Understanding Society User Support - Support #1173

Clustering

03/27/2019 11:21 AM - OLAYIWOLA OLADIRAN

Status:	Resolved	**Start date:**	03/27/2019
Priority:	Low	**% Done:**	100%
Assignee:	OLAYIWOLA OLADIRAN		
Category:	Weights		

Description

I am trying to run some analysis using individual and household characteristics and having to deal with clustering issues. Presently, I am clustering at Pidp level on the premise that errors are unlikely to be uncorrelated within individuals. Do you think clustering at Pidp makes sense? I am also trying to cluster at household level (hidp), but household IDs change every year and it will be difficult to keep track of all these changes for a large number of observations. Do you recommend clustering at hidp level instead (despite that these change every year)?

History

#1 - 03/27/2019 11:41 AM - Alita Nandi

- Status changed from New to Feedback

- Assignee changed from Alita Nandi to OLAYIWOLA OLADIRAN

- % Done changed from 0 to 50

- Private changed from Yes to No

Hello,

Please provide some more information about your analysis.

We can say that the data is clustered and stratified and the variables representing the PSU and Strata are called "psu" and "strata" in the file "xwavedat". These variables with their wave prefixes are also available in each wave file "w_psu" and "w_strata". If you are pooling data across waves then you can cluster at the PIDP level as that will include any PSU level clustering.

The household IDs are only unique within a wave and cannot be used across waves - this is because household composition changes across waves.

Best wishes,
Alita

#2 - 03/27/2019 11:49 AM - OLAYIWOLA OLADIRAN

Alita Nandi wrote:

> Hello,
>
> Please provide some more information about your analysis.
>
> We can say that the data is clustered and stratified and the variables representing the PSU and Strata are called "psu" and "strata" in the file "xwavedat". These variables with their wave prefixes are also available in each wave file "w_psu" and "w_strata". If you are pooling data across waves then you can cluster at the PIDP level as that will include any PSU level clustering.
>
> The household IDs are only unique within a wave and cannot be used across waves - this is because household composition changes across waves.
>
> Best wishes,
> Alita

Thank you for your response.
We are using a multinomial model to analyse the factors that may be responsible for the housing tenure choices (ownership, rental and public housing) of individuals (using both individual and household characteristics. We are using a pooled cross-section (so we use waves 1-8, but not in panel form). Do you feel clustering at pidp level is necessary especially since we are using the pooled cross-sectional setup?

#3 - 03/27/2019 11:57 AM - Alita Nandi

If you use a pooled cross-sectional set up then there will be more than one observation for some individuals and so the error terms will not be independently distributed for all observations in your dataset - clustered on individuals. Hence the need to cluster on pidp.

#4 - 03/27/2019 12:01 PM - OLAYIWOLA OLADIRAN

Alita Nandi wrote:

> If you use a pooled cross-sectional set up then there will be more than one observation for some individuals and so the error terms will not be independently distributed for all observations in your dataset - clustered on individuals. Hence the need to cluster on pidp.

Thank you very much.

#5 - 04/02/2019 10:54 AM - Stephanie Auty

- % Done changed from 50 to 90

#6 - 08/10/2022 11:46 AM - Understanding Society User Support Team

- Status changed from Feedback to Resolved

- Priority changed from Immediate to Low

- % Done changed from 90 to 100

#7 - 08/22/2023 01:14 PM - Understanding Society User Support Team

- Category changed from Data analysis to Weights