

Understanding Society User Support - Support #1141

Weighting

02/11/2019 10:40 AM - Freya Cook

Status:	Resolved	Start date:	02/11/2019
Priority:	Normal	% Done:	100%
Assignee:	Freya Cook		
Category:	Weights		
Description Hi, I am currently working on a project looking at financial security, I am looking at the variable _finnow in every wave and am looking to calculate a mean which is representative of the UK population for each wave. However, I am struggling to work out which is the correct weighting to use. I have currently used _xewght for wave 1 of the BHPS and _indpxui_xw for wave 8 of the UKHLS however I am unsure if this is correct. Please could you advise which is the most appropriate weighting to use for each wave in order gain an output representative of the UK population at the time that wave was carried out. Thanks			

History

#1 - 02/11/2019 10:44 AM - Stephanie Auty

- 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 - 02/11/2019 11:48 AM - Olena Kaminska

Dear Freya,

Thank you for your question. I am not sure how long back you are looking at. But overall, you are right in using the cross-sectional weight that is most inclusive. You can use us weight for wave 1 of UKHLS, ub weight between wave 2 and wave 5, and ui weight starting at wave 6 onwards. If you are looking back at BHPS, you can use cross-sectional weights as well. But overall please be aware of the differences in the population that the data represents. For example us weight represents people who lived in the UK continuously since 2007/2008 (excluding immigrants that came to the country since). But ui weight represents people who lived in the UK continuously since 2012/2013. You may see the jump therefore in your estimates over time between wave 5 and wave 6 estimates (which use su and ui weights). If so, it may be better to use us or ub weight consistently - this will exclude more recent immigrants. If you don't see any jump, then immigrants do not influence your estimates so much - and you may continue to use us, ub and ui weights when appropriate.

Hope this helps,
Olena

#3 - 02/11/2019 04:06 PM - Freya Cook

Dear Olena

Many thanks for your very helpful email. I have a follow-on question which relates to how I use the chosen weight in my logistic regression analyses which I am carrying out using STATA. I am currently specifying the weight using the syntax:
logistic h_finnow ib1.h_age_dv [pw= h_indpxui_xw] , base vce(cluster h_hidp)

This allows for the weighting and clustering by household (h_hidp). However, I do not think it correctly estimates the standard errors. I am wondering if I need to use the svy commands instead? My understanding is that this will provide an appropriate (and larger) estimate of the standard errors. Thus I am proposing to use:

```
svyset h_ppid [pw=h_indpxui_xw] || h_hidp  
svy: logistic h_finnow ib1.h_age_dv , base
```

Many thanks

Freya Cook

#4 - 02/11/2019 04:22 PM - Olena Kaminska

Freya,

Yes, svy command is better suited for your purpose. This style of svyset command may be better, and also include strata in your analysis:

```
svyset psu [pweight=pw], strata(strata)
```

And then use logistic regression with svy:

Does this help? There are more examples in UKHLS training course materials.
Olena

#5 - 02/13/2019 05:23 PM - Stephanie Auty

- *Status changed from New to Feedback*
- *Assignee changed from Olena Kaminska to Freya Cook*
- *% Done changed from 0 to 70*

#6 - 08/10/2022 11:37 AM - Understanding Society User Support Team

- *Status changed from Feedback to Resolved*
- *% Done changed from 70 to 100*