
Subject: Correct use of weights for subsamples

Posted by [Christian Bommer](#) on Thu, 16 Nov 2023 13:06:51 GMT

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

Dear DHS team,

I have a fairly general and probably simple question but I couldn't find the answer by browsing the forum (sorry if I missed anything). I am using multiple Women's Recode (IR) datasets from various countries. For each of them I want to calculate aggregate statistics. I know that the surveys include standard weight variables such as V005 that should be used when trying to estimate aggregate nationally representative statistics. However, I was wondering if additional weights are required when I look at certain subpopulations.

For instance, say, I create a binary indicator that captures whether a women below the age of 18 (at interview) has at least one child (using V201 - number of children, and V012 - age of respondent). The aggregate statistic I want to derive from this variable is the percentage of women below the age of 18 who have at least one child (so all women below the age of 18 are the denominator). Is it sufficient to use the weight V005 for this or do I need an additional weight that accounts for the fact that I only look at a subset of women (those below the age of 18)?

Another example that is slightly different: I want to know the percentage of women who have at least one child (regardless of women's age at interview) but I want to know this percentage for the following subgroups:

- rural households (V102)
- female-headed households (V151)
- poorest-quintile households (based on the asset index) (V190)
- women who are literate (V155)
- women who did not complete primary education (V106)

In case you need a specific survey for the answer: One of the surveys I want to use is the 2015/16 DHS from Tanzania (TZIR7B). I will work with Stata.

Best regards,
Christian

Subject: Re: Correct use of weights for subsamples

Posted by [Bridgette-DHS](#) on Mon, 20 Nov 2023 13:08:51 GMT

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

Following is a response from Senior DHS staff member, Tom Pullum:

The sampling weight coded in the data files is a characteristic of the cases. It is proportional to the inverse of the sampling fraction. The sampling fraction varies according to the survey design, mainly from one stratum (v023) to another, although it includes a small adjustment for nonresponse. Stratum and nonresponse are the only sources of variation in the weights. Weights do not vary by covariates and do not need to be adjusted if you select sub populations. (If your selection involved subsampling with different sampling fractions for different subpopulations, then an adjustment would be required, but I don't think that's ever done.)

Except for the factor of 1,000,000, the weighted and unweighted numbers of cases in a subpopulation is about the same, but never exactly the same, and such differences are not a problem. The purpose of the weights is to adjust the sample so that estimates of means, proportions, etc., are unbiased, and this requires weighting up or weighting down, depending on whether a stratum was under-sampled or over-sampled.

Subject: Re: Correct use of weights for subsamples
Posted by [Christian Bommer](#) on Mon, 20 Nov 2023 14:07:21 GMT
[View Forum Message](#) <> [Reply to Message](#)

Thank you!