

---

Subject: Re: weighting to obtain subnational summary estimates

Posted by [Bridgette-DHS](#) on Fri, 07 Aug 2015 13:59:28 GMT

[View Forum Message](#) <> [Reply to Message](#)

---

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

It's up to you whether you want the units of analysis to be adults or households. You could do the analysis either way. But note that household possessions/assets and source of water, type of sanitation, etc.--the wealth index and all of its components--are inherently household-level characteristics. The two main options with these variables are as follows. One option would be to have one record per household, using the HR file, or the PR file reduced to one record per household, with `hvidx==1` or with `hv101=1`. Then the households are the units. The second option would be to have one record per household but to change the weight from `hv005` to `hv005*hv009`. (`hv009` is the number of people in the household). Then the individuals in the households are the units. The data analysis would give more weight to larger households, etc. This second option could be better than using the IR and MR files, which would limit you to women and men age 15-49 (or some other age range for the men).

Religion, ethnicity, listening to the radio, etc., are not, strictly speaking, household level variables, but in fact almost everyone in a household has the same religion and ethnicity and (less so) media exposure.

If you treat household-level variables as individual-level variables, as you described or as I suggested with weight `hv005*hv009`, it is true that the standard errors will go down, but this is really an artificial inflation of the sample size, due to ignoring household-level clustering. I'd consider the reduction in standard errors to be spurious. We usually ignore household-level clustering, but the true sample size for household-level variables (for calculating standard errors) is the number of households, not the number of individuals in those households, and the effective sample size is reduced even further because such variables tend to be similar within clusters, as you will see if you calculate standard errors with and without the `svy` adjustment for `v001`. I would not base the choice between households as units, or individuals as units, on what happens with the standard errors.