
Subject: Weighting after de-normalization
Posted by [plnep](#) on Tue, 16 Dec 2014 04:48:19 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hi forum users,

I am working on the pooled datasets of Nepal DHS for 2001 and 2006. As discussed in this thread - <http://userforum.dhsprogram.com/index.php?t=msg&th=1189&start=0&S=dac787ddfcaa55c72987b9d7b09759fa>, I have already de-normalized and treated the cluster. My confusion is on how to weight after pooling as I will be only using the births for the last two years of 2001 so should I continue to use the same weights or how should I weight it??

Many thanks,

Subject: Re: Weighting after de-normalization
Posted by [Bridgette-DHS](#) on Tue, 16 Dec 2014 15:07:36 GMT
[View Forum Message](#) <> [Reply to Message](#)

Your post is being reviewed by a DHS moderator, and someone will respond soonest.

Thanks for your patience.

Subject: Re: Weighting after de-normalization
Posted by [Bridgette-DHS](#) on Tue, 20 Jan 2015 14:38:05 GMT
[View Forum Message](#) <> [Reply to Message](#)

Following is a response from Senior DHS Sampling Specialist, Ruilin Ren
After pooling the data and de-normalizing the weight, you can use the de-normalized weight for any kind of analysis, restricted to a domain or not, except for estimating totals, because the weight is not in the right scale for totals. If your analysis is restricted to one survey, then you have the choice to use either the original weight or the de-normalized weight, you should get the same results. If your analysis crosses surveys, you must use the de-normalized weight.

Hope this is helpful.

Subject: Re: Weighting after de-normalization
Posted by [plnep](#) on Wed, 21 Jan 2015 01:18:13 GMT
[View Forum Message](#) <> [Reply to Message](#)

Many thanks !!!

Subject: Re: Weighting after de-normalization
Posted by [Bridgette-DHS](#) on Wed, 21 Jan 2015 13:20:31 GMT
[View Forum Message](#) <> [Reply to Message](#)

You're welcome.

Subject: Re: Weighting after de-normalization
Posted by kinsukmanisinha@gmail.com on Mon, 16 Mar 2015 15:39:00 GMT
[View Forum Message](#) <> [Reply to Message](#)

Hi,

Many thanks for the discussion, it helped me understand few critical points.

I am new to survey analysis and hence, have a very basic naive question,

In the discussion you guys mention:

After pooling the data and de-normalizing the weight, you can use the de-normalized weight for any kind of analysis, restricted to a domain or not, except for estimating totals, because the weight is not in the right scale for totals.

What do you mean when you say "estimating totals"? What does "totals" stand for? From what I understand, is it mean?

In that case, does it imply that pooled data should not be used to perform cross-country (over time) descriptive analysis?

I would really appreciate any clarification.

Many thanks..!!!!

Regards
Kinsuk

Subject: Re: Weighting after de-normalization
Posted by [Reduced-For\(u\)m](#) on Mon, 16 Mar 2015 19:08:05 GMT
[View Forum Message](#) <> [Reply to Message](#)

I think Ruilin means that if you wanted to, say, count the number of stunted children, you would get the wrong answer, because in that calculation the value of the weights has a meaning itself (instead of just the relative values of the weights). You will get the same mean stunting rate with original or de-normalized weights, but you wouldn't get the same total number of stunted children.

Subject: Re: Weighting after de-normalization

Posted by kinsukmanisinha@gmail.com on Mon, 16 Mar 2015 21:16:28 GMT

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

Thanks a lot..!!!!

Subject: Re: Weighting after de-normalization

Posted by kinsukmanisinha@gmail.com on Tue, 17 Mar 2015 09:57:19 GMT

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

Hi,

Thanks a lot for the previous explanations.

As I mentioned I am new to survey analysis and DHS database. Consequently, I have few more questions and I would appreciate any help.

Please find attached with this msg an excel sheet which contains the list of countries, respective years and surveys that I intend to pool for my analysis.

I am interested in child birth and health variables, women empowerment variables, household living conditions. I know that for the child health variables, I need to look into the child file. However, I found that women data file is not available for all the countries always. Am I missing out on searching somewhere?

Then, I learned that in order to pool the datasets I need to de-normalize the sampling weights. (<http://userforum.dhsprogram.com/index.php?t=msg&th=1189&start=0&S=dac787ddfcaa55c72987b9d7b09759fa>)

And, also change the PSU variable. So, if I understand well because I pool surveys of different phases from different countries, I will have two PSU. First, at country level and then at household level, right?

