Understanding Society User Support - Support #537

Merged File Request

03/29/2016 05:55 PM - zoe young

Status:	Closed		**Start date:**	03/29/2016
Priority:	Normal		**% Done:**	100%
Assignee:	zoe young			
Category:	Data analysis			

Description

I am wondering if there is a merger file or a way to obtain a merged file for waves 1-5 including the files _egoalt and _youth for wave a-e?

thanks
zoe

History

#1 - 04/04/2016 08:58 AM - Alita Nandi

- Assignee set to zoe young

- % Done changed from 0 to 50

Hi Zoe,

Which software are you using for Data Management? We offer guides to merging data files across many waves for Stata, SAS and SPSS. If you are using Stata or SPSS you can register for the relevant online course and in those courses there is a module titled "Merging data into wide and long formats" that walks you through this and shows you how to do this efficiently and quickly. To find out more and register for the course: https://www.understandingsociety.ac.uk/documentation/training/online
The SAS material is available here:https://www.understandingsociety.ac.uk/documentation/training/online/sas

But given the files you have asked for - Egoalt and Youth - I have to warn you that these are at different levels. Each row in the youth file is uniquely identified by pidp while each row in the egoalt file is identified by pidp apidp. Each row of indresp and indall files are also uniquely identified by pidp but while indall includes all individuals in an enumerated household including youths, indresp only includes adults. So, the individuals in the youth and the indresp files are a subset of the individuals in the indall file and there is no overlap of individuals between the youth and indresp. Thus we would recommend against a general merge of all these different files. If you tell us what it is you are trying to achieve we can tell you how to most efficiently merge the files.

Best wishes,
Alita

#2 - 04/04/2016 01:02 PM - zoe young

Hi alita,
I am trying to examine those who experienced care versus those who have not by examining educational attainment and current economic status. I have decided to recode the relationship_dv variable so that those who identify as being in foster/adopt and those who identify as children foster adopted are the experienced care while everyone else is not experienced care. I then want to use the indresp file to examine the rest of the variables that I identified.

I wanted to look longitudinally at the 15 people that can be traced from wave1-wave5
Can you advise the best way to merge either the relationship variable or the a_lvag14 from wave 1 to wave 5.
I have looked at the Indall file, if i used that i would need to merge variables from e_indresp still and the ego_alt.

Would it be easier to use the a_lvag14 and a_lvag16 and merge it to e_indresp?

thanks
zoe

#3 - 04/04/2016 03:46 PM - Alita Nandi

About whether you should use lvag14 or relationship_dv - depends on who you are analysing. relationship_dv tells you about the current relationships between household memmbers while lvag14 tells you who the adults in the sample (16+ above) were living with at age 14. From the first paragraph of your latest post it seems like you are interested in the current living conditions. In that case using relationship_dv is ok. If you are using the relationship_dv variable please note that this variable represents the relationship of the EGO to ALTER. So, make sure you identify the foster/adopted children and not the foster/adoptive parents. Also note that data from indresp is only available for adult respondents (16+ years). So many of those identified as currently living with their foster or adopted parents using relationship_dv in egoalt will be <16 years and so will not have any data in indresp. In the online course I mentioned you will also find a module devoted to helping users understand the structure of the EGOALT files and how to use it to identify different relationships.

General rules of merging files. If you want to merge data from the same set of files across different waves then first merge these files separately for each wave and save those files. Then simply append those files. You can make this more efficient by using Stata's foreach loops. To know more about this technique see the course material I suggested in my last post.

Best wishes,
Alita

#4 - 04/06/2016 12:59 PM - zoe young

Thank you for the above information. I have examined the EGOALT and I have a question regarding pidp and apidp variables. I was under the impression that the crosswave identifier was the pidp variable. However when examining the _egoalt file it appears that this is not the case that in fact there are two the pidp and the apid are both crosswise identifiers, as the pidp is the EGO identifier and the apidp (which needs to be uses in conjunction with the gender variable) is the ALTER.

1. Assuming the above is correct...I am looking at those who identified as being in local authority care at the age of 14 (a_lvag14). Given this in order to merge this variable with the e_indresp file do I need to merge it once on pidp and again on apidp?

2. If I examine the relationship_dv variable and recode it so that 5-6 and 17-18 is (1) meaning experienced some form of care and (0) is 1-4,7-16,19-30 Not experienced care, when merging this file to the e_indresp file would i need to again merge once with pidp and again with apidp?

