

Understanding Society User Support - Support #351

longitudinal vs. cross-sectional weights

03/19/2015 11:32 AM - Carolina Zuccotti

Status:	Closed	Start date:	03/19/2015
Priority:	Normal	% Done:	100%
Assignee:			
Category:	Weights		
Description			
Hello,			
<p>I have been doing analysis on the impact of parental employment (measured with the retrospective question) on individuals' labour market outcomes. So far, I have used Wave 3 only, and I have used the other waves to replace missing data in Wave 3, such as age, gender, ethnicity, parental information, etc. I have been using therefore a cross-sectional weight:</p>			
<pre>svyset c_psu [pweight = c_indinub_xw], strata (c_strata) singleunit(centered)</pre>			
<p>With this weight I have done all my descriptive tables and my regression models so far. Now, I would like to add to my regression models a variable that measures employment status of the individual in Wave 2, since I am planning to start looking at transitions into and out of employment. Also, I would like to make crosstabs with this variable as well. This means, I believe, that I need to start using longitudinal weights. But how do I do that?</p>			
<p>1) Descriptive tables: Should all my tables use the longitudinal weight? Or should I use it only when I am studying, for example, employment status in two different waves? (i.e if I make a table that relates employment in Wave 3 with parental background, should this table also have a longitudinal weight, even if I am not studying transitions?)</p>			
<p>2) Regression tables: Imagine that I make two regression models, where I estimate the probability of employment by age, gender and parental background; and a second one where I also add employment in t-1 as control variable. In the first model there are no changes over time, while in the second there are. Should I still use the longitudinal weight for both?</p>			
<p>Of course, changing the svyset all the time does not seem like the logic solution here, so I assume I will just stick to one. I just want to understand the implications of using one or another weight.</p>			
<p>Thanks in advance for your help!</p>			
<p>Regards,</p>			
<p>Carolina</p>			

History

#1 - 03/24/2015 09:37 AM - Redmine Admin

- Target version set to X M

- % Done changed from 0 to 50

I can pass on the following advice:

1. Descriptive statistics: Using longitudinal weights, c_indinub_lw, means the sample statistics can be used as the population estimates for the 2010/11 UK population who have survived until 2012/13 (to be used with the General Population Sample + Ethnic Minority Boost + BHPS samples). If using wave 3 cross-sectional weights then the sample statistics can be used as the population estimates for the 2011/12 UK population. So, which weights you use depends on which population estimates are required for. But in general, if employment transitions are a central part of the research question one should stick to using the same sample for all descriptive statistics, that is, those who responded in wave 2 and wave 3, throughout and use the appropriate longitudinal weights, c_indinub_lw. NB if sample statistics is meant by descriptive statistics then weights should not be used in general.
3. Regression analysis: same answer as in 1 - if you want to compare results for both models it makes sense to use the same sample and hence the same weight for both models.

On behalf of the team,
Jakob

#2 - 03/27/2015 08:42 AM - Carolina Zuccotti

Thanks Jakob for your response. It is much clearer now; however, what do you mean by "NB if sample statistics is meant by descriptive statistics then weights should not be used in general". Do you refer to statistics of the sample itself? (that is, when you are not looking at the representativeness of results?)

Thanks!

#3 - 03/27/2015 05:25 PM - Redmine Admin

- % Done changed from 50 to 90

Basic description of the sample as opposed to estimation of population parameters.
On behalf of the team,
Jakob

#4 - 03/30/2015 08:30 AM - Carolina Zuccotti

Hi Jacob,

Thanks again for your response. I was re-reading your previous answer and I wanted to double-check.

1) When you say that "Using longitudinal weights, c_indinub_lw, means the sample statistics can be used as the population estimates for the 2010/11 UK population who have survived until 2012/13 (to be used with the General Population Sample + Ethnic Minority Boost + BHPS samples)", you actually mean people who have survived until 2011/12 (Wave 3), right?

2) If instead of c_indinub_lw I used d_indinub_lw, would then the weight represent people who have survived in Waves 2 (2010/11), 3 (2011/12) and 4 (2012/13)?

3) Is it possible to use a weight for people who participated in Waves 3 and 4, no matter whether they have participated in Wave 2 as well?

Thank you!

Carolina

#5 - 04/13/2015 10:43 AM - Alita Nandi

1. Yes. Using c_indinub_lw means that the sample statistics can be used as the population estimates for those who are alive and living in the UK in waves 2 and 3 (2010/11 and 2011/12)

2. Yes. Using d_indinub_lw means that the sample statistics can be used as the population estimates for those who are alive and living in the in waves 2, 3 and 4 (2010/11, 2011/12 and 2012/13)

3. Not as yet.

4. Response to an earlier query of yours: ""NB if sample statistics is meant by descriptive statistics then weights should not be used in general". Do you refer to statistics of the sample itself? (that is, when you are not looking at the representativeness of results?)"

Yes. Do not use weights if all you want to do is describe the sample. Use weights when you are using the sample statistics as population estimates.

#6 - 04/13/2015 10:47 AM - Carolina Zuccotti

Thanks a lot, much clearer now!

#7 - 04/13/2015 10:50 AM - Redmine Admin

- Status changed from New to Closed

- % Done changed from 90 to 100