

Understanding Society User Support - Support #1044

Merging children to parents in Wave 2 of Understanding Society survey

09/10/2018 11:38 PM - Joseph Williams

| | | | |
|---|-----------------|--------------------|------------|
| Status: | Resolved | Start date: | 09/10/2018 |
| Priority: | Normal | % Done: | 100% |
| Assignee: | Joseph Williams | | |
| Category: | Data management | | |
| Description Hi, I am very new to Stata and I have been trying without success to merge young children to their parents using in the b_youth file and b_indresp file. I am aware that I can potentially use Example 3 under the section 'example Stata code for matching files' from the Understanding Society user guide : https://www.understandingsociety.ac.uk/sites/default/files/downloads/documentation/innovation-panel/user-guides/ip_user_guide.pdf , to get my result, but I am unsure how to tailor the inputs to get my desired result. Would it be possible for someone to give me a step by step instruction on how to do this. I have searched the forum and all similar questions with solved answers make large leaps in regards to knowledge of Stata. Thank you in advance, any assistance would be greatly appreciated. | | | |

History

#1 - 09/11/2018 09:04 AM - Gundi Knies

- Category set to Data analysis
- Status changed from New to Feedback
- Assignee set to Joseph Williams
- Priority changed from Immediate to Normal
- Target version set to M2
- % Done changed from 0 to 90

Dear Joseph,
please see this query. It has a step-by-step guide on how to link info child records with their mother's: <https://userswww.essex.ac.uk/support/issues/971>. As to the Stata code, the User Guide, Section 3.11 has a couple of very basic examples - the third one provides the code to link spouses' information using the variable _ppno. In your case, to link parents' info you want to use b_mnpno (to link mother info) or b_fnpno (to link father's info) on b_youth, do the renaming tasks and then link to b_indresp.
Hope htis helps,
Gundi

#2 - 09/11/2018 09:34 AM - Joseph Williams

Dear Gundi,

Thank you for the speedy response, following the instructions I tried to merge the data, but I am unsure if I did so successfully, as I did not reshape the data to wide and some of the responses do not make sense. Below is the code I used.

```
use "/Users/user/Desktop/dissertation/data/Understanding Society/UKDA-6614-stata/stata/stata11_se/us_w2/b_youth.dta",  
  
clear  
  
. renpfix b  
  
. keep _yp2uni _country _ypsex _yptvvidhrw _yptvvidhrs _mnpid  
  
. rename _mnpid pidp  
  
. sort pidp  
  
. save youth_moth  
file youth_moth.dta saved  
  
. use "/Users/user/Desktop/dissertation/data/Understanding Society/UKDA-6614-stata/stata/stata11_se/us_w2/b_indresp.dta"
```

```
. sort pidp
```

```
. merge 1:m pidp using "/Users/user/Desktop/dissertation/data/Understanding Society/UKDA-6614-stata/stata/stata11_se/us_w2/youth_moth.dta"  
(label b_country already defined)  
(label b_mnpid already defined)
```

| Result | # of obs. | |
|-------------|-----------|----------------------|
| not matched | 51,430 | |
| from master | 51,073 | (<u>_merge</u> ==1) |
| from using | 357 | (<u>_merge</u> ==2) |
| matched | 4,663 | (<u>_merge</u> ==3) |

When I look through the data, observations which are specific for children, e.g. whether they would like to go to higher education are given as observations for adults.

#3 - 09/11/2018 09:55 AM - Gundi Knies

Hi Joseph,
the suggestion is to rename all child-level variables of interest, including the child's pidp, so they are clearly marked out as the child's information before merging. Obviously, all variables that come from the b_indresp data file relate to adults, not to the child. For those with _merge==3, however, the dataset is at the level of mother-child pair. As you have not kept and renamed the child pidp you cannot see this (but if you had kept and renamed the youth's pidp to youth_pidp the command "duplicates report youth_pidp pidp" should return only zero duplicates as each mother-child pair is unique). To more easily stay on top of the multilevel nature of the information you could rename all data relating to the mother so they are clearly marked out. This is particularly true if you later on want to also add the father's information - his information cannot have the same variable name as the mother's.
Gundi

