
Subject: weighting to obtain subnational summary estimates

Posted by [fabienne](#) on Wed, 05 Aug 2015 14:58:50 GMT

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

Hello everyone

I am working with the DHS VI for Liberia.

Since I need household-level information on wealth and religion, I have appended the IR (women) and MR (men) datasets and merged them with the PR (household member) dataset and I have then kept only the observations for the heads of the household.

My aim is to obtain summary estimates of variables such as hv207 (has radio), hv270 (wealth index) or v130 (religion) at subnational (COUNTY) level. According to the strata variable (hv022), the sampling is self-weighted within county and urban/rural.

I have the following questions:

1. Is my stata syntax for survey-setting appropriate and do I still have to use the weights?

```
gen weight=hv005/1000000  
svyset hv021 [pw=weight], strata(hv022)
```

2. Can I safely ignore the individuals weights ((m)v005) because I look at (heads of) households?

Apologies for another questions on the subject, but previous threads were not specific for my dataset.

Thank you and best regards
Fabienne

Subject: Re: weighting to obtain subnational summary estimates

Posted by [Bridgette-DHS](#) on Wed, 05 Aug 2015 16:22:20 GMT

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

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

Yes, you do need to use weights. You will find that there is some variation in the weights from one cluster to another, even within the same stratum.

The general rule is to use individual-level weights rather than household weights, when you have a choice. I would expect virtually no difference in results, whether you used (a) hv005 or (b) v005 and mv005, especially for the household head, but I recommend that you use (b) because it may have a more complete adjustment for nonresponse.

Here is what I would do.

* open the PR file; save if hv101==1; sort hv001 hv002 hvidx; save as PRtemp.dta

* open the IR file; keep the variables you want ; save as IRtemp.dta

* open the MR file; rename mv* v*; keep the variables you want; append to IRtemp.dta; save as IRMRtemp.dta

* open IRMRtemp; rename v001 hv001; rename v002 hv002; rename v003 hvidx; sort hv001 hv002 hvidx; merge with PRtemp; keep if _merge==3;

* gen weight=v005/1000000; svyset as you have it, etc.

Note that in DHS surveys the household head (hv101=1) is usually on line 1 (hvidx=1) BUT NOT ALWAYS.

Subject: Re: weighting to obtain subnational summary estimates
Posted by [fabienne](#) on Fri, 07 Aug 2015 09:23:43 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you for the fast reply. I may need to rethink my analysis strategy, which is why I have some follow-up questions:

1. Would it be more appropriate to use the whole dataset of the individual IR/MR records to establish district-level summary estimates of religion and ethnicity (and not restrict it to the heads of the households)? The percentage point estimates are probably very similar but the standard errors will probably be lower with the whole data? I assume the weights applied for the svyset will then be the individual weights.
2. For the district-level summary estimates of household-level variables such as hv207 (has radio) or hv270 (wealth index), do I have to restrict the dataset to 1 observation per household and use the household weights? Or can I keep the whole dataset with all records of all household members and apply the individual weights, because they already take account of the household cluster?

Thank you very much.

Best regards
Fabienne

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.