Understanding Society User Support - Support #176

Missing parental data

07/31/2013 12:26 PM - Sam Baars

Status:	Closed		**Start date:**	07/31/2013
Priority:	Normal		**% Done:**	100%
Assignee:	Redmine Admin			
Category:	Data analysis			

Description

Hi,

I'm merging data from the USoc Wave 1 youth questionnaire (a_youth) and the Wave 1 adult individual questionnaire (a_indresp) so that I can consider the role of parents' characteristics in my analysis of young people's occupational aspirations. I'm using the variables a_mnspid (to identify mothers) and a_fnspid (to identify fathers) in a_youth and the variable pidp in a_indresp to match young people with their parents. I'm using the /TABLE command in SPSS to carry out a one-to-many match so that siblings in a_youth (who have the same value for a_mnspid or a_fnspid) all receive parental data from a_indresp.

After the matching there is quite a lot of missing data. 360 cases have missing data on mothers: 190 of these are coded -8 for a_mnspid ("natural/adoptive/step mother not in household"), and the remaining 170 have a value for a_mnspid which does not match any of the cross-wave person identifiers (pidp) in a_indresp.

Likewise, 1840 cases have missing data on fathers. 1349 of these are coded -8 for a_fnspid ("natural/adoptive/step father not in household"), and the remaining 491 have a value for a_fnspid which does not match any of the cross-wave person identifiers (pidp) in a_indresp.

My three questions are:
1) How should I interpret a value of -8 for a_mnspid or a_fnspid?
2) Why are there parent identifiers in the youth dataset which do not exist in the main adult dataset?
3) Is there a better way to go about matching data on young people with data on their parents?

Thanks for your help,

Sam

History

#1 - 08/11/2013 08:30 PM - j petersen

I don't have access to the data right now to reproduce your results, but can offer a few hints...
The relationships are recorded on the household grid with reference to household and person numbers (can be found on w_INDALL for all household members). The datasets also holds some cross-wave personal identifiers or pids, see e.g.;

https://www.understandingsociety.ac.uk/documentation/mainstage/dataset-documentation/wave/1/datafile/a_indall/variable/a_mnspid

Pids are only assigned to individuals that have been enumerated at some point during the study and -8 values represent the opposite case. w_INDRESP only holds data on adult respondents and is a subset of w_INDALL. Non-response can occur if the person is temporarily away, refuses or for other reasons is unable to take part (see w_IVFIO for interview outcome info).
Your description sounds fine. The user guide should also has an example of creating a dataset based on relationship variables.

Jakob

#2 - 08/19/2013 04:35 PM - Redmine Admin

- % Done changed from 0 to 40

#3 - 08/27/2013 08:59 AM - Redmine Admin

- Status changed from New to Closed

- % Done changed from 40 to 100

#4 - 11/10/2015 02:30 PM - Gundi Knies

- Category set to Data analysis

- Assignee set to Redmine Admin

- Target version set to X M