

Using design weights

In many studies, data arise from sample surveys in which units have been selected with known but unequal probabilities. In these cases, it will often be desirable to weight observations in order to produce unbiased estimates of population parameters. According to standard practice in such cases, the information from each unit is weighted inversely proportional to its probability of selection.

Suppose, for instance, that in a pre-election poll, ethnic minority voters are over-sampled to insure that various ethnic groups are represented in the sample. Without weighting, the over-sampled groups would exert undue influence on estimates of the proportion of voters in the population favoring a specific candidate. Use of design weights can yield unbiased estimates of the population parameters.

Design weights are also commonly used to correct for differential non-response of sub-groups. Response rates are estimated for relevant sub-groups, and information from each respondent is weighted inversely proportional to the probability of response. That way, respondents who are over-represented in a sample as a function of non-response are appropriately weighted down.

Design weighting in the hierarchical context

Hierarchical data can be described as arising from a multi-stage sampling procedure. For example, schools might be sampled from a national frame of schools and then, within each school, students might then be sampled from a list of all students attending the school. Probabilities at each level might be known but unequal. For example, one might over-sample private schools and then over-sample minority students within each school. Weights might be constructed at each level to be inversely proportional to the probability of selection at that level. In some cases, weights might be available at only one level. For example, in a two-level design with students nested within schools, one might compute the marginal probability that a student is selected as the product of the probability that student's school is selected multiplied by the conditional probability that the student is selected given that his or her school is selected. In another context, suppose persons are selected with known probability and then followed longitudinally over time. In this case, we have occasions at level 1 nested within persons at level 2. The only weight may be a level-2 weight, inversely proportional to the probability of selection of that person. It is, of course, possible to include level-1 weights as well, but it is common to have weights only at level-2 in such longitudinal studies.

HLM 6 uses a method of computation devised by Pfefferman et al. (1998) for hierarchical data. This method, based on weighting the information of each case

in the framework of maximum likelihood, is more appropriate than the method of weighting in earlier versions of HLM, which used a more conventional approach of weighting observations.

Weighting in two-level designs

In the two-level context, weights might be available at level 1, at level 2 or at both levels. If weights are available at level-1 only, the methodology used in HLM 6 assumes that these weights are inversely proportional to P_{ij} , the marginal probability of that student i in school j is selected into the sample. HLM 6 will then normalize the weight to have a mean of 1.0. Thus we have

$$w_{ij} = \frac{N / P_{ij}}{\sum_{j=1}^J \sum_{i=1}^{n_j} 1 / P_{ij}} \quad (2.1)$$

in which case

$$\sum_{j=1}^J \sum_{i=1}^{n_j} w_{ij} = N \quad (2.2)$$

where N is the total sample size of level-1 units. In contrast, if weights are available only at level 2, the methodology assumes that these weights are inversely proportional to P_j the probability of selection of the level-2 unit. In this case, HLM 6 will again normalize the weight to have a mean of 1.0, yielding

$$w_j = \frac{J / P_j}{\sum_{j=1}^J 1 / P_j}, \quad (2.3)$$

in which case

$$\sum_{j=1}^J w_j = J. \quad (2.4)$$

where J is the total number of level-2 units. If weights are available at both level-1 and level-2, the methodology assumes that the level-1 weight is P_{ij} , the conditional probability of selection of unit i given that unit j was selected, so that $P_{ij} = P_{ij} / P_j$. The level-2 weight is assumed to be inversely proportional to P_j . In this case, HLM will normalize the level-1 weight within level-2 units:

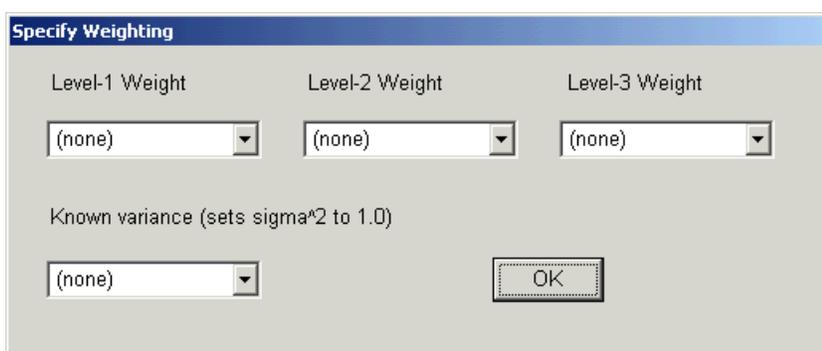
$$w_{ij} = \frac{n_j / P_{ij}}{\sum_{i=1}^{n_j} 1 / P_{ij}} \quad (2.5)$$

so that the sum of these weights within a level-2 unit will be

$$\sum_{i=1}^{n_j} w_{ij} = n_j \quad (2.6)$$

where n_j is the sample size of level-1 units in level-2 unit j .

In HLM6, weights are selected at the time of analysis, not when the MDM file is made. To select weights for an HLM2 analysis, select the **Estimation Settings** option from the **Other Settings** menu, and use the pull down menus to select the weighting variables at any level.



Weighting in three-level designs

In the three-level context, weights might be available at any one of the three levels, at any pair of them, or at all three levels. Normalization proceeds in a fashion completely analogous to that in the case of two levels. If weights are available only at level 1, we assume these are inversely proportional to P_{ijk}^1 , the marginal probability of selection of unit ijk . Similarly, if weights are available only at level 2 or only at level 3, the corresponding probabilities are P_{jk}^2 or P_k^3 , respectively. If the weights are at levels 1 and 2 but not 3, the corresponding probabilities are P_{ijk}^1 and P_{jk}^2 ; if at levels 2 and 3 (but not 1), the corresponding probabilities are P_{jk}^2 and P_k^3 ; if the weights are at levels 1 and 3 (but not 2), the corresponding probabilities are P_{ijk}^1 and P_k^3 . If weights are present at all three levels, the probabilities are P_{ijk}^1 , P_{jk}^2 and P_k^3 .

In HLM6, weights are selected at the time of analysis, not when the MDM file is made. To select weights for an HLM3 analysis, select the **Estimation Settings** option from the **Other Settings** menu, and use the pull down menus to select the weighting variables at any level.

Specify Weighting

Level-1 Weight	Level-2 Weight
(none) ▼	(none) ▼
Known variance (sets σ^2 to 1.0)	
(none) ▼	OK

No weighting options are available for the multivariate and cross-classified models fitted using HMLM, HMLM2, and HCM2.