

Understanding Society User Support - Support #1368

Unmatched households when merging hhresp with Isoa data

06/23/2020 05:01 PM - Natalie Bennett

Status:	Resolved	Start date:	06/23/2020
Priority:	Normal	% Done:	100%
Assignee:			
Category:	Special license		
Description			
<p>Hi I am trying to distribute LSOA identifiers to household level data using wave 3. I have an issue when I merge the c_hhresp file with the c Isoa11_protect file using the below stata code:</p> <pre>use c_hhresp.dta, clear merge m:1 c_hidp using c Isoa11_protect.dta</pre> <p>When performing this command I find 7 households from the hhresp file which are not found in the LSOA file, and 6,507 households from the LSOA file not found in the hhresp file.</p> <p>I haven't been able to identify a reason from looking at the data as to why these data are only found in one file, rather than both. I was wondering if you could confirm if the above merge is correct and if so, what the reason is for these data which cannot be matched?</p> <p>Many thanks</p> <p>Natalie</p>			

History

#1 - 06/25/2020 05:28 PM - Gundi Knies

- Category set to Special license
- Assignee set to Natalie Bennett
- Target version set to X M
- % Done changed from 0 to 80
- Private changed from Yes to No

Hi Natalie,

the universe of cases in the geographical identifier datasets is sampled households with a valid postcode on the ONS Postcode Directory, see the user guidance accompanying the file. Not all sampled households have a valid postcode, and not all sampled households participate in the household interview. The hhresp data file only has households with a household interview record.

Hope this explains the non-matches you find.
Best wishes,
Gundi

#2 - 06/26/2020 01:19 AM - Alita Nandi

- Status changed from New to Feedback

To follow up on what Gundi has said,

All households in c_hhsamp who have a valid postcode for their addresses are available in c Isoa11_protect file. c_hhresp comprises of only those sampled households who have completed a household questionnaire, so the 6507 cases in c Isoa11_protect are those households who are present in c_hhsamp but not in c_hhresp. The 7 cases in c_hhresp but no in c Isoa11_protect are households without a valid postcode.

You can check this by merging c_hhsamp with c Isoa11_protect using c_hidp. You will find that there are no cases in the c Isoa11_protect which are not in c_hhsamp, but there are 2898 cases in c_hhsamp which are not in c Isoa11_protect.

Note, c_hhsamp c_hhresp c Isoa11_protect are all at c_hidp level, so m:1 merging is not needed. You can use 1:1 merging.

Best wishes,
Alita
On behalf of Understanding Society User Support Team

#3 - 06/26/2020 02:18 PM - Natalie Bennett

Hi Alita and Gundi

Thanks both for your help, I understand the source of the non-matches now.

Thanks so much.

Best wishes

Natalie

#4 - 06/26/2020 04:38 PM - Natalie Bennett

Sorry just to follow up so I can justify this later, could you confirm why those 2,898 cases are not in the LSOA file?

Many thanks

Natalie

#5 - 06/27/2020 11:07 AM - Alita Nandi

Hi Natalie,

These are the households without valid postcodes.

Best wishes,
Alita

#6 - 10/13/2021 11:35 AM - Understanding Society User Support Team

- *Status changed from Feedback to Resolved*

- *Assignee deleted (Natalie Bennett)*

- *% Done changed from 80 to 100*