Subject: Questions about weighting when using merged data sets
Posted by DHS user on Wed, 20 Feb 2013 17:07:48 GMT
View Forum Message <> Reply to Message

I am trying to merge the HIV test and Household members data. I am confused as to how to apply the weights--which are different for the different data sets--to dichotomous variables, when comparing HIV occurrence from one data set with another variable (wealth index or anemia level are examples) from the other data set.

Subject: Re: Questions about weighting when using merged data sets
Posted by Bridgette-DHS on Wed, 20 Feb 2013 17:10:02 GMT
View Forum Message <> Reply to Message

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

I will insert some lines that show the correct way to do the merge. I assume you are using Stata. I am using the "old" version of the merge syntax, which I prefer. Merges are always done with cluster number, household number, and line number.

```
* use the AR file

ren hivclust hv001
ren hivnumb hv002
ren hivline hv003
sort hv001 hv002 hv003
save ARtemp.dta, replace

* use the PR file

sort hv001 hv002 hv003

merge hv001 hv002 hv003 using ARtemp.dta
tab _merge
keep if _merge==3
drop _merge

* the correct weight variable will now be hiv05
```

The general rule about which weight to use is that hiv05 takes precedence over hv005 or v005 or mv005. If your file includes hiv05, that would be the correct weight.

The wealth index, hv270, is already coded with numbers 1 through 5. If you enter "tab hv270" you will see the labels, not the codes. If you enter "tab hv270, nolabel", you will see the codes. For more on the syntax of labels, enter "help label".

I hope this helps.