Understanding Society User Support - Support #1196

Adapting syntax from website ("Matching individuals within a household") - queries

06/06/2019 12:36 PM - fabiana macor

Status:	Resolved		**Start date:**	06/06/2019
Priority:	High		**% Done:**	100%
Assignee:	fabiana macor			
Category:				

Description

Hi there

I'm using the syntax from Understanding Society website called "Matching individuals within a household" (attachment 1) - https://www.understandingsociety.ac.uk/documentation/mainstage/syntax. I have two questions.

> 1. I have used UKHLS before and have used the 'indresp' datafile(s) in each case. The syntax file above uses the 'indall' datafile instead; which has less information. My question is: can I replicate the steps for 'indresp'?

> 2. I then (after linking personal and spouse info) want to pool waves 1-8 together. For this I have previously used syntax from file "Merging individual files across waves into long format" (attachment 2). The question here is: after doing step 1, can I simply follow the steps from this second syntax document?

The two questions relate to the two steps (part 1 and part 2) of the attached do-file (attachment 3), where I've tried carry them out.

I would be extremely grateful for any thoughts on the process: on the above questions and the resulting do-file (attachment 3).

I am marking the query as urgent in case any part of my question is unclear. I hope this is OK and I appreaciate your assistance on the matter.

Kind regards
Fabiana

History

#1 - 06/06/2019 08:06 PM - Stephanie Auty

- Status changed from New to Feedback

- Assignee set to fabiana macor

- % Done changed from 0 to 70

- Private changed from Yes to No

Dear Fabiana,

The reason w_indall is used in the example for matching individuals within a household is that w_indall contains basic information on all of the individuals within a household, while w_indresp only contains those who completed a full interview or had a proxy interview completed for them. You can use this process with w_indresp but you will have fewer matches.

The answer to your second question is yes, you can use the syntax to create a long file format after the first step. You just need to be aware that in the files you have created you will have one row per individual, that is two rows per couple, so depending on how you use the data you may have a double counting issue.

Also, in your syntax you seemed to be asking what `w' represents. It represents each of the values you are using in the loop, so if you write foreach w in a b c d {...} then the first time `w' will represent a, the second time b and so on.

Best wishes,
Stephanie

#2 - 06/06/2019 09:40 PM - fabiana macor

fabiana macor wrote:

> Hi there
>
> I'm using the syntax from Understanding Society website called "Matching individuals within a household" (attachment 1) -

https://www.understandingsociety.ac.uk/documentation/mainstage/syntax. I have two questions.

    1. I have used UKHLS before and have used the 'indresp' datafile(s) in each case. The syntax file above uses the 'indall' datafile instead; which has less information. My question is: can I replicate the steps for 'indresp'?

    2. I then (after linking personal and spouse info) want to pool waves 1-8 together. For this I have previously used syntax from file "Merging individual files across waves into long format" (attachment 2). The question here is: after doing step 1, can I simply follow the steps from this second syntax document?

The two questions relate to the two steps (part 1 and part 2) of the attached do-file (attachment 3), where I've tried carry them out.

I would be extremely grateful for any thoughts on the process: on the above questions and the resulting do-file (attachment 3).

I am marking the query as urgent in case any part of my question is unclear. I hope this is OK and I appreaciate your assistance on the matter.

Kind regards
Fabiana

Stephanie Auty wrote:

Dear Fabiana,

The reason w_indall is used in the example for matching individuals within a household is that w_indall contains basic information on all of the individuals within a household, while w_indresp only contains those who completed a full interview or had a proxy interview completed for them. You can use this process with w_indresp but you will have fewer matches.

The answer to your second question is yes, you can use the syntax to create a long file format after the first step. You just need to be aware that in the files you have created you will have one row per individual, that is two rows per couple, so depending on how you use the data you may have a double counting issue.

Also, in your syntax you seemed to be asking what `w` represents. It represents each of the values you are using in the loop, so if you write foreach w in a b c d {...} then the first time `w` will represent a, the second time b and so on.

Best wishes,
Stephanie

Hi Stephanie

Thank you for the fast reply, that all makes sense. I just have one more query.

Regarding the potential double-counting issue/two rows per couple: thank you for pointing this out. I do actually need the partner and main respondent info on the same row for the long file. Is there a way that this can be done?

Would the following work, for example?
1. Create a file with partner info from the original level datafile (as with Step 1 of previous attachment 3) - e.g. call this 'sp1'
2. Turn 'sp1' file into long format - e.g.call this 'long1'
3. turn individual level datafile into long format - e.g. call this 'long2'
4. put long1 and long2 together
If this is correct, what is the best way to merge long1 and long2?

Many thanks in advace.

Fabiana

#3 - 06/07/2019 11:26 AM - Stephanie Auty

- Status changed from Feedback to In Progress

Dear Fabiana,

The steps in syntax for matching partner info does put the partner and main respondent info on the same row, but if we say that pno 1 and 2 within a household are partners, there will be a row where pno 1 is the main respondent and pno 2 is the partner, and another row where pno 2 is the main respondent and pno 1 is the partner.

If you just want to keep one of those rows then you will need to decide which one to keep. It will not be possible to keep for example rows where the main respondent is male and the partner is female as this will not work for same sex couples. You could use age or something else that is useful for your research. If you need help with this then let me know which way you would like it to work.

Best wishes,
Stephanie

#4 - 06/07/2019 12:00 PM - fabiana macor

Stephanie Auty wrote:

> Dear Fabiana,
>
> The steps in syntax for matching partner info does put the partner and main respondent info on the same row, but if we say that pno 1 and 2 within a household are partners, there will be a row where pno 1 is the main respondent and pno 2 is the partner, and another row where pno 2 is the main respondent and pno 1 is the partner.
>
> If you just want to keep one of those rows then you will need to decide which one to keep. It will not be possible to keep for example rows where the main respondent is male and the partner is female as this will not work for same sex couples. You could use age or something else that is useful for your research. If you need help with this then let me know which way you would like it to work.
>
> Best wishes,
> Stephanie

Dear Stephanie

Thank you ever so much for explaining in detail as it has allowed me to better visualise the data .
Because I am focusing on heterosexual couples (main female + male as spouse), I will proceed by keeping rows where main respondent is female and the partner is male.
Thank you for your thoughts and solution.

Kind regards
Fabiana

#5 - 06/07/2019 01:55 PM - Stephanie Auty

- Due date deleted (06/06/2019)

- Status changed from In Progress to Resolved

- % Done changed from 70 to 100

Files

stata-merging-individual-files-across-waves-long-format (3).do	1.54 KB	06/06/2019	fabiana macor
stata-matching-individuals-within-household (1).do	2.3 KB	06/06/2019	fabiana macor
stata-matching-individuals-within-household - do file from UKHLS website adapted for T.do	4.42 KB	06/06/2019	fabiana macor