http://faculty.chass.ncsu.edu/garson/PA765/multilevel.htm

Linear Mixed Models

Hierarchical Linear, Random Effects, Multilevel, Random Coefficients, and Repeated Measures Models

Overview

Linear mixed models (LMM) handle data where observations are not independent. That is, LMM correctly models correlated errors, whereas procedures in the general linear model family (GLM) usually do not. (GLM includes such procedures as t-tests, analysis of variance, correlation, regression, and factor analysis, to name a few.) LMM is a further generalization of GLM to better support analysis of a continuous dependent for:

Random effects: where the set of values of a categorical predictor variable are seen not as the complete set but rather as a random sample of all values (ex., the variable "product" has values representing only 5 of a possible 42 brands). Through random effects models, the researcher can make inferences over a wider population in LMM than possible with GLM.

Hierarchical effects: where predictor variables are measured at more than one level (ex., reading achievement scores at the student level and teacher-student ratios at the school level; or sentencing lengths of offenders, gender of judges, and budgets of judicial districts).

Repeated measures: where observations are correlated rather than independent (ex., before-after studies, time series data, matched-pairs designs)

It is true that GLM has been adapted to handle these models also, but problematically so and therefore LMM is preferred. For instance, GLM in SPSS does support random effects but estimates their parameters as if they were fixed, calculating variance components based on expected mean squares; LMM, in contrast, uses maximum likelihood estimation to estimate these parameters. GLM in SPSS supports repeated measures but the LMM module supports more variations and data options. Hierarchical models in SPSS require LMM implementation. Linear mixed models include a variety of multi-level modeling (MLM) approaches, including hierarchical linear models, random coefficients models (RC), and covariance components models. Differences between LMM and GLM are discussed further in the FAQ section.

Note that multi-level mixed models are based on a multi-level theory which specifies expected direct effects of variables on each other within any one level, and which specifies cross-level interaction effects between variables located at different levels. That is, the researcher must postulate mediating mechanisms which cause variables at one level to influence variables at another level (ex., school-level funding may positively affect individual-level student performance by way of recruiting superior teachers, made possible by superior financial incentives). Multi-level modeling tests multi-level theories statistically, simultaneously modeling variables at different levels without necessary recourse to aggregation or disaggregation. It should be noted, though, that in practice some variables may represent aggregated scores.

In SPSS, select Analyze, Mixed Models, Linear; if there are repeated measures, enter the repeated variables and the subject variable; click the Fixed Effects button and fill out that dialog; click the Random Effects button and do likewise; click the Statistics button to select output; click OK. (There are also other options).

See also variance components analysis (VARCOMP). Note that the linear mixed models procedure includes all VARCOMP models, but output options differ somewhat.

Contents

Key concepts and terms

Data for mixed models

Variables for mixed models

Covariance structure type

Estimation methods

Types of mixed models

Unconditional random effects (null) models

One-way ANCOVA random effects models

Random intercept regression models

Random coefficients regression models

Full random coefficients models

Three-level hierarchical linear models

Repeated measures models

Assumptions

FAQ

Bibliography

Key Concepts and Terms

Rationale. Why use linear mixed models (LMM) rather than OLS regression, logistic regression, or GLM when modeling?

OLS regression. Compared to OLS regression, linear mixed models handle random effects, thereby handling the problem of correlated error terms. In general, the sampling unit is a random effect. In a study of the federal bureaucracy, "agency" might be the sampling unit, for example. Error terms may cluster by agency, violating OLS assumptions. Unlike OLS regression, linear mixed models take into account the fact that over many samples, different b coefficients for effects would be computed. That is, mixed models treat the b coefficients as random effects drawn from a normal distribution of possible b's, whereas OLS regression treats the b parameters as if they were fixed constants (albeit within a confidence interval). Treating "agency" as a random rather than fixed effect will alter and make more accurate the ensuing effect parameter estimates. Put another way, the misestimation of standard errors in OLS regression inflates Type 1 error, whereas mixed models handle this potential problem. In addition, LMM can handle a random sampling variable like "agencies" even when there are too many agencies to make into dummy variables in OLS regression and still expect reliable coefficients.

Logistic regression also does not provide for random effects variables, nor (even in the multinomial version) does it support near-continuous dependents (ex., test scores) with too large a number of values. Binning such variables, as is sometimes done, loses information and attenuates correlation.

GLM models do support random effects independent variables but assume independent observations, whereas it is common for the sampling unit (ex., cities, schools, agencies) to display intraclass correlation - that is, to have an effect on the dependent. Put another way, individual-level observations from the same upper level group will not be independent but rather will be more similar due to such factors as shared group history and group selection processes. LMM does not assume such data independence. This is discussed further in the assumptions section.

Linear mixed models in some disciplines are called "random effects" or "mixed effects" models. In economics, the term "random coefficient regression models" is common. In sociology, "multilevel modeling" is common, alluding to the fact that regression intercepts and slopes at the individual level may be treated as random effects of a higher (ex., organizational) level. And in statistics, the term "covariance components models" is often used, alluding to the fact that in linear mixed models one may decompose the covariance into components attributable to within-groups vs between-groups effects. All these terms are closely related, albeit emphasizing different aspects of linear mixed models.

The application of mixed models such as hierarchical linear models can lead to substantially different conclusions compared to conventional regression analysis. Raudenbush and Bryck (2002: 9-10), citing their 1988 research on the increase over time of math scores among students in grades 1 through 3, wrote that with hierarchical linear modeling, "The results were startling - 83% of the variance in growth rates was between schools. In contrast, only about 14% of the variance in initial status was between schools, which is consistent with results typically encountered in cross-sectional studies of school effects. This analysis identified substantial differences among schools that conventional models would not have detected because such analyses do not allow for the partitioning of learning-rate variance into within- and between-school components."

Data

Hierarchical data involve measurement at multiple levels such as individual and group as, for example, a study of certain variables studied in terms of individual students' opinions, their classes, and their schools. In fact, much early work on multi-level modeling focused on educational settings. In general, hierarchical data are obtained by measurement of units grouped at different levels, such as a study of children nested within families; employees nested within agencies; soldiers nested within platoons, divisions, and armies; or subjects nested within studies.

Cross-classified data also involve multileve data but there can be overlap rather than strict hierarchy at higher levels. This is discussed in the separate section on HLM software, which supports mixed models for cross-classified data.

Groups. Individuals are clustered within groups. In two-level mixed models, the base layer (level 1) is individuals (ex., students) who are clustered within the groups formed by the upper level layer (ex., level 2 = schools).

Group effects and intraclass correlation. Consider the example of student test scores grouped by school. There is a group effect, also called the "between-groups" effect, if the grouping variable (school) has an effect on the intercept of test score over and beyond whatever other test-level (individual-level) variables are used to predict test score (ex., student age, student IQ, student socioeconomic status, etc.). The residual variance in test score not explained by the between-groups effect represents variation within the set of individual students. This effect is the "within-groups" effect. The "total effect" is, of course, the sum of the between- and within-groups effect.

The "intraclass correlation" (ICC) is the between-groups effect divided by the total effect. If ICC approaches zero, there is no between-groups effect. And if there is no between-groups effect, there is no need to model individual-level regression parameters as random effects of a higher or grouping level. That is, the lower the ICC, the less difference hierarchical linear modeling or linear mixed modeling will make in predicting the dependent (ex., test score) compared to traditional regression techniques. Put another way, ICC is a test of whether hierarchical linear modeling is needed.

Multistage sampling. Hierarchical data are normally obtained by multistage sampling. For instance, one might sample schools within school districts, then sample students within sampled schools.

Centering. It is customary to center data prior to running LMM or HLM. Centering means subtracting the mean, so means become zero. Two main types of centering are group mean centering and grand mean centering. For instance, in a study of PerformanceScore, there might be a PerformanceIndividualScore at level 1 and a PerformanceAgencyScore at level 2, where the latter was a mean score for all employees in an agency. The researcher might center PerformanceIndividualScore for individuals by centering on their group (agency) means, where groups were agencies, on the theory that group performance influenced individual performance and differences from the group means should therefore be the variable of interest. Or one could center each PerformanceIndividualScore on the grand mean of all such scores across agencies. Grand mean centering is preferred over group mean centering unless there is theoretical justification for the latter.

Grand mean centering often improves the interpretability of coefficients because "0" now has a meaning (ex., 0 income is mean income, whereas before centering, 0 income might be out of the range of actual observations). Group mean centering, in contrast, changes the meaning of coefficients in complex ways which make coefficients hard to interpret, as different mean values are subtracted from different sets of raw scores. As a result, with group mean centering it is not possible to recalculate output back to raw score interpretations. In essence, one is dealing with a different variable after group mean centering. Grand mean centered income, for instance, will yield different slopes but the same deviance and residual errors as uncentered raw data. Group mean centered income does not. Group mean centered income is no longer simple income but rather measures income deviation from group means. The researcher must examine his or her theoretical model and decide if that is really what was wanted for the "income" variable. As noted by Kreft, de Leeuw, and Aiken (1995), the choice of centering must be made on a theoretical rather than statistical basis, and "centering around the group mean amounts to fitting a different model from that obtained by centering around the grand mean or by using raw scores" (p. 1). Most LMM/HLM software packages support various types of automatic centering. Centering considerations are further discussed in Burton (1993) and Hoffman & Gavin (1998).

Data formats. Usually data are in 'multiple variable' (MV) format, with an individual ID variable, a nesting group ID variables, and a series of other variables, with one row per record (case). A 'multiple record' (MR) format has one row per variable. Each row has the individual ID variabl;e, the nesting group (level) variables, and one other variable of interest. Any given individual may be represented by as many rows as there are variables of interest. Repeated measures data often are in MR format.

Variables

Subject variables. The subject variable is one which is used to define groups such that each group is independent of the others. The subject variable ordinarily is not the id of the individual subject of the study. For instance, in a study of individual test scores grouped by school, if "ID" were the subject variable, it would be the ID of the school, not the student. An exception would be repeated measures studies (ex., where any given student is tested, say, in each of 6 months). In repeated measures, the six monthly scores are grouped by student. Since individual student is the grouping variable, "StudentID" would be the subject variable.

The subject variable, for ex. school, is seen as a random sample of all possible schools. Because it is random it is associated with a random error term and each school will have a different effect on the level 1 variable being predicted, such as student test score. That is, test score is a random effect of the subject variable (school) plus possibly other school-level or higher-level predictors that may be in the model, plus possibly other level 1 predictors. By making school a random effect variable, one regression will be run for each school. The intercept of the dependent, test score, is modeled as the mean of all the intercepts in the separate regressions for each school (each subject).

In some cases more than one variable is needed as a subject. That is, more than one variable defines the groups. For instance, if subject variables are Gender and JobStatus, then the groups might be Male-Working, Male-NotWorking, Female-Working, and Female-NotWorking. The researcher would be hypothesizing that these two subject variables explained variance in whatever dependent variable was measured. Observations would be similiar within groups but each group is assumed independent of the others.

SPSS: The initial "Linear Mixed Models: Specify Subjects and Repeated" dialog screen allows the researcher specify one or more subject variables. If there are repeated measures or random effects a subject variable is usually entered. Note that if an observation has a missing value on any of the subject variables, it will be dropped from analysis.

Repeated measures variables. Similar to a subject variable, this variable or variables constitute the id for the repeated observations. For instance, one might have a single repeated measure "Year" to indicate each of four years in which a measurement was taken. Alternatively, there could be two repeated measures variables, Month and Day, which together would constitute the id for the repeated measurements. This, too, is specified on the initial "Linear Mixed Models: Specify Subjects and Repeated" dialog screen in SPSS.

The dependent variable is assumed to be a normally distributed quantitative variable which is linearly related to the fixed and random factors and covariates in the model. Do not use a multinomial variable as a dependent, for example.

Fixed factors are categorical variables where all possible category values (levels) are measured, even if one of the values is "other." (Ex., religion = Protestant, Catholic, Jewish, Other). Fixed factors may be the primary variables of interest in a research study. Fixed factors have different, varying intercepts for each group, but the regression slope is the same for each group. Fixed factors are used to model the mean of the dependent variable.

Random factors are categorical variables where only a random sample of possible category values are measured. (ex., city = Philadelphia, Miami, Denver, Detroit, Los Angeles). Random factors often reflect the sampling and data collection design of a research study. In a multilevel study random factors may be grouping factors (ex., individual-level voting data grouped by city, so that voter attributes are level 1 fixed factors and city is a level 2 random factor). Random effects model the level 1 intercept of the dependent variable and/or the level 1 slope of level 1 predictors of the dependent variable.

However, random factors may also be any categorical variable whose levels are conceived as a sample of possible levels. Thus in a simple one-level individual sample of soldiers, soldierID could be a random factor representing a soldier effect.

It is important to note that a given effect may be modeled as a random or as a fixed factor. For instance, SES (socioeconomic status) in a study of employee performance scores grouped by agency might be modeled as a random covariate if it is thought its regression coefficient varied randomly by agency, but if the regression coefficient is assumed to be constant across agencies, SES might be modeled as a fixed factor. Designating a level 1 (ex., employee) variable as a random factor means the researcher assumes its coefficient varies randomly across level 2 (ex., agency) groups.

Covariates are continuous predictor variables. (ex., income). SPSS syntax is dependentvariable BY factors WITH covariates. Factors and covariates are entered in the main "Linear Mixed Models" dialog box in SPSS, then either the Fixed or Random buttons or both are clicked to specify how factors and covariates are to be modeled.

Residual weight variable (REGWGT). Optionally the researcher can specify the name of a variable containing the regression weights. These are applied (and only applied) to the residual covariance matrix. Residual weights are used for models with unequal variances between groups formed by the subject variable(s) and are akin to weighted least squares analysis in regression models.