#4 - 09/11/2018 10:24 AM - Joseph Williams

Hi Gundi,

I tried renaming the variables, but seem to come out with the same outcome. Please see below the code I used:

```
. use "/Users/user/Desktop/dissertation/data/Understanding Society/UKDA-6614-stata/stata/stata11_se/us_w2/b_youth.dta",  
  
clear  
  
. renpfix b  
  
. keep _yp2uni _country _ypsex _yptvvidhrw _yptvvidhrs _mnpid _dvage  
  
. rename _yp2uni youth_yp2uni  
  
. rename _country youth_country  
  
. rename _ypsex youth_ypsex  
  
. rename _yptvvidhrw youth_yptvvidhrw  
  
. rename _yptvvidhrs youth_yptvvidhrs  
  
. rename _dvage youth_dvage  
  
. rename _mnpid pidp  
  
. sort pidp  
  
. save youth_mother1  
file youth_mother1.dta saved  
  
use "/Users/user/Desktop/dissertation/data/Understanding Society/UKDA-6614-stata/stata/stata11_se/us_w2/b_indr  
esp.dta"  
  
clear  
  
. keep b_ukborn b_jbsemp b_single_dv b_fimngrs_dv b_scend b_jbsoc00_cc b_tuin1 b_dvage pidp  
  
. rename b_ukborn mother_ukborn  
  
. rename b_scend mother_scend
```

```
. rename b_jbsemp mother_jbsemp

. rename b_tuin1 mother_tuin1

. rename b_dvage mother_dvage

. rename b_single_dv mother_single_dv

. rename b_jbsoc00_cc mother_jbsoc00_cc

. rename b_fimngrs_dv mother_fimngrs_dv

. sort pidp

. merge 1:m pidp using "/Users/user/Desktop/dissertation/data /Understanding Society/UKDA-6614-stata/stata/stata11_se/us_w2/youth_mother1.dta"
(label b_country already defined)
(label b_mnpid already defined)
(label b_dvage already defined)
```

```
Result                                     # of obs.
-----
not matched                             51,430
  from master                           51,073  (_merge==1)
  from using                             357    (_merge==2)

matched                                 4,663  (_merge==3)
-----
```

Was I supposed to rename the child's pidp aswell, I was unsure as I do not intend to use it as one of my key variables of interest. I am unsure what you meant by 'As you have not kept and renamed the child pidp you cannot see this (but if you had kept and renamed the youth's pidp to youth_pidp the command "duplicates report youth_pidp pidp" should return only one zero duplicates as each mother-child pair is unique).' Was I supposed to include youth's pidp in my variable of interest and rename it to youth_pidp. The part which states 'the command "duplicates report youth_pidp pidp" should return only one zero duplicates as each mother-child pair is unique' is what I am confused about, if I keep youth_pidp and run the command "duplicates report youth_pidp pidp" what exactly will happen.

Thank you again for the speedy response and apologies if you for think I am being dense, as explained earlier I am very new to Stata and am trying to understand why I am excluding or including certain variables.

#5 - 09/11/2018 11:17 AM - Gundi Knies

Joseph,
that is fine - the data structure is a bit complex.

Remember, that your dataset is now at the level of adults and because mothers may have more than one child responding to the youth questionnaire PIDP in your data does no longer uniquely identify rows in your data: you will have multiples of pidp in the data file for mothers with >1 youth respondent. The pidp of a mother with, say, 4 youth respondents will appear 4 times in your data file. Keeping the child's pidp or pno allows you to identify mother-child pairs. You may not need the youth's pidp for your analysis but it is good practice to keep the variables in the data set that uniquely identify its rows.

Do you have to rename the child's pidp: Yes. If you use pidp (a child's pidp) to link to pidp (an adult's pidp) in b_indresp you will not have any matches: pidp uniquely identify individuals and no individual can be both an adult respondent and a youth respondent at the same time. If you keep the child's pidp and do not rename it, you cannot rename the _mnpid pidp as pidp already exists (as the child's pidp).

