

## Understanding Society User Support - Support #1091

### Sampling weights youth self-completion survey

11/12/2018 02:23 PM - Melanie Luhrmann

<b>Status:</b>	Resolved	<b>Start date:</b>	11/12/2018
<b>Priority:</b>	Normal	<b>% Done:</b>	100%
<b>Assignee:</b>	Olena Kaminska		
<b>Category:</b>	Weights		
<b>Description</b>			
<p>I would like to conduct a panel analysis of teenagers responses in the youth self-completion survey based on waves 1-6 (using all answers), and I am struggling to understand the sampling weights supplied. In my understanding, I should use ythscus_xw as weights for the first two waves, and ythscub_xw as weights for the following waves. Is this correct?</p> <p>If this is correct, were these weights constructed the same way? It seems that the weights in the first 2 waves are systematically lower than in the latter waves, hence my uncertainty.</p> <p>Thanks for any support!</p>			

#### History

##### #1 - 11/12/2018 02:43 PM - Olena Kaminska

Melanie,

Thank you for your question. Could you tell us a little bit more about your analysis? Are you interested in longitudinal analysis? Are you planning to analyse children who were 10 in wave 1 and became 15 in wave 6? Are you planning to use only youth.dta or any other dataset? Or are you thinking of pooling information from all the teenagers across years to study events?

Thank you,  
Olena

##### #2 - 11/12/2018 02:51 PM - Melanie Luhrmann

Dear Olena,

thanks for your fast response!

I am interested in longitudinal analysis, but am using an unbalanced pannel for the following reason: I am particularly interested in a difference-in-difference analysis of a policy change that happens roughly between waves 2 and 3, where the outcome variables are form the youth survey. My main focus is on kids who were observed at least once (in wave 1 or 2) pre-policy change, and in (at least one of the) waves 3 to 6.

I am merging in some household characteristics and parental variables from the main survey, but my object of interest are the teenagers who fill in the self-completion survey.

P.S.: I will estimate one specification as a robustness check where I will just pool, so if you could let me know the appropriate weights for this as well, that would be marvellous.

##### #3 - 11/12/2018 03:09 PM - Olena Kaminska

Melanie,

Thank you. This is very helpful. For your policy analysis the responding group will be very specific: it is not only non-monotone response (so you don't need response in each wave) but also it depends on responses of parents. This is very unique and we don't have a specific weight for this. The suboptimal weight that you should use is enumeration weight from the last wave that you have observations. It will be f\_psnenuus\_lw. Theoretically you may want to have additional step of nonresponse correction which you can run yourself. I suggest that you look at your results and if you think they are against your expectations I will be happy to help you with this extra step.

For pooled analysis if you are using only youth dataset you should use W\_ythscus\_xw weight or W\_ythscub\_xw weight for each wave respectively. the 'ub' weight includes BHPS sample so it is larger but BHPS joined UKHLS only at wave 2 - so 'ub' weight is available later (it is available from wave 2 in the most recent release). You will need to scale the weight such that each year has a similar share - and thus each wave contributes evenly to your analysis.

The scaling factor is:

Let X= the total weighted N (number of people that you have for your nalaysis after weighting in each year)

Create X\_av - the average of the weighted total N

Calculate scaling factor SC=X\_av/X - for each wave (you will have one value per wave - and it will be larger for years with smaller sample size)

Multiply the appropriate weight by SC = new\_weight

Use the new\_weight in your pooled analysis.

Hope this helps. And please let me know if you have further questions,  
Olena

**#4 - 11/12/2018 03:30 PM - Melanie Luhrmann**

Perfect! Thanks so much, Olena, this was immensely helpful! I'll implement this and see where it takes me.

**#5 - 11/12/2018 03:45 PM - Melanie Luhrmann**

Just one additional question: When I use `f_psnenus_lw`, all teenagers who are not observed in that wave have no weight, so I'd lose more than half my sample.

Do you think I could follow a similar procedure to the one you suggest for the pooled analysis, i.e. taking long run averages of the wave-specific `psnenus_lw` weights? I guess my underlying question is whether this weight changes much over time, or not very much?

Thanks again and sorry for all these questions,  
Melanie

**#6 - 11/12/2018 04:41 PM - Stephanie Auty**

- *Status changed from New to In Progress*

- *% Done changed from 0 to 60*

- *Private changed from Yes to No*

**#7 - 11/13/2018 01:10 PM - Olena Kaminska**

Melanie,

`f_psnenus_lw` excludes two samples: BHPS that joined in wave 2 and IEMB that joined in wave 6. You could increase the numbers if you start your analysis at wave 2 (for example). For this you will use `f_psnenub_lw` weight - but you can not use wave 1 information as it is not available for BHPS sample. I don't think you can use IEMB - because they only started at wave 6, and the effective sample size (the sample size that matters in your analysis) is much smaller than raw numbers anyways - so this shouldn't add much statistical power to your analysis.

My suggestion is to try your analysis with `f_psnenus_lw` - the sample size matters only if you have borderline p-value (for example around 0.05).

Hope this helps,  
Olena

**#8 - 11/13/2018 01:57 PM - Melanie Luhrmann**

Dear Olga,

ok, I got it. Then I'll do that and can present some robustness checks using `f_psnenub_lw` weight.

Have a good day - and many thanks again!  
Melanie

**#9 - 12/21/2018 10:31 PM - Stephanie Auty**

- *Status changed from In Progress to Resolved*

- *% Done changed from 60 to 100*