

Understanding Society User Support - Support #106

weights

01/08/2013 03:41 PM - Carolina Zuccotti

Status:	Closed	Start date:	01/08/2013
Priority:	Normal	% Done:	100%
Assignee:	Redmine Admin		
Category:	Survey design		
Description In the UKHLS course in Essex, we were told that in order to account for the complex survey design, we needed to use this formula: svyset psuvar [pweight = weightvar], strata (tratavar) However, the variables "psu" and "strata" are not in the currently released version of the data. Will these be available in the new release? And for both waves? How should we use the above formula when the Waves are used together? Thank you for your help! Carolina			

History

#1 - 01/14/2013 11:34 AM - Redmine Admin

- Status changed from New to In Progress
- % Done changed from 0 to 50

Carolina,

There is some advice on this in the new user guide, "If your analysis uses data from both Wave 1 and Wave 2, select the "lw" (longitudinal) version of the weight" (to appear in the section, Weighting adjustments for the Wave 2 release).

Jakob

#2 - 01/14/2013 11:34 AM - Redmine Admin

- Target version changed from M1 to M2

#3 - 01/15/2013 03:14 PM - Carolina Zuccotti

Ok, my confusion was with the PSU and STRATA, but I have understood how it works now. Thank you!

#4 - 01/15/2013 04:57 PM - Carolina Zuccotti

Having said this, I have a further question. Currently I am only working with Wave 1 (and with individuals/no proxy), for which I am applying this formula:

```
svyset a_psu [pweight = a_indinus_xw], strata (a_strata)
```

However, when running some regressions I get no standard errors, and this Note: "missing standard errors because of stratum with single sampling unit".

And I assume that the more variables I add to the regression, the more likely I will have this problem of missing standard errors, is this correct?

In the ESSEX course we were explained a way to merge strata and then avoid this error. But this example simple calculates a mean age by gender. What about more complex analysis such as a regression with 4 or 5 independent variables? Should I use the "pweight" at the end of the regression commands, instead? Or should I merge the strata every time? In which other way could I deal with this?

Thank you very much in advance!

Carolina

#5 - 01/16/2013 01:29 PM - Olena Kaminska

Carolina,

Thank you for your question. This is a common issue which is related to missingness in the variables included in your regression. The issue occurs when there is at least one strata value with only one cluster (psu). If you tabulate psu and strata variables excluding all cases with missing values for all the variables in your model, you will find which strata values cause the issue. Theoretically, because strata values are ordinal you should combine the adjacent values of the strata. The course example should have given you a code to do this. In your situation, instead of female variable you should use var1 described below.

There are two simpler alternatives:

First, (not recommended) is to drop strata variable. The values will be unbiased, but conservative, i.e. there may be situation that you would have

enough power to detect significant difference but it would appear non-significant because you omit strata. Omitting psu or weight is wrong and will introduce bias to your estimates. The point is that if you omit strata (the code will run then) and you find significance, then you are safe. If you have marginal significance or nonsignificance, then you may still find significant difference with strata. Second (recommended) is to use subpop option within svy command. I suggest that first, you create a variable indicating whether there is no missing value on any of the variables in your model (var1=1 and 0 otherwise). Now, keeping all the cases in the dataset (make sure you don't delete the ones that are not used in the model), run the usual svyset command. When running regression use the following syntax:
svy, subpop(var1): reg y1 x1 x2

This also will work if you are interested in a subgroup, e.g. only female. In this situation, make sure to keep all people in the dataset, and var1 will indicate females with no missing values on any of the variables.

Hope this helps,
Olena

#6 - 01/18/2013 09:20 AM - Carolina Zuccotti

Thank you this helps a lot!

#7 - 01/24/2013 08:54 AM - Redmine Admin

- *Status changed from In Progress to Closed*
- *Assignee set to Redmine Admin*
- *% Done changed from 50 to 100*