

Understanding Society User Support - Support #393

Weights for unbalanced panel

07/28/2015 12:46 PM - Gareth Hagger-Johnson

Status:	Closed	Start date:	07/28/2015
Priority:	Normal	% Done:	100%
Assignee:			
Category:	Weights		

Description

I want to use all available data from BHPS and US, rather than restricting my analysis to those who took part in every wave (which would lower statistical power). This produces an unbalanced design with no suitable weight. Are there any strategies I can use to address the complex survey design problem, such as controlling for variables that were used to create weights? The model I will be using is xtlogit in Stata which does not allow svy options.

History

#1 - 07/28/2015 04:22 PM - Peter Lynn

Could you clarify what kind of analysis you intend to undertake? Are you thinking of some kind of longitudinal model in which you can include, and deal with appropriately, censored data of all kinds? Or something more like a cross-sectional analysis or pooled CS?

#2 - 07/28/2015 04:25 PM - Gareth Hagger-Johnson

Peter Lynn wrote:

Could you clarify what kind of analysis you intend to undertake? Are you thinking of some kind of longitudinal model in which you can include, and deal with appropriately, censored data of all kinds? Or something more like a cross-sectional analysis or pooled CS?

It is a random intercept logistic regression model, e.g.

```
meqrlogit y age i.x c.age#i.x || pid:
```

This deals with nested observations within individuals and uses all available data from all waves. But it doesn't address the complex survey design.

#3 - 07/28/2015 05:08 PM - Peter Lynn

So, it looks as though you are effectively pooling across waves. Your model structure should deal with the clustering of observations within persons, as you say. But I don't think it deals with higher levels of the hierarchy of clustering (i.e. that persons are clustered within households and households are clustered within PSUs). Neither does it deal with either stratified sampling or unequal inclusion probabilities.

The last of these (unequal inclusion probabilities) is what weights are designed to deal with. I suggest that for each observation you use the relevant cross-sectional weight. That should correct for design probabilities and non-response. But it does not deal with stratified sampling or with the remainder of the clustering.

#4 - 07/29/2015 10:50 AM - Gundi Knies

- Category set to Weights

- Status changed from New to In Progress

- Target version set to X M

- % Done changed from 0 to 90

#5 - 07/29/2015 04:06 PM - Gareth Hagger-Johnson

Peter Lynn wrote:

So, it looks as though you are effectively pooling across waves. Your model structure should deal with the clustering of observations within persons, as you say. But I don't think it deals with higher levels of the hierarchy of clustering (i.e. that persons are clustered within households and households are clustered within PSUs). Neither does it deal with either stratified sampling or unequal inclusion probabilities.

The last of these (unequal inclusion probabilities) is what weights are designed to deal with. I suggest that for each observation you use the relevant cross-sectional weight. That should correct for design probabilities and non-response. But it does not deal with stratified sampling or with the remainder of the clustering.

How then can I additionally account for strata and psu, in addition to the cross-sectional weight and allowing for clustering of individuals using || pid?

#6 - 08/11/2015 04:23 PM - Peter Lynn

You could account for psu-level clustering by running multilevel (hierarchical) analysis, with psu as the higher level units. Or alternatively use some kind of replication method (e.g. BRR or Jackknife) to estimate standard errors. The latter would in theory allow you to take into account stratification too, though you would probably have to make some simplifying assumptions like merging strata. Or you could just (fairly safely) ignore the stratification, on the grounds that a) this is a conservative approach, and b) effects are anyway likely to be small (and least compared to clustering and weighting effects).

#7 - 08/24/2015 09:22 AM - Redmine Admin

- *Status changed from In Progress to Closed*

- *% Done changed from 90 to 100*