Once I de-normalize the weights, fix the PSU and append the datasets, the database is ready for analysis, right? As in I don't need to do something else to the weights or the PSU after I pool in the database. I am sorry if you have already answered this question, I am not confident with how to proceed.

My analysis will consist of descriptive statistics. I intend to perform descriptive analysis with the country level databases (these will be country level for multiple years) and then a regression analysis for the pooled dataset.

Do, I need to take into account some special treatment for the weight and PSU for the above two analysis, apart from what I already mentioned?

Once again many thanks, this forum has been very helpful, thanks a lot..!!!

Regards
Kinsuk

File Attachments

1) [Country_year_DHS.csv](#), downloaded 765 times

Subject: Re: Weighting after de-normalization

Posted by [Reduced-For\(u\)m](#) on Wed, 18 Mar 2015 01:14:19 GMT

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

"if I understand well because I pool surveys of different phases from different countries, I will have two PSU. First, at country level and then at household level, right?" - actually you just want 1 PSU number for each sample PSU, it just needs to be unique across years and countries. So for example, you could generate your PSU numbers by giving each country a number between 10 and 99 and then generating a PSU by "PSU*1000000 + Year*100 + CountryNumber" ... something like that (you could also probably concatenate variable in some way, you just need to create a unique number of each country-X-survey-X-PSU

"Once I de-normalize the weights, fix the PSU and append the datasets, the database is ready for analysis, right" - yep. You just need to use svyset and the svy: prefix before your regressions (so as to actually use the weights and PSUs).

"Do, I need to take into account some special treatment for the weight and PSU for the above two analysis, apart from what I already mentioned?" - nope, just set the svyset*.

This is just a technical note: you will be implicitly weighting regressions here by not just probability weight, but also by the sum of the total weights for the country (that is, by the number of survey rounds and the size of the country, which determine the values of the de-normalized weights). If you have big countries or countries with many more survey rounds than other countries, they will get higher weight in your regression. Maybe they should, maybe they shouldn't - that is just an issue regarding interpretation of the regression coefficients.

Subject: Re: Weighting after de-normalization

Posted by [Bridgette-DHS](#) on Wed, 18 Mar 2015 12:56:42 GMT

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

Following is a response from Senior DHS Sampling Specialist, Ruilin Ren:

If you de-normalized the weight properly, you can apply the de-normalized weight in your analysis, either restricted to certain domains or to certain sections of the questionnaire. Define a new variable name for the de-normalized weight in your pooled data set, and then declare that variable as your weight variable.

As for the problem of estimating "totals", here "totals" means estimates of population totals, such as the estimation of "total number of children under 5 years of age who had fever in the last two

weeks in a country". Because the sampling weight in the data file is a relative weight without a scale, it is not valid for estimating population totals. While the de-normalized weight provides a weight that can be used for your analysis, the weight produced may be user specific, depending on the de-normalization procedure used, so different users may produce different estimates of the same indicator. Therefore we do not recommend using the de-renormalized weight for estimating population totals. However the scale of the weight has less to no effects on other analyses such as for estimating means, proportions, ratios and rates, and for correlation analysis.

Subject: Re: Weighting after de-normalization

Posted by kinsukmanisinha@gmail.com on Tue, 24 Mar 2015 17:00:17 GMT

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

Many thanks for the replies and the clarification..!!!

"actually you just want 1 PSU number for each sample PSU, it just needs to be unique across years and countries. So for example, you could generate your PSU numbers by giving each country a number between 10 and 99 and then generating a PSU by "PSU*1000000 + Year*100 + CountryNumber" ... something like that (you could also probably concatenate variable in some way, you just need to create a unique number of each country-X-survey-X-PSU"

So, for every country for every year I will have one PSU. And, this is not affected by the survey. Hence when you say country-X-survey-X-PSU, you basically mean country and year.

Thanks once again, it has been very helpful..!!!

Subject: Re: Weighting after de-normalization

Posted by [Reduced-For\(u\)m](#) on Tue, 24 Mar 2015 20:59:35 GMT

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

Suppose you have datasets from two countries and two years for each country:

County A Year 1 - 20 PSUs

Country A Year 2 - 20 PSUs

Country B Year 1 - 30 PSUs

Country B Year 2 - 35 PSUs

Here you would want a total of 105 PSUs. You want the number of PSUs in the final dataset to equal the SUM of the number of PSUs in EACH dataset. This is NOT one PSU per dataset (I'm calling a dataset a county-round or country-year - that is, a point in time and space where a DHS is conducted), it is N PSUs per dataset, where N is the number of PSUs in the original sampling design.
