

---

Subject: When/How to use (de-normalized) weights for pooled data analysis (four waves of BD child anthro data)

Posted by [annadinny](#) on Thu, 13 Mar 2014 14:49:03 GMT

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

---

Hello,

I am interested in getting region-level means for the child anthropometric variables from the Bangladesh DHS 1999/2000, 2004, 2007, and 2011. As my base, I am using the PR files for 2004, 2007, and 2011 and the KR file for 2000 (since only children of interviewed women were measured/weighed). Also, for 2004 and 2007, I am using the HW files to get the anthro variables related to the new WHO standards, which I merge with the base files.

Once I pool the data, I understand (from the note by Rulin Ren) that I need to de-normalize the weights (hv005) using the following formula:  $hv005\_denorm = hv005 \times (\#residential\ HH\ in\ country\ at\ time\ of\ survey) / (\# HH\ interviewed\ in\ survey)$ .

At this point, I am not sure how to use the survey weights in calculating the region-level means. My plan is to do the following: first calculate household means; then calculate region-level means. Do I use the de-normalized weights when calculating the household means or only when calculating the region-level means?

Also, I want to confirm that this is the correct code to use to set the survey design in STATA, where hv005\_denorm is the de-normalized weight, hv021 is the primary sampling unit variable for BD, and hv023 is the strata id variable for BD (note: even though stratification changes over time, my understanding is that these variables capture the correct survey design for each survey year):

```
svyset [pweight=hv005_denorm], psu(v021) strata(v023)
```

Any suggestions would be much appreciated. Thank you.

Anna

---

---

Subject: Re: When/How to use (de-normalized) weights for pooled data analysis (four waves of BD child anthro data)

Posted by [Reduced-For\(u\)m](#) on Thu, 13 Mar 2014 18:50:05 GMT

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

---

Hi Anna,

Well... this is a tough one. I have a couple of thoughts, but there isn't really an agreed upon answer as best as I can tell.

First - by switching recodes, you might be accidentally switching weighting schemes. Why can't you use the child recode for all? (I forget, I thought I had used all 4 child recodes from Bangladesh).

Second - what 'weight' do you want each total survey to have? Equal? you could just sum up the weights from each survey round, divide the old weight by the sum-of-survey-weights, and have within-survey-preserving-pweights that will sum to 1 for each survey.

Third - for regions, you could do the above, but at the region level (sum up regionXsurvey, divide, so each region sums to 1). Then do a weighted summarize (or collapse, with weights, by regionXsurvey) to recover weighted region means. Not sure the value of household means as an intermediate - at least not if you use the child recode for everything.

Fourth - I would create new strata and new PSUs for each survey round - I usually, say, multiply the survey year by 10,000 and add the psu to get a unique psu for each survey round, and something similar for strata. I don't think you want to lump them together.

Some small things: you can use a new stata package "zscore06" to get the new WHO standards fro all surveys using raw height/age/gender data. Why would you only want those for the 04/07? Or are the newer variables merged into the older files already? Seems weird to use two different standards.

Running a bit out of time (train WiFi!), but if this isn't clear, write back and I'll try again. Also, you don't have to take my advice. This is just one way to think about it and do it. Like I said, I don't think there is full agreement on this.

---

Subject: Re: When/How to use (de-normalized) weights for pooled data analysis (four waves of BD child anthro data

Posted by [Sarah-DHS](#) on Fri, 14 Mar 2014 12:35:34 GMT

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

Dear Anna,

If the previous post has not answered your question or you require further assistance, please let us know.

Thanks,  
Sarah at DHS

---

Subject: Re: When/How to use (de-normalized) weights for pooled data analysis (four waves of BD child anthro data

Posted by [annadinny](#) on Mon, 17 Mar 2014 19:53:48 GMT

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

Thanks for your response.

Here are some responses/thoughts:

1. I want data on all children in the HH (in PR file), not only those whose mothers completed interviews (in KR child recode file).

2. and 3. I want population estimates so each regional mean is representative of the population of that region in that year. I see two ways to update the weights: (1) updating them so that they sum to 1 for each survey year (as you suggest in 2 below) and (2) de-normalizing the weights so that they can be used to get population estimates. I guess that I'm not sure that I understand when to do (1) and when to do (2). From what I see on the Forum, when pooling survey data, one should do (2).

