

Understanding Society User Support - Support #398

Identifying drop outs and re joiners

08/06/2015 02:54 PM - Gretta Mohan

Status:	Closed	Start date:	08/06/2015
Priority:	Normal	% Done:	100%
Assignee:	Redmine Admin		
Category:	Data analysis		
Description			
It feels like a basic question but I cannot identify respondents of the BHPS and Understanding Society who drop out year on year and those who re-join. I do not think there is a variable to identify these people. I am using stata and the data is in long format. I want to do an attrition analysis to see if attrition is health related - this drop out, re-joining attrition health analysis has been done by Jones, Koolman and Rice (2005). Many thanks			

History

#1 - 08/11/2015 04:08 PM - Redmine Admin

- Category set to Data analysis
- Assignee set to Gretta Mohan
- Target version set to X M
- % Done changed from 0 to 50

You may find the file, XWAVEID, useful. It holds the interview outcomes (ivfio/ivfho) for all sample members and all waves in a wide format. Let us know if you have got any further queries.
On behalf of the team,
Jakob

#2 - 08/12/2015 04:43 PM - Gretta Mohan

I have the variable(s) ivfio from xwaveid merged into my master dataset (indresp) in both long and wide formats.

In terms of when I change the xwaveid data into the long format and merge it to the master (so that the variables b_ivfio, c_ivfio, d_ivfio become a single variable) what I just get when I tabulate wave and ivfio is just full, proxy and telephone interviews.
e.g.

tabulate wave ivfio

gives us:

wave of BHPSUS	full inte	proxyint	Total
W1 2010/11	1,964	54	2,018
W2 2011/12	1,805	62	1,867
W3 2012/13	1,700	38	1,738

On the other hand when I simply merge in the xwaveid data (in its wide format) to the master and i tabulate wave by each of the ivfio's e.g. tabulate wave b_ivfio , I get for the current wave (_b) the full and proxy interviews as well as subsequent waves with other reasons e.g

tabulate wave b_ivfio (wave b is wave one)

gives us:

wave of BHPSUS	full inte	proxy int	refusal	other non	moved	ill/away	youtint youth	non refusal/n	Total	
W1 2010/11	1,964	54	0	0	0	0	0	2,018		
W2 2011/12	1,765	16	6	24	3	5	31	4	9	1,868
W3 2012/13	1,605	14	5	25	1	3	67	7	5	1,736

I dont really understand what the ones in the wide format mean - in wave 2, the wave b variable has 1,765 full interviews, 6 refusals etc.

Getting to the point.... I want to identify those who drop out and those who rejoin.
How do I use this ivfio variable to do that? I think I use the long formatted merge? Then to get the drop outs what do I do?

Many thanks,
Gretta

#3 - 08/12/2015 05:25 PM - Gundi Knies

Gretta Mohan wrote:

It feels like a basic question but I cannot identify respondents of the BHPS and Understanding Society who drop out year on year and those who re-join. I do not think there is a variable to identify these people.
I am using stata and the data is in long format.

I want to do an attrition analysis to see if attrition is health related - this drop out, re-joining attrition health analysis has been done by Jones, Koolman and Rice (2005).

Many thanks

Hi Gretta,
have you tried appending the datafiles in long format, xtset the dataset, and then use the lag (L.) and lead (F.) functions in Stata to generate the variables of interest? You could also use the _merge variables that are created when you merge data files on pidp in wide format to generate an indicator for whether or not a person was present before. you'll probably need to play around with this until you have the exact indicators you are looking for but should be doable.
hope this helps,
Gundi

#4 - 08/12/2015 05:53 PM - Redmine Admin

- Assignee changed from Gretta Mohan to Redmine Admin
- % Done changed from 50 to 80

more long format data handling tips: <http://www.stata.com/support/faqs/data-management/first-and-last-occurrences/>

#5 - 08/24/2015 09:21 AM - Redmine Admin

- Status changed from New to Closed
- % Done changed from 80 to 100