## Types of models available in HLM

Using HLM2 and HLM3, a number of nonlinear (so-called HGLM) 2- and 3-level models may be fitted. The HMLM and HMLM2 modules allow estimation of multivariate normal models for incomplete data, while HCM2 is used for 2-level cross-classified random effects models. A brief description of the models available is given below.

## Models identified by type of outcome variable

### *Normal outcome variable*

In general it is possible to transform the level-1 predicted value, $\mu_{ij}$, to $\eta_{ij}$ to insure that the predictions are constrained to lie within a given interval. Such a transformation is called a link function. In the normal case, no transformation is necessary. However, this decision not to transform may be made explicit by writing

$$E(Y_{ij} \mid \mu_{ij}) = \mu_{ij} \qquad Var(Y_{ij} \mid \mu_{ij}) = \sigma^2.$$

The link function in this case is viewed as the "identity link function."

### *Binomial outcome*

When the level-1 sampling model is binomial, HGLM uses the logit link function

$$\eta_{ij} = \log\left(\frac{\phi_{ij}}{1 - \phi_{ij}}\right).$$

In words, $\eta_{ij}$ is the log of the odds of success. Thus if the probability of success, $\phi_{ij}$, is 0.5, the odds of success is 1.0 and the log-odds or "logit" is zero. When the probability of success is less than 0.5, the odds are less than one and the logit is negative; when the probability is greater than 0.5, the odds are greater than unity and the logit is positive. Thus, while $\phi_{ij}$ is constrained to be in the interval (0,1), $\eta_{ij}$ can take on any real value.

### *Bernoulli outcome*

Let $Y_{ij}$ be the number of "successes" in $m_{ij}$ trials. Then we write that

$$Y_{ij} \mid \phi_{ij} \sim B\left(m_{ij}, \phi_{ij}\right),$$

to denote that $Y_{ij}$ has a binomial distribution with trials and probability of success

$\phi_{ij}$. According to the binomial distribution, the expected value and variance of $Y_{ij}$ are then

$$E(Y_{ij} \mid \phi_{ij}) = m_{ij}\phi_{ij} \quad Var(Y_{ij} \mid \phi_{ij}) = m_{ij}\phi_{ij}\left(1 - \phi_{ij}\right).$$

When $m_{ij} = 1$, $Y_{ij}$ may take on values of either zero or unity. This is a special case of the binomial distribution known as the Bernoulli distribution. For the Bernoulli case, the predicted value of the binary $Y_{ij}$ is equal to the probability of a success, $\phi_{ij}$.

## *Poisson outcome*

Let $Y_{ij}$ be the number of events occurring during an interval of time having length $m_{ij}$. The time-interval of $m_{ij}$ units may be termed the "exposure." Then we write that

$$Y_{ij} \mid \lambda_{ij} \sim P(m_{ij}, \lambda_{ij})$$

to denote that $Y_{ij}$ has a Poisson distribution with exposure $m_{ij}$ and event rate $\lambda_{ij}$. According to the Poisson distribution, the expected value and variance of are then

$$E(Y_{ij} \mid \lambda_{ij}) = m_{ij}\lambda_{ij} \quad Var(Y_{ij} \mid \lambda_{ij}) = m_{ij}\lambda_{ij}.$$

A common case arises when, for each *i* and *j*, the exposure is the same. In this case, we set $m_{ij} = 1$ for simplicity. HGLM allows estimation of models in which $m_{ij} = 1$ or $m_{ij} >= 1$.

HGLM uses the log link function when the level-1 sampling model is Poisson, that is

$$\eta_{ij} = \log(\lambda_{ij}).$$

In words, is the log of the event rate. Thus, if the event rate, $\lambda_{ij}$, is one, the log is zero. When the event rate is less than one, the log is negative; when the event rate is greater than one, the log is positive. Thus, while $\lambda_{ij}$ is constrained to be non-negative, $\eta_{ij}$ can take on any real value.

## *Multinomial outcome*

HGLM uses the logit link function when the level-1 sampling model is

multinomial. Define $\eta_{mij}$ as the log-odds of falling into category $m$ relative to that of falling into category $M$. Specifically

$$\eta_{mij} = \log\left(\frac{\phi_{mij}}{\phi_{Mij}}\right)$$

where

$$\phi_{Mij} = 1 - \sum_{m=1}^{M-1} \phi_{mij}.$$

In words, $\eta_{mij}$ is the log odds of being in $m$-th category relative to the $M$-th category, which is known as the "reference category."