4. Yes, this makes sense - hadn't realized it before. So before setting the survey design in STATA, I will create new strata and psu variables that uniquely identify each strata-psu-year combination. Thanks.

Small things: I am going to use the new WHO vars - they are included in the HW files for earlier years and I have merged them with the other data.

Thanks again for your thoughts.

---

Subject: Re: When/How to use (de-normalized) weights for pooled data analysis  
(four waves of BD child anthro data)

Posted by [Reduced-For\(u\)m](#) on Mon, 17 Mar 2014 21:00:00 GMT

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

Hi Anna,

Not a whole lot I can offer, but as for your options 1/2 on weighting: what I suggest is a kind of de-normalizing, just one where you only want to preserve relative within-survey probability of selection. The other version you mentioned earlier tries to re-adjust the weights from the household weights to get a "fraction of total households in the area that were surveyed" and would, I think, allow you to use a fully pooled sample and make that "population representative" by (sort of) accounting for changing population over time (supposing you had different Nhouseholds estimates for each survey year). But since you want separate estimates for each region-X-survey (yes?), you shouldn't need to do that.

Like I said, I tend to think of each survey as population-representative, and usually conceive the population as being static (I don't want to weight my 2000 results less than my 2007 just because the population was larger in 2007). But that is a preference, not a rule.

Question - do you really need to pool these surveys at all to do what you want? Why not just collapse each survey down into regional cells, weighting by the given weights? Oh, except that 2004 claims (in the coding, but not the documentation) it is not regionally representative, which is odd.

Also, a last comment on stratification (which obviously I'm still not great on): accounting for strata shouldn't affect point estimates, only standard errors. In fact, accounting for strata should

decrease the size of your se's a little bit. So you could be "conservative" with your p-values/CIs and not account for strata at all. You could even, to be more conservative, skip the "svyset" part, specify the weight in the regression itself, and use the cluster robust se option clustered on PSU. Those SEs will be, maybe, slightly too big (conservative), but not by much, and if they are significant you should know that regardless of how you stratify they will remain significant (I know, this is a somewhat non-compelling argument for many reasons, but I find stratification specification doesn't matter too much here).

Can I ask what you want region-X-survey-round estimates for? You'll end up with like  $N=10\text{regions}\times 3\text{-surveys} = 30$  mean anthropometric measurements. Not a lot to do further analysis on, right? I ask because I have special interest in Bangladesh, but would totally understand if you had some really cool idea you didn't want to share yet.

---

Subject: Re: When/How to use (de-normalized) weights for pooled data analysis (four waves of BD child anthro data)

Posted by [annadinnyc](#) on Tue, 18 Mar 2014 14:56:04 GMT

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

Thanks for your response.

I would like to understand DHS data better for this project (looking at food security in Bangladesh over time) and for future projects. For my current needs, you are right, I can just collapse each year by region and then append the collapsed data sets. But I want to make sure that I understand the sample design for each round (hence the questions on svyset, strata, etc.

Question - what 2004 coding are you talking about? I didn't know that DHS posted code anywhere, except in some responses on this forum.

Anna

---

Subject: Re: When/How to use (de-normalized) weights for pooled data analysis (four waves of BD child anthro data)

Posted by [Reduced-For\(u\)m](#) on Tue, 18 Mar 2014 19:40:13 GMT

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

Hey - I just meant (by coding) the values for V023 (and repeat variables). My data has that variable equal to 0 for 2004, and the documentation says that 0 means non-representative for any sub-national region. The Bangladesh report, though, seems to say it is representative, but that looks a bit like copy/paste from other reports. So I really have no idea what to believe...

---

Subject: Re: When/How to use (de-normalized) weights for pooled data analysis (four waves of BD child anthro data)

Posted by [Liz-DHS](#) on Wed, 19 Mar 2014 02:41:21 GMT

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

---

Dear Ana,  
Do you still need assistance or have questions? Let us know if you do?  
Thanks!

---