents), then for the GPS my current understanding of the structure is: 1. Postal codes are translated into sectors which are sorted into 103 strata. 2. PSU's are drawn (first clustering level) with proportionate probability, 3. Addresses/delivery points are drawn at random from PSU (second level of clustering?) with correction for multiple household at the same address.			
I would be thankful for any help you could provide Andreas W. Andersen			

History

#1 - 04/01/2020 10:25 AM - Andreas Wejs Andersen

2. Question implies I want a definitive advice on how to cluster in practice. What I meant is: Is there theoretically/strictly speaking 2 levels of clustering.

I see from various examples of stata "svy set"-functions, that you often apply only one level of clustering (PSU) and I fully intend to do so myself.

#2 - 04/01/2020 03:07 PM - Stephanie Auty

- Due date deleted (04/10/2020)

- Assignee set to Alita Nandi

- Estimated time deleted (0.25 h)

- Private changed from Yes to No

#3 - 04/03/2020 05:15 PM - Stephanie Auty

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

#4 - 04/06/2020 02:23 PM - Stephanie Auty

- Assignee changed from Alita Nandi to Olena Kaminska

#5 - 04/06/2020 05:43 PM - Olena Kaminska

Dear Andreas,

Thank you for your questions.

1. The strata variable is correct and correctly reflects the sample design. Trust it. The details on the stratification design are probably hidden somewhere in documentation.

2. Unless you use multilevel analysis or pooled analysis, you should only use PSU variable as your cluster variable. The higher geographies to psu do

not matter as they did not influence our sample design clustering. But technically indeed we have waves nested within individuals nested within households nested within psu's. In this situation taking into account clustering within psu (in other words the highest level of clustering) will take into account clustering at lower levels as well - read more on this in statistical books.

Hope this helps,
Olena

#6 - 04/07/2020 05:29 PM - Stephanie Auty

- *Category set to Weights*
- *Status changed from New to Feedback*
- *Assignee changed from Olena Kaminska to Andreas Wejs Andersen*
- *% Done changed from 0 to 60*

#7 - 10/13/2021 11:55 AM - Understanding Society User Support Team

- *Status changed from Feedback to Resolved*
- *Assignee deleted (Andreas Wejs Andersen)*
- *% Done changed from 60 to 100*