Subject: Multilevel Weights
Posted by swinter on Tue, 11 Aug 2015 17:16:43 GMT
View Forum Message <> Reply to Message

Greetings DHS experts!

I recently posted a question on the forum about the domestic violence weights that is somewhat related to this post, but is focused on different questions:

(http://userforum.dhsprogram.com/index.php?t=tree&th=4410
&goto=6918&S=85ccc984d9dd623f59e3088f0d1e73c9#msg_69 18)

This post is meant to focus on the multilevel nature of the DHS weights (or, seemingly, the lack thereof). I am conducting a number of two and three-level multilevel modelling using both individual-level country data (two-level models) and multi-country data (three-level models). I have been doing the modeling largely in Stata 14 (using melogit) and MLwin 2.34 (using PQL and MCMC). Unfortunately, both of these softwares provide very specific instructions in their manuals to use multilevel sampling weights instead of a single-level weight when conducting multilevel analyses (except MCMC in MLwin or crossed mixed effects models, which don't allow weights). For an example of these instructions, see following passage from Stata 14 manual:

"it is not sufficient to use the single sampling weight wij , because weights enter the log likelihood at both the group level and the individual level. Instead, what is required for a two-level model under this sampling design is wj , the inverse of the probability that group j is selected in the first stage, and wi|j, the inverse of the probability that individual i from group j is selected at the second stage conditional on group j already being selected. You cannot use wij without making any assumptions about wj. Given the rules of conditional probability, wij = wjwi|j. If your dataset has only wij , then you will need to either assume equal probability sampling at the first stage (wj = 1 for all j) or find some way to recover wj from other variables in your data; see Rabe-Hesketh and Skrondal (2006) and the references therein for some suggestions on how to do this, but realize that there is little yet known about how well these approximations perform in practice. What you really need to fit your two-level model are data that contain wj in addition to either wij or wi|j. If you have wij--that is, the unconditional inclusion weight for observation i, j--then you need to divide wij by wj to obtain wi|j" (Stata 14 Manual - "meglm -- Multilevel mixed-effects generalized linear model, p.21 available at: http://www.stata.com/manuals14/memeglm.pdf#memeglmMethodsand formulas)

From my reading of the DHS sampling literature, the multilevel nature of the DHS sampling is particularly important in the domestic violence sampling weights because, unlike the other weights (v005 and hv005), individual women sampled for the dv module do not have the same weight as the households. So, it seems to me, that for a two-level model, there should be, at a minimum, a PSU-level (level 2) weight and an individual-level (level 1) dv weight that incorporates the dv sample design and non-response. Does anyone have suggestions about how to tackle the multilevel weighting issue? Should individuals interested in multilevel modeling just assume the sampling probability at the first-stage (e.g. the PSU weight) is equal for all PSUs (e.g. wj=1 for all j)? Or, should we try to "recover" the multilevel weights using the technique cited in Rabe-Hesketh and Skrondal (2006) or another technique? Will DHS provide methodology for extracting the multilevel weights (I did try doing this based on the equations in the DHS sampling manual, but I

have more unknowns than equations, particularly in regards to back-calculating the domestic violence weights). Any other thoughts?

Thanks so much!

Subject: Re: Multilevel Weights
Posted by Liz-DHS on Tue, 18 Aug 2015 16:28:52 GMT
View Forum Message <> Reply to Message

Dear User,
Here is a response from on of our technical experts, Dr. Tom Pullum:
Quote:We agree that multi-level analysis should include weights at each level. Until Stata 14, multi-level models in Stata could not use weights at all. Beginning with Stata 14, weights (multi-level weights) are allowed and should be used (as in MLwin). The problem is that DHS does not have cluster-level weights. The clusters are sampled with probability proportional to size, within strata. The sizes of the clusters, usually census enumeration areas, are part of the sampling frame, usually the most recent census. The sampling frame is not public information. DHS only has access to it within the country and is not allowed to make a copy of it. Statistical offices typically do not want to share it.