Moderator variables in a multi-level model are the level 2 or higher level independent variables. Level two variables are ordinarily modeled as random factors. For instance, school budget at level 2 may moderate the effect of socio-economic status (SES) on test performance at the base (student) level. The theory, presumably, would be that school budget compensates for some of the educational resources implicit in SES at the individual level. The regression coefficients connecting the moderator variables at level 2 to the regression slopes and intercepts at the individual level are assumed not to vary across groups (schools) and hence these are fixed coefficients, in contrast to the random coefficients at the base level.

Cross-level interactions. Multi-level modeling will identify significant cross-level interactions (the joint effect of a variable at a base level in conjunction with a variable at an upper level). When such an interaction is present, the estimated coefficient of the direct variable estimates the effect of that variable when the other variable in the interaction is controlled (is zero). Sometimes, of course, a zero value is impossible for that variable because it is outside the permissible range. Interpretation may be improved in such cases by centering the variable: subtract the mean value from all cases for that variable. Interactions can also be visualized by plotting separate scatterplots of one of the interacting variables with the dependent, getting a separate plot for each value of the other interacting variable.

Covariance Structure Type

What covariance structure type is. Linear mixed modeling uses an iterative algorithm to estimate coefficients. The estimate will be more reliable if the algorithm uses as a starting point an accurate assumption about the nature of the covariance matrix for the variables in the model. Covariance structure type may be specified for repeated measures or for random effects. It is possible to need to specify both repeated measures and random effect covariance matrix types and if both, the specified types may differ.

For repeated measures one may create a table in which both rows and columns are the measurement times (ex., pretest and postest; or time\_1, time\_2, time\_3,..., time\_n). Cell entries may be the covariances of the residuals when predicting the dependent variable, using all subjects, for the row time period and the column time period. In the iterative algorithm by which LMM calculates repeated or random effects (with output displayed in the "Estimates of Covariance Estimates" table discussed below), it makes a difference what pattern of covariance is assumed as the starting point for the iterative algorithm. The researcher must tell LMM what covariance structure type to assume. There are many possible patterns but some of the most common are diagonal structure (the default for repeated measures), compound symmetry, and unstructured - all discussed below. For repeated measures, the covariance structure type is specified in SPSS in the opening "Linear Mixed Models: Specify Subject and Repeated" dialog. This specifies what is called the repeated measures type "R" variance/covariance matrix. It is this matrix which is the basis for creating the within-subjects correlation matrix

For random effects one may likewise create a similar matrix in which the rows and columns are levels of the factor which is to be treated as a random effect (ex., particular cities treated as a random sample of all cities). As with repeated measures, the researcher must specify the type of covariance structure (covariance of the row and column groups) of the residuals when predicting the dependent. This specifies what to assume as a starting point for the iterative maximum likelihood (ML or REML) algorithm used to compute parameter estimates. For random effects, the default is variance components structure, discussed below. Specification is accomplished under the "Random" button dialog in LMM which specifies what is called the random effects type "G" variance/covariance matrix. This matrix is the basis for estimating between-groups effects.

In random effects models, the groups formed by the subject variable(s) are assumed to be independent and are assumed to each have the same covariance structure, so only one covariance type is specified by the researcher per random effect variable and only one estimated by LMM algorithms, regardless of how many groups there are. The available covariance types are the same as for repeated measures covariance types, except there is an additional type - Variance Components - which is the default for random effects models. "Variance Components" structure means that the variances of the random effects are assumed to be independent and sum to the variance of the dependent variable. The covariance structure specified for random effects need not be the same as specified for repeated measures.

Multiple random effects. Ordinarily there will be just one random effects model and just one covariance type specified for it. However, under the Random button in SPSS, it is possible to have multiple random effects models, each with its own specified covariance structure, which may be the same or different from the others. If there are multiple random effects models, they are assumed to be independent of one another (terms within the model for any random effect may be correlated, however). If the same random effects variable is listed in more than one random effects model, it must refer to a different combination of Subject variables. When there are multiple random effects models, a separate covariance structure is estimated for each.

When to specify the covariance structure type? If there are random effects or repeated measures, the researcher specifies the covariance structure type.

Why specify the covariance structure type?. The covariance structure type specified by the researcher is used as a starting point in the iterative REML or ML algorithms for estimating parameters. Assuming a too-simple covariance structure will increase Type I errors, while assuming a too-complex structure will increase Type II errors.

Selecting the best covariance structure assumption. Goodness of fit statistics (-2LL, AIC, AICC, BIC - discussed in the section on structural equation modeling, for instance) can be used to select the best covariance structure type to assume. Simulation studies by Ehlers (2004) suggest BIC is preferred generally (compared to AIC, BIC penalizes for lack of parsimony), but AICC is better for sample sizes < 20 and when there is only one group. Usually, however, this is a moot question because the criteria will agree in most cases. The researcher runs the model under the desired covariance structure assumptions (ex., under variance components, unstructured, and AR(1), or others listed below), then looks in the "Information Criteria" table in the output to take note of the BIC, AICC, or other fit statistic for each run of the model. The model with the lowest value on BIC (or other chosen fit criterion) is the one with the best-fitting covariance structure assumption. Note that if one or more fixed effects is dropped from the model, BIC (or other fit criteria) must be reestimated and reevaluated.

Likelihood ratio tests. It is also possible to run a likelihood ratio test on the difference between two models, one under a given covariance structure assumption and another model under a different assumption. Likelihood ratio tests are described in the section on structural equation modeling.

Data-driven vs. theory-driven selection. Note, however, that both the information criteria or the LR test methods are data-driven and thus prone to over-fitting, particularly is cross-validation is not part of the design. Covariance structure assumptions should start with theoretical selection based on covariance expectations described below.

Examples of types of covariance structure. For the following types of covariance structure, the figures reflect the example, discussed below in the section on basic mixed models, of city as a random effect in a city tax study of the relation of appraisal value to selling price. Figures are the "G matrix," reflecting the City random effect covariance structure.

Diagonal (Simple). This is the default for repeated measures covariance structure type but it may be used in other contexts. This covariance structure has heterogenous variances (unlike variance components structure discussed below) and zero correlation between elements. For instance, if the repeated measures elements are time periods, the diagonal structure assumption is that there is no correlation of the residuals of Time 1 with the residuals of Time 2 when predicting the dependent variable. Each element (ex., Time Period) may have a unique variance for its residuals, but the covariance between time periods is zero. This implies that the measures for an individual in one time period are independent of (uncorrelated with; hence the diagonal model may be called the independence model) measures for that individual in any other time period. This may well be an unrealistic assumption. Alternative common structures that might be assumed for repeated measures are AD(1) and AR(1), discussed below. While diagonal structure is the default for repeated measures, it may also be specified for random effects, as illustrated for cities as a random effect in the figure below.

Variance Components, also called "simple structure" or "the independence model" (because residual variance is independent of effect variances) . This is the default for random effects covariance in both SPSS and SAS linear mixed models, and is only available for random effects and not repeated measures. Each element (here, city) has the same variance of residuals as each other element (contrast diagonal structure, where this may be unique by element), and the covariance of residuals (error terms) is assumed to be 0 (ex., residuals are not correlated between any pair of cities). The variance components assumption is usual in random effects models. If the assumptions of a variance components model can be met, it means that effects are additive: the ratio of effect covariance estimates to residual covariance estimates is the ratio of the importance of between-subject effects to within-subject effects in accounting for the variance of the dependent variable.

Others. Many other types of covariance structure are possible, discussed below in the FAQ section.

Estimation

Estimation algorithm. The default algorithm used to compute coefficients for the predictor variables in SPSS is "restricted maximum likelihood" (REML), but by clicking the Estimation button in SPSS the researcher may choose "maximum likelihood" (ML) instead. REML estimation has better bias characteristics (Diggle, 1988). handles high correlations more effectively, and is less sensitive to outliers than ML, but cannot be used for model comparison of fixed effects, as noted below in the section on likelihood ratio tests. ML estimations ignores the degrees of freedom used up by fixed effects in mixed models, leading to underestimation of variance components. However, ML may nonethless be preferred when comparing two models with different parameterizations of the same effect (ex., simple variable vs. quadratically transformed version of the variable), because ML is invariant to different parameterizations of a fixed effect but REML will treat different parameterizations as different models and compute different likelihood ratios. If parameterization is not an issue, however, one may use -2RLL in likelihood ratio tests of difference between models. The Estimation button also allows the researcher to set the number of iterations and various convergence options though normally the defaults suffice. LMM estimation, whether by REML or ML, contrasts with the ANOVA methods utilized in GLM.

Covariance type. If the model includes random factors, the covariance type must be specified. The default in SPSS is "variance components" but under the "Random" button of the Linear Mixed Models dialog, there are over a dozen alternatives. Variance components models treat the variance of the dependent variable as the sum of variances for each level of the random factor(s), giving the advantage that the percent of variance in the dependent variable attributable to a variance component (a random effect or the interaction of a random effect with a fixed effect) is computed as the parameter estimate for that component as a percentage of the total of parameter estimates for all components (including the residual component).

Convergence . Occasionally the maximum likelihood estimation algorithm, which is iterative, fails to converge on a solution. SPSS will print an error message, "Iteration was terminated but convergence has not been achieved." Lack of convergence generally indicates the model is badly misspecified and must be thrown out, but it may also be due to too small a sample. Misspecification is often associated with trying to estimate random (base-level) coefficients which are close to or equal to zero, which in turn leads to lack of convergence. Several remedial actions may be possible:

Check for and remove any redundant variables, where the correlation approaches 1.0.

Under the Estimation button, increase the maximum iterations above the default (100).

Under the Estimation button, increase step-halvings above the default (5).

Under the Estimation button, increase the singularity tolerance value (a drop-down list of choices is provided).

Under the Estimation button, increase the number of scoring steps above the default (1).

Under the Estimation button, increase the parameter convergence value to be larger than the default (.000001).

Types of mixed models

Fixed effects models are models with only fixed factors and optional covariates as predictors. Most models in analysis of variance, regression, and GLM are fixed effects models, which are by far the most common type in social science. An example would be a study of job satisfaction by gender, controlling for salary level. Job satisfaction would be the dependent variable, gender the fixed factor, and salary the covariate, treated as fixed. Both GLM and LMM can model fixed effects models with very similar if not identical estimates and similar but not identical output tables. However, where LMM is clearly superior to GLM lies in handling other types of models discussed below.

Random effects models are models with one or more random factors and optional covariates as predictors. If there are covariates, they are treated as fixed effect variables. An example would be a study of job satisfaction by city, controlling for salary level. Job satisfaction would be the dependent variable, city the random factor (assuming only a random sample of cities was studied), and salary the covariate.

The level 1 intercept of job satisfaction may be modeled as a random effect of city. Likewise, the level 1 slope of salary might be modeled as a random effect of city. If only the intercept is modeled, it is a random intercept model. If only the slope is modeled, it is a random coefficients model. If both slope and intercept are modeled, some authors still call it a random coefficients model while others use hierarchical linear model to distinguish it.

Mixed models in their full form have both fixed and random factors as well as optional covariates as predictors. An example would be a study of job satisfaction by gender by city, controlling for salary level. Job satisfaction would be the dependent variable, gender the fixed factor, city the random factor, and salary the covariate.

Hierarchical linear models (HLM) are a type of mixed model with hierarchical data - that is, where data exist at more than one level (ex., student-level data and school-level data). In explaining a dependent variable, HLM models focus on differences between groups (ex., schools) in relation to differences within groups (ex., among students within schools).

Random intercept models are models where only the intercept of the level 1 dependent is modeled as an effect of the level 2 grouping variable and possibly other level 1 covariates and possibly other level 2 random effect predictors.

The null model (one-way ANOVA with random effects), sometimes called the unconditional model, predicts the level 1 intercept of the dependent as a random effect of the level 2 grouping variable, with no other predictors at level 1 or 2. The same logic is applied if there are three or more levels. For instance, differences in mean performance scores (the intercepts) may be analyzed in terms of the between-groups effect of agency.

One-way ANCOVA with random effects models. It is also possible to have a level 1 covariate but still predict the level 1 intercept (and not the slope of the level 1 covariate) as a random effect of the level 2 grouping variable with no other level 2 predictors. For instance, differences in mean performance scores (the intercepts) may be analyzed as predicted by salary at level 1, predicting only the level 1 intercept in terms of the between-groups effect of agency.

Random intercept regression models are also called "means-as-outcomes regression models.". Another variant of the random intercept model is to predict the level 1 intercept on the basis of the level 2 grouping variable and also on the basis of one or more level 2 random effect predictors. For instance, differences in mean performance scores (the intercepts) may be analyzed, predicting the level 1 intercept in terms of the between-groups effect of agency and the level 2 random effect variable EquipmentBrand (a factor representing a sample of some of many brands of equipment, where different agencies used different brands).

Random intercept ANCOVA models are also called "means-as-outcomes ANCOVA models." This is simply a random intercept regression model in which there is also a level 1 covariate treated as a fixed effect (slope not predicted by level 2). Some authors would label this another type of "random intercept regression model."

Random coefficients models (RC), also called multi-level regression models, are a type of mixed model with hierarchical data. The level 1 dependent is predicted by at least one level 1 covariate. The slope of this covariate and possibly also the intercept are predicted by the random effect of the grouping variable at level 2. That is, each group at the higher level (ex., school level) is assumed to have different regression slopes as well as different intercepts for purposes of predicting a level 1 dependent variable.

