Subject: Weighted data and population size
Posted by shopnobaz on Tue, 14 May 2019 00:29:25 GMT
View Forum Message <> Reply to Message

Hello DHS experts,

I want to do a pooled analysis of BDHS 2007, 2011 and 2014 (KR file for children). As per forum discussion, during regression analysis of pooled data, I need to de-nomalize the sampling weight. I did this using

gen wgt = weight_all *** weight_all = v005/1000000(Total number of households during each survey year/sample households in each survey) and append three surveys data
gen psu = cluster *** each survey clusters are unique eg. 2007_1, 2011_1 and so on
svyset psu, weight(wgt) strata(strat) , singleunit(centered) || _n *** each survey strata are unique

When I fitted weighted logistic regression, after adjusting weight this way, I found:

svy: logit y x

Number of strata = 63 Number of obs = 19,896
Number of PSUs = 1,561 Population size = 44,882,311

Could anyone please suggest that the process is correct? Is the population size reliable or not?

Thank you very much.

Subject: Re: Weighted data and population size
Posted by Bridgette-DHS on Fri, 17 May 2019 16:59:35 GMT
View Forum Message <> Reply to Message

Following is a response from our Research & Data Analysis Director, Tom Pullum:

Is "weight_all = v005/1000000(Total number of households during each survey year/sample households in each survey)" a verbal description of a Stata command? If so, what command? It's not clear what you are doing to the weights, other than dividing by 1000000, which will not affect the results at all.

The "Population size" in the results has been distorted by svyset, particularly by the components other than the weight. You can ignore it.