Understanding Society User Support - Support #127

variables used to construct survey weights

02/28/2013 11:58 PM - Laura Kudrna

Status:	Closed		**Start date:**	02/28/2013
Priority:	Immediate		**% Done:**	100%
Assignee:	Redmine Admin			
Category:	Data analysis			

Description

Hello,

Where can I find out which variables were used to construct the survey weights available for Wave 2 of US?

This will be essential information for persuading my colleagues to use them on an imminent analysis.

Many thanks.

Kind regards,
Laura

History

#1 - 03/01/2013 08:42 AM - Redmine Admin

- Category set to Data analysis

- Status changed from New to In Progress

- Assignee set to Redmine Admin

- Target version set to M2

- % Done changed from 0 to 50

Laura,
I wonder whether this is a question about whether or not to use survey weights? The User Guide p.26 has a section on that question. The chapter on weighting refers to a working paper on the overall weighting strategy that might also be of interest. If this does not solve you problem, I think it would be useful if you would provide more detail on the specific issue at hand.
Jakob

#2 - 03/01/2013 12:49 PM - Laura Kudrna

Hi Jakob,

Thank you for your reply. I am not seeking advice about whether or not to use survey weights. I have read the working paper and the User Guide.

I want to know if the variables I am using in my models are the same variables that were used to construct the weights. If they are, this may affect the estimates.

Where can I find out how the weights were constructed? Many thanks.

Best regards,
Laura

#3 - 03/01/2013 05:03 PM - Olena Kaminska

Laura,

The answer to your question is - it depends. Each weight uses different set of variables, but the wave 2 weights will combine correction based on roughly between 50 to 100 variables of different kinds. I hope the details below will give you a better understanding of what this may mean for your estimation.

The weights in wave 2 consist of a number of corrections, all of which are important. The first part is design weight, which reflects the design and is not calculated based on any of the variables in the dataset. The second part corrects for household nonresponse at wave 1. This is a large part of the weight value, and the variables used for modelling this part are not released in the dataset - most of these variables are neighbourhood characteristics from Census and other administrative datasets. The next part, the correction for the individual within household nonresponse, is modelled using information from household grid and household questionnaire at wave 1. My understanding is that you are mainly worried about information from an individual questionnaire used for weighting. If you are an advanced user, you may want to model nonresponse yourself (possibly as part of your model). In this situation I would suggest to use wave 1 cross-sectional relevant weight, and correct for attrition between wave 1 and

wave 2 as part of your model.

Just a warning note: some users think that one may control for nonresponse (or attrition) by putting control variables related to nonresponse into the model. This isn't true. Such approach corrects only for nonresponse in the intercept (mean or proportion estimate), and does not correct for the difference between respondents and nonrespondents in other estimates (for example regression coefficients), neither it reflects the correct standard errors of such estimates.

In my opinion you should not worry if the weights use the same variables as are present in your model. If the relationships between variables are influenced by weighting, they are changed in the direction closer to the relationship in the population - i.e. to reflect that nonrespondents tend to be with a different relationship between these variables.

I hope this is helpful,
Olena

#4 - 03/04/2013 08:51 AM - Redmine Admin

- Status changed from In Progress to Closed

- % Done changed from 50 to 100