## Ordinal outcome

The level-1 structural model assumes "proportional odds",

$$\eta_{mij} = \beta_{0j} + \sum_{q=1}^{Q} \beta_{qj} X_{qij} + \sum_{m=2}^{M} \delta_m.$$

Under the proportional odds assumption, the relative odds that $R_{ij} \leq m$, associated with a unit increase in the predictor, does not depend on $m$.

Here $\delta_m$ is a "threshold" that separates categories $m$ -1 and $m$. For example, when $M = 4$,

$$\eta_{1ij} = \beta_{0j} + \sum_{q=1}^{Q} \beta_{qj} X_{qij}$$

$$\eta_{2ij} = \beta_{0j} + \sum_{q=1}^{Q} \beta_{qj} X_{qij} + \delta_2$$

$$\eta_{3ij} = \beta_{0j} + \sum_{q=1}^{Q} \beta_{qj} X_{qij} + \delta_2 + \delta_3$$

## Models identified by type of covariance structure:

### Unrestricted model

This model is appropriate when the aim of the study is to collect $T$ observations per aprticipant according to a fixed design. However, one or more observations may be missing at random. We assume a constant but otherwise arbitrary $T$ x $T$

covariance matrix for each person's "complete data".

The level-1 model relates the observed data, $Y$, to the complete data, $Y^{*}$:

$$Y_{hi} = \sum_{t=1}^{T} m_{thi} Y_{ti}^{*}$$

where $Y_{hi}$ is the *r*-th outcome for person *i* associated with time *h*. Here $Y_{ti}^{*}$ is the value that person *i* would have displayed if that person had been observed at time *t*, and $m_{thi}$ is an indicator variable taking on a value of 1 if the *h*-th measurement for person *i* did occur at time *t*, 0 if not. Thus, $Y_{ti}^{*}$, $t = 1, ..., T$, represent the complete data for person *i* while $Y_{hi}$, $h = 1, ..., T_{i}$ A are the observed data, and the indicators $m_{thi}$ tell us the pattern of missing data for person *i*.

### *Homogeneous model*

Under the special case in which the within-person design is fixed , with *T* observations per person and randomly missing time points, the two-level HLM can be derived from the unrestricted model by imposing restrictions on the covariance matrix, A. (Note: regressors  having varying designs may be included in the level-1 model, but coefficients associated with such values must not have random effects at level 2). The most frequently used assumption in the standard HLM is that the within-person residuals are independent with a constant variance, $\sigma^{2}$.

The level-1 model has a similar form to that in the case of the unrestricted model

$$Y_{i}^{*} = A\pi_{i} + \varepsilon_{i}, \quad \varepsilon_{i} \sim N(0, \Sigma)$$

with $\Sigma = \sigma^{2} I_{T}$.

### *Heterogeneous model*

One can model heterogeneity of level-1 variance as a function of the occasion of measurement. Such a model is suitable when we suspect that the level-1 residual  variance varies across occasions. The models that can be estimated are a subset of the models that can be estimated within the standard HLM2. The level-1 model is the same as in the case of homogenous variances, except that now

$$Var(e_{i}) = \Sigma = diag\{\sigma_{t}^{2}\},$$

that is, $\Sigma$ is now diagonal with elements $\sigma_t^2$, the variance associated with occasion $t$, $t = 1, \ldots, T$.

The number of parameters estimated is $f + r(r+1)/2 + T$. Now $r$ must be no larger than $T - 1$. When $r = T - 1$, the results will duplicate those based on the unrestricted model.

### *Log-linear model*

The model with varying level-1 variance, described above, assumes a unique level-1 variance for every occasion. A more parsimonious model would specify a functional relationship between aspects of the occasion (e.g. time or age) and the variance. We would again have $\Sigma = diag\{\sigma_t^2\}$, but now

$$\log\left(\sigma_t^2\right) = \alpha_0 + \sum_{k=1}^{H} \alpha_k c_{kt}.$$

Thus, the natural log of the level-1 variance may be a linear or quadratic function of age. If the explanatory variables $c_k$ are $T - 1$ dummy variables, each indicating the occasion of measurement, the results will duplicate those of the previous section.

