Subject: Re: Using weights in regression analysis
Posted by Bridgette-DHS on Thu, 11 Apr 2013 21:31:10 GMT
View Forum Message <> Reply to Message

Here is a response from one of our STATA experts Tom Pullum, that should answer your question.

At DHS we mostly use Stata, so I will answer in terms of Stata. The weighting of the data is done as part of the estimation (e.g. regression) command. There is no other sense in which you would "weight the data". For example, instead of "regress y x" you would say "regress y x [pweight=v005]".

You will get the same result if you first say "gen pwt=v005/1000000" and then "regress y x [pweight=pwt]", which some users would prefer to do, but as I said it makes no difference, because Stata always automatically normalizes the weights. Without weights, the estimates are biased toward the oversampled subpopulations and away from the undersampled subpopulations.

The adjustments for clusters and strata will affect the standard errors but not the estimates. If you want to test the significance of the coefficients, you must make those adjustments. For the clusters you expand the above statement to "regress y x [pweight=pwt], cluster(v001)". If you use strata, you must use "svyset" and "svy: regress". The svyset can specify the pweights, clusters, and strata, and then apply them with "svy: regress y x". The svyset command differs slightly across different versions of Stata, e.g. between 11 and 12, so just enter "help svyset" to get the syntax for your version. The strata variable is usually either v022 or v023. However, it is not always labeled correctly. As a general rule, the strata are all combinations of urban/rural and region (the first subnational unit). If the variable labeled "strata" is not consistent with that rule, you should ask someone at DHS to check it.