

Understanding Society User Support - Support #1301

Volume change between releases

01/17/2020 02:32 PM - Laura M

Status:	Resolved	Start date:	01/17/2020
Priority:	Normal	% Done:	100%
Assignee:	Laura M		
Category:	Data releases		
Description Comparing the wave g datasets in the most recent release to the first time it was released 2 years ago I see a difference in raw volumes. I was expecting weights and some other variables to change but thought that the overall volumes might be the same. The effect on a statistic I'm making is small (only in the second decimal place is there a difference without weights) but I wanted to double check this was normal.			

History

#1 - 01/22/2020 01:26 PM - Alita Nandi

- Status changed from New to In Progress
- Assignee changed from Stephanie Auty to Laura M
- % Done changed from 0 to 20
- Private changed from Yes to No

Hello Laura,

Sorry for the delay in getting back to you.

What do you mean by "raw volumes"? Do you mean that the number of cases/observations is different between the current and a previous release? If so, please tell me the date of the previous release version you are referring to and which specific datafiles are you comparing - INDALL, INDRESP, YOUTH,... ?

Best wishes,
Understanding Society User Support Team

#2 - 01/23/2020 09:44 AM - Laura M

Hello,

Further information:

I've made a variable using relationship questions in INDRESP and combining this at the child level to see how many children have parents with relationship difficulties. The dataset with the variable in combines several datafiles.

if &wave._screlparir in (1,2) OR &wave._screlparar in (1,2) OR &wave._screlpards in (1,2) OR &wave._screlparrg in (1,2) then &wave._distress = 1;
When I run this on the new release of wave g I get 11,662 responses including missings, when I run this on the first release of wave g (sorry I can't be more specific, it's stored on our servers so I can only tell it was from the time it first came out) I get 11,317. These are the unweighted volumes. Is it normal for these volumes to change?

Thanks,
Laura

#3 - 01/23/2020 10:24 AM - Alita Nandi

- % Done changed from 20 to 40

Hi Laura,

One reason I can think of that you would get a difference in the number of cases in the two created datafiles is due to changes in the parent ID variables which would have been used to match parent-child information. These ID variables are produced from relationship information collected every wave, and in a few cases the relationship between household members could change across waves - one of which could have been mis-reported, or miscoded by interviewer. When new wave data comes, the data team does longitudinal checks and so may go back and change relationship variable in past waves.

As I don't know exactly which files you used and which you matched etc, I will suggest you compare each file from the two releases and check that the number of cases are the same. If that is the case, then look at the variables used in this program and compare those from the two sets of files to see if there are differences in values.

If you tell me the dates of the two releases, I can check if the number of cases in all wave g files are the same across the two.

Best wishes,
Understanding Society User Support Team

#4 - 01/24/2020 10:08 AM - Laura M

Hello,

Just looking at G_INDRESP I see different volumes between the 2 versions I am comparing. As I said before I don't know the exact version as it was put on our servers, I work for a large govt. department. I am comparing the version that was released in November and, I assume, the first release of wave G as it is in a folder from that time period. I get 42,217 for the original and 42,168 for the most recent. This doesn't account for the full difference I see in my variable, I will be looking into this, but should I be worried about the discrepancy in INDRESP?

thanks,
Laura

#5 - 01/24/2020 06:17 PM - Alita Nandi

- % Done changed from 40 to 80

We have checked and you are correct there is a difference in number of observations between the g_indresp files in the 2017 and 2019 releases. This is because individuals in a TSM only households are not eligible for interviews. After the 2017 release we found that the sample status of some individuals was not correctly assigned and after that was done, these individuals turned out to be in TSM only households. So, they were dropped. You can still find them in the g_indsamp files which records everyone who was enumerated in the last wave and shows their current interview status. You will find that their interview outcome says they are in TSM only households.

Best wishes,
Understanding Society User Support Team

#6 - 01/24/2020 06:30 PM - Alita Nandi

To clarify, if in a future wave these individuals appear in a household with at least one OSM or PSM they will be eligible for interviews.

If you don't know who are OSM, TSM, are then please look at our FAQ. Go here <https://www.understandingsociety.ac.uk/help/faqs> Then click "Study" then click on "Who is in the sample? Who is followed?" Basically these are sample status which determines who is eligible for interviews every year.

You can also find this information in our User Guide
<https://www.understandingsociety.ac.uk/sites/default/files/downloads/documentation/mainstage/user-guides/mainstage-user-guide.pdf>

#7 - 01/28/2020 10:39 AM - Laura M

Perfect, thank you for looking into it.

#8 - 08/14/2020 05:29 PM - Alita Nandi

- Status changed from In Progress to Resolved

- % Done changed from 80 to 100