

Understanding Society User Support - Support #886

Zero weights and statistical power

12/04/2017 05:45 PM - Eric Emerson

Status:	Closed	Start date:	12/04/2017
Priority:	Normal	% Done:	100%
Assignee:	Olena Kaminska		
Category:	Weights		
Description			
Hi			
<p>I'm interested in data contained the harassment modules (in Waves 1, 3, 5 and 7), but am concerned about the significant reduction in statistical power arising from the increasing proportion of respondents who are assigned values of 0 in w_ind5mus_xw. I understand from a previous thread (#877) that 'The provision of weights requires the ability to estimate probabilities of continuing to respond over multiple waves. This is true of cross-sectional weights as well as longitudinal ones, as they are derived from the longitudinal ones (how this was done is described in section 3.8.3.10 of the User Guide). In consequence, a person in a household where there is no person who has been enumerated at every wave up to wave w will get a weight of zero. Such people should not be given a weight, as the weights for all other sample members are calculated in a way that compensates for these "missing" people.'</p>			
<p>However, the 'compensation' appears to also result in a significant loss of statistical power. Taking as base the unweighted number of respondents who provide a valid answer to the 'attacked' items, the weighted population size has reduced from 92% of actual respondents in W1 (7418/8072) to just 27% in W7 (2711/9973). The resulting reduction in power is of concern and given the rationale outlined above, will continue to increase over time as the % of households in which someone has been enumerated at every wave will continue to diminish. It also seems rather wasteful of people's time that the responses of the majority of participants is, through the weighting process, assigned to a statistical waste bin!</p>			
<p>Be very grateful if you could suggest any ways round this problem.</p>			
<p>Many thanks</p>			
<p>Eric</p>			

History

#1 - 12/04/2017 09:38 PM - Stephanie Auty

- Private changed from Yes to No

#2 - 12/06/2017 10:45 AM - Olena Kaminska

Eric,

You are correct about the 'loss' of statistical power in the analysis of extra five minutes. This loss is almost exclusively not due to attrition and nonresponse but due to design. Here are a few details about it.

Firstly, this 'large' loss of power is only for extra 5 minutes analysis - and really only for when you want an estimate of the population for extra 5 minute questions. It results from the fact that we oversample by a big factor the minority groups and the fact that the non-minority group is relatively small (note - it is sufficient for most of analysis though). This is because the extra 5 minute questions are not really designed for overall population estimates (although you can use our data to obtain them). The main purpose of these are to obtain good estimates for minorities - 5 specific groups, and to compare them to non-minority group. Please note that if you look at the statistical power for the estimates of each of the separate group and of any estimate that compares them to non-minority group - it will be much higher. The 'loss' is only apparent when you combine these together - and even that will give you enough statistical power to detect most important differences and obtain good estimates for social statistics - in effective sample size it is equivalent to a good quality survey.

I hope this helps,
Olena

#3 - 12/06/2017 02:20 PM - Eric Emerson

Hi Olena

Many thanks for this. I am aware of the purposive oversampling of specified minority ethnic groups. However, the main issue appears to be the increasing proportion of extra 5 minute respondents being given zero weights (presumably as they live in a household where there is no person who has been enumerated at every wave up to wave w). Hence, the reduction in the proportion of the weighted to unweighted extra 5 min sample over waves (0.92 in W1, 0.81 in W3, 0.72 in W5, 0.27 in W7). So, although the oversampling strategy has remained constant across waves the weighted sample size has systematically dropped (7418 in W1, 4549 in W3, 3599 in W5, 2711 in W7). This is my major concern as, given the procedure for allocating zero weights we can assume continuing decreases in future waves. Given that some of the variables include important but relatively rare

events (being assaulted in specific settings) a sample size of 2711 is rather small. By W9 it could well be unusable. I am aware that we can proceed without using the weights, but this always seems a less desirable course of action.

Best wishes

Eric

#4 - 12/06/2017 03:30 PM - Stephanie Auty

- Status changed from *New* to *In Progress*

- % Done changed from 0 to 60

#5 - 12/11/2017 04:43 PM - Olena Kaminska

Eric,

Thank you for your kind response and for providing the numbers. Most importantly, I would like to warn you that especially with extra 5 minute analysis one can not run it unweighted – it won't represent the population simply because the 5/6 of the sample comes from minority groups. But one can use either design weight for extra 5 minute or (I suggest) wave 1 extra 5 minute weight and then apply a nonresponse correction (just to correct for attrition between wave 1 and wave 7 if the data used in analysis is from wave 7 for example). In the latter situation the weights will take care for all the different probabilities of selection and for wave 1 nonresponse – and you can gain numbers through a more tailored nonresponse correction.

There are two things to mention about the weights: the currently released cross-sectional extra five minute weight relies on a continuous longitudinal household participation in a survey. I can see now that this would not be ideal for your analysis. There will be released (planned to be released with the next release) another extra five minute weight. This one will incorporate IEMB (which started at wave 6) – which will add more numbers to cross-sectional analysis, it will also not rely on continuous response – just on wave 1 response.

I hope this answers your question,
Olena

#6 - 12/14/2017 09:24 AM - Eric Emerson

Hi Olena

Many thanks for this suggestion and news regarding a new extra 5 min weight. Any idea when the next release will be? If it's some time away, I will pursue your suggestion about using the W1 weight and modelling non-response.

Thanks again

Eric

#7 - 12/14/2017 10:47 AM - Olena Kaminska

Eric,

My understanding is that re-release is not going to be before May next year - and likely later.

Hope this helps,
Olena

P.S. In modelling nonresponse, please pay attention to rising 16 (those who became 16 year old between wave 1 and wave 7).

#8 - 12/15/2017 09:27 AM - Eric Emerson

Will do. Many thanks again.

Eric

#9 - 01/09/2018 01:51 PM - Stephanie Auty

- Status changed from *In Progress* to *Resolved*

- % Done changed from 60 to 100

#10 - 03/07/2018 05:50 PM - Stephanie Auty

- Status changed from *Resolved* to *Closed*