Random coefficients regression models are a type of RC model in which the level 1 model is a typical regression model in which a dependent is predicted by one or more level 1 covariates. The level 1 slope(s) as well as the level 1 intercept is predicted on the basis of the level 2 grouping variable as a random effect, but there are no other level 2 predictors. For instance, differences in mean performance scores (the intercepts at level 1) may be analyzed as predicted by salary at level 1, predicting both the level 1 intercept and the slope of salary in terms of the between-groups effect of agency at level 2. Random coefficients regression models are unconditional at level 2.

Intercepts-and-slopes-as-outcomes models are a type of RC model in which the level 1 slopes and intercepts are modeled not only by the level 2 grouping variable as a random effect but also by one or more other level 2 random effect variables. For instance, differences in mean performance scores at level 1 may be analyzed, predicting the level 1 intercept and the slope of the level 1 predictor salary in terms of the between-groups effect of agency and the level 2 random effect variable EquipmentBrand. Intercepts-and-slopes-as-outcomes models are conditional at level 2.

Nonrandomly varying slopes models. It is possible that a level 2 random effect predictor so fully explains the slopes of the level 2 predictor that the random effect of the level 2 grouping variable is not significantly different from zero. For instance, differences in mean performance scores at level 1 may be analyzed, predicting the level 1 intercept and the slope of the level 1 predictor salary in terms of the level 2 random effect variable EquipmentBrand (and not also the between-groups effect of agency). This type of model may also be considered conditional at level 2.

Unconditional models are often used as a comparison baseline. They focus on the variability in the dependent when there are no predictors at level 1 or level 2 (or higher) which condition (control, moderate) the variability of the level 1 intercept in random intercepts models or the variability of the level 1 slopes and intercepts in random coefficients models.

A model is "conditional" by the presence of a level 1 predictor and/or a level 2 predictor. Since the researcher almost always employs predictor variables and is not simply interested in the null model, most mixed models are conditional. In fact, the central point of mixed models/hierarchical linear modeling is to assess the difference between the researcher's conditional model and the null model without predictors. The likelihood ratio test (model chi-square difference test) can be used to assess the difference in fit between a conditional model and the corresponding unconditional model.

Steps in multi-level modeling. Hox (1995) suggests the following procedure for multi-level analysis:

Compute deviance for the baseline (null) model which includes only the intercept.

Compute deviance for the model with the base level independent variables included and the variance components of the slopes constrained to zero (that is, for the fixed model).

Use a chi-square difference test to see if the fixed model has a significantly better fit than the baseline model. If it does, then the researcher proceeds to investigate higher level modifier variables using random effects and/or RC models. Also, at this second step the researcher can assess the relative contribution of the base level independent variables. Drop non-significant base level independents and covariances from the model.

Identify which base-level regression slopes have significant variance across upper level groups. Compute -2LL for the model with the variance components of the base-level slopes constrained to zero only for the slopes which do not have signficant variance across upper level groups.

Add upper level modifier variables, determining which improve model fit. Drop modifier variables which do not improve model fit.

Add cross-level interactions between explanatory modifier variables and base level independent variables that had slope variance (in step 3). Drop interactions which do not improve model fit.

The null or unconditional random effects model

Overview. Also called the intercept-only model, the unconditional model, or the one-way ANOVA model with random effects. This model, which is a type of random intercept model, predicts the dependent based on an intercept term and an error term at level 1 (ex., the student level where the dependent is test scores). There are no predictors, either factors or covariates, at either level 1 (ex., student level) or level 2 (ex., school level). However, the level 1 intercept is predicted as a function of the level two grouping variable (ex., school). Specifically, the level 1 intercept is predicted on the basis of the level 2 intercept (representing mean test score among schools in this example) plus a level 2 error term (representing unmeasured factors unique to each school).

Null models are mainly used to obtain baseline -2LL, also called deviance. This baseline is then used in measuring significant improvements with models which do include fixed and random predictors other than the grouping variable. The -2LL deviance value appears in the "Information Criteria" table of SPSS output when the researcher requests descriptive statistics under the Statistics button.

Why LMM?. One could approach such an example with a GLM procedure such as fixed-effects analysis of variance, testing differences in mean scores across agencies. However, fixed-effects ANOVA assumes independent observations, whereas in fact it is possible, even likely, that valuew of the dependent variable are biased up or down depending on the nature of the level 2 grouping units (ex., schools). This can be handled by declaring the level 2 unit (school) to be a random factor in LMM.

SPSS. For the example where the level 2 nesting factor is SchoolID in a study of causes of TestScore, then in SPSS select Analyze, Mixed Models, Linear; in the "Linear Mixed Models: Specify Subjects and Repeated" dialog box, assign SchoolID to the Subjects listbox (the Repeated box is left blank). Click Continue. In the "Linear Mixed Models" dialog, assign the dependent variable (TestScore) to the Dependent Variable listbox. List SchoolID in the Factor(s) box. The Covariate(s) and Residual Weights boxes are left blank. Click the Random button. In the Linear Mixed Models Random Effects dialog box, enter SchoolID as the Combinations variable under Subject Groupings. In the same dialog, check "Include intercept". Set the Covariance Type: Variance Components is the default for random effects models, but Unstructured will give a solution characterized by fewer constraints and thus less parsimony. Click the Statistic button and select the desired output, such as Click Continue, OK to run the model.

Fixed effects. The only fixed effect in an unconditional random effects model is the constant (intercept). Thus in SPSS output there is a table labeled "Tests of Fixed Effects" with one row for the intercept as the only fixed effect. In this example, the estimate column for the Intercept row gives the overall school mean math achievement score. Also listed is a t-test of the significance of the mean, and 95% confidence intervals. For a significant mean, zero will not be within the lower and upper confidence bounds. If the intercept is significant (it almost always is, as here) then we can be 95% confident the intercept is different from 0. Usually this is of no interest.

If under the Statistics button we checked "Parameter estimates," then a second table of "Estimates for Fixed Effects" will appear, where the "Estimate" column will contain the actual estimated value of the intercept along with its confidence interval. If the intercept is significant, the value 0 will not be within the confidence interval.

In later models, when additional fixed factors or covariates are added (making the random effects model conditional rather than unconditional), fixed effect parameter estimates include regression slopes. These fixed effects parameters are interpreted as a regression coefficients. For instance, consider a model similar to the example but adding student socio-economic status (SES) as a fixed covariate. If the SES parameter estimate is 2.5, then for a unit increase in SES the dependent variable (ex., test score) will increase 2.5 units - the same interpretation as in ordinary regression. If the fixed effect is a 0, 1 dichotomy such as men=0, women=1, then a coefficient of 2.5 would mean the mean value for women is 2.5 units higher than for men on the dependent variable. The intercept is interpreted as the overall mean of the dependent variable when other factors and covariates are zero.

Random effects. In the "Estimates of Covariance Parameters" table, for the row "id [subject=id] Variance", the Estimate column gives the estimated variance of the intercepts between schools. The Estimate column for the Residual row gives the estimated variance within schools (in this case, among students). Comparing the individual-level variance with the group-level variance informs the researcher about whether the variation in test scores is primarily within or between schools. In this case it is primarily within schools, by a ratio of over 4:1. (Note Singer (1998) and Singer & Willett (2003) state that the t-test of significance for the estimated variance may not be reliable.)

In this example, although within-school variation is more important, since the school variance component is significant we conclude that math scores do vary by school. This further means that a fixed-effects analysis of scores ignoring the agency effect would violate the assumption of independence of observations since observations are biased up or down depending on school. If the value 0 appeared within the confidence limits for school, the random factor (school in this case) would not be significant and an analysis with just fixed effects factors (ex., SES or race) might be possible.

Wald test. With MLE estimation, the significance of coefficients is assessed using the Wald statistic. The Wald statistic is the ratio of a coefficient to its standard error, resulting in a Z-value which is looked up in a table of the standard normal distribution to generate a corresponding probability (p) value, the normal cut-off for which is .05. The Wald test is discussed further in the section on logistic regression, including a warning about its unreliability for small samples.

Intraclass correlation: interpreting random effects variance components. Let the variance component estimate for the random factor School ID = 8.61. Let the variance component estimate for Residual = 39.15. In this null model, since the agency variance component then is 18% of the total of both variance components in this simple example, we would say that the school effect accounts for 18% of the variance in math scores. We can also say that scores cluster by school, meaning that two students randomly selected from the same school are more likely to have similar scores than a pair of randomly selected students representing different schools.

This ratio of the between-school variance component to the total of variance components is the "intraclass correlation coefficient". Note ICC as computed in this manner applies only to the null model or other random intercept models.

Goodness of fit. Goodness of fit measures for the model appear in the "Information Criteria" table of SPSS LMM output. This table contains five goodness of fit measures: -2 Restricted Log Likelihood (assuming REML rather than ML estimation was chosen), Akaike Information Criterion (AIC), Hurvich and Tsai's Criterion (AICC, meaning AIC Corrected, for finite sample corrected AIC), Bozdogan's Criterion (CAIC, consistent AIC), and Schwarz's Bayesian Criterion (BIC).

For all five measures, the lower the value the better the model. When comparing models, the model with the lower values is the best-fitting. These measures are discussed further in the section on structural equation modeling. The null or unconditional model serves as a useful baseline model to compare with other models discussed below.

Likelihood ratio tests for model differences. When considering whether to drop a term from the model, the likelihood ratio test is preferred over the Wald test. The model is run with and without the term in question and the difference taken between the two -2LL coefficients (from the "Information Criteria" table above). This is the model chi-square difference. The degrees of freedom are 1 (assuming one term is dropped, otherwise df = number of terms dropped, which is the difference in df between the two models). The df is thus the difference in the number of model parameters, which is listed in the default "Model Dimensions" tables at the top of SPSS output. The probability of a model chi-square difference that large or larger with given degrees of freedom can be looked up in a chi-square table, or can be obtained in SPSS under Transform, Compute, then entering the fomula sig.chisq2(d, df), where d is model chi-square difference and df is the degrees of freedom. If the computed probability is > .05, then the covariance parameter for the dropped term cannot be assumed to be different from 0 and the researcher proceeds with dropping the term in question.

Likelihood ratio tests for model fixed effects. The likelihood ratio test may also be used when considering dropping a fixed effect from the model. A computed probability > .05 means the population coefficient for the fixed factor cannot be assumed to be different from 0 and the researcher proceeds with dropping the fixed factor. Note: Whereas the likelihood ratio test for model covariance parameters may be used under either ML or REML estimation, the likelihood ratio test for fixed effects assumes ML estimation.

Saving new variables. Click the "Save" button in the Linear Mixed Models dialog to have SPSS save to your dataset new variables representing predicted values, standard errors, degrees of freedom, and/or residuals for each case.

A one-way ANCOVA model with random effects.

Overview. As discussed previously, this type of random intercept model is similar to the null model discussed above but includes at least one level 1 covariate. It is also a type of "conditional random effects model" since the variance in the dependent is conditional on a level 1 predictor. There are no level 2 predictors but the level 1 intercept is still predicted as a random effect of the level 2 grouping variable.

Example: For instance, differences in mean math scores may be analyzed as predicted by student socioeconomic status (SES) at level 1, predicting the level 1 intercept but not the SES slope using the level 2 grouping variable, school, as a random effect. There will be one regression of math score on SES for each school, generating as many intercepts as schools. This enables estimation of the between-schools versus within-schools variability in math scores.

SPSS. For the example where the level 2 nesting factor is ID (school ID) in a study of causes of math scores, then in SPSS select Analyze, Mixed Models, Linear; in the "Linear Mixed Models: Specify Subjects and Repeated" dialog box, assign ID to the Subjects listbox (the Repeated box is left blank). Click Continue. In the "Linear Mixed Models" dialog, assign the dependent variable (mathach) to the Dependent Variable listbox. List ID in the Factor(s) box. List SES in the Covariates box. Click the Fixed button and make SES a fixed variable. Click the Random button and set ID as the grouping variable in the Combinations area. In the same Random dialog, check "Include intercept". Set the Covariance Type: Variance Components is the default for random effects models, but Unstructured will give a solution characterized by fewer constraints and less parsimony. Click Continue, OK to run the model.

Fixed and random effects are interpreted largely as described in the previous section on the null model. For this example, student SES is shown to be a significant predictor at level 1 of the level 1 dependent, math achievement. The "id[subject=id] Variance" estimate reflects the between-schools variation in math scores. The "Residual" estimate reflects the within-schools variation in math scores.

Looking at the variance components, which total to 41.80, the between-schools component (4.77) is 11% of the total. We may say that the between-school effect accounts for 11% of the variance in math scores once student-level SES is controlled. Likewise we may say that the within-school effect accounts for 89% of the variance in math scores, controlling for SES. We may note that controlling for student-level SES reduced the between-school effect from 18% in the null model discussed above to 11% in the ANCOVA model with random effects. This is a 7% reduction. Since 7/18 rounds to .40, we may say that controlling for student-level SES reduced the between-school effect by about 40%.

Likelihood ratio test. The deviance (-2 log likelihood above) is 46644.23 for this model compared to 47116.79 for the null model. Subtracting, we can say the ANCOVA with random effects model is a better fit because its deviance is approximately 472.5 lower. From the "Model Dimension" tables (not shown), we can read that in these examples, the null model had 3 parameters and the ANCOVA model with random effects had 4 parameters, a difference of 1. Looking in a chi-square table for 1 df, at the .05 significance level the critical value is 3.84; at .01 it is 6.64; at .001 it is 10.83. Obviously, 472.5 is far greater than even the .001 level. We may say that the ANCOVA model with random effects is different from and better than the null model at a significance level better than .001.

