

Understanding Society User Support - Support #985

Weights for pooled cross-section over all waves

06/22/2018 11:51 AM - Nhat An Trinh

Status:	Resolved	Start date:	06/22/2018
Priority:	Urgent	% Done:	100%
Assignee:	Olena Kaminska		
Category:	Weights		
Description			
Hello,			
Although this issue has already been discussed a couple of times, I would like to address the selection and use of the appropriate weights when pooling across all waves of Understanding Society once again to avoid any mistakes. I'm very much appreciating the guidance that has been provided so far, but haven't found a clear answer to my question and thus be extremely grateful if someone could help me out.			
For my analysis of intergenerational social mobility across labour market entry cohorts, I am using all waves including all samples of Understanding Society in a pooled cross-section. Obviously, I have dropped all duplicates as I want to have each observation only once in my dataset and take the first interview in which the individual has indicated both her first occupation and year of leaving school/further education as my observation of interest. In line with [#758] , I have constructed the individual cross-sectional weight as follows:			
gen xweight = .			
replace xweight = a_indpxus_xw if wave == 1			
foreach x in b c d e {			
replace xweight = `x'_indpxub_xw if inlist(wave,2,3,4,5)			
}			
repalce xweight = f_indpxui_xw if wave 6			
replace xweight = g_indpxui_xw if wave 7			
Is this the correct way of selecting the cross-sectional weights? And do I need to do anything else such as rescaling to correctly apply them for my pooled cross-sectional analysis (i.e. calculating social mobility rates and proportions of class of origin and destination by labour market entry cohorts)?			
Thank you very much!			
Nhat An			

History

#1 - 06/22/2018 05:10 PM - Stephanie Auty

- Category set to Weights
- Assignee set to Olena Kaminska
- Target version set to X M
- 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.

Best wishes,
Stephanie Auty - Understanding Society User Support Officer

#2 - 06/25/2018 11:49 AM - Olena Kaminska

Dear Nhat An,

Thank you for your question. You suggest that it is a pooled cross-sectional analysis, yet you mention that you are studying intergenerational social mobility. Could you clarify whether you take information for each person only from one wave or from a few waves? If you take information from a few waves you will need longitudinal weights.
Also, can you clarify what you mean by deleting duplicates? Do you mean each person has only one entry in your analysis? This will not influence weights but may influence PSU.

Thanks,

Olena

#3 - 06/25/2018 12:10 PM - Nhat An Trinh

Dear Olena,

Yes, you are right: for each individual information is only taken from one wave and each individual only figures once in my analysis. I study intergenerational social mobility by using the information of the individuals' first occupation and the occupation of their parents. I hope this makes everything clear know.

Many thanks for your kind help!

Nhat An

#4 - 06/25/2018 12:30 PM - Olena Kaminska

Nhat An,

Thanks. Yes, your selection of weights looks fine. Ideally you would want to scale your weights such that sample sizes per year don't vary. For this find the average of total sample size for waves 1-7. Calculate the scaling factor for each wave $sc = (\text{average sample size for waves 1-7}) / (\text{sample size in wave X})$. For wave X calculate new weight $= xweight * sc$.

The sample size may relate to the overall total of all people in that wave or to the subgroup that you are interested in. In any case it should have the same definition for the denominator and the numerator.

I hope this helps,
Olena

#5 - 06/25/2018 02:26 PM - Nhat An Trinh

Great, thank you very much, Olena! This is indeed very helpful.

Nhat An

#6 - 06/25/2018 02:39 PM - Stephanie Auty

- *Status changed from New to Resolved*

- *% Done changed from 0 to 100*