Subject: Weights not normalized?
Posted by MiFoo on Sat, 01 May 2021 20:36:20 GMT

Hello,

in my dataset (DHS from Bangladesh, PR file), the weighted number of observations is always lower than the unweighted one. Shouldn't the weights be normalized in DHS surveys for each survey year, so that the sum of the normalized weights equals the sum of the cases over the entire sample?

Even when using the full data in the PR file from a single year such as 2017,
data %>% summarize(n=survey_total())
or equivalently
sum(dataPR$hv005/1000000))

I get a lower number than the number of observations in the dataset. Have I misunderstood the normalization in DHS surveys?
Note: I am using R Studio

Thank you!

Subject: Re: Weights not normalized?
Posted by Bridgette-DHS on Thu, 06 May 2021 18:27:43 GMT

Following is a response from DHS Research & Data Analysis Director, Tom Pullum:

In the HR, the mean value of hv005 is 1 (or 1000000 if you keep the factor of one million). In that file, the units are households and hv005 is a household weight. The PR file lists all the individuals in all the households but keeps the same value of hv005 as the HR file. That is, the PR file is an individual-level file but the weight variable is the household weight. If you calculate the average value of hv005 in the PR file, restricted to hvidx=1 or hv101=1 (one person per household) you will get a mean of 1.

The mean of v005 is 1 in the IR file, but the children in the BR file are given the mother's weight (v005), so the mean of v005 in the BR file is not 1.

We always use hv005 when analyzing the PR file and v005 when analyzing the BR file (or KR file) and do not re-normalize to a mean of 1. If you want to go through that step, you certainly can, but it won't make much difference. You are not using Stata, but in Stata, with the pweight option, the weights are automatically normalized to have a mean of 1.