Gundi

#6 - 09/11/2018 12:07 PM - Joseph Williams

Hi Gundi,

Thank you for the explanation, I believe i have now been able to merge the data successfully using the following code:

```
. use "/Users/user/Desktop/dissertation/data/Understanding Society/UKDA-6614-stata/stata/stata11_se/us_w2/b_youth.dta",clear

. renpfix b

. keep _yp2uni _country _ypsex _yptvvidhrw _yptvvidhrs _mnpid _dvage pidp

. rename _yp2uni youth_yp2uni

. rename _country youth_country

. rename _ypsex youth_ypsex

. rename _yptvvidhrs youth_yptvvidhrs

. rename _yptvvidhrw youth_yptvvidhrw
```

```
. rename _dvage youth_dvage

. rename _mnpid youth_mnpid

. rename pidp youth_pidp

. rename youth_mnpid pidp

. sort pidp

. save youthmother
file youthmother.dta saved

. use "/Users/user/Desktop/dissertation/data/Understanding Society/UKDA-6614-stata/stata/stata11_se/us_w2/b_indresp.dta", clear

. keep b_ukborn b_jbsemp b_single_dv b_fimngrs_dv b_scend b_jbsoc00_cc b_tuin1 b_dvage b_sex pidp

. rename b_ukborn mother_ukborn

. rename b_scend mother_scend

. rename b_jbsemp mother_jbsemp

. rename b_tuin1 mother_tuin1

. rename b_dvage mother_dvage

. rename b_single_dv mother_single_dv

. rename b_jbsoc00_cc mother_jbsoc00_cc

. rename b_fimngrs_dv mother_fimngrs_dv

. rename b_sex mother_sex

. sort pidp

. merge 1:m pidp using "/Users/user/Desktop/dissertation/data/Understanding Society/UKDA-6614-stata/stata/stata11_se/us_w2/youthmother.dta"
(label b_country already defined)
(label b_mnpid already defined)
(label b_dvage already defined)
```

| Result | # of obs. | |
|-------------|-----------|----------------------|
| not matched | 51,430 | |
| from master | 51,073 | (<i>_merge</i> ==1) |
| from using | 357 | (<i>_merge</i> ==2) |
| matched | 4,663 | (<i>_merge</i> ==3) |

```
. sort pidp

. save mergedmotherchild
file mergedmotherchild.dta saved
```

The data actually seems to make sense now following your explanation. However, I did have some questions regarding the setup. Because I only want mother's data would it make sense to drop the gender female only. I noticed for the *_merge*==3 cases they are all female but I have data on *_merge*==1 that are both male and female. Also in regards to adding father merged with their child's data I understand I would do the same steps but how would I merge the mother and father data.

Thank you again Gundi for all your assistance and help in this matter, it has been greatly appreciated.

#7 - 09/11/2018 12:26 PM - Joseph Williams

Hi Gundi,

I also had a question about the ethnic minority boost sample, how would one go about removing this sample ?

Thank you again,

#8 - 09/11/2018 01:31 PM - Gundi Knies

Hi Joseph,

to decide which cases to keep you need to understand who the mismatches on *_mnpid* on *b_youth* and *pidp* on *b_indresp* are. Some of these will be

mothers but their children have not participated in a youth interview (some of the cases with _merge==1). Likewise, some mothers will not have participated in an adult interview (_merge==2). If your population of interest is mother-child pairs then you keep only _merge==3 if it is another population you may want to do something else. Only you can decide what is appropriate as depends on your research question.

Your research question will also guide which is the most efficient/"correct" way to link the father's info. If you are interested in a child-mother-father-level dataset, you might want to create the mother-child level dataset and the father-child level datafile separately and then merge the two together using the youth_pidp.

As to removing the ethnic minority sample, we strongly recommend not to remove cases from any design samples but to use the appropriate population weights for your analysis as this is will correct for unequal selection and response probabilities. Dropping the boost samples will not do this trick. The user guide section on sampling and weighting is a good place to read up on this.

Gundi

#9 - 09/11/2018 01:39 PM - Joseph Williams

Dear Gundi,

Thank you again for all your help in this matter, my research question is based on a child-mother-father-level dataset, so I think I will follow your advice on creating a mother-child and then father-child separately then merge them together.

With regards to the merge would it be a 1:1 merge using the youth_pidp variable as each child only has one mother or father.

Thank you again for all your assistance.

#10 - 09/11/2018 01:57 PM - Gundi Knies

Why not just try and see! If it does not work out as a 1:1 match, it may be useful to think again about which variable(s) uniquely identify observations in the specific datasets you are trying to merge. :)

G

#11 - 09/11/2018 02:35 PM - Joseph Williams

Thank you Gundi,

I will let you know how I get on with the merging, your input and explanations have been invaluable.

#12 - 09/11/2018 04:57 PM - Stephanie Auty

Dear Joseph,

I just wanted to add that we run a training course here at Essex and online, which could be helpful for you to get to know our data and some more data management techniques. <https://www.understandingsociety.ac.uk/help/training>

Best wishes,
Stephanie

#13 - 10/12/2018 04:21 PM - Stephanie Auty

- Status changed from Feedback to Resolved

- % Done changed from 90 to 100

#14 - 08/22/2023 01:07 PM - Understanding Society User Support Team

- Category changed from Data analysis to Data management