Subject: Logit regressions using two different weights
Posted by Elisa on Sat, 12 Apr 2014 17:30:40 GMT
View Forum Message <> Reply to Message

Hello there:

I want to estimate the impact of domestic violence on the degree of malnutrition among children. For this case, I have taken the IR file for Bangladesh. Because I have two different weights: one for domestic violence variables and other for the rest of the data, I would like to know how I should properly estimate the regressions in STATA considering different weights for each variable.
Shortly, this is what I did:
gen psu=v021
gen strata=v023
gen sampwt=v005/1000000
egen double totw=total(sampwt)
gen dvsampwt=d005/1000000
svyset psu [pw=dvsampwt], strata(strata)

svy: logistic underweight1 pIPV_ i.v024 i.v025 i.v155 i.v130 i.v190
The problem is that pIPV_ and the other variables have different weights. I really appreciate any help.
Best regards

Subject: Re: Logit regressions using two different weights
Posted by Reduced-For(u)m on Sat, 12 Apr 2014 21:13:55 GMT
View Forum Message <> Reply to Message

I'm pretty sure you use the domestic violence weights, because you are only using the domestic violence sub-sample. Weights are, as best as I know, related to the sample, and not the outcome.

Subject: Re: Logit regressions using two different weights
Posted by Sarah B on Mon, 13 Apr 2015 02:57:21 GMT
View Forum Message <> Reply to Message

Hi Elisa,

As Reduced wisely said, the weight is defined for the sample, not for a variable/outcome. The DV weights handle the differential sample selection of women into the DV module and non-response to the DV module (see other posts in this section on these issues), in addition to the sample design.

Are you trying to include women who did not respond to the DV module, that is, for whom v044 !=

1? v044 = 1 means the woman was selected for the DV module and interviewed. All of the women who didn't respond to the DV module will (and should) be missing on any DV-related variables. Depending on how you've created your pIPV_ variable, your regression will either drop all women for whom v044!=1, or give you uninterpretable regression results for pIPV_ because you are including women who did not answer the DV questions.

Note that using your regression as specified will drop women who did not respond to the DV module, because d005 is missing for all women for whom v044 !=1. I think restricting your analysis to these women makes sense for this analysis.