

Understanding Society User Support - Support #44

Weighting on a restricted sample (outcome=110)

04/30/2012 02:15 PM - Rigmor Nielsen

Status:	Closed	Start date:	04/30/2012
Priority:	Normal	% Done:	100%
Assignee:			
Category:	Data analysis		
Description			
Dear User Support			
I am using the Understanding Society data for poverty analysis and would like to ask you if you can confirm whether I am using the weights correctly. These are the steps of my analysis:			
<ul style="list-style-type: none">- Restrict the sample to only those households where all individuals have responded(a_outcome=110 in HHSAMP).- Create poverty flag for households below 60% of median income using following stata code:			
cap drop poor generate byte poor = .			
summarize eq_inchh [aw = a_hhdenus_xw], detail replace poor = eq_inchh < .6 * r(p50) if !missing(eq_inchh)			
<ul style="list-style-type: none">- When calculating the median income for household, I used the household level weight (a_hhdenus_xw) from datafile HHRESP.			
<ul style="list-style-type: none">- Merge household level data with individual level datafile INDALL to obtain variable for dependent children (a_depchl_dv)- Calculate proportion of dependent children living in poor households. When doing this, I use the individual level weight (a_psnenus_xw) from datafile HHRESP.			
svyset [pw=a_psnenus_xw] svy: tab poor a_depchl_dv, col			
Many thanks, Rigmor			

History

#1 - 04/30/2012 03:00 PM - Redmine Admin

- Category changed from Survey design to Data analysis

#2 - 04/30/2012 06:36 PM - Olena Kaminska

This is a summary of a longer discussion via email. The process has two stages and will require two weights: first stage is to create a 60th percentile cut point for household income; second stage is to calculate proportion of kids in poor households.

1) If you are using all households with a_outcome 110, you are looking at households where all adults gave either full or proxy interview. We do not have a specific weight for this subgroup of households. One could predict nonresponse for this specific group separately and multiply it by design weight, but this process is limited as little information is provided in the released data on nonresponding households. Instead, we suggest to create a suboptimal weight (newweight). The newweight is equal to the highest adult proxy weight in the household. This is because we are looking for those households where all members (even people with the lowest probability of response) responded. We then use this lowest probability of response, which corresponds to the highest within household weight, as an approximation of having complete households. In practice, one would go to indresp file, delete everyone with any other outcome than 110, select the highest proxy weight a_indpxus_xw and infer it for the household. This is newweight.
2) One would then create an aggregate measure of household income; and create a dataset which has one entry per household (delete duplicate entries from each household). This dataset then will be used to look at the distribution for income and pick the 60th percentile.
3) One would now merge this dataset to indall, where children's information is present and would use a_psnenus_xw (which comes from indall) to infer the proportion of children in poor households.

#3 - 05/03/2012 02:19 PM - Redmine Admin

- % Done changed from 0 to 90

#4 - 05/14/2012 09:07 AM - Redmine Admin

- Status changed from New to Closed

- Target version set to M1

- % Done changed from 90 to 100