
Subject: Re: Pooling men, women and household DHS from Haiti

Posted by [Trevor-DHS](#) on Wed, 04 Jan 2017 19:16:30 GMT

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

Your thinking as you laid it out seems valid. A few notes, though:

1a) For the de-normalizing, you can find estimates of population age 15-49 from the UN World Population prospects in the following links:

https://esa.un.org/unpd/wpp/DVD/Files/1_Indicators%20

(Standard)/EXCEL_FILES/1_Population/WPP2015_POP_F08_2_TOTAL_

POPULATION_BY_BROAD_AGE_GROUP_MALE.XLS and

https://esa.un.org/unpd/wpp/DVD/Files/1_Indicators%20

(Standard)/EXCEL_FILES/1_Population/WPP2015_POP_F08_3_TOTAL_

POPULATION_BY_BROAD_AGE_GROUP_FEMALE.XLS

1b) Merging men and women - I would generate a variable for sex for the men's and women's files, and then you can combine things like the weights into a single variable, rather than two separate variables (e.g. by renaming the men's variable before merging the two datasets).

1c) Once you have combined the women's and men's data, you can then merge the PR dataset to the combined data using the household ID and the line numbers.

2a) In your appended dataset, I would create a variable for the phase (or, in this case just use V000 as it is unique for each phase - not that for other countries they may not be unique as v000 tells you the recode structure being used and that might be the same for two separate surveys)

2b) Your strata variable should probably include the survey year or phase so that you have unique strata for each year/phase.

3) Unless there is a compelling reason to re-normalize the weights I would not bother re-normalizing. DHS has, by convention, always normalized the weights such that the total weighted N matches the total unweighted N, but there is no statistical reason for doing this.