A random intercept regression model

Overview. Random intercept regression models, also called "means-as-outcomes regression models," are a type of random intercept model in which the level 1 intercept is predicted on the basis of the level 2 grouping variable and also on the basis of one or more level 2 random effect predictors. There are no level 1 covariates, and accordingly no slopes are estimated. For instance, differences in mean math achievement scores (the intercepts) may be analyzed, predicting the level 1 intercept in terms of the between-groups effect of school and the level 2 random effect variable meanses (representing mean socioeconomic status of schools).

SPSS. In SPSS, select Analyze, Mixed Models, Linear; in the Specify Subjects and Repeated dialog, enter id (which is school id) as the Subjects variable; in the next "Linear Mixed Models" dialog, enter math achievement as the Dependent Variable and enter meanses as the Covariate; click the Fixed button and set Model to meanses, making sure the Include Intercept checkbox is checked. Then click the Random button and make id the Subject Grouping in the Combinations area, making sure the Covariance Type is Variance Components and Include Intercept is checked. Click the Statistics button to select output options such as Parameter Estimates and Tests for Covariance Parameters. Click OK.

Example notes: The level 2 grouping variable, school id, is entered as the Subjects variable (signifying individuals are independent observations within each school) and under Random effects as the Subject Grouping variable (signifying it is level 2 random effect on the level 1 intercepts, which represent mean math achievement). This causes LMM to compute separate regressions for each school, with the intercepts treated as random effects. School id is not entered as a random effect factor as that is already assumed by making it a Subject Grouping variable. Meanses is a fixed factor and, being level 2, is not a random effect of some higher level. Therefore meanses is not entered under the Random button as a random effect to be modeled.

Information Criteria table. Goodness of fit measures for the model appear here, primarily of use when comparing models, with lower log likelihood or lower information criteria values reflecting better fit. For these data, fit is better than the null model but not quite as good as the ANCOVA with random effects model (where level 1 ses was a predictor and meanses was not). Likelihood ratio tests of the differences of the random intercept regression model with either the null model or with the ANCOVA with random effects model. How such tests are computed is described in the section above on the ANCOVA with random effects model.

Tests of Fixed Effects table. For the intercept and fixed effects (here, meanses), this table shows the significance level. If Sig. &le .05 for the effect, as it is in this example, it is retained in the model and means math achievement score and meanses are significantly related within school id, the grouping variable.

Estimates of Fixed Effects table. For the intercept and fixed effects (only meanses in this example), this table shows the parameter estimate, its standard error, its significance level, and its confidence intervals. These estimates will differ from the b coefficients in the corresponding OLS regression because intercept by school id is treated as a random effect.

The estimate for the intercept is the estimate of the dependent variable, math achievement score, when meanses is controlled (is zero). This is more easily interpretable by centering meanses prior to analysis, so its mean is zero, in which case the intercept becomes the value of math achievement score when meanses is at its mean.

The estimate for meanses is interpreted the same as in regression: math achievement score increases by 5.86 for each unit increase in meanses (here coded -1, 0, +1).

Estimates of Covariance Parameters table. This table gives the covariance parameter estimates for Residual and for Intercepts within Subject=id, along with a Wald test of significance.

Intercept [subject=id] Variance. If significant, this means that intercepts of math achievement score predicted by meanses vary significantly between schools. That is, this is the between-subjects effect, where school id is the subject variable in this example. We may say there is a significant between-school effect.

For the HLM model above, the covariance component is Intercept [Subject = id] and is conditional, controlling for meanses as a covariate. That is, its share of the total of parameter estimates in the HLM model is the percent of variance in math achievement scores attributable to differences between schools after meanses is controlled.

Residual. If significant, this means that math achievement scores vary significantly within schools. That is, this is the within-subjects effect, where school id is the subject variable.

Comparison of HLM with random effects models (REM). For the same example, HLM enters school id as a subjects variable on the opening SPSS LMM dialog and as the subject groupings/combinations variable under the Random button dialog. REM does neither, but instead enters school id as a factor and under the Random button dialog enters school id as a random effects variable to model, in the model section. One will get the same effect (parameter) estimates under either model, but the standard errors of the intercept will differ somewhat, as will the information criteria goodness-of-fit measures.

Adding additional factors and covariates. The researcher may add additional level 1 (ex., individual) and/or level 2 (ex., school) factors or covariates to the model, and may model their main effects and interactions. These additional variables may be entered as fixed effects and/or random effects (the same variable may be considered both a fixed and a random effect).

A typical strategy would be to enter a full factorial model for fixed effects, then drop the effects (higher level interactions are often non-significant, for example) found non-significant in the "Estimates of Fixed Effects" table, then re-run the analysis. Likewise, if the variance of slopes involving a random effect are found to be not significant in the "Estimates of Covariance Parameters" table, that variable may be removed as a random effect (not necessarily as a fixed effect, which it may be also). Information criteria measures (ex., AIC, BIC) may be used to compare models, with lower being better fit.

A random coefficients regression model

Overview. Random coefficients regression models are a type of random coefficients (RC) model in which the level 1 slope(s) as well as the level 1 intercept is predicted on the basis of the level 2 grouping variable as a random effect, but there are no other level 2 predictors. For instance, differences in mean math scores (reflected in the level 1 intercept) may be analyzed as predicted by the level 1 covariate student socioeconomic status (SES), predicting the slope of SES as well as the level 1 intercept as a function of the between-groups effect of the grouping variable school at level 2. Such random coefficients regression models are unconditional at level 2 but conditional at level 1.

SPSS. Select Analyze, Mixed Models, Linear. On the initial dialog, let Subjects=id, where id is the level 2 school id. On the next dialog page, set the Dependent to be mathach and set SES (socioeconomic status of student; this variable is pre-centered in the data) as a covariate. Click the fixed effects button and model SES, meaning SES is a level 1 predictor of mathach. Click the Random button and move id to the Combinations area. Check "Include Intercept" (not the default). Select "Variance Components" as the model assumption (this is the default). Also under the Random button dialog, model SES as a random effect, meaning the slope of the level 1 variable SES will be modeled as a random effect of the level 2 grouping variable (school id), in which a separate regression will be computed for each school. Note SES is thus modeled both as a fixed and a random effect. Continue. Under Statistics, select Parameter Estimates and Tests for Covariance Parameters. Continue. OK.

Output. The "Information Criteria" table illustrated below shows the deviance as "-2 Restricted Log Likelihood" = 46640.66. This is lower than the 47116.79 value in the null model, indicating the model with ses as a predictor is somewhat better. Likewise, the information theory measures (ex., BIC) are lower, which is better.

The fixed effects tables also illustrated above show that SES is a significant predictor of math achievement at level 1.

In the "Estimates of Covariance Parameters" table below, the "Residual" row represents within-school variance in math achievement. The "Intercept[subject=id] Variance" row represents between-school variance in intercepts, which represent mean math achievement. The ses[subject=id] Variance" row represents between-school variance in slopes, which represent the strength of the relationships between SES and math achievement.

Partition of variance components. If a variance components model has been assumed, as for this example, then the total variance is the sum of the three values: 36.82 + 4.85 + .42 = 42.10. Of this total, most is still attributable to the variance of student scores within schools (36.82/42.10 = 87%). Another 11.5% (=4.85/42.10) is attributable to differences in intercepts (mean achievement scores) between schools, controlling for level 1 SES. This compares with a between-schools effect of about 18% in the uncontrolled null model. About 1% (=.424/42.10) of the variation in math achievement is attributable to between-school differences in slopes (representing the strength of the relation of SES to math scores), controlling for level 1 SES.

R2 estimate. The variance of math achievement within schools after SES is controlled is 36.82 in this example. In the null model it was 39.14. The difference is about 2.32. Adding individual level ses to the model thus reduces within-schools variance of math achievement by 2.32/39.14 = .06. This value is an estimate of R2 for the random coefficients model with level 1 SES as the only predictor. We may say that 6% of the within-schools variance in math scores in the null model is attributable to between-school effects when SES as a level 1 predictor is controlled. This compares with 18% of within-school variance in math scores explained by the

A second random coefficients regression example

Overview. In this second example, let agencies be the sampling unit used as a random effect in a study of performance ratings as impacted by seniority. Apart from Agency as a grouping variable, there are no other level 2 predictors. The random coefficients regression model runs a level 1 seniority-score regression for each agency, with each regression reflecting the seniority-score relationship within that agency. If the intercepts (reflecting mean rating) do not covary with the seniority slopes (reflecting strength of relationship to the dependent, score), then there is no within-group effect: agencies with higher mean scores do not have stronger seniority-score effects (slopes) on a within-group basis.

SPSS. In SPSS, select Analyze, Mixed Models, Linear; in the Specify Subjects and Repeated dialog, enter Agency as the Subjects variable. In the next "Linear Mixed Models" dialog, enter Score as the Dependent Variable and enter Seniority (in this example, C\_Seniority is centered Seniority) as the Covariate. Click the Fixed button and set Model to C\_Seniority, making sure the Include Intercept checkbox is checked. This means Seniority is a level 1 predictor of Score. Click the Random button. setting Model to C\_Seniority and make Agency the Subject Grouping in the Combinations area. Set the Covariance Type to Unstructured and check Include Intercept. Click the Statistics button to select output options such as Parameter Estimates and Tests for Covariance Parameters and Covariances of Random Effects; click OK.

Notes: Agency is not entered as a random effect factor as that is already assumed by making it a Subject Grouping variable. Variance Components vs. Unstructured covariance structure assumptions are compared in this example, below. The VC assumption assumes variance components are independent, which means their random effect terms are uncorrelated. Unstructured is preferred when the researcher determines this assumption is unwarranted or if the researcher simply does not know. As the latter is often the case, Unstructured is a common assumption.

Information Criteria table gives the -2 restricted log likelihood (a.k.a. model chi-square or deviance), plus the AIC, AICC, CAIC, and BIC goodness of fit measures as before, used when comparing models with lower being better. For instance, for the example data, fit is better (information criteria are lower) for the model run under the assumption of an unstructured covariance structure than a variance components covariance structure, as will be discussed below. Model chi-square can be used to test if the given model is significant overall (that is, significantly different from the model chi-square reported for the null model).

Tests of Fixed Effects table gives the significance level for the intercept and for C\_Seniority, the only fixed effect in the model in this example. If C\_Seniority is significant, the researcher concludes seniority does impact score. (Note df is 11 since with 12 agencies, df = number of groups minus 1).

Estimates of Fixed Effects table, illustrated above, gives the coefficients for the intercept and for C\_Seniority, the only fixed effect, along with the standard errors, significance, and confidence intervals. The coefficients computed in this table are the same whether an unstructured or a variance components covariance structure is assumed.

The coefficient for C\_Seniority is the average slope across all agencies (recall RC models run a regression for each of the agencies) and is interpreted as in regression: for each year increase in C\_Seniority, Score increases by 2.5 units, on the average. Since agency is the grouping variable, if the estimate/coefficient/slope for C\_Seniority is significant, as it is here, then performance Score and C\_Seniority are related within Agencies.

The coefficient for the intercept is the average Agency mean on performance Score, here a Score of 68. In a model where additional individual or group-level fixed factors and their interactions had been entered in the model unter the Fixed button, this table would show the significance of each fixed effect and each interaction term, and the researcher would drop non-significant terms from the model, re-running the RC model.

Estimates of Covariance Parameters table. This table, which might be said to be the "payoff" of RC analysis, gives the covariance parameter estimates for the random effects, which in this example are the Residual and "Intercept + C\_Seniority [subject = Agency]", plus the Wald statistic and its significance level.

Above, the Residual is interpreted as the within-agencies variance in performance Score controlling for C\_Seniority. The magnitude of the C\_Seniority effect could be estimated by how much the residual covariance estimate was reduced compared that in the null model without C\_Seniority as a covariate. Since the residual is unexplained variance, if C\_Seniority has a large effect on the variability of performance scores within Agencies, the residual variance should drop appreciably in the RC model compared to the null model without C\_Seniority at level 1 but with Agency as a random grouping effect and no other level 2 predictors.

When the Unstructured covariance assumption is selected, estimates are also computed for three values, listed under the "Intercept + Seniority [subject = agency]" rows:

The estimated variability of agency intercepts, labeled UN(1,1). The more the intercept variance (the higher UN(1,1)), the more the effect of group-level variables on the dependent - in this example, the more the effect of Agency on Score. This value is identical to the Intercept [subject = agency] effect for variance components models described below.

The estimated variability of Agency slopes, labeled UN(2,2). If significant, the researcher concludes that the variability is not 0 and agencies indeed differ in the slopes, which reflect the strength of the relationship of C\_Seniority and Score in this example. The larger the UN(2,2) estimate, the more the agencies differ in how much C\_Seniority affects Score. This value is also identical to the Intercept [subject = agency] effect for variance components assumptions below.

The estimated covariance of the Agency slopes and intercepts, labeled UN(2,1). If non-significant, the covariance of the intercepts (plural because the regression is done for each Agency, with intercepts reflecting mean Scores in this example) and the slopes (reflecting the strength of the C\_Seniority-Score relationship) cannot be said to differ from 0. That is, agencies with higher mean Scores do not tend to be ones with stronger (or weaker) C\_Seniority-Score relationships on a within-Agency basis. This value has no counterpart for variance components models because VC models assume random effects (in this example, Intercepts and Seniority are random effects) are independent of each other, thus having zero covariance of slopes and intercepts. In this example, UN(2,1) is not significant under the assumption of an unstructured covariance structure, thus warranting use of a VC model. Accordingly, when the Variance Components covariance type is selected under the Random button estimates are presented for three effects in the manner illustrated below:

In a VC model, variance components are uncorrelated and additive, enabling the percentage calculations below:

Intercept [subject = agency]. Same as unstructured model above, for UN(1,1). The estimate of 145.23 is 96.67% of the total variance, which is 150.23. The between-agency effect of C\_Seniority on Score accounts for over 96% of the variance in Score. That is, differences in Score are largely accounted for by differences in Agency. (For the artificial data in the example, agencies varied widely in average Score). This is the between-agency effect.

