Subject: Weighting Data Question
Posted by nholla on Mon, 23 Nov 2015 14:23:43 GMT
View Forum Message <> Reply to Message

Hello DHS,

I'm currently trying to define my weighting and strata procedures in STATA and had a couple of questions. I'm appending 18 African countries over 3 decades (1980-2010) for a total of 54 total surveys for this analysis (I'm looking at nutrition and child mortality):

I'm using the BR file for each country and year, and am running the following code (before pooling) to denormalize the weights:

summarize v005
scalar T = r(sum)
gen weight = v005*fempop/T

Where fempop is the female 15-49 population for a particular country. My question is, when summing up the weights, should I be summing up the weights over the BR file or should I be summing them up over the observations in the IR file (since v005 is duplicated for each child of a women in the BR file I wasn't sure)?

I'm also a bit confused on how to define my strata variable. I'm using the following code:
gen strata = v101 * v102
egen strata_pooled=group(survey strata)

where "survey" is an identifier of the country and year of the survey. Assuming v101 always represents region and v102 always urban/rural, can I safely do this for the 18 countries over three decades that I am working with? It seems like I should be looking at Appendix A in the final report for the sampling methodology, but several of the reports are in French so it's hard for me to discern what's happening in these particular countries.

Thank you for all your help!!

Subject: Re: Weighting Data Question
Posted by Bridgette-DHS on Wed, 02 Dec 2015 13:08:53 GMT
View Forum Message <> Reply to Message

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

The BR files include all the children in the birth histories, even children who have died, grown up, left home, etc. Child survival is relevant for all of them, but nutrition data are only available for children under five. Perhaps you should restrict to children 0-4, in which case the BR file becomes equivalent to the KR file. Incidentally, this would greatly reduce the size of your combined file.

If you are working with children, why would you want to re-scale to the number of women 15-49? Another possibility would be to construct the sum of v005 limited to children who are alive at the time of the survey, and then scale up according to estimated number of living children under five in the UN Population Division spreadsheets on July 1 of the calendar year which contains the median date of date collection. (The UN spreadsheets give the number of living children; it would be a little harder to estimate the number of children born in the past five years, including those who died.) This would mean replacing "fempop" with "childpop", say, and you get the sum of v005 from the BR file, not the IR file.

You should also check the age range of the children for whom you have height and weight. It's usually 0-4, as I said above, but not always. You could get this as the maximum value of b8 in the KR file. If by "nutrition" you mean something other than anthropometry, such as recent consumption of foods and liquids, you need to check the age range on those variables.

The stratum variable is usually combinations of region and urban/rural, as you say. If it was something else, that rule should give a very good approximation. These are v023 (usually) and v025, respectively. Don't combine them with a product. Note, for example, that "10" could either urban region 10 or rural region 5. Within a single survey, use "egen stratum=group(v023 v025)". Better, wait until you have appended all the files, and assigned them survey numbers (1, 2, ..., 54), and then enter "egen stratum=group(survey v023 v025)".

Subject: Re: Weighting Data Question
Posted by nholla on Wed, 16 Dec 2015 22:21:49 GMT
View Forum Message <> Reply to Message

Thank you for your detailed reply, and I apologize for not replying sooner.

We definitely can use the number of living children under 5 instead. Part of the confusion lies in the fact that we are doing a cox regression for one of our analyses, and as such are looking at the entire history of children born (both dead and alive) to a particular mother and are restricting our samples to different age groups (1-5 in one analysis and 1-15 in another). Would you recommend the # of children currently alive (assuming this would need to be changed depending on the age group I'm restricting to) to be a better way to rescale the data?

Did you mean grouping v024 and v025 ?(not v023, which I thought already represents the strata in some countries). That's what I've seen elsewhere on this forum, so I wanted to make sure. I also was wondering if I could use v101 and v102 in lieu of v024 and v025. In the earlier surveys that use phase I, these variables seem to be missing much of the time.

Thanks for all your help!

Subject: Re: Weighting Data Question
Posted by Bridgette-DHS on Tue, 12 Jan 2016 18:35:15 GMT
View Forum Message <> Reply to Message

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

For the cox regression you describe, I would agree that children born would be more appropriate than children surviving. That would correspond with the initial size of the birth cohort (real or synthetic). And yes, I should have recommended the grouping of v024 (NOT v023!) and v025. Sorry about that!