

Understanding Society User Support - Support #1348

Merging

05/14/2020 06:24 PM - Laura Silva

Status:	Resolved	Start date:	05/14/2020
Priority:	Normal	% Done:	100%
Assignee:			
Category:	Data management		
Description I am merging info from the child and the indresp datafiles of wave 9 (and I have to do it for all the other waves). As a first step I did the following: use i_child.dta merge m:m i_hidp using i_indall drop _merge merge m:m i_hidp using i_indresp I end up with about 1,081 observation not matched from the master. Doesn't it mean means that there are observation in the indall and child files which are not in the indresp? Is it possible? More generally it is ok to proceed this way to put together info from the indresp and child datafiles?			

History

#1 - 05/15/2020 09:33 AM - Annette Pasotti

- % 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.

We aim to respond to simple queries within 48 hours and more complex issues within 7 working days. While we will aim to keep to this response times due to the current coronavirus (COVID-19) related situation it may take us longer to respond.

Best wishes,

Annette

#2 - 05/15/2020 11:20 AM - Annette Pasotti

- Status changed from New to Feedback
- Assignee set to Laura Silva
- % Done changed from 10 to 80

Dear Laura

Cases in the child data file are a subset of the indall file (i.e. all children in enumerated households), and the cases in the indresp file are a subset of the indall files as well, i.e. all adults who participated in an interview. Moreover, not all households have children (and, in principle it is possible that not all households with children have adult interviews). Thus, we would expect the number of cases to drop when we merge all three data files on w_hidp keeping only the matches.

We'd also point out that it is unorthodox to link individual level data such as indall, child and indresp on the household identifier. We wondered whether you wanted to add the parents' information to the child's record or vice versa?

If you haven't been on our online courses we are running an Introduction to Understanding Society next week 20-21 May and again in November, sign up here <https://www.understandingsociety.ac.uk/help/training>. Our online moodle is also available here <https://moodlex.essex.ac.uk/course/view.php?id=221>.

You can also browse previous Q&As in the user forum as well and read the FAQs on our website <https://www.understandingsociety.ac.uk/help/faqs>.

We also have information on our new online user guide <https://www.understandingsociety.ac.uk/documentation/mainstage/user-guides/main-survey-user-guide/>. We're always pleased to receive feedback on our user support materials. If you would like to send feedback on the new user guide, please email us at usersupport@understandingsociety.ac.uk.

I hope this helps.

Best wishes

Annette

#3 - 05/15/2020 03:46 PM - Laura Silva

Hi Annette,

thank you very much for the reply.

Just to be sure, even if unorthodox (why though?) is it correct? Which is the alternative in order to match info about parenting styles (eg i_homewk415, Parent helps their children with homework) which are available in the child file with parents?

Best wishes,

Laura

#4 - 05/15/2020 03:57 PM - Laura Silva

Moreover, I forget to say:

if I start from the child file, after I merge with the indall I have unmatched observations from the using, and that makes perfect sense to me (households without children I guess). The problem I fail to understand is when, after keeping only the matched, I merge with the indresp: I understand to have many not-matched from the using (individuals with no children) but I do not understand why I have 651 not matched from the master. Are they children without any matching parent?

Apologies for the split message.

Best regards,

Laura

#5 - 05/15/2020 04:25 PM - Annette Pasotti

- Assignee changed from Laura Silva to Gundi Knies

Dear Laura

Thank you for your reply. I am referring you to my colleague Gundi who will be able to explore this further with you.

Best wishes.

Annette

#6 - 05/18/2020 10:51 AM - Gundi Knies

- Category set to Data analysis

- Assignee changed from Gundi Knies to Laura Silva

- Target version set to X M

Laura,

if you look at the interview outcome variable _ivfio or _hhresp_dv, you can see that there are households in _indall data file that do not have any adult interviews, and some of them are households with members that are below age 16 and will therefore also be in the _child data file. That is why you have mismatches when merging _child, _indall, _indreasp data files.

Merging individual level records on _hidp is unorthodox because _hidp does not uniquely describe the rows in either data file. It is not 'wrong'; many routes lead to the same results just some seem more efficient than others. In our training course we have a worked example for matching spouses' records using the spouse identifier, and on this forum we have described how to match child and parent records.

For example, when we want to merge child records with their biological mothers's record, we'd first select only children who have a biological mother in the household (i.e. non-missing *mnpid*). Then rename the child *pidp* and other variables so they have a child prefix (e.g., *ch*), then rename the *mnpid* to *pidp*, then merge on *pidp* using the *_indresp* data file. This way it is very clear who the mismatches are: children (with biological mothers in the hh) whose biological mothers did not provide an interview. All variables that have the *ch* prefix refer to the child, all other variables refer to the biological mother.

Best,
Gundi

#7 - 05/19/2020 10:06 AM - Laura Silva

Hi Gundi,

thank you very much for your tremendous help, really appreciated.

Have an awesome day,

Laura

#8 - 05/21/2020 02:08 PM - Gundi Knies

- *Status changed from Feedback to Resolved*

- *% Done changed from 80 to 100*

You are welcome!

Best,

Gundi

#9 - 10/13/2021 11:43 AM - Understanding Society User Support Team

- *Assignee deleted (Laura Silva)*

#10 - 08/22/2023 01:28 PM - Understanding Society User Support Team

- *Category changed from Data analysis to Data management*