It is conceivable that there is some way to approximate the cluster-level weights or at least to get away from the invalid assumption that the first stage sampling probabilities are the same for all clusters. We will look into this. Thanks for raising the issue.

Thank you!

Subject: Re: Multilevel Weights
Posted by mmr-UMICH on Wed, 27 Jan 2016 04:05:36 GMT
View Forum Message <> Reply to Message

Dear Survey Sampling Expert,

I often come to this user forum to see if you please finally gave any recommendation about to calculate/adjust both the cluster level and the household/individual level weights for multilevel analysis. Thanking you with great appreciation for this matter.

Moshiur Rahman

Subject: Re: Multilevel Weights
Posted by Liz-DHS on Wed, 27 Jan 2016 17:39:54 GMT
View Forum Message <> Reply to Message

Dear User,
There is nothing new to report in this area. Thank you!

Subject: Re: Multilevel Weights
Posted by ttuti on Thu, 16 Jun 2016 09:35:08 GMT
View Forum Message <> Reply to Message

Hello.

Please explain to me what "...all sample weights are normalized such that the weighted number of cases is identical to the unweighted number of
households when using the full dataset with no selection...." means for DHS column V005. Can I use these weights in MLwiN as 'raw weights' at household level or are they to be treated as standardised weights?

Thank you for your help.

Subject: Re: Multilevel Weights
Posted by Liz-DHS on Wed, 29 Jun 2016 15:15:29 GMT
View Forum Message <> Reply to Message

Dear User,
A response from sampling expert, Dr. Mahmoud Elkasabi:
Quote:
This means that V005 is the normalized version of the sampling weight. The main purpose of the normalization process is to avoid the large values for the number of weighted cases in the tables in DHS survey final reports. This applies for all the DHS standard weights, including weights for households, such as HV005, and individuals. The V005 was calculated by multiplying the individual sampling weight by a normalization factor at the national level. The normalization factor is the total number of completed cases divided by the total number of weighted cases. In case of the V005, it is the total number of completed women divided by the weighted total number of completed women. In case of the HV005, it is the total number of completed households divided by the weighted total number of completed households. Therefore the standard weights in the DHS data files are relative weights. Relative weights can be used to estimate means, proportions, rates and ratios because the normalization factor is cancelled out when used in both numerator and denominator, so it has no effect on the calculated indicator values. However, the standard weights are not valid for estimating totals. Also the normalized weight is not valid for pooled data, even for data pooled for women and men in the same survey, because the normalization factor is country and sex specific.

Subject: Re: Multilevel Weights
Posted by e_lee on Fri, 23 Sep 2016 18:16:29 GMT
View Forum Message <> Reply to Message

Dear DHS,

I am following up to the earlier thread in this exchange. I am interested whether a work-around for the lack of cluster weights in the household survey data has been identified, to be used in

multilevel modeling in Stata 14.

Many thanks for your assistance on this!

Subject: Re: Multilevel Weights
Posted by Liz-DHS on Mon, 26 Sep 2016 15:37:12 GMT
View Forum Message <> Reply to Message

Dear User,
Thank you for your post. There is still nothing new to report in this area.

Subject: Re: Multilevel Weights
Posted by soumava on Fri, 12 May 2017 20:01:24 GMT
View Forum Message <> Reply to Message

Dear DHS,
I am interested in any update on the issue raised in this thread: ways to approximate the cluster level weights.
Thanking you,
Soumava Basu

Subject: Re: Multilevel Weights
Posted by Liz-DHS on Wed, 28 Jun 2017 20:55:12 GMT
View Forum Message <> Reply to Message

A response from Senior Sampling Expert, Dr. Ruilin Ren:
Quote:
Unfortunately, there is nothing new. With the confidentiality requests from the DHS protocol, we cannot provide the selection probabilities (we cannot keep them) of different levels which are the components of the multi-level weights. So the question is not a technical one, but rather confidential obligations.
Thanks
Ruilin

Subject: Re: Multilevel Weights
Posted by ab803 on Fri, 13 Oct 2017 15:18:36 GMT
View Forum Message <> Reply to Message

Dear DHS experts,

What is the best way to do a multilevel analysis which includes individuals within households within clusters? Should we be including the weight at the individual level of this model?

