
Subject: Weights when aggregating survey results
Posted by [oyinkansola](#) on Fri, 12 Jun 2020 09:19:03 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hi, I am using the Nigerian DHS for the years 1990, 2003, 2008, 2010 and 2013 for my analysis. I make use of the household recode and the children's recode data sets. I am analyzing the living conditions in Nigerian cities, and the cities data comes from another data set. Since the geographical location for the DHS clusters are provided, I was able to aggregate the surveys for each year to the city level. This means that my data set for analysis is in the form of panel data with my units of observation as city, and I have an unbalanced panel.

I aggregate the surveys to city level by using the Stata command collapse I am allowed to take the weights into account with the collapse command, which I do following the DHS recommendation to use [iweights = weight/1000000] When I do this the summary statistics are slightly different than when I do not take the weights into account.

It is important to my analysis that I make use of fixed effects estimation model, which specifies that the weights must be the same within a panel. This means that each city must carry the same weight for all years.

My questions:

1. How can I re-weight the data in this scenario? There are 2 different weights for each unit of observation- the weight from the household and the weight for the child recode.
2. Would my results be valid to all cities in Nigeria if I do not use weights in my regressions?
3. If I am not able to re-weight my data for my regression analysis, is it sensible to report the summary statistics when it is weighted or without weights?

Thank you for your anticipated response.

Subject: Re: Weights when aggregating survey results
Posted by [Bridgette-DHS](#) on Tue, 30 Jun 2020 12:51:32 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from DHS Senior Sampling Specialist, Mahmoud Elkasabi:

Thanks for posting such interesting use of DHS data on the forum. I went over your message several times to make sure that I completely understand your design. In the bullets below, I'm documenting my understanding and providing answers to several points. Please let me know if I misunderstood any of the design aspects or if you can add more details about the design and the analysis that might be helpful for the discussion:

- You used a weighted-collapse to estimate survey estimates, so now we have a dataset of

Nigerian cities by survey years as unit of analysis (I would call it city-year), and weighted survey estimates for each city-year. Speaking with the multilevel models language, cities are in Level-1 and years are in Level-2.

- In your analysis, to take cities population into account, you should consider the city population as analysis weight. However, I'm not sure whether such data might be available or not. If it is available, the percentage distribution of population in each city can be declared as a weight. For each city, these percentages might change across years, and therefore each city-year might have its unique weight value.
- As an alternative for the actual cities population, the weighted percentage distribution of population in each city can be used as a proxy. I'm assuming that all these surveys have enough sample size to produce reliable population percentage distribution on the city level, including the old surveys. This can be done in the collapse step. These percentages can be used as they are as a weight for the city-year data.
- Now with a weight calculated for each city-year, is it true that the weight has to be the same within each panel? I'm not sure this is true. As in several publications (Rabe-Hesketh and Skrondal 2006; Carle 2009), the fixed weights is desirable for Level-1 weight, not for the final weight. In our case, this is the cities, which means that within each year, Level-1 weight should be the same. This can be done using several re-scaling approaches, such as using the average weight or the average of the squared weights. However, in our case, we don't even have Level-1 and Level-2 weights. So it makes sense to declare the calculated weight as Level-2 weight and declare Level-1 as 1.