Subject: Weight about pooling 2003,2008,2013,2018 BR dataset in Nigeria
Posted by geass on Sat, 21 Mar 2020 22:43:22 GMT
View Forum Message <> Reply to Message

I did the following de-normalization.
In 2003,2008,2013,2018 BR data,according to"Note on DHS standard weight de-normalization"
V005*=V005×(total females age 15-49 in the country at the time of the survey)/(number of women age 15-49 interviewed in the survey)
For example, in 2018 BR dataset, I did
gen wgt=v005/1000000
gen wgt_new=(wgt*44911147)/41821
gen survey=1
egen clusterid=group(survey dhsclust)
egen stratumid=group(survey v023)

I did the same progress in 2013,2008,2003.
Then I pooled four rounds into one dataset.
I use the pooling data to do regression.
reg Y X [pw=wgt_new]

My question:
(1) Is the above weighting and pooling data right?
(2) When I do the regression, do I need to do "reg Y X [pw=wgt_new]" , OR just do "reg Y X" without "[pw=wgt_new]"

Thank you very much
I hope to receive your reply.

Subject: Re: Weight about pooling 2003,2008,2013,2018 BR dataset in Nigeria
Posted by Bridgette-DHS on Mon, 30 Mar 2020 19:54:15 GMT
View Forum Message <> Reply to Message

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

I don't really see what is gained by pooling these surveys into one file. As a conceptual matter, rather than a technical one, a consolidation of four surveys from the same country but at different times does not seem to me to be a meaningful population. It might make sense if nothing was changing, but because most characteristics ARE changing over time in Nigeria, wouldn't you want to analyze each survey separately?

Sometimes I do pool surveys into one file, but only in order to make it easier to make comparisons between the surveys and to graph trajectories over time. For that purpose, no re-weighting is needed. It helps to number the surveys 1, 2, 3, 4, although v007 can be used to distinguish the surveys. Again, this would be done to make data processing easier; I would not run models that ignored the identifiers for the four specific surveys.

For what I think you want to do, I believe the re-weighting is correct. To check it, you want to see whether the total weight in each round is proportional to the population in each round.

The number of women age 15-49 in Nigeria at the time of the survey would be appropriate for re-weighting the IR file, but I see that you are using the BR file. The BR file has one record for every child that appears in a birth history, whether living or dead. It would be better to re-weight using the IR files and then reshape (wide) the b variable to get one record for each child.

Subject: Re: Weight about pooling 2003,2008,2013,2018 BR dataset in Nigeria
Posted by geass on Mon, 30 Mar 2020 23:52:35 GMT
View Forum Message <> Reply to Message

Thank you for you reply.

What I want to know is the factor which will affect pregnancy termination. I need the variable v228(ever had a terminated pregnancy).
In IR file, women who have never had a child is also included. I think It will underestimate the results.
In BR file, each women at least has a child, wether the child is alive or not. So I consider to use BR file.
First, I restrict each BR file to women level. i.e., each caseid is unique(each caseid means one women).
Then I used population to de-normalized the weight and pooled all rounds together.
Is it right?

In you reply, you mean, after I pooled the de-normalized data together, I can simply to do "reg Y X" rather than "reg Y X [pw=wgt_new]",right?

Subject: Re: Weight about pooling 2003,2008,2013,2018 BR dataset in Nigeria
Posted by geass on Wed, 01 Apr 2020 05:13:24 GMT
View Forum Message <> Reply to Message

geass wrote on Mon, 30 March 2020 19:52Thank you for you reply.

What I want to know is the factor which will affect pregnancy termination. I need the variable v228(ever had a terminated pregnancy).
In IR file, women who have never had a child is also included. I think It will underestimate the results.
In BR file, each women at least has a child, wether the child is alive or not. So I consider to use BR file.
First, I restrict each BR file to women level. i.e., each caseid is unique(each caseid means one women).
Then I used population to de-normalized the weight and pooled all rounds together.

Is it right?

In you reply, you mean, after I pooled the de-normalized data together, I can simply to do "reg Y X" rather than "reg Y X [pw=wgt_new]",right?
Sorry.Let me add some thing.
When I do "reg Y X", I also include survey fixed effect in regression. But in this case, does it means I did not include the de-normalized weight?
OR, I should use "reg Y X i.surveyround [pw=wgt_new]".
But I also see some papers which use the pooled sample including several years in one country, they haven't include the weight.

How can I do?

Subject: Re: Weight about pooling 2003,2008,2013,2018 BR dataset in Nigeria
Posted by Bridgette-DHS on Wed, 03 Jun 2020 20:52:33 GMT
View Forum Message <> Reply to Message

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

The BR file has one record for every live birth. The mother is identified with caseid and the child with bidx. The unique ID for each record is caseid plus bidx. I do not see how or why you plan to convert the BR file to a file of women. If you want a file of women, I would recommend that you use the IR file. "Mothers" are women who have had children, i.e. who have v201>0.

There have been endless postings on pooling successive rounds from the same country. This makes sense and can be helpful if you want to describe changes from one round to another. For this purpose you do not need to change the weights at all.

If your goal is to synthesize the surveys, in the sense of calculating some kind of rate that ignores which round the case comes from, I would recommend against doing that. Analyzing successive cross-sections would be better than pooling the cross-sections. If you take the latter approach then you have to take account of the fact that the successive rounds have different sample sizes and possibly take account of the fact that the population size was changing. If you look through earlier postings you will find some relevant procedures.
