Subject: Weighting across surveys when only including youth in analysis
Posted by cgreenba on Wed, 15 Nov 2017 19:10:52 GMT

Hello,

I have pooled DHS surveys from over 30 countries, looking at 2 or 3 surveys per/country. I am interested in looking at percent demand satisfied for modern method of family planning by wealth quintile among youth 15-19 and 20-24 for each survey and am using the individual recode. I then am hoping to use a regression model to determine whether wealth quintile is a significant predictor of percent demand satisfied for family planning.

First, I want to ensure that breaking the data down in this manner still provides an adequate sample size to estimate demand satisfied by wealth quintile and age group for each survey. Does the DHS have any guidance on this?

Second, when pooling the data, in order to denormalize the weights, I sued the following procedure recommended by Bridgette with the DHS:

"When pooling multiple surveys, I would first re-scale the weights (e.g. hv005) in each survey by a factor. For example, if you have S surveys, Ni total (weighted=unweighted) cases in survey i, and a total of N cases in all S surveys (N=sum Ni) then you could decide to give equal weight to each survey. You then want the weights in survey i to add to N/S, rather than to Ni. To do that, you multiply the weights in survey i by the ratio (N/S) / Ni. (I think of this as the target total divided by the original total.) "

However, in my case, should the Ni should be the 15-24 sample for each survey that I am using or the full 15-49 sample (even though I am only looking at the 15-24 age group)? Also, is this weighting procedure more appropriate in my case than instead weighting by the country 15-24 or 15-49 population at the time of the survey?

Finally, for my regression model, I am trying to understand whether using the svyset function is sufficient at will account of clustering at the country-level. Is there another type or cluster or adjustment I need to do? Does anyone have experience with the difference between using the svy function with melogit and using the melogit regression with svy, but including weights and clustering?

Any help would be greatly appreciated! Thank you so much!

Best,
Charlotte G

Subject: Re: Weighting across surveys when only including youth in analysis
Posted by Bridgette-DHS on Fri, 17 Nov 2017 19:01:24 GMT

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

Yes, you could equalize the weighted number of cases in age group 15-24 rather than the number of cases in the full sample. That would make sense.

You can re-weight in proportion to the population sizes, but if you do that you will invariably find that one or two large countries are dominating the results. I would recommend giving equal weight to each survey, but that's really up to you.

You should be cautious about producing, say, an "estimate" that refers to a nonexistent population. A pooling of surveys from many different countries, with surveys conducted at different times, is hard to justify or interpret. At DHS we avoid doing that unless the outcome is so rare that this is the only way to get any statistical power.

Yes, you should use svyset, but the strata in different surveys have to have unique id codes and the clusters in different surveys also need unique id codes. If, say, the surveys are numbered 1, 2, 3, etc., you can have commands such as "egen stratumid=group(survey v023)" and "egen clustered=group(survey v021)". Here I am using v023 as the stratum code for each survey, but that's not always correct.

Subject: Re: Weighting across surveys when only including youth in analysis
Posted by cgreenba on Mon, 20 Nov 2017 14:46:31 GMT
View Forum Message <> Reply to Message

Thank you very much for the response. That is helpful. Just to clarify, I am not attempting to produce an "estimate" for the pooled data or any nonexistent population. The only estimates I want to produce are for levels of wealth quintile among the 15-19 and 20-14 age groups separately for each survey. The pooled data will only be used for the regression analysis to determine the relationship between wealth quintile and demand satisfied for modern family planning.

One other question I had is whether there is a need to modify the the svyset code to account for clustering in countries. I am looking at multiple surveys per country, so would using the standard "svyset psu [pw=weight], strata(strata)" take this into account? Is there a way to add a country-level strata to the code?

Any help would be appreciated. Thank you!

Subject: Re: Weighting across surveys when only including youth in analysis
Posted by Reduced-For(u)m on Tue, 21 Nov 2017 04:03:37 GMT
View Forum Message <> Reply to Message

I think I answered some of these in a PM, but maybe you didn't get it.