Many thanks!

Subject: Re: Multilevel Weights
Posted by Liz-DHS on Wed, 08 Nov 2017 19:52:26 GMT
View Forum Message <> Reply to Message

Dear User, A response from Dr. Tom Pullum:
Quote:
Individuals in households are not sampled. In the survey design, clusters are sampled and then households are sampled within clusters. After the household has been selected, all eligible respondents (based on age, sex, and de facto residence) are selected.

Ideally we would provide separate sampling fractions (or their inverse, the weights) for the clusters and then the households. At this time it is not possible to do a full multilevel model because we can only provide the product (hv005, etc.) after an adjustment for nonresponse. As has been stated in other responses, for privacy reasons we do not save the more detailed information.

Subject: Re: Multilevel Weights
Posted by dflood011 on Mon, 13 Nov 2017 22:12:33 GMT
View Forum Message <> Reply to Message

Thank you to Tom for his incredibly helpful comments about the challenges in constructing multilevel models with DHS data, on this thread and others.

If I may ask a follow-up question to Tom:

I am interested in building a model to explore the association between maternal contraception and child stunting in a single-country DHS dataset. I originally thought about constructing a mixed model using cluster, household, and individual levels, but given Tom's comments about the unavailability of disaggregated household + cluster weights, I am thinking it would be better to carry out a classic regression in Stata using the examples set forth in Heeringa's text, "Applied Survey Data Analysis."

Yet I'm still drawn to the elegance of a hierarchical model, and when I search PubMed, I find numerous examples of papers in high-quality journals using mixed models for DHS data -- often asserting PSU weight-adjustment. (See examples below.) How is this possible? Are these analysts making errors?

https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5291969/
http://onlinelibrary.wiley.com/doi/10.1111/mcn.12463/full
https://www.ncbi.nlm.nih.gov/pubmed/26378858
http://journals.plos.org/plosone/article?id=10.1371/journal. pone.0037905

Subject: Re: Multilevel Weights
Posted by Liz-DHS on Wed, 04 Apr 2018 19:50:27 GMT
View Forum Message <> Reply to Message

Dear User, Do you still need input from Dr. Pullum?

Subject: Re: Multilevel Weights
Posted by Hassen on Mon, 30 Apr 2018 18:53:43 GMT
View Forum Message <> Reply to Message

Dear DHS Experts, Can Multilevel Modeling be conducted Using DHS Data set by using SPSS? Can I make Weighting at each levels (community,household and individual levels)? I already planned to use Multilevel modeling to investigate factors associated with childhood nutritional status using 2016 Ethiopia DHS Data set. What are your recommendation to conduct it using SPSS?
I need your reply.
Thank you in advance!!

Subject: Re: Multilevel Weights
Posted by Liz-DHS on Wed, 02 May 2018 03:28:59 GMT
View Forum Message <> Reply to Message

Dear User, A response from Dr. Tom Pullum:
Quote:
The DHS analysis group uses Stata, sometimes R. We do not use SPSS and we cannot help you with the syntax of multilevel commands in SPSS. Regarding weights, we do not have separate weights for the different levels, only the net weight, which is proportional to hv005 or v005. We intend to develop recommendations for how to partition this into cluster-level and household-level weights but are not yet prepared to suggest anything specific.