Seniority [subject = agency]. Same as unstructured model above, for UN(2,2). Between-agency differences in how much C\_ Seniority affects Score accounts for only 1% of the total variance. That is, the relation of C\_Seniority to Score is similar across agencies, even if the mean Score differs substantially by Agency. This is the C\_Seniority effect.

Residual: Same as unstructured model above, for Residual. The estimate of 3.79 is 2.5% of the total of all estimates, which was 150.23. The within-Agency effect of C\_Seniority explains 2.5% of the variance in Score. That is, only 2.5% of the variance in Score remains after C\_Seniority and Agency are controlled. This is the within-agency effect.

Random Effects Covariance Structure (G) table shows the same information as the Estimates of Covariance Parameters table but in a different format. For this example it is a 2x2 table for which the rows and columens are illustrated below for both the unstructured model and the variance components model. The "Intercept|Agency" row by "Intercept|Agency" column cell displays the variability of Agency intercepts (what above was labeled UN(1,1)). The "C\_Seniority|Agency" row by "C\_Seniority|Agency" column displays the variability of agency slopes (what above was labeled UN(2,2)). The other two cells (the off-diagonal) both display the slope-intercept covariances, labeled UN(2,1) in the Estimates of Covariance Parameters table. Similar information is presented for the VC model.

Is an RC model really needed?. If the variability of the intercept (labeled UN(1,1)) is low, there is little variability between agencies in mean value on the dependent. Further, if the variability of the slopes and of the interaction of slopes and intercepts is non-significant (that is, UN(2,2) and UN(2,1) respectively are non-significant), this suggests that it may not be necessary to model the dependent variable for within-group effects using an RC model. That is, if in the regressions for each agency conducted in the random coefficients regression model, it turns out that the slope is the same for all regressions, then the same conclusions will be reached without modeling the level 1 slope as a random effect of the level 2 grouping variable.

This could be confirmed by using a likelihood ratio (model chi-square difference) test. Take the -2LL from the "Information Criteria" table for the random coefficients regression model, then re-run the model removing C\_seniority from the model under the Random button dialog (thus leaving no random effects to model apart from the grouping variable in the Combinations area). To be comparable one would leave the covariance structure type as Unstructured. The degrees of freedom for the chi-square difference between the two -2LL's is 2 in this example (the removed slope for seniority as a random effect, and the removed covariance of this slope with the intercept). As above, the probability of a model chi-square difference this large or larger with 2 degrees of freedom can be looked up in a chi-square table, or can be obtained in SPSS under Transform, Compute, then entering the fomula sig.chisq2(d, df), where d is model chi-square difference and df is the degrees of freedom. If the computed probability > .05, then the HLM model is not significantly different from the RC model and on the basis of parsimony, one need not model seniority as random coefficients (but one would still use agency as the grouping variable and have whatever individual level and agency level covariates were in the analysis).

A third random coefficients regression example.

Overview. This additional random coefficients regression model is used to compare linear mixed models with related techniques. The example is a study of city real property appraisal values ("Appraise") to see how well they predict actual selling price of homes ("Price"), for each of three cities ("City"). The dependent variable in this example is Price. City is the grouping factor, conceptualized as a random effect. Appraise is a fixed covariate at level 1. There are no other level 2 predictors.

SPSS. In SPSS, select Analyze, Mixed Models, Linear; Continue past the initial dialog since there are no repeated measures; in the next "Linear Mixed Models" dialog, enter Price as the Dependent Variable, City as Factor, Appraise as covariate; click the Fixed button and set Model to Appraise (so we are modeling one main fixed effect); click the Random button and set Model to City (again modeling one main random effect); click the Statistics button to select output options such as Parameter Estimates and Tests for Covariance Parameters; click OK.

Output. Output is similar to that discussed for the first two models, including estimated coefficients for fixed and random effects. Interpretation is parallel to previous examples.

Comparison with Variance Components. Linear mixed models in SPSS has the functionality of the variance components procedure described separately. That is, VARCOMP is mostly a subset of MIXED. For the same property appraisal example discussed above for linear mixed models, one may enter Price as dependent, Appraise as a covariate, and City as a random factor, specifying REML as the method so as to be comparable, to obtain the "Variance Estimates" below, which are the same as in the "Estimates of Variance Components" table above for linear mixed modeling. However, variance components will not generate most of the additional output tables available via linear mixed models. While mostly a subset of MIXED, VARCOMP does have a few options not present in MIXED: it offers ANOVA and MINQUE estimation, not just ML and REML; and with ANOVA estimation comes sums of squares, mean squares, and expected mean squares not produced by the ML and REML estimation methods used by MIXED.

Comparison with GLM. While one may obtain similar estimates using the general linear model (GLM), they will not be identical to the variance components and linear mixed model implementations illustrated above. When the design is unbalanced, as it is in this example (there are unequal numbers of observations in Cities 1, 2, and 3), the difference may be substantial. In GLM, one may enter Price as Dependent, Appraisal as a covariate, and City as a random factor, with the model set to main effects for Appraise and City, and selecting "Parameter estimates" under the Option button. GLM will generate these tables:

GLM produces Type III sums of squares for fixed effects only. Evem though City is entered as a random factor, the table above treats City as if it were a fixed effect for purposes of computing the sums of squares used to compute the F statistic. GLM estimates variance parameters for City (or any random effect) indirectly as described below, using expected mean squares. Linear mixed models and variance components analysis, in contrast, estimate variance parameters directly, using maximum likelihood (ML) or restricted maximum likelihood (REML) methods. For unbalanced designs (unequal n's in the groups formed by the random effect), as in this example, the GLM method will return estimates different from the methods used by linear mixed models or variance components analysis. Thus where in the linear mixed model run, the F value for the main effect Appraisal was 3.010E3, for the GLM run it is 2.994E3 in the table above.

Above it is seen that the GLM method generates coefficient estimates for the fixed effect Appraise similar to that for linear mixed model. It also generates coefficients for the random effect City, which is not part of linear mixed model output due to the LMM not being based on sums of squares methods of estimation. However, the variance estimate for City, which was 38,180,000 in linear mixed modeling and in variance components analysis, is only 21,376,633 in GLM. The GLM variance estimate is computed as Var(City)=[MS(City)-MS(Error)]/EMS(City), where MS(City) = 4.215E9 and MS(Error)=6.519E8 (both from the "Between Subjects Effects" table in GLM) and EMS(City)=166.682 (from the "Expected Mean Squares" table in GLM). Even when the LMM and GLM variance estimates are the same as they will be in balanced designs, GLM has the drawback that the standard error of estimate for the variance of random factor(s) (ex., City) that appear in the "Estimates of Covariance Parameters" table in LMM, cannot be computed in GLM.

A full random coefficients model

Overview. A full random coefficients model, also called an intercepts-and-slopes-as-outcomes model, is an extension of a random coefficients regression model to include level 2 predictors in addition to the grouping variable and any level 1 predictors. Thus the level 1 slope(s) as well as the level 1 intercept is predicted on the basis of the level 2 grouping variable as a random effect and also by one or more other level 2 random effect predictors.

Example: For instance, differences in mean math scores (reflected in the level 1 intercept) may be analyzed as predicted by the level 1 covariate student socioeconomic status (SES), predicting the slope of SES as well as the level 1 intercept as a function of the between-groups effect of the grouping variable school at level 2 and by the level 2 random effect of Meanses (a centered variable representing whether a school is below, near, or above average in mean socioeconomic status of its students. Such full random coefficients regression models are conditional at both levels 1 and 2.

SPSS. Select Analyze, Mixed Models, Linear. On the initial dialog, let Subjects=id, where id is the level 2 school id. On the next dialog page, set the Dependent to be mathach and set SES (socioeconomic status of student; this variable is pre-centered in the data) as a covariate. Also set Meanses (a -1/0/+1 coded variable signifying if a school is below, near, or above average in mean SES) as a covariate. Click the fixed effects button and model SES, meaning SES is a level 1 predictor of mathach. Also enter as fixed effects both Meanses and Meanses\*SES. Click the Random button and move id to the Combinations area, meaning it is the grouping variable. Also under the Random button dialog, model SES as a random effect, meaning the slope of the level 1 variable SES will be modeled as a random effect of the level 2 grouping variable (school id), in which a separate regression will be computed for each school. Check "Include Intercept" (not the default). Select "Variance Components" as the model assumption (this is the default). Continue. Under Statistics, select Parameter Estimates and Tests for Covariance Parameters. Continue. OK. Note SES is thus modeled both as a fixed and a random effect, but Meanses is modeled only as a random effect.

Notes. The SPSS interface is less than intuitive, but note the following:

The level 1 intercept is modeled as a random effect of having school id as the grouping variable.

The level 1 slope of SES is modeled as a random effect of having school id as the grouping variable and declaring SES to be a random effect to be modeled, under the Random button dialog.

The level 1 intercept is also modeled as an effect of level 2 Meanses by including Meanses as a fixed effect.

The level 1 slope of SES is also modeled as an effect of level 2 Meanses by including the Meanses\*SES interaction term as a fixed effect.

Thus both the level 1 intercept and the level 1 slope are modeled as effects of both the level 2 grouping variable (school) and the level 2 covariate (Meanses).

The fixed effects output will have four coefficients for this example, reflecting the two level 2 effects (school as grouping variable, Meanses as covariate) on the two level 1 parameters (intercept and slope of SES).

The random effects output will have three coefficients for this example, reflecting the three variance components: (1) the Residual, reflecting within-school random effects; (2) the SES component, reflecting between-school random effects on the slope of SES; and (3) the Intercept component, reflecting between-school random effects on the level 1 intercept.

Managing a full random coefficients model with multiple predictors at each level can become a daunting logical puzzle in SPSS. The HLM software interface, presented separately, is far more intuitive and thus much less prone to human errors in constructing multivariate full random coefficients models.

Output

Fixed effects.

In the table above, there are four fixed effects:

Intercept: This represents the effect of the school-level intercept on the level 1 intercept for math score. The mean math score is 12.62 when predictors are 0.

SES: When student-level SES increases by 1 unit, math score is predicted to increase by 2.19 units, controlling for other variables in the model. That is, this is the predicted slope of SES when Meanses is 0 (which means at-average as coded in the current example).

Meanses: When Meanses increases by 1 unit (recall Meanses is coded -1, 0, 1), math score is predicted to increase by 3.77 units, controlling for other variables in the model (that is, when SES is 0, which, if centered as it is in this example, means at its mean).

SES\*Meanses: This interaction effect is the effect of the level 2 variable Meanses on the slope of SES at level 1. As Meanses increases by 1 unit, the slope of SES is predicted to increase by .18.

Random effects / variance components.

In the table above, there are three fixed effects:

Residual: This is the largest variance component and represents within-school variance in math scores.

Intercept[subject=id] Variance: This much smaller variance component represents the between-school effect on the intercept, which is mean math score.

ses[subject=id] Variance: This small and marginally significant variance component represents the between-school effect on the slope of SES. For these data, modeling SES as a random effect of level 2 makes little difference and a more parsimonious model might be judged better.

Model fit.

In the table above, the deviance (-2LL) is 46563.17, with 7 parameters (number of parameters is displayed in the "Model Dimension" table, not shown). In the random coefficients regression model discussed previously, the deviance was 46640.66, with 5 parameters. This is a model chi-square difference of 77.43 and a degrees of freedom difference of 2 parameters. In a chi-square table with 2 degrees of freedom, the critical value is 5.99. As 77.43 is far larger, we can say that the full random coefficients model is significantly better fit than the random coefficients regression model. This method of comparison is called the "likelihood ratio test" or the "model chi-square difference test."

Of course, the full random coefficients model might be compared to yet other models. One obvious model for these data would be a modified full random coefficients model, identical to the one just run, but dropping the interaction term (ses\*meanses) as a fixed effect. Dropping it means the level 1 slope is still made a function of the level 2 grouping variable, school, but is no longer also estimated by the level 2 covariate, meanses. The deviance for this modified model, which has 6 parameters (one less, having dropped the interaction term), is 46562.91. Comparing the two models, the chi-square difference is .25 and df = 1. The critical value of chi-square of 1 degree of freedom is 3.84, much larger than .25. We conclude that there is no significant difference in model fit and therefore choose the modified model as the better one on parsimony grounds.

A second full random coefficients example. This second example is similar the the one above but adds one additional level 2 predictor, the dichotomous variable "sector," (coded 0=public, 1=parochial). The remaining aspects of the model remain the same, except the level 1 variable ses is renamed "cses" to indicate that it is centered.

SPSS. In SPSS, select Analyze, Mixed Models, Linear; in the Specify Subjects and Repeated dialog, enter id (which is school id) as the Subjects variable; in the next "Linear Mixed Models" dialog, enter math achievement as the Dependent Variable; enter cses, meanses, and sector as Covariates; click the Fixed button and set Model to cses, meanses, sector, meanses\*cses, and sector\*cses, making sure the Include Intercept checkbox is also checked. Then click the Random button and make id the Subject Grouping in the Combinations area, with the Covariance Type as Variance Components and making sure Include Intercept is checked. Also under the Random button, model cses. Click the Statistics button to select output options such as Parameter Estimates and Tests for Covariance Parameters. Click OK.

