

Understanding Society User Support - Support #1802

Missing data for pasoc10 in xwavedat

11/07/2022 11:38 AM - Robert de Vries

Status:	Resolved	Start date:	11/07/2022
Priority:	Normal	% Done:	100%
Assignee:	Understanding Society User Support Team		
Category:	Data documentation		
Description Hi all, I'm conducting research which involves looking at social mobility - so I'm looking at data on parental occupation. I'm primarily focusing on Wave H and I am curious why there is so much missing data for (in particular) father's occupation. If you look at Wave A, around 15% of the sample are 'inapplicable' for pasoc10. This level of missingness corresponds to respondents who were not living with their fathers at age 14, or whose fathers were not working (as given in paju). If we look at Wave H, around 20% of respondents are given in pasoc10 as 'no data from BHPS', and another 24% (9,441) are given as 'missing'. Around half of this 24% is explainable for the same reason as above (father not working/not lived with). However, the remaining half (around 4,500 respondents) is marked as 'missing' in paju. My questions are: 1. Why do so many BHPS respondents have no data for father's occupation? 2. What explains the additional missing 4,500 respondents from paju in Wave H? The only explanation I can see in the routing is that these would be rising 16 year olds. However, 4,500 seems to many for this to be the explanation.			

History

#1 - 11/08/2022 01:50 PM - Understanding Society User Support Team

- Status changed from New to In Progress
- Assignee changed from Alita Nandi to Understanding Society User Support Team
- % Done changed from 0 to 10
- Private changed from Yes to No

Many thanks for your enquiry. The Understanding Society team is looking into it and we will get back to you as soon as we can.

We aim to respond to simple queries within 48 hours and more complex issues within 7 working days.

Best wishes,
Understanding Society User Support Team

#2 - 11/09/2022 11:44 AM - Understanding Society User Support Team

- Status changed from In Progress to Feedback
- % Done changed from 10 to 80

Hello,

Currently not all reports of parents' occupation are available using SOC 2010 coding frame. Parents occupation information which were collected during the BHPS (when SOC 90 was available) were coded using SOC 90 coding frame. Later when SOC 2000 became available, those collected were coded using this frame. And the recent reports were coded using SOC2010 coding frame. UKHLS sample members pasoc reports were recoded using all 3 coding frames, but not for the existing BHPS members.

For those with paju=1 in xwavedat (N=77405), 93.2% (N=72105) have a positive value for either pasoc90_cc, pasoc00_cc or pasoc10_cc. If we split this by UKHLS and BHPS sample members -

For UKHLS sample members with paju=1: for 90.2% pasoc90-00-10 is available, for 9.6% none are available.

For BHPS sample members with paju=1: for 84.8%, only pasoc90 is available, for 13.4% only pasoc00 is available, for 1.1% pasoc90-00-10 is available, 0.6% none are available.

Best wishes,
Understanding Society User Support Team

#3 - 11/10/2022 12:20 PM - Robert de Vries

Thank you for the prompt and helpful response.

That makes sense, and if I incorporate pasoc90 responses, there are only 2,353 remaining missing values for this variable in Wave H respondents. I presume these are the rising 16 year olds?

One more question however, is why pasoc90 is only given to 3 digits in this data (whereas pasoc00 and pasoc10 are given to 4 digits)?

Best,
Rob

#4 - 11/14/2022 01:00 PM - Understanding Society User Support Team

It may not be clear from my response above, pasoc90_cc, pasoc00_cc or pasoc10_cc are available in the End User License (EUL) version of the data and pasoc90, pasoc00 or pasoc10 are available in the Special License (SL) version of the data. Access to SL versions of the data is more restrictive as these are considered to be more disclosive. Basically, it includes all the variables in the EUL version plus a few additional ones (e.g., month of birth) or more detailed versions of some variables (e.g., pasoc90 instead of pasoc90_cc).

pasoc90 is 3-digit but pasoc90_cc is 2-digit
pasoc00 is 4 digit but pasoc00_cc is 3 digit
pasoc10 is 4-digit but pasoc10_cc as 3-digit

If you want to know more about these different levels of access please see <https://www.understandingsociety.ac.uk/documentation/access-data>

#5 - 08/22/2023 01:38 PM - Understanding Society User Support Team

- *Category changed from Data analysis to Data documentation*

#6 - 08/22/2023 01:38 PM - Understanding Society User Support Team

- *Status changed from Feedback to Resolved*

- *% Done changed from 80 to 100*