

History of HLM

Multilevel models deal with the analysis of data where observations are nested within groups. Social, behavioral and even economic data often have a hierarchical structure. A frequently cited example is in education, where students are grouped in classes. Classes are grouped in schools, schools in school districts, etc. We thus have variables describing individuals, but the individuals may be grouped into larger or higher-order units. In the case of repeated measurement data, the individual serves as the group, with multiple measurements nested within the individual.

Traditionally, fixed parameter linear regression models are used for the analysis of such data, and statistical inference is based on the assumptions of linearity, normality, homoscedasticity, and independence. Ideally, only the first of these assumptions should be used. It has been shown by Aitkin & Longford (1986), that the aggregation over individual observations may lead to misleading results. Aggregation of, for example, student characteristics over classes facilitate a class analysis, but in the process all individual information is lost. As within-group variation frequently accounts for most of the total variation in the outcome, this loss of information can have an adverse effect on the analysis and lead to distortion of relationships between variables. The alternative, disaggregation, implies the assignation of all class, school, and higher-level characteristics to the individual students. In the process, the assumption of independent observations no longer holds. Both the aggregation of individual variables to a higher level of observation and the disaggregation of higher order variables to an individual level have been somewhat discredited (Bryk & Raudenbush, 1992). It has also been pointed out by Holt, Scott and Ewings, (1980), that serious inferential errors may result from the analysis of complex survey data if it is assumed that the data have been obtained under a simple random sampling scheme.

In hierarchical data, individuals in the same group are also likely to be more similar than individuals in different groups. Due to this, the variations in outcome may be due to differences between groups, and to individual differences within a group. Thus, variance component models, where disturbance may have both a group and an individual component, can be of help in analyzing data of this nature. Within these models, individual components are independent, but while group components are independent within groups, they are perfectly correlated within the groups. Random regression models have been developed to model continuous data (Bock, 1983), and also dichotomous repeated measurement data (Gibbons & Bock, 1987) where certain characteristics of the data preclude the use of traditional ANOVA models. In random regression models, however, there is still no possibility of including higher level variables. In order to accommodate both random coefficients and higher order variables, multilevel models should be used.

Multilevel analysis allows characteristics of different groups to be included in models of individual behavior. Most analyses of social sciences data entail the analysis of data with built-in hierarchies, usually obtained as a sequence of complex sampling methods. Thus, the scope for application of multilevel models is very wide. The formulation of such models and estimation procedures may be seen as an effort to develop a new family of analytical tools that correspond to the classical experimental designs. These models are much more flexible in that they are capable of handling unbalanced data, the analysis of variance-covariance components and the analysis of both continuous and discrete response variables. As the characteristics of individual groups are incorporated into the multilevel model, the hierarchical structure of the data is taken into account and correct estimates of standard errors are obtained. The exploration of variation between groups, which may be of interest in its own right, is facilitated. Valid tests and confidence intervals can also be constructed, and stratification variables used in the sample design can be incorporated into the model.

The use of multilevel models has been hampered in the past by the fact that closed form mathematical formulas to estimate the variance and covariance components have only been available for perfectly

balanced designs. Iterative numerical procedures must be used to obtain efficient estimates for unbalanced designs. Among the procedures suggested are full maximum likelihood (Goldstein, 1986, and Longford, 1987), and restricted maximum likelihood estimation as proposed by Mason et. al. (1983) and Bryk and Raudenbush (1986). Another approach is the procedure of Bayes estimation (Dempster et.al., 1981). Some other procedures include the use of Iteratively Reweighted Generalized Least Squares (Goldstein, 1986), and a Fisher scoring algorithm (Longford, 1987). Increased interest in these models, which are known by various names in the literature (hierarchical linear models, multilevel models, mixed-effects models, random-effects models, random coefficient regression models, covariance components models) have led to new developments in this field in recent years.

