
Subject: Re: Sample weight/Survey design
Posted by [user-rhs](#) on Fri, 07 Feb 2014 06:04:38 GMT
[View Forum Message](#) <> [Reply to Message](#)

Better late than never

Reduced-For(u)m wrote on Thu, 26 December 2013 14:48 - why is v001 before the [pweight=weight] bit? The DHS FAQ lists this code (following) and looking at the svy help file for STATA it doesn't seem like it should be there.

DHS FAQ code: svyset [pweight=weight], psu(v021) strata(strata)

What Kusum had in his/her -svyset- specification is correct. The primary sampling unit (in this case, the EA/cluster-->v001) is specified before the pweight. See Stata documentation for -svyset-: <http://www.stata.com/help.cgi?svyset>

kusum wrote on Tue, 24 December 2013 15:13 When I do run the analysis, the population size is much small (see below). I just wanted to confirm you that I am using the sample weight correctly. Perhaps someone has encountered similar problem?

Thanks,
Kusum

svy, subpop (sample2):logit stunting i.female
(running logit on estimation sample)

Survey: Logistic regression

Number of strata = 25 Number of obs = 5306
Number of PSUs = 289 Population size = 5391.3722

Not sure what "small" is in relation to the total # of children in this dataset, but the svy: logit you ran was done only on the subset of your dataset where "sample2" == 1. I agree with Reducedform that you should do a svy: tab sample2,count to see what the # should be for sample2==1 and check against your weighted pop'n size from the regression output. From what I can see of your -svyset- command, you have set it correctly.

An important thing to note is that the weighting sometimes causes the pop'n size from your regression to be lower than the # of obs'ns. For example, if people living in Kathmandu were overrepresented in your sample relative to actual proportion of pop'n living in Kathmandu, their sampling weights would probably be <1 whereas ppl living in underrepresented regions would probably have sampling weights >1. Therefore, if you have many people from Kathmandu in the subpop you're running the regression on, your pop'n size may be < the # of obs'ns.

HTH,
rhs