3. I have noticed from previous communication the _indall file, however since I could only identify 15 youth that could be examined longitudinally without BHPS permission this meant that for the added complexity, not real value would be added and I have decided to examine those who reported to be "in-care" (my created variable from a_lvag14) only in the _indresp and look at the NS-Sec as well.

Thank you for the advice and look forward to your response. Also just so you are aware I am using SPSS for analysis not STATA, I am assuming it is possible to do the above using SPSS, please advise if this is not the case.

#5 - 04/07/2016 02:22 PM - zoe young

Hi Alita,
I have further examined the EGOALT file and believe that I can not use this file for the intended purpose. Please can you correct me if I am wrong but the INDRESP file contains 77,320 and the EGOALT contains 182,802. I am assuming this is due to the fact that the INDRESP only looks at those 16+, while is why you suggested the INDALL file.

However, i intend on using the INDRESP file as explained in point 3 above. Given this is there a way to merge the EGOALT file with the INDRESP file? I have looked at model however they suggest identifying the key variable relationships, which I can recode the relationship to 1 in care and 0 no in care but then I am confused with how I can examine the once I merge the files to the INDRESP file.

I apologise if it is not clear it is very difficult to explain this via text.

thanks
zoe

#6 - 04/12/2016 04:55 PM - Gundi Knies

- File ukhls_support_20160412_zoe_young.do added

- Category set to Data analysis

- Status changed from New to In Progress

- Priority changed from Immediate to Normal

- Target version set to X M

- % Done changed from 50 to 90

Hi Zoe,
you are right; it is very difficult to explain all this in text. The process of identifying those in care/foster children is not that difficult once you understand the data structure and how to tag observations with certain characteristics, and how to merge files (which we describe in the training course materials).

I am not familiar with SPSS but I have written Stata code to identify all those respondents in Waves 1-5 enumerated households that have ever lived in a household where they were somebody else's foster child (w_relationship_dv=6 in data file w_egoalt) or who reported, at their interview in Wave 1, that they lived in care when they were aged 14 (a_lvag14=7 in data file a_indresp). Looking at the cross-wave response pattern, across Waves 1-5, you will have 492 such individuals overall, including children and non-interviewed individuals (see w_ivfio on data file w_indall). For some of them you will have data from their interviews: 351 interviews with adults (w_ivfio==1) and 63 interviews with youths (w_ivfio==21). [For current foster children aged 0-15 there may also be some information on their outcomes from interviews with the adult responsible for them in the child data file.]

The steps involved in putting such a data set together are as follows:

1. load a_egoalt data file, keep only observations with a_relationship_dv=6, keep only pidp, removing duplicate observations. Save, and repeat for all other waves.
2. append data files created in (1.), removing duplicates. Save.
3. use a_indresp, keep if a_lvag14==7, keep only pidp. merge on pidp using data file created in step (2.). remove duplicate pidps. The resulting file has all the pidp for everybody who ever reported to be in care/foster child. Save.
4. Create a long format file of indall files for Waves 1-5. Save. Merge on pidp using data file created in step (3.). Those who were in the data file

you added will be individuals with foster/care experience. Create a dummy to tag them.
5. You can now use the pidp to merge in information on from the indresp, [child] or youth data files depending on what it is that you want to look at specifically.

We hope this helps you with the data management task. We cannot advise on your specific research question/sample selection but hope that the indicator of who experienced care or not helps you to get started.

Best wishes,
Gundi

#7 - 04/14/2016 04:55 PM - zoe young

Hi Gundi,
Can you please tell me where I can get UKHLScombine.sps?

thanks
Zoe

#8 - 04/14/2016 05:17 PM - Alita Nandi

It has now been uploaded and you will find it in the online Moodle course "Introduction to Understanding Society Using SPSS" module "Merging data into wide and long formats".
Best wishes,
Alita

#9 - 04/14/2016 06:18 PM - zoe young

- File indallmerge.sps added

Hi Alita,
I am having trouble running the macro. I keep getting line 45 REN Var - Undefined variable name. I have attached the sass syntax for you to look at.
I changes the a b c d to temp_a temp_b temp_c temp_d temp_e and I am still getting the error. Any advice would be appreciated.

