Understanding Society User Support - Support #1926

Parent-reported variables

06/22/2023 04:05 PM - Anna Dearman

Status:	Resolved	**Start date:**	06/22/2023
Priority:	Normal	**% Done:**	100%
Assignee:	Understanding Society User Support Team		
Category:	Data management		

Description

I would like to check whether I have merged some files correctly, please? My aim is to gather parent-reported variables relating to the respondents' children, and merge these with the children's own data.

I used the indall files to create a concatenated, wave-specific "household-person" ID using hidp and pno for the "children". I then went through every indresp file which contained variables such as pythh, pywhr, etc, and created three wave-specific "household-person" IDs using hidp and either pypno1, pypno2 or pypno3. I have merged the indall file with the indresp file using these variables (within each wave), on the assumption that the pypno IDs will identify the same individuals as the corresponding pno ID. Is this correct?

History

#1 - 06/23/2023 10:24 AM - Understanding Society User Support Team

- Category changed from Youth to Data management

- Status changed from New to Feedback

- % Done changed from 0 to 50

- Private changed from Yes to No

Hi Anna,

We provide some of this linking information so you don't need to do this yourself. In the w_indall files, for each person, there are variables available which are the PIDP of their bio/step/adop father and mother who are living in the same household: w_mnspid and w_fnspid. You can use these to link information of children to that of their parents. The syntax for doing this matching is available on our syntax pages (https://www.understandingsociety.ac.uk/documentation/mainstage/syntax). Click on the "UKHLS deposited syntax" tab and then search for "Matching co-resident parents' information"

Is this what you were looking for? If not, please let us know.

(Alita)
Best wishes,
Understanding Society User Support Team

#2 - 06/23/2023 10:43 AM - Anna Dearman

Hi Alita/team,

Thanks for getting back to me. This information will be useful for me later on, but I'm not sure it answers my specific question this time. I'm not a Stata user so I can't fully understand the syntax.

For the specific variables I'm interested in, it isn't sufficient to know who was co-resident, as the parents answer questions about up to three different children, each with separate variables (e.g. pythh1, pythh2 and pythh3). The child is identified using "person number" type variables pypno1, pypno2 and pypno3, and my question is whether these variables are the same as the pno in the indall file for the corresponding wave? If so, this will allow me to create a hidp-pno variable in indall and a hidp-pypno variable in indresp which will link together. The questions are asked in order of age of child, but I thought it might be nicer to link them using person number and household number, rather than using the children's ages/YOBs.

Best wishes,

Anna

#3 - 06/23/2023 03:19 PM - Understanding Society User Support Team

- % Done changed from 50 to 80

I see. What you are suggesting is correct.

These questions were asked only in the BHPS. Based on the questionnaires and the variable labels these are the PNOs of the 3 youngest children in the household. To check that and do what you are suggesting I wrote a small piece of Stata code - it seems to work. Please note that as more than

one parent may provide information about the same child, the hidp+pypno1 will not be unique. You will need to reshape the data to create multiple variables - one reported by one parent, and the other by the other parent. Here is the Stata code to do that. I know you said you use R but this is to provide an idea of the data management steps required.

```
use bd_indresp, clear
keep bd_hidp bd_pypno1 pidp bd_pythh1 // Here the pidp is the pidp of the adult reporting
bys bd_hidp bd_pypno1 (pidp): g num=_n // this creates a running counter of the rows with the same combination of hidp & pypno1
// Next reshape the data to produce a wide format file where each row is unique on bd_hidp bd_pypno1 but the variable bd_pythh1 & pidp will have more than one version if more than one parent reported about that child
reshape wide bd_pythh1 pidp, i(bd_hidp bd_pypno1) j(num)
// Next rename bd_pypno1 into bd_pno
rename bd_pypno1 bd_pno
// Then match it to indall. For this case I found 636 matches
merge 1:1 bd_hidp bd_pno using bd_indall
```

Best wishes,
Alita

#4 - 11/30/2023 01:10 PM - Understanding Society User Support Team

- Status changed from Feedback to Resolved

- % Done changed from 80 to 100