

## Understanding Society User Support - Support #43

### Weighting-related questions

04/30/2012 12:39 AM - Sung Park

<b>Status:</b>	Closed	<b>Start date:</b>	04/30/2012
<b>Priority:</b>	Urgent	<b>% Done:</b>	100%
<b>Assignee:</b>			
<b>Category:</b>	Ethnicity		
<b>Description</b>			
Hello: I have various weighting-related questions, which I have tried to list by similarity below:			
1) What are the distinctions between the analysis weight and the design weight, as described in the user manual ? For example, what is the distinction between a_ind5mus_xw and a_ind5mus_xd as it relates to individual-level weights? I ask in order to determine whether the design weight is relevant to my analysis (some of which may be inferred from the subsequent questions below).			
2) If I am conducting an individual-level analysis of respondents in the ethnic boost sample, but ALL the variables (both independent and dependent variables) I am including in my analysis are NOT specific to the extra 5 minutes of questions (such as sex, # children, household size, employment), would I still use the weight a_ind5mus_xw?			
3) I notice that the psu and strata variables (a_psu_dv and a_strata_dv) are available only in the household file (hhresp). If I am doing individual-level analyses, if I just append these variables (psu and strata) to the individual-level (indresp) by merging on a_hidp, is this a valid way to be able to adjust for the sampling design in my analysis which is at an individual-level?			
4) If I am combining ethnic respondents from the ethnic boost sample, as well as from the general comparison group (those ethnic minorities who live in low-density areas) for an individual-level analysis, is it sufficient to adjust for sampling design by selecting the a_ind5mus_xw weight, as well as adjusting for sampling, stratification and clustering by a_psu_dv, a_strata_dv and a_hidp, respectively?			

#### History

##### #1 - 04/30/2012 12:12 PM - Olena Kaminska

Thank you for your questions.

1) you should use analysis weights. These also correct for nonresponse. Design weights do not correct for nonresponse and are meant only for advanced users who correct for nonresponse on their own (e.g. as part of their own model).

2) As you mention in point 4), you don't want to use only ethnic boost sample - indeed you must be interested in ethnic minority groups. If you are not using any of the extra-five minutes questions, use the usual weights, e.g. a\_indinus\_xw. These will correctly represent each of the ethnic minority group in the population (including people who live in HDA and LDA in correct proportions).

3) Yes, please infer psu and strata from hhresp file for individual analysis. These are the same for all members of a household.

4) As mentioned before, the only way to correctly analyse ethnic groups is to combine ethnic boost with the rest of the sample. If fact, it is already combined. As a rule of thumb, your syntax will be correct if you /never/ use emboost variable in it. And yes, a\_psu\_dv, a\_strata\_dv and a\_indinus\_xw used with svyset command should correctly account for complex sample design in your analysis.

Best of luck :)

##### #2 - 05/03/2012 02:20 PM - Redmine Admin

- % Done changed from 0 to 90

##### #3 - 05/04/2012 10:32 PM - Sung Park

Thank you so much for your response!

I have one last question: I AM interested in studying ethnic minorities (those that are in the EM boost sample, as well as those from the low-density areas) only, and plan to run a regression. My dependent variable is constructed from a question that is NOT associated with the 5 extra minutes. However, I have one independent variable (out of maybe 10 or so independent variables) in my regression that IS based on a question that is part of the extra 5 minutes. Would I use the weight a\_ind5mus\_xw?

I think so, but I just wanted to confirm with you. Thanks very much.

##### #4 - 05/08/2012 11:57 AM - Olena Kaminska

Yes, you will need to use the weight for extra five minutes a\_ind5mus\_xw, simply because by using at least one variable from extra 5 minutes part

you will use only the part of a sample who were asked these questions.

**#5 - 05/09/2012 10:02 AM - Sung Park**

Thank you for your help!

One general follow-up question to one of your prior responses that "a\_psu\_dv, a\_strata\_dv and a\_indinus\_xw used with svyset command should correctly account for complex sample design in your analysis".

If I am conducting a multivariate analysis at the individual-level, will using a\_psu\_dv, a\_strata\_dv and a\_indinus\_xw (assuming they were not asked the extra 5 min questions and not including proxy respondents) with the svyset command adjust for the fact that I am using clustered data where I have multiple respondents per household? In other words, do I need to include a\_hidp at all in my svyset command for weighting purposes to account for similarities among respondents who are from the same household?

Thank you again for your response!

**#6 - 05/10/2012 06:14 PM - Olena Kaminska**

No, you don't need to include a\_hidp in addition to a\_psu. Households are nested within psu, so controlling for psu will control for within household clustering as well.

**#7 - 05/11/2012 07:20 PM - Sung Park**

Thank you!

**#8 - 05/14/2012 08:47 AM - Redmine Admin**

- *Status changed from New to Closed*
- *Target version set to M1*
- *% Done changed from 90 to 100*