

Understanding Society User Support - Support #1081

Youth and individual respondents datasets - merging info

10/26/2018 02:22 PM - Theodora Kokosi

Status:	Feedback	Start date:	10/26/2018
Priority:	High	% Done:	80%
Assignee:			
Category:			
Description			
Dear all,			
I would like to merge data from the "indresp" file into the youth file. Which would be the best way to do that?			
I am assuming that using the pidp as a key variable is not ideal since they are different respondents and their cases wouldn't match. Is the household identifier a better solution?			
To be more specific, I would like to use the variable for the maternal highest qualification as a covariate in models using data from the youth questionnaire.			
Thank you in advance.			
Kind regards, Dora			

History

#1 - 10/26/2018 02:39 PM - Stephanie Auty

- Status changed from New to In Progress
- Assignee set 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 - 10/28/2018 03:20 PM - Stephanie Auty

- Status changed from In Progress to Feedback
- Assignee changed from Stephanie Auty to Theodora Kokosi
- % Done changed from 10 to 80

Dear Dora,

The w_youth files contain the mother's ID in the variables w_mnpid (for natural mothers) and w_mnspid (natural, step and adoptive mothers). This is the variable which will match the pidp in w_indresp.

The simplest way to match the mother's highest qualification into the youth file would be to take the highest qualification and pidp from w_indresp, rename pidp to w_mnpid or w_mnspid depending on which you want to use, then merge with w_youth using w_mn(s)pid as the merge variable.

Best wishes,
Stephanie

#3 - 10/31/2018 04:40 PM - Theodora Kokosi

Dear Stephanie,

This is really helpful. Thanks a lot!

Best wishes,
Dora

#4 - 09/10/2019 03:37 PM - Marina Fernandez Reino

Hi,

I have a question regarding Support [#1081](#).

I am following Stephanie's advice because I also want to merge mother's information from indresp with the youth datafile. However, in the youth datafile there are duplicates of mother's id because there are sometimes more than 1 children interviewed in each household. When I try to merge it I get an error saying h_hidp h_mnspid do not uniquely identify observations in the master data (i.e. youth data file). What should I do?

Thanks

Theodora Kokosi wrote:

Dear all,

I would like to merge data from the "indresp" file into the youth file. Which would be the best way to do that?

I am assuming that using the pidp as a key variable is not ideal since they are different respondents and their cases wouldn't match. Is the household identifier a better solution?

To be more specific, I would like to use the variable for the maternal highest qualification as a covariate in models using data from the youth questionnaire.

Thank you in advance.

Kind regards,
Dora

#5 - 09/10/2019 04:06 PM - Gundi Knies

- Assignee deleted (*Theodora Kokosi*)

Hi Marina,

I think you might want to look up the `merge` command in Stata. You can do a m:1 or 1:m merge on mnspid. In this case, you have many youths in the youth data file who have the same mother in the indresp data file.

Hope this helps.

Gundi

#6 - 09/10/2019 04:15 PM - Marina Fernandez Reino

Thanks, Gundi. I don't know how I didn't realise it could be done that way

#7 - 09/10/2019 04:24 PM - Marina Fernandez Reino

Hi Gundi,

Just to make sure I am doing things right: there are 743 children who have a mother pidp identifier that cannot be matched with the mother's data from indresp because there are no such identifiers there. I assume these are non-respondent mothers, aren't they?

Thanks