Subject: Re: Multilevel Weights

Posted by Hassen on Wed, 02 May 2018 07:14:48 GMT
View Forum Message <> Reply to Message

Thank you very much for your Attractive Response!
I am waiting to get your reply on How to partition the samples into cluster-level and household-level!! I will come up with My Challenges after I have seen all issues regarding Multilevel analysis etc.
Again,Thank you in advance!!

Subject: Re: Multilevel Weights
Posted by dflood011 on Fri, 29 Jun 2018 20:26:33 GMT
View Forum Message <> Reply to Message

Hi Liz, I would still be very interested in Dr. Pullum's response.

Subject: Re: Multilevel Weights
Posted by Liz-DHS on Wed, 04 Jul 2018 03:04:18 GMT
View Forum Message <> Reply to Message

A response from Dr. Shireen Assaf:
Quote:

Dear user,

To be able to use multilevel modeling with DHS data and the svy command, a weight must be applied for each level. Since we only have one weight, we can make the assumption that all individuals in the household have the same weight. The Stata code to run a multilevel model for a binary outcome is below using the melogit command. However, you can use other mixed model commands if your outcome had a different distribution. The svyset should be the same.

```
gen wt=v005/1000000
gen wt2=1
svyset v001, weight(wt) strata(v023) , singleunit(centered) || _n, weight(wt2)

*this is a random intercept model
svy: melogit outcome var1 var2 || v001:

*random intercept and random slop for var3
svy: melogit outcome var1 var2 || v001: var3
*for this model you can also add the covariance(unstructured) option. Please read the stata documentation for the melogit command on this
* you can also check the Stata documentation for svyset, they discuss how to construct this for multistage sample design.
```

Hope this helps.

Best,

Shireen Assaf
Technical Specialist
The DHS Program

Subject: Re: Multilevel Weights
Posted by teketo on Thu, 04 Oct 2018 10:35:16 GMT
View Forum Message <> Reply to Message

Dear DHS,

I am doing analysis on maternal health service use (modern contraception among married women, antenatal care, health facility delivery and caesarean section). I am analysing the 2016 Ethiopia DHS data using SAS software.
I am just wondering how I can use Multilevel weights for this analysis. It will be great if I can get solutions how it will be done using SAS.

I am using a two level random effects model (Level 1: individuals, and Level 2: regions) for modern contraception among married women, antenatal care and health facility delivery. Moreover, for caesarean section, a three level random effects model (Level 1: individuals, Level 2: clusters and Level 3: regions) using MCMC method.
The other question I have is what should be the maximum number of a grouping (membership) variable? To make it clear, just have a look the following SAS Proc glimmix program:

```
Proc glimmix data = care;
Class region V001;
Model y (event = last) = a b c d e f g h i / solution;
Random intercept e f g / solution subject = region;
Random intercept a b c d / solution subject = V001 (region);
Run;
```

The grouping variables showing how one is nested within the other on the above SAS code are region level 3 (there are 11 regions) and V001 level 2 (there are 622 clusters). We are asking the program to produce random effects for each of the 11 regions and the 622 clusters. It will have a convergence issue it will take long time to process and even might stop processing.

Regards
Teketo

Subject: Re: Multilevel Weights
Posted by aheto on Sat, 04 Apr 2020 16:00:33 GMT

Hi Dr. Shireen Assaf,
Thanks for this response.

How can I implement your suggestions below in R for the standard multilevel you presented, and also for Bayesian multilevel model? Any R codes to work with?

gen wt=v005/1000000
gen wt2=1
svyset v001, weight(wt) strata(v023) , singleunit(centered) || _n, weight(wt2)
*this is a random intercept model
svy: melogit outcome var1 var2 || v001:

Thanks

Subject: Re: Multilevel Weights
Posted by Liz-DHS on Fri, 17 Apr 2020 13:17:22 GMT

A response from Dr. Shireen Assaf:
Quote:
Hello,
Unfortunately I do not have any R code for this. However, please study the survey package in R and you may find an answer.
Thank you.
Best,
Shireen