Subject: Pooled datasets - weighting data problem
Posted by lukassg on Wed, 16 Apr 2014 12:10:23 GMT
View Forum Message <> Reply to Message

Hey all,

I have already extensively searched the forum but couldn't find a definite answer that's why I would like to raise my question:

I am pooling DHS survey, always two rounds per country, but all countries still separate. However since I pool two years together, I still need to de-normalize m weights before I do the actual pooling as far as I understood from various sources.

I am using birth recode files and thus I should proceed like this:

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)

Now I read the one-pager by Ruilin Ren but I still don't know where to find "total number of females age 15-49 in the whole country at the time of the survey".
I guess the only way is to go and check all the censuses that have been used for my DHS surveys at hand and look for that specific number there? Is there a database at DHS maybe where all that information is stored already?

Subsequently I have a question for the procedure. I open both datafiles, let's say Ethopia 2000 and Ethopia 2011. For both datasets I de-normalize the weight according to the formula above (before that I divide v005 by 1,000,000 of course). Then I assume I can pool the two datasets by appending one to the other.
But now that I have them pooled and want to specify my svyset, do I just use the weight variable as it is, e.g. "svyset psu [pweight=v005], strata(strat_id)" ?
Or do I have to make any other adjustments?

Thanks a lot for your help, this forum is really helpful always!

Lukas

Subject: Re: Pooled datasets - weighting data problem
Posted by Bridgette-DHS on Thu, 17 Apr 2014 18:14:47 GMT
View Forum Message <> Reply to Message

Following are comments by our Senior Sampling Specialist, Ruilin Ren:
Quote:Now I read the one-pager by Ruilin Ren but I still don't know where to find "total number of females age 15-49 in the whole country at the time of the survey".

Comment: You need to find this piece of information from other sources, such as the country's statistical office or census office, or from the UN's World Population Prospects: http://esa.un.org/wpp/unpp/panel_indicators.htm

Quote:I guess the only way is to go and check all the censuses that have been used for my DHS surveys at hand and look for that specific number there? Is there a database at DHS maybe where all that information is stored already?

Comment: DHS does not store the population information for any country.

Quote:...now that I have them pooled and want to specify my svyset, do I just use the weight variable as it is, e.g. "svyset psu [pweight=v005], strata(strat_id)" ?
Or do I have to make any other adjustments?
Comment: The way you pool the data together is fine. You do not need to rename the weight variable, but you do need some treatment for the PSU/cluster (HV001 or V001 or MV001) and the stratification variables (HV022, V022, MV022, or HV024, V024, MV024 if you use region cross urban/rural for stratification). The idea is that each PSU/cluster and each stratum from each survey must stand alone. If you do not treat them properly, for example, the software cannot distinguish cluster 001 (HV001=001) from survey one and survey two, the system will simply merge them as one cluster. This is not true. The easiest way to do is to add 10000 to HV001 from survey one, add 20000 to HV001 from survey two. So cluster 10001 and 20001 will be treated as two different PSUs/ clusters. Similarly you need to treat HV022, or HV024 if you use HV024 cross HV025 for stratification.

Subject: Re: Pooled datasets - weighting data problem
Posted by user-rhs on Thu, 17 Apr 2014 19:33:00 GMT
View Forum Message <> Reply to Message

Bridgette-DHS wrote on Thu, 17 April 2014 14:14
Comment: DHS does not store the population information for any country.
Hi Bridgette,
I wonder if this is info that can be stuck in the sampling appendix of the DHS final reports moving forward. I had a discussion with Reduced-For(u)m the other day about this. Since the use of DHS data has evolved since its WFS/CPS days, it might be a good idea to keep info on the denominators on record for the purpose of wight rescaling.

Just a thought. Of course we can always use the UN pop'n prospects/Wikipedia too.

RHS

Subject: Re: Pooled datasets - weighting data problem
Posted by geoK on Fri, 30 May 2014 10:38:25 GMT
View Forum Message <> Reply to Message

Hi all and thanks for all these interesting info! May I please ask how should I treat weights for 2 pooled sub sample datasets? In my case: I am pooling together 2 years for 1 country (and as far as I have seen I have to de-normalise weights) but only including women who had at least 1 child (so, not all women 15-49).
Thanks a lot!

Subject: Re: Pooled datasets - weighting data problem
Posted by Reduced-For(u)m on Sun, 01 Jun 2014 19:36:35 GMT
View Forum Message <> Reply to Message

I think one issue we haven't dealt with (yours) is re-normalizing weights when using a sub-population. For a simple solution, I'd suggest going with the recommendation from the DHS staff (buried somewhere in this thread) for when you have multiple survey rounds from the same country - just use the regular weights*.

The idea is that, if the sample sizes are similar across survey rounds, the re-normalizing shouldn't matter much (you are implicitly weighting each survey by sample size when using the regular weights). And with the sub-pop command, I'm not sure how you would want to re-normalize anyway (that bit of survey design inference I'm not real teched up on and the Stata documentation isn't super helpful to me - I think the right re-normalizing might somehow relate to the ratio of the prevalence of the sub-population to the full population, probably across strata or something really difficult and nuanced).

*Note - you want to use the "subpop" command and you want to create new identifiers for "cluster" and "strata" that are survey-round specific (say, replacing cluster "10" with cluster "svy2011_10" or something like that.

**Additional option: just take out all the sub-pop observations from both rounds, sum (within round) all the weights, and divide the old weights by the sum of the new weights. Then you have sub-populations only in each survey round, and each round has the same total weight, but nothing is "population representative", it's just corrected for selection probability across the sub-population.

Any thoughts from the DHS staff on either of these options?

Subject: Re: Pooled datasets - weighting data problem
Posted by geoK on Mon, 02 Jun 2014 16:50:01 GMT
View Forum Message <> Reply to Message

O.K. thank you! I'll try what you suggest...Just a last quick question please: is there a SUBPOP command in SPSS? What If I have only selected the subpop cases and create a new dataset (then pooled the two together?). How can I apply subpop? Still creating a Country-specific variable (like 0;1) ? Thanks!!!

Subject: Re: Pooled datasets - weighting data problem
Posted by Reduced-For(u)m on Tue, 03 Jun 2014 01:16:19 GMT
View Forum Message <> Reply to Message

If you only have a dataset with the sub-population observations, you don't need to use any special "sub-pop" command, in SPSS or in anything else. I would just do it unweighted, and then weighted with regular (probability) weights, and then by the "survey round weights sum to 1" method, and compare the three. If they are similar, you should be fine (and they will probably all be similar).

Also - "country specific variable?" I thought you only had one country.

Subject: Re: Pooled datasets - weighting data problem
Posted by Bridgette-DHS on Wed, 04 Jun 2014 20:45:20 GMT
View Forum Message <> Reply to Message

Comment from Ruilin Ren, Senior DHS Sampling Expert:

Yes, it is true that even when pooling two different surveys from the same country, you need to de-normalize the sampling weight because the sampling fractions of the two surveys might be very different. But it does not matter if your study units are all women 15-49 or a sub-set of them, or even if your study units are children under five from the women's data file. You need to de-normalize V005 using the population of women 15-49 at the time of the survey, or the best estimates you can get.

Subject: Re: Pooled datasets - weighting data problem
Posted by geoK on Fri, 06 Jun 2014 16:19:40 GMT
View Forum Message <> Reply to Message

Thank you for your kind replies.
Also, even in the case of pooled dataset, is it always advisable the use of svyset, when I perform any regression analysis? Should I use as STRATA 'country specific strata' or, given that I am working with the same country, can I merge the strata of the two?
thanks
regards.