Understanding Society User Support - Support #25

Merge the files

02/05/2012 11:55 PM - Anita Staneva

Status:	Closed		**Start date:**	02/05/2012
Priority:	Normal		**% Done:**	80%
Assignee:	Redmine Admin			
Category:	Data analysis			

Description

Dear Jakob,
I merged all the files: (hhsamp+hhresp+a_empstat+indresp+a_income+a_marriage+a_indall+a_egoalt+a_child) in one resulting file.
Are all these files designed to be merged in this way?
However, when I did try to set the resulting file by pidp and a_isyear and get some descriptive stat,
I have repeated time values within panel error
r(451);
Could you help me please with some advice here?
Best
Anita

History

#1 - 02/06/2012 11:22 AM - Redmine Admin

- Category set to Data analysis

- Status changed from New to In Progress

- Assignee set to Redmine Admin

- % Done changed from 0 to 80

So far we have released the data from the first wave of interviews carried out 2009-2010.
When will future waves be released? The data from the full Wave 2 (data collected 2010-2011) will be released by the end of this year, but the data from households interviewed during 2010 are due for release very soon – hopefully by the end of this month.
The Wave 1 data that you've got can in the meantime be exploited for cross-sectional analyses.
To see examples of how you can set the data up for longitudinal analysis, I suggest you have a look at the BHPS course notes. The BHPS data structure is similar and there are now 18 waves worth of data.
hth Jakob
PS If you've got a follow-up question simply click the "update" button on this page to continue the thread
Related issue: #21

#2 - 02/06/2012 01:29 PM - Anita Staneva

Thank you very much for your response. If it is a cross section each observation should refers to a different individual (different pidp). In this case what should be the total approximate number of individual records?
When I merge the a_hhsamp+ a_hhresp files, and then merge with the a_empstat file, however what I can see is individuals within the sample pidp report in the employment history more than 1 records, so for the same individual I have more than one observation. Therefore I decide I can exploit the panel element of the data, but maybe I am wrong.

#3 - 02/06/2012 03:40 PM - Redmine Admin

NB some files are in a long format with several lines for each individual to report on events, e.g. employment spells in the last year;

use a_empstat, clear
isid a_hidp a_pno a_spellno // Check for unique identifiers

They are, in other words, different from 'waves' - a fact that you would have to consider when you design your study.

hth
Jakob

#4 - 02/07/2012 11:23 AM - Anita Staneva

So shall I reshape the empstat file before merging it with the other files and if so what should be the code. I did try with the following:
reshape wide varlist, i(a_hidp) j(a_spellno),
However the a_spellno is not unique within a_hidp; there are multiple observations at the same a_spellno within a_hidp.

#5 - 02/09/2012 09:35 AM - Redmine Admin

Anita, Here is an example of how you could reshape a data set from long to wide (I've use the marriage file instead for simplicity - most people go through fewer marriages than jobs!)

```
EXAMPLE: reshape year-of-marriage from long to wide
*********************************************************************
use a_hidp a_pno a_marno a_lmary4 using a_marriage, clear
isid a_hidp a_pno a_marno // to confirm unique ID
sort a_hidp a_pno a_marno
l in 1/10,sepby(a_hidp a_pno) noobs // snippet of the long data set
ds a_hidp a_pno a_marno,  not // create list of variables to flatten
reshape wide `r(varlist)', i(a_hidp a_pno) j(a_marno)
l in 1/10,sepby(a_hidp a_pno) // snippet of the wide data set
*********************************************************************
```

Jakob

#6 - 02/23/2012 01:38 PM - Redmine Admin

- Status changed from In Progress to Closed

#7 - 11/10/2015 03:07 PM - Gundi Knies

- Target version set to M1

```
EXAMPLE: reshape year-of-marriage from long to wide
*********************************************************************
use a_hidp a_pno a_marno a_lmary4 using a_marriage, clear
isid a_hidp a_pno a_marno // to confirm unique ID
sort a_hidp a_pno a_marno
l in 1/10,sepby(a_hidp a_pno) noobs // snippet of the long data set
ds a_hidp a_pno a_marno,  not // create list of variables to flatten
reshape wide `r(varlist)', i(a_hidp a_pno) j(a_marno)
l in 1/10,sepby(a_hidp a_pno) // snippet of the wide data set
*********************************************************************
```