

---

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\\_6918](http://userforum.dhsprogram.com/index.php?t=tree&th=4410&goto=6918&S=85ccc984d9dd623f59e3088f0d1e73c9#msg_6918))

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  $w_{ij}$ , 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 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  $w_j$  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  $w_{ij}$ " (Stata 14 Manual - "`meglm` -- Multilevel mixed-effects generalized linear model, p.21 available at: <http://www.stata.com/manuals14/memeglm.pdf#memeglmMethodsandformulas>)

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.  $w_j=1$  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!

---