Example notes: In this illustration, cses is a fixed effect as it would be in OLS regression. The level 2 variables, meanses and sector, are fixed effects also: they are not random effects of some higher level. The interaction effects, meanses\*cses and sector\*cses, must also be added as fixed effects as they are used to estimate the effects of meanses and sector on the slope of cses.

NOTE: When there are both level 1 and level 2 predictors, and when both level 1 intercepts and level 1 slopes are to be modeled as random effects, the interactions of the level 2 predictors with the level 1 predictors must be modeled.

Output.

Information Criteria table. Again, this table displays goodness of fit measures, where lower is better fit. For these data, fit is better than the null model or previous models. This could be tested using the previously-described likelihood ratio test.

Tests and Estimates of Fixed Effects tables. These tables show all fixed effects are significant.

In the "Estimates of Fixed Effects table" above, there are two sets of coefficients, though SPSS does not clearly list them in sets.

Coefficients pertaining to the level 1 intercept of math achievement predicted by student's centered ses. The higher the intercept, the higher the mean math score. The "Intercept" means that average math score is 13.33 when predictors are 0. The "meanses" means that this intercept increases by 5.34 units when meanses increases 1 unit (recall meanses is coded -1, 0, 1 and that one regression is run for each school in a random coefficients model), controlling for other variables in the model. The "[sector=0]" value means that public schools (sector=0) compared to parochial schools (sector = 1) lowers the intercept (representing mean math scores) by 1.21, controlling for other variables in the model. That is, parochial schools have higher mean math scores.

Coefficients pertaining to the level 1 slope of cses as a predictor of math achievement. The higher the slope, the stronger the relation of cses to math score. The "cses" value is the estimated mean slope of cses across all schools when sector and meanses are 0, and under these conditions, when cses increases 1 unit, math score increases by 1.3 units. The "cses\*meanses" value means that when meanses increases by 1 unit, the slope of cses increases by 1.03 units, controlling for other variables in the model. The "[sector=0]\*cses" value means that public schools (sector = 0) compared to parochial schools (sector = 1) increase the slope of cses by 1.64 units, controlling for other variables in the model. That is, the relation of cses to math score is stronger in public schools.

All these estimates are tested to be significant, as indicated in the "Sig" column.

Estimates of Covariance Parameters table. This table gives the covariance parameter estimates for Residual and for Intercepts within Subject=id, along with a Wald test of significance.

Intercept [subject=id] Variance. If significant, this means that intercepts of predicted math achievement score vary significantly between schools. That is, this is the between-subjects effect, where school id is the subject variable. We may say there is a significant between-school effect on mean math scores. In a variance components model, its proportion of the total of parameter estimates is the percent of variance in math achievement scores attributable to differences between schools after other predictors in the model are controlled. The between-group variance component divided by the total of variance components is .06. We may say 6% of the total variance in math achievement is due to the between-schools effect on mean math scores (on the level 1 intercepts).

Cses[subject=id] Variance. This coefficient reflects the between-schools effect on the slope of the level 1 predictor, cses. For these data, this coefficient is not significant, indicating that the slopes of cses do not vary significantly when, under a random coefficients model, separate regressions are run for each school. Put another way, we may say there is no variance between schools in the slope of cses. Put a third way, the strength of relationship between cses and math achievement does not vary significantly between schools.

Residual. If significant, this means that math achievement scores vary significantly within schools. That is, this is the within-subjects effect, where school id (not student id!) is the subject variable. For these data, almost 94% of the variance in math achievement scores is explained by variance among students within schools, assuming a variance components model.

Three-level hierarchical linear models

Example: a three-level ANCOVA model with random effects. Let level 1 = employee, level 2 = agency, level 3 = department, such that employees are nested within agencies which are nested within departments. Let performance score be the level 1 dependent, and let femalepercent be the covariate. Let prescore be an earlier performance score used as a control variable so as to focus on change in score. Let gender be a level 1 factor. The general approach for three-level HLM is to make the level 3 grouping variable the subject variable (here, Department). Department is also listed as the grouping variable in the Combinations area of the Random Effects dialog. The dependent, of course, is a level 1 variable (individual level, here performance score). The level 2 grouping variable (here, agency) is modeled as a random effect.

SPSS for three-level models. In SPSS, select Analyze, Mixed Models, Linear; in the Specify Subjects and Repeated dialog, enter Department as the Subjects variable; in the next "Linear Mixed Models" dialog, enter Gender (a level 1 factor) and Agency (the level 2 grouping variable) as a Factors, and also enter Score as the dependent variable (which is always level 1) and Prescore as the level 1 Covariate. Click the Fixed button and set Model to Gender Prescore, making sure the Include Intercept checkbox is checked. Click the Random button and set Model to Agency and then in the Subject Grouping in the Combinations area let Department be the grouping variable, selecting the Covariance Type as Variance Components (the default) and checking Include Intercept; click the Statistics button to select output options such as Parameter Estimates and Tests for Covariance Parameters; optionally, in the EM Means button set "Display Means for" to Gender (or other factors of interest) and check "Compare main effects"; click OK.

Note: This is a three-level ANCOVA model because unlike random intercept models there are no higher-level predictors other than the level 2 and level 3 grouping variables, but unlike random coefficient regression models or full random coefficient models, the slopes of the level 1 predictors are not modeled at higher levels. ANCOVA models have level 1 predictors whose slopes are not modeled but employ grouping variables (indicating higher levels) as random effects on the level 1 intercept.

Information Criteria table. Goodness of fit measures for the model appear here, interpreted as discussed for previous two level models.

Tests of Fixed Effects table. For the intercept and fixed effects (here, gender and prescore), this table will show the significance level. If Sig. &le .05 for the effect, it is retained in the model and means performance score and either or both of gender and prescore are significantly related to score (the dependent). In this example, it appears neither is a significant effect.

Estimates of Fixed Effects table. For the intercept and fixed effects (gender and prescore in this example), this table shows the parameter estimate, its standard error, its significance level, and its confidence intervals. Because prescore is continuous, it will have a single estimate. For a categorical variable there will be one estimate per category, except the last category is redundant and is set to 0. Thus for gender, there will be an estimate for gender=0 (F) but not for gender=1 (M). Interpretation of fixed effects parameters was in previous examples. A significant estimate means it can be assumed to be different from 0 and that effect is contributing to the model. Here they do not (though Prescore comes close and might be significant with a larger sample).

Estimates of Covariance Parameters table. This table will give the covariance parameter estimates for Residual, for Intercept with Subject=department, and agency with Subject=department, along with a Wald test of significance for each.

Intercept [subject = Department]. The significance of the intercept tests the level 3 between-groups effect: if it is not significant, as is the case here, then the variablility associated with department cannot be assumed to differ from 0, by the Wald test. Mean Score, controlling for other variables in the model, does not differ between Departments.

Agency [subject = Department]. The significance of the agency estimate tests the level 2 between-groups effect (recall Agency was entered as the variable to model under the Random random effects button): if it is not significant, as here, then the variability associated with agencies within departments cannot be assumed to differ from 0, by the Wald test. Note, however, that the preferred likelihood ratio test will not necessarily come to the same conclusion.

Residual. The residual, which is significant in this example, is the within-subjects effect. There is significant variance in Score within departments and agencies, even after controlling for other effects in the model.

EMM Estimates table. If selected, the EMM Means button will generate this and the following two tables. Recall in the example above, we clicked the EMM Means button and set "Display Means for" to gender and checked "Compare main effects." The Estimates table gives the mean, standard error, degrees of freedom, and 95% confidence level for each level of gender (male=0, female=1). Below we see males and females differ very little on mean Score.

Pairwise Comparisons table. Also generated under the EMM Means button, this table is of greater interest, especially when there are more categories than the two reflecting Gender. This table gives the mean difference between categories, the standard error of the difference, its significance level, and the 95% confidence interval for the difference. For gender as an example, if the difference is non-significant, as it is here, then the mean score (the dependent) difference between men and women cannot be assumed to differ from 0. Of course, if the comparison were done on a variable with more than two categories (ex., ethnicity), then there will be more than one pairwise comparison to assess.

Univariate Tests table. Also generated under the EMM Means button, this table gives and F test significance for the effect of the selected variable (here, gender) on the dependent variable, based on the linearly independent pairwise comparisons among the estimated marginal means. For gender, which has only two categories and thus only one pairwise comparison, this significance level will be the same as that reported in the Pairwise Comparisons table. However, for a variable with more than two categories, the Univariate Tests significance level tests the significance of the predictor variable overall (ex., ethnicity), whereas the Pairwise Comparisons table will report a significance level for each pair of categories (ex., Irish vs. Italian ethnicity) but no overall significance test of ethnicity.

Hierarchical linear and multilevel models in HLM. HLM software is perhaps the leading software for multilevel modeling. Extensive discussion is provided separately: click here.

Repeated Measures Models Using LMM

Repeated measures as mixed models:. The Repeated option in LMM models within-subject variance (ex., variance in the same subjects over time, as opposed to between-subject variance modeled by the Random option discussed above). Example. In LMM, one may create a model in which level 1 is within-individuals (for the variance among repeated measures for given individuals, on the average) and level 2 is between-individuals with the individuals considered a random effect. For example, employees might be measured at three times, obtaining three performance scores rather than one. There might also be covariates, such as seniority, a baseline performance score, or a time variable - all as discussed below. The time variable baseline is coded 0 and subsequent measurement times are 1, 2, 3, etc. If the time codes represent equal metric intervals, the time variable may be used as a covariate as well as fixed effect.

Repeated measures data setup. The data format for repeated measures LMM is one row per person per measurement time ("time as case format"). Thus if each person is measured at three times, there may be three data rows for person 1, then three rows for person 2, etc. A major advantage of LMM repeated measures over GLM repeated measures is that if a given subject was not measured at a given time, that data row may be simply omitted from the data set: the values of the "time" variable may be different for different subjects. Each data row thus contains an id variable for the person, a time variable (0, 1, 2 in this example), and whatever effects are being modeled. If the researcher's raw data is of the one row per person type, with repeated measures being different columns for each person, then the data must be converted to rows per person per time format. This conversion can be done with the SPSS Data Wizard, under Data, Restructure in the menu. Restructuring is illustrated in SPSS, Inc. (2005: 1-4).

SPSS. Select Analyze, Mixed Models, Linear and under "Specify Subjects and Repeated" enter the id variable as "Subjects"; in the "Linear Mixed Models" dialog, enter performance score as the "Dependent Variable" and make the time variable the "Factor(s)"; click the Fixed button and Model the time variable, making sure to check "Include Intercept"; click the Random button and set the "Covariance Type" to Variance Components, make the id variable the "Subject Groupings" in the Combinations area, and make sure to check "Include Intercept"; click the Statistics button and select the desired output; click OK.

Note that employee id is thus both the Subjects variable and the grouping variable (Combinations). Time is modeled as a fixed effect. In this example employee is level 1 as Subjects variable and is level 2 as Combinations variable.

Covariance structure assumptions. The initial LMM (Mixed) dialog in SPSS provides a pulldown menu of choices of assumptions to make about the covariance structure of residuals in repeated measures models. The default is "Diagonal". Options are discussed above in the "Key Concepts and Terms" section.

Output

Tests for Fixed Effects table. If the F test for the time variable is significant, then performance score varies by time of measurement within the same individuals. That is, there is a "time effect" by which subjects perform better at later testing times (assuming a positive relationship).

Estimates of Covariance Parameters table. This is the key output table for repeated LMM modeling.

Intercept-only model. In the simplest model, with subjects (id) measured on performance (score) at multiple measurement periods (time), id is the subject variable, the dependent is score, and time is a fixed effect entered as a factor. The only random effect is the intercept (assuming the "Include intercept" check box is checked), grouped by subject (Combinations=id in the SPSS dialog). Time is not modeled as a random effect. In the Estimates of Covariance Parameters table, the only parameter will be "Intercept[subject=id]", which is the intercept as a random effect. If this term is significant by the Wald test, between-subjects effects do impact the score variable. In a variance component model, the variance of this term divided by total variance (intercept + residual variance) is the percent of variance in score explained by between-subject effects.

For example, if the estimate for the "Intercept[subject=id]" intercept were 750 and the estimate for the Residual parameter was 250, with no other parameters, then 75% of the variance in performance score would be attributable to variability between subjects if one is using the default variance components model, which supports additivity of components.

If more complex models below are run, one can look at (1) how much adding terms to the model reduces the residual covariance estimate, and (2) from the "Information Criteria" table in SPSS output, how much AID, BIC, or other information goodness-of-fit measures are reduced.

Intercept + time model. If the time codes 0, 1, 2 are not be arbitrary but reflect a real scale, such as 1=elapse of one month, 2=elapse of two months, etc., then time can be modeled as a covariate predictor. In SPSS, id is the subject variable, the dependent is score, and time is a fixed effect entered as a covariate. Assuming the "Include intercept" checkbox is checked in the Random effects dialog and one selects to model time, and Unstructured is selected as the covariance matrix type, the random effects will be "Intercept + time[subject=id]" in this example.

In this intercept + time model, the estimate in the "Estimates of Fixed Effects" table is a slope (regression coefficient) indicating the number of performance points employees change on the average for each unit increase above the mean (the intercept) in time. In the "Estimates of Covariance Parameters" table there will be a parameter for "Intercept+time[subject=id]" with rows for:

UN(1,1): variance estimate for the intercepts, the significance test for which tests if the intercept of performance score (the mean value for employees) varies significantly between subjects.

UN(2,2): variance estimates for the slopes, the significance test for which tests if the slope varies significantly between individuals.

UN(2,1): estimate of covariance between intercepts and slopes, the significance test for which tests intercept-slope interaction. If positive and significant, in this example, as intercepts go up with time, slopes also go up (time has even more effect on score increase).