#10 - 04/15/2016 11:39 AM - zoe young

Hi Alia,
I have tried throughout the night and I still get an error message when I run the syntax, however the files get produced, but incorrectly as the waves are not identified.
Can you please advice, please see the above message with the attached file.

thanks
zoe

#11 - 04/15/2016 12:37 PM - Alita Nandi

Hello Zoe,

The macro ukhlscombine allows you to combine different wave indresp files. You are using indall files. I think that is why your syntax is not working. If you first modify the macro by replacing indresp with indall and then run your syntax, it should work.

For the future we will modify the macro so that it allows you to choose the file (and you won't have to change the macro every time you use a different file).

Best wishes,
Alita

#12 - 04/15/2016 01:23 PM - zoe young

Hi Alita
I have already changed the macro to say indall as instructed in the documents to merge multiple files

Is there anything wrong with the script?
It' pulls on the line which uses the macro after the insert and says it doesn't exist or the other error is a ren nam error.
I know it's not correct what it produces. Is there anyway you can merge the indall files for wave 1-5 for me and send me the file? I'm only interested in pidp age corrected sex corrected race government geographical location

Thanks
Zoe

#13 - 04/15/2016 02:42 PM - Alita Nandi

I am sorry Zoe we are not allowed to send these data directly. You should be able to do this appending without using the macro. The macro only reduces the amount of syntax you have to write.

At the beginning of the Example 3 syntax file it shows you how to put wave a and wave b files into a long format file called ab_long. Just repeat that

syntax for all five waves and combine them using the append command as shown.

Best wishes,
Alita

#14 - 04/18/2016 01:54 PM - zoe young

- File abcmerge.sps added

- File abmerge.sps added

Hi Alita,
I am having trouble still with the merge. I have been able to do the merge for a and b but when i try to do c d and e i keep getting an error. it seems to work up until line 35 as when i run the select if i get a file with NO values. i have attached the syntax for the abmerge and then the syntax for the abcmerge. if you could advise i would appreciate it

thanks
zoe

#15 - 04/18/2016 09:42 PM - zoe young

Hi Gundi
Re your message:

I am not familiar with SPSS but I have written Stata code to identify all those respondents in Waves 1-5 enumerated households that have ever lived in a household where they were somebody else's foster child (w_relationship_dv=6 in data file w_egoalt) or who reported, at their interview in Wave 1, that they lived in care when they were aged 14 (a_lvag14=7 in data file a_indresp). Looking at the cross-wave response pattern, across Waves 1-5, you will have 492 such individuals overall, including children and non-interviewed individuals (see w_ivfio on data file w_indall). For some of them you will have data from their interviews: 351 interviews with adults (w_ivfio==1) and 63 interviews with youths (w_ivfio==21). [For current foster children aged 0-15 there may also be some information on their outcomes from interviews with the adult responsible for them in the child data file.]

Can you tell me what files you used to find those numbers (493 overall (i am assuming use merged the file you created with indall??) I couldnt follow which files you merged the overall foster file (created) with to find those numbers. Also can you explain how to Create a dummy to tag the foster in as you suggested below:
Create a long format file of indall files for Waves 1-5. Save. Merge on pidp using data file created in step (3.). Those who were in the data file you added will be individuals with foster/care experience. Create a dummy to tag them.

Thanks
Zoe

#16 - 04/19/2016 09:09 AM - Alita Nandi

- Private changed from Yes to No

#17 - 04/19/2016 04:29 PM - Victoria Nolan

Dear Zoe,

Many thanks for your email.

The status of all posts is changed to public after we have verified that there is nothing in the post that refers to any Understanding Society data that can only be released under special license. As there was no such issue with your post we should have set the status to public earlier in the process. We have just realised this and so changed the status. We will be responding to you about your query iteself soon.

Best wishes, Victoria.

#18 - 04/20/2016 07:07 AM - Alita Nandi

Dear Zoe,

Our remit at the User Forum is to answer queries related to Understanding Society data and provide general advice about how to manage the data. Given the number of users we have I'm afraid we cannot advise on individual users' analysis syntax specifically.

We provide online and in person training to use the data set and for setting up datasets for different kinds of analysis. See here:
https://www.understandingsociety.ac.uk/documentation/training

Attending a course in person or via moodle may help you understand the data more effectively to undertake your analyses.

We do not provide training in specific statistical software or statistical methods but there are a wide range of courses available - NCRM provide a use list of course held across the country http://www.ncrm.ac.uk/training/.