An interesting development was in terms of the type of outcome variables considered: where previously interest was confined to continuous outcome variables, statistical theory has been extended and implemented in software such as HLM to appropriately handle binary outcomes, ordered categorical outcomes, and multi-category nominal scale outcomes within the hierarchical framework. Goldstein (1991) and Longford (1993) both developed software that allowed the use of several types of discrete outcomes for two- and three-level models, followed by an improved second-order approximation by Goldstein in 1995. Hedeker and Gibbons (1993) and Pinheiro and Bates (1995) contributed accurate approximations to maximum likelihood using Gauss-Hermite quadrature that has since been implemented in software such as MIXOR and SAS PROC NL MIXED. The use of a high-order Laplace transform as an alternative approximation was introduced by Raudenbush, Yang & Yosef in 2000 and is implemented in the HLM program.

In the preceding discussion, the assumption was that each lower-level unit, for example an individual student, was nested within an unique higher-level unit such as a school. In other words, a one to one relationship was assumed to define the nesting of units within groups. Excluded from this were hierarchies in which cross-classification occur, for example where students from multiple neighborhoods may end up going to multiple schools; a situation where students are "cross-classified" by neighborhoods and schools. To address this situation and allow for the inclusion of predictors for more than one "classification" variable where the coefficients of a individual level model describing the association between individual-level variables and the outcome for groups defined by the "classification" variables, cross-classified random-effect models were developed (Raudenbush, 1993; Goldstein, 1995). A two-level cross-classified random-effect model has been implemented in HLM 6.

Due to increased interest in multivariate outcome models, such as repeated measurement data, contributions by Jennrich & Schluchter (1986), and Goldstein (1995) led to the inclusion of multivariate models in most of the available hierarchical linear modeling programs. These models allow the researcher to study cases where the variance at the lowest level of the hierarchy can assume a variety of forms/structures. The approach also provides the researcher with the opportunity to fit latent variable models (Raudenbush & Bryk, 2002), with the first level of the hierarchy representing associations between fallible, observed data and latent, "true" data. An application that has received attention in this regard recently is the analysis of item response models, where an individual's "ability" or "latent trait" is based on the probability of a given response as a function of characteristics of items presented to an individual.

In HLM7, unprecedented flexibility in the modeling of multilevel and longitudinal data was introduced with the inclusion of three new procedures that handle binary, count, ordinal and multinomial (nominal) response variables as well as continuous response variables for normal-theory hierarchical linear models. HLM 7 introduced four-level nested models for cross-sectional and longitudinal models and four-way cross-classified and nested mixture models. Hierarchical models with dependent random effects (spatial design) were added. Another new feature was new flexibility in estimating hierarchical generalized linear models through the use of Adaptive Gauss-Hermite Quadrature (AGH) and high-order Laplace approximations to maximum likelihood. The AGH approach has been shown to work very well when cluster sizes are small and variance components are large. The high-order Laplace

approach requires somewhat larger cluster sizes but allows an arbitrarily large number of random effects (important when cluster sizes are large).

In HLM8, the ability to estimate an HLM from incomplete data was added. This is a completely automated approach that generates and analyses multiply imputed data sets from incomplete data. The model is fully multivariate and enables the analyst to strengthen imputation through auxiliary variables. This means that the user specifies the HLM; the program automatically searches the data to discover which variables have missing values and then estimates a multivariate hierarchical linear model ("imputation model") in which all variables having missed values are regressed on all variables having complete data. The program then uses the resulting parameter estimates to generate M imputed data sets, each of which is then analyzed in turn. Results are combined using the "Rubin rules".

Another new feature of HLM8 is that flexible combinations of Fixed Intercepts and Random Coefficients (FIRC) are now included in HLM2, HLM3, HLM4, HCM2 and HCM3.

A concern that can arise in multilevel causal studies is that random effects may be correlated with treatment assignment. For example, suppose that treatments are assigned non-randomly to students who are nested within schools. Estimating a two-level model with random school intercepts will generate bias if the random intercepts are correlated with treatment effects. The conventional strategy is to specify a fixed effects model for schools. However, this approach assumes homogeneous treatment effects, possibly leading to biased estimates of the average treatment effect, incorrect standard errors, and inappropriate interpretation. HLM 8 allows the analyst to combine fixed intercepts with random coefficients in models that address these problems and to facilitate a richer summary including an estimate of the variation of treatment effects and empirical Bayes estimates of unit-specific treatment effects. This approach was proposed in Bloom, Raudenbush, Weiss and Porter (2017).