Subject: Stratification in pooled datasets - Tanzania
Posted by mariachiara on Mon, 18 Sep 2017 15:54:40 GMT
View Forum Message <> Reply to Message

Dear DHS Forum Users,

My research project will try to estimate the impact of foreign land acquistions on children malnutrition in Tanzania. As such, I am using only children datasets and I pooled together the children datasets for 1996, 1999, 2004-2005, 2010 and 2015-2016.

Reading several threads, I understood that I have to:

1) de-normalize the sample weights as I am using several rounds of survey. This can be done by making sure that the weights for each round sum up to one and then multiplying those weight by the eligible population, i.e. kids under five (which I found here https://esa.un.org/unpd/wpp/Download/Standard/Population/). Is there someone who can confirm this procedure?

2) setting my dataset as survey data using the "svyset" command in Stata. Here is the point where I am stuck.
I have read other threads where they indicate the code for correctly using "svyset" for different surveys round, but my problem is that I can't identify the strata variables for 1996, 1999 and 2004-2005 datasets.
For the 2010 and 2015-2016 datasets, the strata variable is v022 and the stratification is well explained in the final report. In those datasets stratification is done at the regional and urban/rural levels, so for each region there are two strata (urban and rural). So, in 2010 there are 26 regions in Tanzania and V022 lists 52 strata, while in 2015 there are 30 regions and v022 lists 59 strata (one region is considered totally urban here).

In 2004-2005, problems start. Even though the sampling frame is the same as the 2010 survey (the 2002 census), the variable v022 lists 221 strata so I don't think that this variable is correctly identifying strata. However, I read in detail the final report and I found no information on the stratification. The same is true for the 1996 and 1999 surveys, where variable v022 lists 177 and 84 strata, respectively. However, the final report for the 1996 rounds says "The list of PSUs for the 1996 TDHS survey was stratified by each of the 20 regions (for the mainland) and within each region by urban and rural areas".

Therefore, I am not sure how to identify strata for those three surveys and I would greatly appreciate any help that you can give me!
I will be eternally grateful for any prompt reply as I am on a quite tight deadline!
I apologize if the questions are silly but I have never worked with survey data and I am trying to understand how DHS works.

Thanks a lot.

Best,

Mariachiara

Subject: Re: Stratification in pooled datasets - Tanzania
Posted by Bridgette-DHS on Tue, 19 Sep 2017 18:01:34 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS Stata Specialist, Tom Pullum:

For all the surveys, you will be safe if you use the combinations of region and urban/rural as the strata. You need a line such as this: egen stratumid=group(v024 v025), and then put stratumid as the stratum variable in svyset.

In svyset the type of weight is always a pweight. You can easily verify that multiplying the weights by a constant, any constant, will have no effect whatever on the results. You can define the weight as v005 or as v005/1000000, for example, and the results will be exactly the same. That's because Stata always normalizes pweights to have an average value of 1.

What you may be thinking of is scaling v005 up to be fweight expansion weights, which is only useful if you are trying to estimate, for example, the number of children that are stunted, using svy and tabulate. Another consideration if you do that is that fweights must be integers. DHS does not recommend this because there is an implication of more accuracy than is justified.