

Understanding Society User Support - Support #1770

Making best use of the Ethnic Minority Boost Sample

10/02/2022 03:18 PM - Laurence Rowley-Abel

Status:	Resolved	Start date:	10/02/2022
Priority:	Normal	% Done:	100%
Assignee:			
Category:	Survey design		
Description			
<p>Hi Understanding Society team,</p> <p>I am conducting a basic analysis of health outcomes across combinations of ethnic groups and age groups using the Wave 10 data. My issue is that when I produce things like cross-tabulations, once I apply weights, and break down the sample by both broad ethnic group and age group, I end up with a small N in many cells (ie: < 10) and in some cases 0 counts for a cell and so I am ending up with very large confidence intervals. Having read the general User Guide and the Ethnicity and Immigration User Guide, my understanding is that one of the reasons the ethnic minority boost samples were included was to try to tackle this issue of small sample sizes in marginalised subgroups, so I wanted to check that I am not missing something. I think my main issue is that once I apply a weight (such as <code>j_indinui_xw</code>), the N for the ethnic minorities is being scaled down in order account for the oversampling of these groups, and since the weighted N (rather than the unweighted N) is being used to calculate confidence intervals by R, I end up with very large confidence intervals.</p> <p>As a quick illustration: I'm using the <code>j_indresp.dta</code> file and have recoded the <code>j_ethn_dv</code> variable into Asian, Black, Other, White. In the Black category, there are 1314 respondents, but when I apply the weight <code>j_indinui_xw</code> and tabulate by ethnicity using the <code>svytable</code> function in R, this is scaled down to only 458.3841. If I then break this down by 10-year age group and my health outcome variable, I end up with very small Ns (or zeros) which means when using a function such as <code>svyciprop</code> in R, I get very large confidence intervals.</p> <p>This may simply be an unavoidable problem, but given that Understanding Society has put lots of effort into including these extra ethnic minority samples, I wanted to make sure I was making best use of them. And just to double check - I can simply use the normal <code>indresp</code> file in order to draw on this ethnic minority boost sample?</p> <p>Many thanks for your help and the amazing resources you provide!</p> <p>Best wishes, Laurence Rowley-Abel</p>			

History

#1 - 10/04/2022 09:48 AM - Understanding Society User Support Team

- Category set to Survey design
- 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 - 10/05/2022 09:55 AM - Understanding Society User Support Team

- Status changed from New to Feedback
- % Done changed from 0 to 50

Dear Laurence,

A significant reduction of the weighted counts for ethnic minorities (so from 1314 respondents to 458.3841 in your example) is expected and intended - due to oversampling they are overrepresented and the weights bring their proportion in the sample down to the level in the population. However, this in itself does not affect standard errors (though weighting may become slightly inflated when the data is weighted but that happens for different reasons).

Does this help?

Best wishes,
Piotr

#3 - 10/05/2022 01:53 PM - Laurence Rowley-Abel

Dear Piotr,

Many thanks for your reply. The issue is that when I use a function such as svymean in R, it uses the smaller scaled down N (ie: 458) rather than the actual raw N (ie: 1314) to calculate the standard error. For example, if I calculate the mean age of Black respondents while accounting for survey design but do not use weights, the standard error is 0.51. But if I calculate the mean age of Black respondents while accounting for survey design and also use weights, the standard error is 0.98. So the standard error almost doubles.

Is there any way around this? Eg: could weights be calculated that don't take account of the oversampling of ethnic minorities, which could be used in an analysis where ethnicity is one of the explanatory variables anyway (and so the bias due to oversampling would be accounted for in that way)?

Thanks for your help.

Best,
Laurence

#4 - 10/07/2022 02:31 PM - Understanding Society User Support Team

Dear Laurence,

Can I ask which weight you use and with which wave?

Best wishes,
Piotr
Understanding Society User Support Team

#5 - 10/11/2022 01:39 PM - Laurence Rowley-Abel

- *File ethnicity_standard_errors_reprex.R added*

Dear Piotr,

I am using Wave 10 and am using the weight j_indinus_lw as later on in my analysis I will be involving variables from prior waves as well.

If it's helpful, I have attached a minimal reprex for the example I used above of calculating the mean age of black respondents.

Many thanks!

Best wishes,
Laurence

#6 - 10/12/2022 06:40 PM - Understanding Society User Support Team

- *% Done changed from 50 to 70*

Dear Laurence,

Here is the advice from Olena Kaminska, our survey statistician:

"For an analysis with only one wave (e.g. wave 10) use _xw weights. I suggest you use data release 11.1 with better sample sizes. For longitudinal analysis you can either use our weight or create your own tailored weight, you can access the tailored weights course here: <https://open.essex.ac.uk/enrol/index.php?id=301>."

Best wishes,
UKHLS User Support Team

#7 - 10/25/2022 01:37 PM - Laurence Rowley-Abel

Dear Piotr and Olena,

Thank you for your help with this. I believe my understanding of the use of weighted/unweighted N in the calculation of standard errors was incorrect, and my issue is simply that certain members of the ethnic minority boost sample are just not eligible for the weight (regardless of whether I use the longitudinal or cross-sectional weight), which is why my sample size was reducing so much.

Many thanks.

Best wishes,
Laurence

#8 - 10/26/2022 01:59 PM - Understanding Society User Support Team

- *Status changed from Feedback to Resolved*

- *% Done changed from 70 to 100*

Files

ethnicity_standard_errors_reprex.R

2.16 KB

10/11/2022

Laurence Rowley-Abel