

Understanding Society User Support - Support #856

how to deal with contradictory records

09/28/2017 10:19 AM - Min Zhang

Status:	Closed	Start date:	09/28/2017
Priority:	Urgent	% Done:	100%
Assignee:	Min Zhang		
Category:	Data inconsistency		
Description			
Dear Understanding Society team,			
I am writing to seek your suggestions about contradictory data.			
The first example is in the same wave. I ran cross-tabulation between a_qfhigh (highest qualification ever achieved) and a_fenow (age when leaving school / never went to college or university). I assume that the respondents who attained university degree would not be found in the category "never went to college or university" in the variable a_fenow. However, this is not the case.			
. tab a_qfhigh a_fenow if a_dvage>=25			
highest qualification still in further education			
al missing inapplica don't kno write in never wen at colleg Tot			
al			
-----+-----			
--			
missing 1 17 0 0 0 0			
18 refused 0 3 0 2 1 0			
6 don't know 0 18 7 19 38 2			
84 university higher deg 0 16 3 3,805 53 236 4,1			
13 1st degree level inc 0 14 3 5,458 203 296 5,9			
74 diploma in higher edu 0 4 2 2,089 440 142 2,6			
77 teaching qualificatio 0 0 2 574 107 23 7			
06 nursing or other medi 0 1 2 556 335 37 9			
31 a level 0 11 3 1,443 1,156 152 2,7			
65 welsh baccalaureate 0 0 0 0 2 0			
2 international baccala 0 1 0 26 10 5			
42 as level 0 1 1 104 75 13 1			
94 higher grade/advanced 0 1 0 175 173 17 3			
66 certificate of sixth 0 0 0 60 49 6 1			
15 gcse/o level 0 18 10 3,198 5,037 262 8,5			
25 cse 0 1 1 529 1,360 41 1,9			
32 standard/ordinary (o) 0 1 0 183 430 21 6			
35 other school (inc. sc 0 7 0 355 967 35 1,3			
64 none of the above 1 392 16 1,676 10,915 287 13,2			
87			
-----+-----			

--									
	Total		2	506	50	20,252	21,351	1,575	43,7
36									

The category that is shown as "never wen" is actually "never went to college/university".

As can be seen from the above table. there are 203+53 respondents who had university degrees, yet reported that they never went to university.

The second example is related to cross-wave youth data. I assume that if a respondent who reported yes to "ever smoked cigarettes" in wave 1 would not say no to the same question in wave 2. However, this is what I got:

```
tab a_ypevrsmo b_ypevrsmo
```

ever smoke cigarettes at all	ever smoke cigarettes at all			Total
	missing	yes	no	
missing	0	1	15	16
yes	2	74	39	115
no	27	209	2,416	2,652
Total	29	284	2,470	2,783

39 respondents who said that he/she has ever smoked in wave 1 said never smoked in wave 2.

This is getting more complicated when I link wave 1-6 together.

I understand that these are measurement errors/recall errors. I could have simply recoded these contradictory records as missing. But the numbers of these recodes may seem trivial but as they add up over waves, they are not small numbers. I am not sure what I am supposed to deal with them.

Many thanks for your time,

Regards,
Min

History

#1 - 09/28/2017 10:54 AM - Stephanie Auty

- Status changed from New to In Progress
- % 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.

Best wishes,
Stephanie Auty - Understanding Society User Support Officer

#2 - 10/10/2017 10:32 AM - Stephanie Auty

- Status changed from In Progress to Feedback
- Assignee changed from Alita Nandi to Min Zhang
- % Done changed from 10 to 70

Dear Min,

In the first case, it is possible that some of these respondents gained their degree through distance learning and so did not go to university. However, at least some will be a data inconsistency based on interviewer or respondent error. These two questions are not asked together in the questionnaire and there is no check implemented between them in the CAPI software.

In your second example, these differences will be due to respondent error as the youth questionnaire is self-completion. It could be that they had only smoked once and then forgotten about it by Wave 2, for example. You might decide that it's more likely that someone would smoke and then forget than make up that they had smoked, or think that they had when they hadn't, but you will need to decide which assumptions you are willing to make based on your research question.

Best wishes,
Stephanie Auty - Understanding Society User Support Officer

#3 - 10/30/2017 03:47 PM - Stephanie Auty

- *Status changed from Feedback to Resolved*
- *% Done changed from 70 to 100*

#4 - 11/20/2017 02:51 PM - Stephanie Auty

- *Status changed from Resolved to Closed*