Understanding Society User Support - Support #1865

Changes to USOC wave data downloaded from UK Data Service compared to previous downloads of the waves

02/23/2023 04:42 PM - William Shufflebottom

Status:	Resolved		**Start date:**	02/23/2023
Priority:	High		**% Done:**	100%
Assignee:	Olena Kaminska			
Category:	Weights			

Description

Hi,

QUESTIONS

Q1: indscub_xw weight from wave 6 of USOC is present in our historical download of the wave 6 data but appears to be missing in the version of wave 6 we downloaded from UKData Service a few months ago and is also not listed as being in wave 6 on the USOC variable search page - can we confirm why only the indscui_xw weight is in the latest Wave 6 version, confirm it was in the original release, and if/when (and if so why) it was removed?

Q2: Our estimates run on the latest download of wave 1 to 12 of USOC are producing different numbers from the estimates we ran at the time of the previous wave's releases. Has there been a change to the data or weights (beyond wave 6 having a different weight) or how the weights work that could explain the difference we are seeing for all waves (bar wave 1 and wave 12) in a recent download of the data from all the waves. We are using the same weight (bar wave 6) and the same variable (sclfsat_7 in this case - but we use a range of USOC variables in our analysis).

BACKGROUND

We are producing estimates for the OECD and just discovered some differences for the estimates and CIs for the sclfsat7 variable when we re-ran historical estimates for all USOC waves 1 to 12. We run breakdowns for this variable (and others) by various domains when we update our publications and a new USOC wave has been released so we have the estimates from previous runs made at the time of USOC wave data release. We only ran the sclfsat7 variable again recently so there may be other changes.

We have a document for the weights to use for each variable which states that the indscub_xw weight is the correct weight to use for the sclfsat_7 variable in wave 6 but we noticed it was "missing" in the wave 6 data we downloaded around November from UK Data service (instead indscui_xw is present). As we are getting differences in our estimates and CIs for all waves (bar wave 1 and 12), this has prompted us to check with you if there have been changes made to the versions of the USOC main study wave data currently on the UK Data Service compared to what would have been available at the time each wave's data was released which could explain the differences we are seeing.

Your help is greatly appreciated as this has the potential to impact a lot of our publications and the current ad hoc we are working on

History

#1 - 02/24/2023 08:47 AM - Understanding Society User Support Team

- Status changed from New to In Progress

- Assignee changed from William Shufflebottom to Olena Kaminska

- % 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 - 02/24/2023 08:48 AM - Understanding Society User Support Team

- Category changed from Data releases to Weights

#3 - 02/24/2023 02:17 PM - Olena Kaminska

Dear William,

Your observation is completely correct. We have taken ub_xw weight out from after wave 6 and have made a number of updates on the weights since wave 1.

ub_xw weight: this has been taken out as we ub_xw became not very representative of the cross-sectional population due to missing new immigrants since wave 1 of UKHLS. Instead, for cross-sectional analysis users should use ui_xw weight starting from wave 6 onwards, which includes new immigrants at wave 6 of UKHLS. Currently a new boost is in the field, so with its release a new weight would follow. To avoid confusion, only one type of cross-sectional weight will always be provided, the weight that best represents cross-sectional population, given our data.

Change in estimates: weights have been updated on many aspects, including:
- update of sample status based on ethnic minority groups learned after wave 1 for EMB and after wave 6 for IEMB households (small numbers, but this resulted in increase of OSMs and improved estimates);
- poststratification at wave 1 to mid-population ONS data now updated given 2011 Census (previously mid-year estimates were used by ONS based on 2001 Census);
- correction for mortality (reducing of the effect of unidentified previously mortality onto nonresponse correction) - this should show largest impact onto research questions related to death, so elderly age, and health issues;
- nonresponse models became more complex taking into account differences in nonresponse patterns across ethnic groups (all nonresponse predictors are now tested for an interaction with ethnic group, and significant interactions are now presenting in nonresponse models).
- Cross-sectional weights have now no zero values (expect for TSMs from wave 1). This in itself should have no impact on estimates, though tiny changes are inevitable.

Hope this helps,
Olena

#4 - 03/08/2023 12:23 PM - Understanding Society User Support Team

- Status changed from In Progress to Feedback

- % Done changed from 10 to 80

#5 - 11/30/2023 12:52 PM - Understanding Society User Support Team

- Status changed from Feedback to Resolved

- % Done changed from 80 to 100