---

1. The nonexistent population problem comes in the second step when you have to decide how to weight the regression for how the wealth quintile affects family planning needs met...are you doing that separately for each survey? If not, that is where the question of population weighting comes in.

2. What does "level of wealth quintile" mean? Level of what? The underlying wealth index? That is not comparable across countries.

3. For clustering on country, generate a "country" variable, and replace "PSU" with "country" in your command line. Or, drop that part, and in your regression include " , cluster(country)". If you mean for the first part where you are estimating "levels of quintiles" then it doesn't matter at all to have the clustering command since it sounds like you are just plugging those estimates into the RHS of a second regression (note: you'd mess up your standard errors in your second stage regression (on family planning needs) that way by not accounting for the uncertainty in your first step estimates, but the only obvious way to deal with that is to use a two-stage bootstrap..that's another post or you can PM me about it).

4. I'm still worried about comparability of "wealth quintile" across time and space. You can't really compare it across countries...but that is beyond the scope of these questions.

Subject: Re: Weighting across surveys when only including youth in analysis
Posted by cgreenba on Wed, 06 Dec 2017 22:10:04 GMT
View Forum Message <> Reply to Message

Hi,

Thank you so much for your response. I apologize for getting back so late, but I am picking the analysis back up and wanted to respond and verify a few things.

1) About the weighting and non-existent population, what I am doing in pooling together the data from the different countries and producing both country-level estimates for the effect on wealth quintile on demand satisfied for family planning as well as a pooled regression analysis with data from all countries. This isn't with the aim for producing a precise estimate for how wealth quintile impact demand satisfied for family planning, but to illustrate in general term what was found when the collection of surveys that were used. I weight each survey by sample size, and not relative size of the population, since I am not attempting to produce some sort of global or regional estimate and instead what to represent what was found across the different surveys included. Does this seem like the right approach?

2) By level of wealth quintile, I simply mean poorest, poorer, middle, richer, and richest quintiles.

3) To cluster the survey data by country, I have used the follow svyset: svyset survey_id, strata(country_id) weight (survey_weight) || psu, strata(strata_id) weight (individual_weight). My goal in using this was to set the survey design so that each survey is weight by in sample size (i.e. survey_weight) and clustered by country (i.e. country_id), which still preserving the psu and strata

breakdown for the individual surveys, as well as the individual weights (i.e. v005). So far it seems to be functioning properly, but I would love to verify with others. Does this seem to make sense?

4) As far as the comparability of wealth quintile, while being in the poorest quintile in one country may be extremely different from being in the poorest quintile in another country, using wealth quintile can still tell us something about the effect of being in the poorest or richest quintile of any country on family planning. Isn't this correct? The overall question is more about equity in family planning across countries than have any specific assets or income, which would be harder to compare. I would love to hear any other thoughts on this though.

Thank you so much!

Subject: Re: Weighting across surveys when only including youth in analysis
Posted by Bridgette-DHS on Fri, 08 Dec 2017 16:10:31 GMT
View Forum Message <> Reply to Message

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

In my view, there is no single "correct" way to set up this kind of analysis. What you have described, including the multi-level svyset specification, is within the range of what I would consider acceptable alternatives and justifications.

Subject: Re: Weighting across surveys when only including youth in analysis
Posted by cgreenba on Mon, 11 Dec 2017 02:25:44 GMT
View Forum Message <> Reply to Message

Great! Thank you for the response. I really appreciate it.

Subject: Re: Weighting across surveys when only including youth in analysis
Posted by Reduced-For(u)m on Mon, 11 Dec 2017 21:54:01 GMT
View Forum Message <> Reply to Message

I don't have much to add other than to agree with Tom above. Not that I'd suggest you listen to me over Tom anyway, but I do agree - it sounds like you have a pretty good handle on what your regressions can and cannot tell you and what you are trying to get done. Everything else sounds totally reasonable - and yes, it does make some sense to compare within-country wealth differences, even averaging those within-coutnry estimates up. I just wanted to make sure you were interpreting that right, and it sounds like you've thought about it very clearly.