

Understanding Society User Support - Support #156

Variables and weights

06/09/2013 11:51 PM - Giulia Montresor

Status:	Closed	Start date:	06/09/2013
Priority:	Low	% Done:	100%
Assignee:	Redmine Admin		
Category:	Data analysis		
Description Hello, I would like to ask a few questions: - In wave 2, what is the difference between the variable b_hcondn and b_hcondno? There is no description for b_hcondno in the documentation section of the Understanding Society website. - I would like to use the variable ivintleng but I don't know what weight to apply, all the weights seems to have zero value for this variable. - When I compute estimates using weights, e.g. indinus_xw, I see that I get different standard errors depending on whether I use the command: svy, subpop (if varname=...): mean varname or the command: svy: mean varname, over(varname) Why does this happen? What command is reliable? Thanks and regards Giulia			

History

#1 - 06/14/2013 11:36 AM - Redmine Admin

Giulia,
Looks like it would be useful to have a short guide to the health condition module. This might take a little bit of time. In the short term, you would have to study the questionnaires to see how we ask this of new vs existing respondents.
The course materials here can give you some ideas of how to take the complex survey design into account in your estimates:
<https://www.iser.essex.ac.uk/bhps/courses>

...,over(varname)

I guess Stata would protest if you ask for the mean of a variable over the same variable?
The data sets come with several weights each for a very specific purpose. The user guide is the best starting point to find out more about the different weights.
Hope some of these leads can be useful at least in the short term.
Jakob

#2 - 06/27/2013 11:55 AM - Olena Kaminska

Giulia,

With regard to your question on subpopulation use in estimation of standard errors, I feel this is a question for a theoretical statistician. Within Understanding Society dataset we suggest using subpopulation option mainly because it helps with good correction for clustering (psu's).

Yet, here is a discussion why subpopulation option is better for other estimation:
http://www.ats.ucla.edu/stat/stata/faq/svy_stata_subpop.htm

The difference arises because with 'if' option standard errors are calculated using only the subset of the dataset. With subpop option stata uses the whole dataset for calculating standard errors. The article suggest that the latter is a correct way of calculating standard errors.

Hope this helps,
Olena

#3 - 06/27/2013 03:07 PM - Redmine Admin

- % Done changed from 0 to 50

#4 - 07/02/2013 05:23 PM - Redmine Admin

- *Status changed from New to Closed*
- *% Done changed from 50 to 100*

#5 - 11/10/2015 02:37 PM - Gundi Knies

- *Assignee set to Redmine Admin*
- *Target version set to X M*