The number of parameters estimated is now $f + r(r+1)/2 + H + 1$. Again, $r$ must be no larger than $T - 1$ and $H$ must be no larger than $T - 1$.

When this option is selected, the **Predictors of level-1 variance** button is activated. Click on this button to open a dialog box that can be used to select the predictors of the level-1 variance.

### *1st-order auto-regressive model*

This model allows the level-1 residuals to be correlated under Markov assumptions (a level-1 residual depends on previous level-1 residuals only through the immediately preceding level-1 residuals). This leads to the level-1 covariance structure

$$Cov(e_{ti}, e_{t'i}) = \sigma^2 \rho^{|t-t'|}.$$

Thus, the variance at each time point is $\sigma^2$ and each correlation diminishes with the distance between time points, so that the correlations are $\rho, \rho^2, \rho^3, \ldots$ as the

distance between occasions is 1, 2, 3,  ....

The number of parameters estimated is now $f + r(r+1)/2 + 2$. $r$ must be no larger than $T - 1$.

Note that level-1 predictors are assumed to have the same values for all level-2 units of the complete data. This assumption can be relaxed. However, if the design for $a_{pti}$ varies over $i$, its coefficient cannot vary randomly at level 2. In this regard, the standard 2-level model (HLM2) is more flexible than HMLM.

## Other models:

### *Cross classified random effects models*

A general random cross-classified model consists of two sub-models: level-1 or within-cell and level-2 or between-cell models. The cells refer to the cross-classifications by the two higher-level units. For example, if the research problem consists of data on students cross-classified by schools and neighborhoods, the level-1 or within-cell model will represent the relationships among the student-level variables, the level-2 or between-cell model will capture the influence of school- and neighborhood-level factors. Formally, there are $i = 1, 2, \ldots, n_{ijk}$ level-1 units (e.g., students) nested within cells cross-classified by = 1,..., $J$ first level-2 units (e.g., neighborhoods), designated as rows, and k = 1,..., $K$ second level-2 units (e.g., schools), designated as columns.

We represent in the level-1 or within-cell model the outcome for case $i$ in individual cells cross-classified by level-2 units $j$ and $k$.

$$Y_{ijk} = \pi_{0jk} + \pi_{1jk}a_{1ijk} + \pi_{2jk}a_{2ijk} + \cdots + \pi_{pjk}a_{pijk} + e_{ijk}$$
$$= \pi_{0jk} + \sum_{p=1}^{P} \pi_{pjk}a_{pijk} + e_{ijk}$$

where

$\pi_{pjk}$ ($p = 1, 2, \ldots, P$) are level-1 coefficients,

$a_{pijk}$ is the level-1 predictor $p$ for case $i$ in cell $jk$,

$e_{ijk}$ is the level-1 or *within-cell random effect*, and

$\sigma^2$ is the variance of $e_{ijk}$, that is the *level-1* or *within-cell variance*. Here we assume that the random term $e_{ijk} \sim N(0, \sigma^2)$.

### *Level-2 or "between-cell" model*

Each of the $\pi_{pjk}$ coefficients in the level-1 or within-cell model becomes an outcome variable in the level-2 or between-cell model:

$$\pi_{pjk} = \theta_p + \left(\beta_{p1} + b_{p1j}\right)X_{1k} + \left(\beta_{p2} + b_{p2j}\right)X_{2k} + \cdots + \left(\beta_{pQ_o} + b_{pQ_oj}\right)X_{Q_ok} +$$
$$\left(\gamma_{p1} + c_{p1k}\right)W_{1j} + \left(\gamma_{p2} + c_{p2k}\right)W_{2j} + \cdots + \left(\gamma_{pR_o} + c_{pR_ok}\right)W_{R_oj} +$$
$$\delta_{p1jk}Z_{1jk} + \cdots + \delta_{pS_ojk}Z_{S_ojk} +$$
$$b_{p0j} + c_{p0k} + d_{p0jk}$$
$$= \theta_p + \sum_{q=1}^{Q_o}\left(\beta_{pq} + b_{pqj}\right)X_{qk} + \sum_{r=1}^{R_o}\left(\gamma_{pr} + c_{prk}\right)W_{rj} + \sum_{s=1}^{S_o}\delta_{pgk}Z_{gk} + b_{p0j} + c_{p0k} + d_{p0jk}$$

