

Understanding Society User Support - Support #502

Weights to use to analyse year-to-year transitions within waves 1-5

02/09/2016 11:15 AM - Sanne Velthuis

Status:	Closed	Start date:	02/09/2016
Priority:	Normal	% Done:	100%
Assignee:	Sanne Velthuis		
Description			
Hi,			
I have a question about which weights to use. A similar question (no. 462) has been asked before, but I'm still not 100% sure about what to do.			
I have created a panel dataset of all working age individuals in waves 1-5. At the moment I've just left all observations in the panel so it is an unbalanced panel. I'm interested in low pay transitions, and what I would specifically like to do is calculate, for individuals who are low-paid in any given year, the transition probabilities to higher pay, to unemployment, self-employment and economic inactivity. In other words, I want to know what proportion of workers who are low-paid in one year are still low-paid the following year, what proportion are higher-paid the following year, what proportion are unemployed, and so forth. The syntax I was planning to use for this would look something like:			
<pre>tab paystatus if L.paystatus == 1 (1 being low-paid)</pre>			
I would like to calculate this across all waves. So I am not specifically interested in transition rates from wave 1 to 2, from wave 2 to 3, from wave 3 to 4, etc. I just want to know the 'average' transition rates from low pay to other pay/labour market states within all waves from 1 to 5 (so within the period 2009-2014). But I am not sure what weights to use to do this.			
Should I create a balanced panel and use the longitudinal weight for wave 5, applying this to all waves for each respondent as suggested in the User Guide on p. 61? But I am not necessarily interested only in those individuals who were present in the sample from waves 1 to 5. If someone was only present at waves 1, 2, and 3, but was low-paid at any point during this period, then I'd like to be able to include them. Or will it mess up my estimates if I do this?			
In question 462 it was suggested pooled analysis could be used if the aim is to represent events, rather than people. However, I don't really understand the difference between these, nor which one of these applies to my situation. I am interested in transitions - which are technically events - but the question I'm trying to answer is what proportion of low-paid <i>people</i> have made the transition to higher pay the following year. So would pooled analysis still be suitable? And if so, what weights should I use? In the above-mentioned question the suggestion was, as far as I understand, to use the wave 2 longitudinal weight for transitions between waves 1 and 2, the wave 3 longitudinal weight for transitions between waves 2 and 3, etc. If I understand this correctly, this means that I could just use the longitudinal weight for each wave as it is provided in the data, without having to copy the weights across all waves for each respondent, right?			
A further question is whether or not I should include former BHPS respondents (and therefore use the <code>_indinub</code> , rather than the <code>_indinus</code> weights)? Or will it screw up my estimates to conduct my analysis from wave 1 onwards but include the BHPS sample from wave 2 onwards?			
If you could provide some advice about these matters that would be greatly appreciated.			
Many thanks in advance,			
Sanne			

History

#1 - 02/09/2016 11:58 AM - Olena Kaminska

Sanne,

Yes, your situation is very similar to question 462. Potentially pooled data or balanced panel may be best option for you. But it is very complicated and is easy to make a mistake with. First you need to be clear that you represent events (moves between statuses) over 5 waves rather than people. Importantly, nesting of events within people should be corrected for.

Alternatively you may estimate your statistic 4 times for each wave pair. For each pair use the longitudinal weight from the dataset (of the last wave used). This should be easy and straightforward. You then can take a simple average of the 4 estimates for a point estimate. The confidence intervals may not be easily estimated though. But you have an option of talking about range of estimates, e.g. the proportion A ranges from X to Y across 4 year pairs (and provide confidence intervals for end estimates).

It does not matter much which weight you use: *indinub* or *indinus*, but I would go with *indinub* as it has higher sample size and therefore better statistical power.

Hope this helps,
Olena

#2 - 02/11/2016 01:12 PM - Victoria Nolan

- *Status changed from New to Resolved*
- *Assignee changed from Olena Kaminska to Sanne Velthuis*
- *% Done changed from 0 to 90*

#3 - 02/18/2016 01:04 PM - Victoria Nolan

- *Status changed from Resolved to Closed*
- *% Done changed from 90 to 100*