Residuals. The Residual covariance estimate also appears in this table, reflecting the estimated variability in performance score across time for the average employee. That is, the "Residual" line in the "Estimates of Covariance Parameters" table is the within-person effect after controlling for any covariates (none in an intercept-plus-time model). This can be compared with the corresponding residual coefficient for the intercept-only model to see if there has been improvement (if it is lower). It may also be compared to the residual in an intercept + time + covariates model, discussed below, to assess effects of adding one or more covariates.

Intercept + time + covariates models. It is, of course, possible to add other covariates such as seniority to the model. In SPSS this is done simply by adding seniority to the "Covariate(s)" list in the "Linear Mixed Models" dialog, as well as adding seniority to the Model list under the Fixed button. The fixed effects Model list optionally could also include interactions of covariates (ex., time\*seniority, assuming time is a covariate also).

Additional covariates serve as control variables. For instance, if a baseline performance score, prescore, were added as a covariate fixed effect (with the same value for any given individual across time periods; prescore is added to the covariate list and to the model under the Fixed Effects button), then one would be modeling change in performance score (score controlling for prescore). If the prescore\*time interaction were also added (added to the Model list under the Fixed button but not added to the covariate list), then in the "Tests of Fixed Effects" table a significant positive prescore\*time interaction would indicate that as prescore increases so does the rate of increase in performance score (the dependent).

Example: A three-level model with repeated measures. Consider a situation in which Verbal is a test of verbal ability given in each of three years (Schlyear is the repeated measure), to students (Student is the student id variable) who are grouped in classes (Class is the class id variable) which in turn are grouped in schools (School is the school id variable). Let the researcher be interested in the effects of gender (Sex is the variable) and socioeconomic status (SES is the variable).

Subject and repeated variables. In this example, three variables define the subject (Student, Class, School). Student is not enough since student id's are assumed unique to each class, and class id's assumed unique to each school. Therefore all three variables are entered in the Subjects area of the initial Linear Mixed Models: Specify Subjects and Repeated dialog in SPSS, as illustrated below. Schlyear is the repeated measure (not Verbal, which is the dependent). The covariance structure type for Schlyear is set to AR(1), which is often chosen when there is thought to be a common trend, such as increasing scores over time. The effect of this initial dialog page is to model Schlyear by student (Student, Class, and School are all needed to specify the student subject), testing if residual errors on estimates of the Verbal score (the dependent, to be named on the next dialog page) are correlated within each student. That is, if the repeated measures effect for Schlyear is significant, as shown later in the Estimates of Covariance Parameters table in SPSS output, then for a typical student, knowing error of estimate of Verbal in one year helps in predicting error or estimate in another year.

Dependent and factor variables. In the main Linear Mixed Models dialog, illustrated below, Verbal is entered as the dependent variable and the factors entered are the three subject variables (Student, Class, School), the repeated measure (Schlyear), and the causal factors of interest (SES, Sex). Including Schlyear as a fixed factor controls for trends in Verbal over the three years measured (that is, in effect the parameters for other fixed effects will be for detrended data).

Modeling the fixed factors. Clicking the Fixed button is necessary to specify exactly how the fixed factors are to be modeled (main effects, two-way effects, factorial, etc.). As illustrated below, in this example only main effects are modeled. Type III tests are specfied, which is usual in this and most statistical procedures.

Modeling random effects. Similarly, clicking the Random button allows the researcher to specify random effects to be modeled. The researcher can specify multiple random effects (in syntax, multiple RANDOM statements) by clicking the Next button and filling in the dialog. Each random effect created this way is assumed to be independent from the others. Here there are two independent random effects, with the second one illustrated below. Note, however, that multiple nonindependent random effects could be specified by entering additional variables in the Model box for a single random effect statement (not the case here, which for each of the two random effects specified only SES, once for class and once for student).

The random effect above models SES as a random effect within students. That is, Estimates of Fixed Effects table in SPSS will later show if SES seems to be related to Verbal scores. In contrast, the random effect of SES, shown in the Estimates of Covariance Parameters table, assesses if a student effect due to sampling of students conceived to be at random from a larger population of students, significantly adjusts the variation in SES. A second random effect, not illustrated, does the same thing for Class: it tests if the variation in SES is significantly due to sampling of classes from a random sample of classes. Because the two random effects are modeled separately, the researcher is assuming the sampling of students is uncorrelated with sampling of classes.

Output. Most statistical output is specified under the Statistics button from the Linear Mixed Models dialog, as illustrated below.

Tests and estimates of fixed effects tables. In the example output below, the Type III Tests of Fixed Effects table shows that SES, School, and Schlyear all had a significant effect on Verbal score, but not Sex. That is, at least one category of the significant variables was related to Verbal. The Estimates of Fixed Effects table spells this out by category. In this example, School, Schlyear, and Sex all had three levels (categories). School level 1 and SES level 1 were significantly more related to Verbal than other levels. Schlyear 1 compared to 3 and 2 compared to 3 were both significantly related to Verbal score. The last level in each variable serves as the reference category.

Estimates of covariance parameters table. This table shows the significance of the repeated and random effects in the model. The nonsignificant AR1 rho means that residual error is not correlated by Schlyear and therefore a scaled-identity (ID) covariance structure may be assumed. One random effect is nonsignificant and the other is redundant, so these may be dropped from the model.

Information criteria table. Using AICC, BIC, or one of the other information theory measures of goodness of fit is useful in comparing models. The lower the measure, the better the model fit. For instance, one might use these measures to assess improvement in fit due to dropping a non-significant random effect from the model. In the illustration below, Model 1 is the model described above and Model 2 is the model with the covariance structure for repeated measures specified as scaled identity rather than AR(1), and with no random effects specified. CAIC and BIC indicate Model 2 to be the better fitting model but AIC and AICC do not. The researcher might conclude that the two models were largely equivalent, but might note that between the two models, Model 2 is the more parsimonious.

Assumptions

Random grouping. For levels above the bottom level of individuals as units of analysis, the groups (ex., schools, where students are the bottom level) are assumed to be a random sample of all such groups (ex., all schools) in random coefficients models. This is a critical assumption in multilevel modeling. I

Independent observations are not assumed, which is why multi-level modeling is recommended when intraclass correlation exists. That is, OLS regression and GLM in general assume error terms are independent and have equal error variances, whereas when one has hierarchical data, individual-level observations from the same upper level group will not be independent but rather will be more similar due to such factors as shared group history and group selection processes. This clustering by group increases Type I error if not taken into account, as LMM does. While random effects (such as upper-level effects, which LMM models as random effects) do not affect lower-level population means they do affect the covariance structure of the data and, indeed, adjusting for this is a central point of LMM models and why they are used instead of GLM, which assumes independence.

To check for lack of independence, meaning some form of mixed modeling is needed, the researcher can run an OLS regression and save the residuals. An ANOVA of residuals by group (ex., agency, where agency is the level 2 grouping for level 1 individual data) can be run. If the ANOVA F-test is significant, the researcher rejects the null hypothesis that residuals are independent by group. That is, a significant F means data are correlated, not independent, and LMM should be used instead of OLS.

Intraclass correlation is a measure of the extent to which observations are not independent of a grouping variable (ex., schools). The presence of a significant intraclass correlation is an indicator of the need to employ multi-level modeling rather than conventional regression. To pursue OLS regression modeling anyway in the face of lack of independence and lack of homoscedastic error variance will mean that significance tests will not be accurate. Put another way, OLS significance tests (and standard errors and confidence limits) are not at all robust when the assumption of independence is violated.

Independent blocks. While observations are not assumed to be independent, the groups (blocks) formed by the subject variable(s) are assumed to be independent and to have the same covariance structures, for models that involve random effects or repeated measures.

Properly specified block structures. For random effects and repeated measures, the researcher must specify the assumed covariance structure. Changing the specified covariance structure will change the covariance parameter estimates and tests of fixed effects.

Adequate sample size. Hox (1995) suggests that for MLM regression models, the higher level sample size be at least 20, preferably 50, and if variance components are important, preferably 100. For structural equation modeling approaches, Hox recommended sample size should be at least 50, preferably 100. More recently Maas & Hox (2005: 90), based on simulation studies, concluded, "The standard errors of the second-level variances are estimated too small when the number of groups is substantially lower than 100. With 30 groups, the standard errors are estimated about 15% too small..." In view of this, Maas & Hox (2005: 91-92) suggest that with small samples (as small as 10), that bootstrapping be used to estimate standard errors. However, bootstrapping was only available for multilevel modeling in MLwIN software, at least as of the time of writing.

The efficiency and power of multi-level tests rests on pooled data across the units comprising two or more levels, which implies large datasets. The REML and ML estimation methods used by LMM give asymptotically efficient estimates, meaning efficiency depends on large samples.

For instance, simulation studies by Kreft (1996) found there was adequate statistical power with 30 groups of 30 observations each; 60 groups with 25 observations each; 150 groups with 5 observations each. The number of groups has more effect on statistical power than the number of observations, though both are important. There is a rapid fall-off in statistical power as the number of groups/observations falls below the threshhold needed. With less than adequate power there is an unacceptable risk of not detecting cross-level interactions (ex., between schools and students). However, both adequate number of individual observations and adequate number of groups are needed. Power for individual-level estimates depends on number of individuals observed, and power for second level estimates depends on number of groups.

Specifically with regard to MSEM, Hox & Maas (2001) used simulation studies to show for small group-level sample sizes, coefficient estimates were not stable. They recommended group-level samples of at least 100. However, Cheung & Au (2005: 612) used resampling to test sample size effects and found sample size "can be as small as 50, yet the results are still comparable with other larger sample size conditions." Unbalanced individual-level samples within groups may require larger group samples. Cheung & Au's experiments also disconfirmed the assertion of some that larger individual-level samples could compensate for small group-level samples.

The above cautions notwithstanding, note that the default method of estimation, based on maximum likelihood, requires large sample assumptions. However, if Bayesian estimation is selected, smaller level 2 samples may be tolerated (Raudenbush and Bryk, 2002: 14).

Similar group sizes. The REML and ML estimation methods used by LMM give asymptotically efficient estimates for unbalanced as well as balanced designs (whereas the ANOVA methods in GLM are optimum only for balanced designs). Nonetheless, when sample sizes within groups are unbalanced, tests of parameter estimates and of overall fit will have inflated Type I error (Hox & Maas, 2001). However, FIML estimators are more robust for unbalanced designs (du Toit and du Toit, 2005).

Normal distribution. RC models assume a normal distribution for purposes of empirical Bayes maximum likelihood estimation. However, REML and ML estimates may be assumed to display asymptotic normality for large samples. Also, extensions have been developed for non-normal data (Wong and Mason, 1985; Goldstein, 1991; Morris, 1995).

Frequently Asked Questions

What is the SPSS syntax for linear mixed modeling (MIXED)?

MIXED dependent varname [BY factor list] [WITH covariate list]

[/CRITERIA = [CIN({95\*\* })] [HCONVERGE({0\*\* } {ABSOLUTE\*\*})

{value} {value} {RELATIVE }

[LCONVERGE({0\*\* } {ABSOLUTE\*\*})] [MXITER({100\*\*})]

{value} {RELATIVE } {n }

[MXSTEP({5\*\*})] [PCONVERGE({1E-6\*\*},{ABSOLUTE\*\*})] [SCORING({1\*\*})]

{n } {value } {RELATIVE } {n }

[SINGULAR({1E-12\*\*})] ]

{value }

[/EMMEANS = TABLES ({OVERALL })]

{factor }

{factor\*factor ...}

[WITH (covariate=value [covariate = value ...])

[COMPARE [({factor})] [REFCAT({value})] [ADJ({LSD\*\* })] ]

{FIRST} {BONFERRONI}

{LAST } {SIDAK }

[/FIXED = [effect [effect ...]] [| [NOINT] [SSTYPE({1 })] ] ]

{3\*\*}

[/METHOD = {ML }]

{REML\*\*}

[/MISSING = {EXCLUDE\*\*}]

{INCLUDE }

[/PRINT = [CORB] [COVB] [CPS] [DESCRIPTIVES] [G] [HISTORY(1\*\*)] [LMATRIX] [R]

(n )

[SOLUTION] [TESTCOV]]

[/RANDOM = effect [effect ...]

[| [SUBJECT(varname[\*varname[\*...]])] [COVTYPE({VC\*\* })]]]

{covstruct+}

[/REGWGT = varname]

[/REPEATED = varname[\*varname[\*...]] | SUBJECT(varname[\*varname[\*...]])

[COVTYPE({DIAG\*\* })]]

{covstruct†}

[/SAVE = [tempvar [(name)] [tempvar [(name)]] ...]

[/TEST[(valuelist)] =

['label'] effect valuelist ... [| effect valuelist ...] [divisor=value]]

[; effect valuelist ... [| effect valuelist ...] [divisor=value]]

[/TEST[(valuelist)] = ['label'] ALL list [| list] [divisor=value]]

[; ALL list [| list] [divisor=value]]

\*\* Default if the subcommand is omitted.

† covstruct can take the following values: AD1, AR1, ARH1, ARMA11, CS, CSH, CSR, DIAG, FA1, FAH1, HF, ID, TP, TPH, UN, UNR, VC.

Why use LMM instead of OLS regression?

Data with multiple levels involve group effects on individuals which may be assessed invalidly by traditional statistical techniques. When grouping is present (ex., students in schools), observations within a group are often more similar than would be predicted on a pooled-data basis. That is, simple regression ignores grouping effects and violates the assumption of independence of observations. Mixed (multi-level) modeling handles this by using variables at upper levels (ex., school-level budgets at level 2) to adjust the regression of base level dependent variables on base level independent variables (ex., predicting student-level performance from student-level socioeconomic status scores).

