Subject: Re: Using weights in regression analysis
Posted by Reduced-For(u)m on Sat, 30 Mar 2013 23:14:23 GMT
View Forum Message <> Reply to Message

From the DHS FAQs (under "using data files": http://www.measuredhs.com/faq.cfm):

***First, use the svyset command to tell Stata how your data is set up:

*generate weight
generate weight = v005/1000000

*make unique strata values by region/urban-rural (label option automatically labels the results)
egen strata = group(v024 v025), label
*check results
tab strata

*tell Stata the weight (using pweights for robust standard errors), cluster (psu), and strata:
svyset [pweight=weight], psu(v021) strata(strata)

****Now for a regression - if you prefix regress with "svy:" Stata will now know how to weight your data and compute the right standard errors

svy: reg Y X

***Quick note: computing standard errors in this way is probably not OK for a lot of regressions. Without getting off track or all statsy, a good way to think of this is that this standard error calculation is alright IF the error terms and covariates are independently and identically distributed across observations, other than as operating through the sampling procedure (the stratification and clustering prior to randomization that produces the particular sample you have). I tend to think of these standard errors as the smallest the "true" standard errors could possibly be, but I'm kind of on the conservative/stickler end of this debate, and others would surely disagree.