Subject: Weighting and pooling multicountry datasets
Posted by Denise_kpebo on Thu, 02 Jun 2022 11:32:21 GMT
View Forum Message <> Reply to Message

Hi,

I am working on a multilevel logistic regression for multi-country (n=15) study using the DHS guide (https://www.dhsprogram.com/pubs/pdf/MR27/MR27.pdf) for the calculation of level weight.

Firstly, I have conducted the analysis within each country separately to determine which level of alpha should be used (e.g., allocation of variation in weights to the level-1 and level-2 units).
Then based on alpha for each country, I've weighted each dataset separately before merging the 15 countries datasets

Is this sufficient? May I start my analysis at this point with the pooled database using the svyset command?
Or do I still need to apply additional weights considering the differences in population size, since the population are not the same? And if so, how can I do that?

Many thanks in anticipation of your help

Subject: Re: Weighting and pooling multicountry datasets
Posted by Janet-DHS on Fri, 03 Jun 2022 12:45:15 GMT
View Forum Message <> Reply to Message

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

There have many postings on how to weight pooled surveys. There are basically two options. The first is to re-scale the weight for each country so that the total weight for country X is proportional to the population of country X at the time of the survey. You can get the estimated population size from the UN Population Division website, World Population Prospects 2019. The second option is to rescale so that the total weight is the same for each country (or survey). That is, if you pool 10 countries, you re-scale so that each survey has 1/10 of the total weight. Specific steps for both options have been given on the forum.

The first option has the problem that typically one large country, such as India or Nigeria, will completely dominate the results.

At DHS we often leave the weights alone and pool surveys into a single file just to simplify the data processing. We give results separately for each survey, but do not give results for all the surveys combined. Pooled surveys, from different countries and different years, do not describe a well-defined population. It's very hard to interpret a mean or percentage or coefficient from a mix of different surveys.

Subject: Re: Weighting and pooling multicountry datasets
Posted by Denise_kpebo on Wed, 08 Jun 2022 02:45:13 GMT
View Forum Message <> Reply to Message

Thanks so much for your help on this

I'm going to use the first option and I have the population size for each country but this is my first time to weight data based on country population size and I actually have no idea of the state command to be used, could you please help me with that one last time?

Best

Subject: Re: Weighting and pooling multicountry datasets
Posted by Janet-DHS on Thu, 09 Jun 2022 12:38:37 GMT
View Forum Message <> Reply to Message

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

If you search through previous forum responses you will find examples of how to do this.

Subject: Re: Weighting and pooling multicountry datasets
Posted by Anonymous on Sat, 13 May 2023 16:01:02 GMT
View Forum Message <> Reply to Message

I was searching but maybe those posts have been deleted. :(
coreball

Subject: Re: Weighting and pooling multicountry datasets
Posted by ewa.b on Tue, 22 Oct 2024 11:07:03 GMT
View Forum Message <> Reply to Message

Hello, I'd like to ask a question on the rescaling so that the total weight is the same for each country (or survey), so the second option mentioned by Janet above. I read a lot of posts on how to rescale the weights using UN estimates (first option) but I was not able to identify steps for the second. Could you send a link where it is described how to rescale the weights according to option 2? Many thanks.

Subject: Re: Weighting and pooling multicountry datasets
Posted by Bridgette-DHS on Tue, 22 Oct 2024 19:24:44 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS staff member, Tom Pullum:

I have prepared a Stata program that illustrates how to pool and reweight three PR files, using the latest surveys from India, Nepal, and Bangladesh (see attached). For population sizes I just googled estimates for 2020, rounded to the nearest million. You can be more precise if you want. For different files you can get estimates of the relevant subpopulation from World Population Prospects (UN Population Division).

I often append surveys for computing convenience, but I avoid giving pooled estimates. If you do that, keep in mind that the actual surveys are from different dates. We rarely, if ever, cover an entire region or subregion. The largest country will dominate the results, sometimes overwhelmingly, as in this illustrative group of countries, which includes India. The label from the last file in the append will over-write other labels. Variables such as hv024 (region) are obviously different in different countries. Other variables, such as types of facilities, sources of drinking water, etc., often also vary across surveys and must be reconciled with recodes.

File Attachments

1) pool_and_reweight_surveys_do_22Oct2024.txt, downloaded 181 times

---

Subject: Re: Weighting and pooling multicountry datasets
Posted by ewa.b on Wed, 23 Oct 2024 10:16:41 GMT
View Forum Message <> Reply to Message

Thank you for this reponse, it is very helpful. I'd like to ask a clarifying question. If I understood correctly, the procedure described in the response and the file involves recalculating weights using the information about countries population size at the time of the survey. This way of rescaling means that the largest country will dominate the results.

The response of Janet-DHS (Fri, 03 June 2022 13:45) below suggested that this is one of the options to rescale weights, with the second being "rescale so that the total weight is the same for each country (or survey). That is, if you pool 10 countries, you re-scale so that each survey has 1/10 of the total weight". Could you provide description on how to do that second type of rescaling?

Thank you again.

Subject: Re: Weighting and pooling multicountry datasets
Posted by Bridgette-DHS on Wed, 23 Oct 2024 11:55:43 GMT
View Forum Message <> Reply to Message

Following is a response from Senior DHS staff member, Tom Pullum:

To weight each survey equally, you would run the same program but replace all the "target" numbers with a fixed number such as 1. It should not matter what number you use. If you have ten surveys, say, then the procedure will add up the target values, get 10, and then alter hv005 so that each survey gets a proportion 1/10 of the total of hv005 in all ten surveys.