where

> $\theta_{p00}$ is the model intercept, the expected value of $\pi_{pjk}$ when all explanatory variables are set to zero;
>
> $\beta_{pq}$ are the fixed effects of column-specific predictors $X_{qk}, q = 1,\ldots,Q_p$ ;
>
> $b_{pqj}$ are the random effects associated with column-specific predictors $X_{qk}$ . They vary randomly over rows $j = 1,\ldots, J$;
>
> $\gamma_{pr}$ are the fixed effects of row-specific predictors $W_{rj}, r = 1,\ldots,R_p$ ;
>
> $c_{prk}$ are the random effects associated with row-specific predictors $W_{rj}$ . They vary randomly over columns $k = 1,\ldots, K$;
>
> $\delta_{pgk}$ are the fixed effects of cell-specific predictors $Z_{gk}$, which are the interaction terms created as the products of $X_{qk}$ and $W_{rj}$, s = 1, …,$S_p$ and $S_p \le R_p \times Q_p$ ; and
>
> $b_{p0j}$ , $c_{p0k}$ , and $d_{p0jk}$ are residual row, column, and cell-specific random effects, respectively, on $\pi_{pjk}$, after taking into account $X_{qk}$, $W_{rj}$, and $Z_{gk}$ . We assume that $b_{p0j} \sim N\left(0, \tau_{p b00}\right)$, $c_{p0k} \sim N\left(0, \tau_{p c00}\right)$, and that the effects are independent of each other.

### *V-known models*

The V-known option in HLM2 is a general routine that can be used for applications where the level-1 variances (and covariances) are known. Included here are problem of meta-analysis and a wide range of other possible uses as discussed in Chapter 7 of *Hierarchical Linear Models*. The program input consists of $Q$ random level-1 statistics for each group and their associated error variances and covariances. While this model is usually only run in interactive/batch mode, a V-known analysis can be eprformed using the

Windows interface when $Q = 1$.

### *Analysis of multiply imputed data/plausible values*

Users can apply HLM2 and HLM3 to mutliple-imputed data to produce appropriate estimates that incorporate the uncertainty resulting from imputation. HLM has two methods to analyze multiple-imputed data. They both use the same equations to compute the averages, so the method chosen depends on the data being analyzed.

Plausible values: is usually preferable for data sets that have only one variable (outcome or predictor) for whivch you have several plausible values. In this case, one MDM file, containing *all* of the plausible values plus any other variables of interest, is required as input.
Multiple imputation: is necessary when there is more than one variable for which multiply-imputed data are available. In this case, as many MDM files as there are plausible values are required as input for HLM.

### *Latent variable regression*

Researchers may be interested in studying the randomly varying coefficients not only as outcomes, but as predictors as well. Treating these coefficients as latent variables, the HLM2, HLM3 and HMLM modules allow researchers to study direct as well indirect effects among them and to assess their impacts on coefficients associated with observed covariates in the model. Furthermore, using HMLM with unrestricted covariance structures, one may use latent variable analysis to run regressions with missing data.

The graph suggests that there is significant variation in the rate of acceleration in vocabulary growth in children during the second year of life. For instance, the confidence intervals of the EB estimates of the AGE12S coefficients for the last four children from the left did not overlap with those of the first eleven children.

Users can look at the actual empirical Bayes estimates and their 95% confidence intervals of individual level-2 units by clicking on the confidence interval plots, and can also choose to include a level-2 classification variable when examining the confidence interval plots.

### *Z-structure models*

When modeling longitudinal, repeated measures, it is possible to select a cumulative effect model to allow carry-over treatment effects by specifying a cumulative Z-structure model..

### *Spatial dependence model*

The spatial dependence option in HLM2 allows researchers to handle nested data collected in spatial settings. In addition to the clustering effects, the spatial HLM2 models accommodate dependence induced by contiguity or proximity in geographic locations..

The graph suggests that there is significant variation in the rate of acceleration in vocabulary growth in children during the second year of life. For instance, the confidence intervals of the EB estimates of the AGE12S coefficients for the last four children from the left did not overlap with those of the first eleven children.

Users can look at the actual empirical Bayes estimates and their 95% confidence intervals of individual level-2 units by clicking on the confidence interval plots, and can also choose to include a level-2 classification variable when examining the confidence interval plots.