
Subject: weighting issues of multilevel modelling using the DHS survey data with multiple-stage sampling

Posted by [YUJP](#) on Mon, 03 Sep 2018 09:19:59 GMT

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

Dear DHS expert,

I am reading extensively the historical and current forum discussion on the weighting issues of multilevel modelling using the DHS survey data with multiple-stage sampling. I would appreciate if you can help to enlighten me on the following question:

Basically, I am using a multilevel model t analysis dataset from DHS Cambodia 2014 with the outcome of the children under five diarrhoea and predictors at both the level of the children as well as the level of cluster (PSU). I am using the "melogit" command for this analysis (same results can be produced using the "meglm" command. I plan to use the scaling methods (methods A or B) as proposed by Sophia Rabe-Hesketh (2006) (http://www.gllamm.org/JRSSAsurvey_06.pdf) and Adam C Carle (2009) (<https://bmcmmedresmethodol.biomedcentral.com/articles/10.1186/1471-2288-9-49>). One problem is that in the DHS database we only have the weight (v005 or hv005) that has taking the two stage sampling (cluster (PSU) and women (or household) into consideration. As it was stated in the STATAMULTILEVEL MIXEDEFFECTS REFERENCEMANUAL RELEASE 15 (page 104) (<https://www.stata.com/manuals/me.pdf>), we don't have W_j or W_{ij} but only W_{ij} :

"Now take these same data and fit a two-level model with meglm, it is not sufficient to use the single sampling weight w_{ij} , 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 w_j , the inverse of the probability that group j is selected in the first stage, and w_{ij} , 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 w_{ij} without making any assumptions about w_j .

Given the rules of conditional probability, $w_{ij} = w_j w_{ij}$. If your dataset has only w_{ij} , then you will need to either assume equal probability sampling at the first stage ($w_j = 1$ for all j) or find some way to recover w_j from other variables in your data; see Rabe-Hesketh and Skrandal (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 w_j in addition to either w_{ij} or w_{ij} . If you have w_{ij} --that is, the unconditional inclusion weight for observation i ; j --then you need to divide w_{ij} by w_j to obtain w_{ij} ."

However, when I re-read the DHS report of Cambodia, I found that there are actually information on the distribution of enumeration areas in the sampling by strata. (page 282 Appendix A Table A2, Cambodia Demographic and Health Survey 2014: <https://dhsprogram.com/pubs/pdf/fr312/fr312.pdf>). If I call them C_j (j = strata 1, 2, ... 38)), as we can easily get the number of selected clusters per each strata, which I call them CS_j (j = strata 1, 2, ... 38)), it seems that I would be able to calculate the probability that the clusters in each strata were selected (CS_j/C_j) and thus the weight $W_j = C_j/CS_j$). With W_j , when I can calculate the w_{ij} which is W_{ij}/W_j .

I use the methods and the information in the Appendix of the report and recalculated the scaled weights and got a results which is a bit different from (but still very similar with) the results that was produced by using the wij (v005) and presume that the second level weight to be "1".

I would appreciate if you can guide me whether this is a valid solution to obtain the two level weights for the multilevel analysis using DHS data? Or at least this can provide a better (less biased) estimate of the parameters than the one using wij as the first level weight and presume the second level weight be "1"?

Many thanks in advance.