Best wishes,
Alita

#19 - 04/23/2016 07:06 PM - zoe young

Hi Alita,
Thank you for you help.
zoe

#20 - 04/26/2016 03:03 PM - zoe young

Hi Alita,
I have been trying to do the steps Gundi mentioned was done in stata in spss. Is there a possibility of getting these instructions for spss users? I have attempted to do this in spss, but the data file is not coming out correctly. Below are the instructions given for stata users.

the steps involved in putting such a data set together are as follows:

```
load a_egoalt data file, keep only observations with a_relationship_dv=6, keep only pidp, removing duplicate o
bservations. Save, and repeat for all other waves.
    append data files created in (1.), removing duplicates. Save.
    use a_indresp, keep if a_lvag14==7, keep only pidp. merge on pidp using data file created in step (2.). re
move duplicate pidps. The resulting file has all the pidp for everybody who ever reported to be in care/foster
 child. Save.
    Create a long format file of indall files for Waves 1-5. Save. Merge on pidp using data file created in st
ep (3.). Those who were in the data file you added will be individuals with foster/care experience. Create a d
ummy to tag them.
    You can now use the pidp to merge in information on from the indresp, [child] or youth data files dependin
g on what it is that you want to look at specifically.
```

#21 - 04/29/2016 05:42 PM - zoe young

Hi
I have managed to finally perform steps 1 and 2. In case other people read this feed you need to:
ADD FILES for merging the data files and DROPping unneeded variables
-SORT CASES, COMPUTE with LAG for identifying duplicates and
-CASETOVARS option.

However, I have only managed to locate 452 individuals while you have stated you have found 492 and of that 63 were youth, where i only find 1 youth. Can you please advise?

thanks
zoe

#22 - 05/03/2016 07:50 PM - Gundi Knies

Hi Zoe,
as we have said before, it is bejond the remit of this support forum to work through individual user's code with them. I was hoping that it would be straightforward to create the indicator in SPSS following the steps in Stata and consulting the Stata code. But this does not appear to be the case (which is fair enough as different programmes and users prefer different strategies!). Having said that, while other users on here may be able to help, a specialist SPSS forum (or a colleague with whom you could go over your code?) may come up with a strategy that is more intuitive in SPSS and may provide you with a speedier response.

Hope it works out for you,
Gundi

#23 - 05/04/2016 02:44 PM - Gundi Knies

- File UKHLS_UserSupport_20160504_Zoe.pdf added

- Status changed from In Progress to Resolved

- % Done changed from 90 to 100

Further to my last post, it has puzzled me why your interpretation of Stata code does not yield the same number of inviduals with experience of care in SPSS as does my Stata code. Hence I have taken the liberty to attempt my own SPSS interpretation. It has taken me a while as I have never used SPSS before, but my code produces a data file with the pidp of 492 individuals who have experienced care. I have used our Understanding Society SPSS training course materials (Worksheet 3 provides examples how to merge files, Worksheet 8 includes code to identify duplicates) and no other aids to work this out.

Gundi

#24 - 05/04/2016 06:52 PM - zoe young

Hi Gundi,
can you please post the m_rename.sps macro for exercise 8? i am reworking through the exercises.

thanks
zoe

#25 - 05/04/2016 11:01 PM - Alita Nandi

Thanks for letting us know that m_rename is not available in the Moodle course. It has now been uploaded. Check the folder for Ex3.

Just to clarify - you can use the SPSS syntax without the macros. If there are repetitive tasks you can use macros which basically means you write one set of syntax and run it many times with the help of loops. It saves time, effort and reduces the probability of error when writing code. But you don't need to use them - they just make life easier. Suppose you want to open the file w_indresp for all five waves. You can use a macro or you can do the following:
GET FILE='datadir\a_indresp.sav'.
GET FILE='datadir\b_indresp.sav'.
GET FILE='datadir\c_indresp.sav'.
GET FILE='datadir\d_indresp.sav'.
GET FILE='datadir\e_indresp.sav'.

Best wishes,
Alita

#26 - 05/17/2016 02:46 PM - Victoria Nolan

- Status changed from Resolved to Closed

Files

ukhls_support_20160412_zoe_young.do	1.86 KB	04/12/2016	Gundi Knies
indallmerge.sps	2.96 KB	04/14/2016	zoe young
abcmerge.sps	2.35 KB	04/18/2016	zoe young
abmerge.sps	2.53 KB	04/18/2016	zoe young
UKHLS_UserSupport_20160504_Zoe.pdf	4.81 KB	05/04/2016	Gundi Knies