Multi-level modeling in LMM is particularly helpful in the analysis of covariance when data are sparse. For instance, in a study of a Social Security agency office, there may be too few minority employees to enable valid statistical inferences on performance evaluations, using traditional regression models. However, if multi-level data are available on employees and multiple SSA offices, then multi-level models can use not only the individual data in the SSA office but also information in the pooled data for all offices. The resulting prediction equation applied to the given SSA office will use coefficients reflecting both their own and also pooled data. For agencies with a large number of minorities, the multi-level and ordinary regression models will be similar. For agencies with sparse data -- few minorities -- it is true their estimate will rely considerably on the pooled data, but the advantage is that the pooling involved in multi-level models affords a "borrowing of strength" that supports statistical inference in a situation where no inference would be possible using traditional methods.

Traditional regression models vs. LMM analysis. There were three traditional approaches to regression modeling of multilevel data:

Simple regression, also called naive regression, simply ignored higher level effects (ex., ignored class or school effects in a study of students). This is appropriate, of course, only when the researcher can be sure there are no higher level effects. More often, there are such effects and simple regression leads to too low estimates of standard error, a higher rate of Type I errors and too-narrow confidence limits compared to multilevel modeling of the same data.

Fixed effects regression. A popular traditional approach was to disaggregate data to the base level (ex., each student is assigned various school-level variables such as funding level per student, and all students in a given school have the same value on these contextual variables, and students are used as the unit of analysis). In fixed effects regression, sampling error is taken into account only for level 1 (the base) level, and sampling error at level 2 (or higher) is ignored. That is, information from fewer units at the upper level is wrongly treated as if it were independent data for the many units at the base level (the number of higher level observations is exaggerated), and this error in treating sample size led to over-optimistic estimates of significance. Also, there was the danger of the ecological fallacy: there is no necessary correspondence between individual-level and group-level variable relationships (ex., race and literacy correlate little at the individual level but correlate well at the state level, since Southern states have many African-Americans and many illiterates of all races). Finally, under fixed effects regression, the number of dummy variables increases as the number of clusters increases, making estimation inefficient. While adding crosslevel interaction terms is possible and does represent an improvement, it is inferior to multilevel modeling in LMM, which will model separate intercepts and slopes for individuals in each level 2 (or higher level) group, where the grouping variable is treated as a random effect.

Summary measures regression. Another traditional approach to multi-level problems was to aggregate data to a higher level (ex., student performance scores are averaged to the school level and schools are used at the unit of analysis). Aggregated data was often centered (the mean was subtracted so the average value was zero). Ordinary OLS regression or another traditional technique was then performed on the unit of analysis chosen. A problem with summary measures regression is that under aggregation, fewer units of analysis at the upper level replace many units at the base level, resulting in loss of statistical power. As with simple regression, summary measures regression regression leads to too low estimates of standard error, a higher rate of Type I errors and too-narrow confidence limits compared to multilevel modeling of the same data. Summary measures regression also suffers from the possibility of ecological fallacy. Finally, summary measures regression prevents the valid analysis of covariate interactions due to loss of individual-level information.

Based on a review of the literature and on simulation studies, Ita G. G. Kreft (1996) concluded, "for researchers specifically interested in variance components, and posterior means, RC modeling provides them with separate estimates for separate contexts, and the iteration procedure improves the estimates of the variance components." That is, although effect size as revealed through regression is apt to be similar to effect size in multi-level modeling (see discussion below), multi-level modeling is more helpful in revealing differences in variance among units of analysis in different groups which comprise the levels. An empirical comparison of OLS regression with multilevel modeling by Moerbeek, van Breukelen, & Berger (2003) found that "The treatment effect and especially its standard error, are generally incorrectly estimated by traditional methods, which should, therefore, not in general be used as an alternative to multilevel regression" (p. 341). Also, multi-level modeling may be a preferred method when data are sparse, including studies (ex., twin studies) where groups are sparse.

In defense of OLS. Based on a review of the literature and on simulation studies, Ita G. G. Kreft (1996) concluded, "if researchers in the social sciences are interested in the estimates of the regression parameters, the results of multi-level analysis will be close to the results obtained with more traditional regression techniques. In both cases the fixed effects estimates are unbiased. The main difference is in the standard errors of these parameters, which are estimated too small if intra-class correlation is present in traditional regression analyses. This fact makes the random coefficient model more conservative than the traditional regression." Kreft also raised questions about whether multi-level models are as generalizable as ordinary regression models. Because multi-level models rely on the complex, particular distribution of relationships across and within levels, Kreft concluded, "Outcomes are less general, since each best fitting model may be very specific for that dataset collected at that time and place." The smaller or less random the sample of higher-level units, the more true this would be. See also Kreft and de Leeuw (1991); Kreft, de Leeuw, and van der Leeden. (1994).

What is the multi-level equivalent to R-squared in OLS regression?

There is no equivalent to R-squared. Goodness of fit measures (ex., AICC, BIC) are used instead in multi-level modeling. Analogues to R-square have been proposed but are not widely used.

Why use LMM instead of GLM?

Many research problems might be modeled as general linear models (GLM option in SPSS) rather than as linear mixed models (Mixed option in SPSS). LMM provides a number of advantages, however:

GLM assumes independence but LMM does not, as discussed in the "Assumptions" section.

Handling missing data: GLM will apply listwise deletion to drop cases with missing values, whereas Mixed (LMM) will include incomplete cases in the analysis.

GLM repeated measures assumes all subjects are measured at the same points in time, whereas LMM allows subjects to be measured at different points in time.

Repeated measures GLM requires subjects to have equal numbers of repeated measurements, but LMM allows unequal repetititons. That is, LMM is asymptotically efficient for both balanced and unbalanced designs, but GLM is optimally efficient only for balanced designs.

GLM requires all interactions of within-subjects (repeated measures) and between-subjects factors be included in the model, whereas LMM allows the researcher to just include the interactions of interest.

GLM makes certain assumptions about the covariance matrix and thus data must meet the sphericity test, whereas LMM allows for a wide variety of assumptions about the covariance matrix.

LMM supports hierarchical data (data at one level nested within a higher level), but GLM does not.

LMM estimates are based on maximum likelihood (ML) or restricted maximum likelihood (REML) methods, whereas GLM is based on ANOVA methods.

Why does my GLM model give the same parameter estimates and corresponding significance levels as LMM for the same data, but LMM does not print out sums of squares?

For some models, such as fixed effects with uncorrelated residuals, GLM and LMM will give the same parameter estimates and significance levels (LMM in the "Tests of Fixed Effects" and "Estimates of Fixed Effects" tables; GLM in the "Tests of Between-Subjects Effects" and "Parameter Estimates" tables). However, LMM can fit a wider variety of models than GLM and in some LMM models, test effects cannot be expressed in terms of ratios of sums of squares and therefore there is no "Sum of Squares" column.

Besides variance components and diagonal (simple) covariance structure assumptions, what other assumptions might be made?

SPSS supports the following, listed alphabetically:

Ante-Dependence: First Order, AD(1). This covariance structure has heterogenous variances and heterogenous correlations between adjacent elements. The correlation between two nonadjacent elements is the product of the correlations between the elements that lie between the elements of interest.

AR(1). This is first-order autoregressive structure with homogenous variances. AR(1) is a common assumption for data where there is a common trend, such as where the correlation of any pair of repeated measurements is assumed to decrease according to how far apart they are in time. It is an assumption common to time series analysis. Thus residuals at time 1 are apt to be less similar to residuals at time 4 as they would be with the residuals at time 2. The larger the time lag, the lower the correlation of residuals. This model also assumes the time variable is metric and equally spaced. In the residual covariance matrix, the diagonal variances will be roughly equal, and the off-diagonal covariances will show a pattern, normally decreasing over time. (The particular data for the random effects variable city, illustrated below, do not follow this pattern). An AR(1) patter would exist, for example, if the correlation between any two elements (ex., time periods) is equal to r for time-adjacent elements , then was r2 for elements that are separated by one other element, r3 for elements that are separated by two other elements, etc., within the constraint that the bounds of +/- 1 are not exceeded.

AR(1): Heterogeneous. This is a first-order autoregressive structure with heterogenous variances. The correlation between any two elements is equal to r for adjacent elements, r2 for two elements separated by a third, and so on.

ARMA(1,1). This is a first-order autoregressive moving average structure. It has homogenous variances. The correlation between two elements is equal to f\*r for adjacent elements, f\*(r2) for elements separated by a third, and so on. r and f are the autoregressive and moving average parameters, respectively, and their values are constrained to lie between –1 and 1, inclusive.

Compound Symmetry. This is a common repeated measures covariance structure assumption. Within-subject correlation of error terms is assumed to be equal. That is, if elements are time periods, the correlation of residuals for measurements nearby in time should be the same as for measurements with a large time-distance. This covariance structure has homogenous variances and homogenous covariances between elements (ex., time periods for repeated measures, cities for random effects). This means that the residuals have the same covariance for any pair of time periods (repeated measures) or cities (random effects) in our example. Also, the variance of residuals is the same for any time period or city. Thus, in the residual covariance matrix, the diagonal variances will be roughly equal, and the off-diagonal variances will also be roughly equal to each other. This is the type assumed in univariate Anova models and is the classical approach to repeated measures. The assumption of compound symmetry is more likely to be met with classical experimental data than with longitudinal data where autoregressive or Toeplitz assumptions may be more appropriate.

Compound Symmetry: Correlation Metric. This covariance structure has homogenous variances and homogenous correlations between elements.

Compound Symmetry: Heterogeneous. This covariance structure has heterogenous variances and constant correlation between elements.

Factor Analytic: First Order. This covariance structure has heterogenous variances that are composed of a term that is heterogenous across elements and a term that is homogenous across elements. The covariance between any two elements is the square root of the product of their heterogenous variance terms.

Factor Analytic: First Order, Heterogeneous. his covariance structure has heterogenous variances that are composed of two terms that are heterogenous across elements. The covariance between any two elements is the square root of the product of the first of their heterogenous variance terms.

Huynh-Feldt. This is a "circular" matrix in which the covariance between any two elements is equal to the average of their variances minus a constant. Neither the variances nor the covariances are constant.

Scaled Identity. This structure has constant variance. There is assumed to be no correlation between any elements. This is a common assumption when modeling the interaction of a random factor (ex., City) with a fixed grouping factor (ex., Agency), where it is assumed that the City\*Agency interaction effect is normally distributed around a mean of zero, with unknown variance to be estimated.

Toeplitz. This structure is a generalization of the AR(1) type. Like AR(1), the Toeplitz model assumes metric, equal time intervals and assumes pairs of within-subject correlations are equal for the same measurement time-distance apart. Unlike the special case of AR(1), however, in a Toeplitz model the pattern sequence for off-diagonal covariances does not have to step by some common multiple but may step by some unique multiple associated with that time step. That is, there is a trend as in AR(1) models but there is no common function describing that trend for all time intervals. This covariance structure has homogenous variances and heterogenous correlations between elements. The correlation between adjacent elements is homogenous across pairs of adjacent elements. The correlation between elements separated by a third is again homogenous, and so on.

Toeplitz: Heterogeneous. This covariance structure has heterogenous variances and heterogenous correlations between elements. The correlation between adjacent elements is homogenous across pairs of adjacent elements. The correlation between elements separated by a third is again homogenous, and so on.

Unstructured. This is a completely general covariance matrix. This choice is common when the researcher has no idea what the covariance structure is. It is a common assumption for random effects regression models for this reason. Unstructured models are the assumption in GLM MANOVA. Slope variances and covariances are estimated from the data in an unconstrained manner which allows them to correlate (in a variance components model, in contrast, slopes of random effects, including intercepts, are assumed uncorrelated). This type is complex, requiring the computation of many parameters - that is, it is unparsimonious and so BIC or other goodness of fit measures taking parsimony into account will penalize unstructured models. One strategy is to choose the unstructured covariance type initially, but to request output of a matrix of covariances of residuals to discern possible patterns which would warrant re-running the model with a simpler covariance type assumption. Sometimes dropping outliers may make such a pattern more discenrable.

Unstructured: Correlations. This covariance structure can have heterogenous variances and heterogenous correlations.

What about multilevel modeling in structural equation models rather than linear mixed models?

Multilevel structural equation models (MSEM) are generalizations of path analysis and as such are part of the structural equation modeling (SEM) family. MSEM models can have multiple dependent variables or latent variables (constructs measured by indicator variables) as well as multiple levels of measurement. See McDonald (1994) and Muthén (1994). While MLM regression models (MRM) and MSEM will perform similarly, compared to MRM, the SEM approach can more easily incorporate complex path models and multiple group models. Multilevel regression modeling (MRM) is compared with multilevel structural equation modeling (MSEM) by Tomarken & Waller (2005: 38). They note the advantages of MSEM to be more and more interpretable measures of goodness of fit, better modeling of residuals, and better capacity to model latent variables. They note the advantages of MRM to be easier model specification, fewer estimation problems, and ability to handle certain types of analysis difficult to handle within MSEM. For instance, in MSEM you cannot model between-group variablility in factor loadings or path coefficients (Hox, 2002). They note, however, that the two approaches are more similar than different, and as techniques and software evolve, functional similarity is increasing. On the relation of multilevel modeling to SEM, see Curran (2003).

By modeling a grouping factor (ex., cities) as a random effect, may the researcher generalize conclusions to all cities, not just those in the sample?

While one sometimes encounters this statement in literature on random effects models, it is incorrect if stated in an unqualified manner. There is no way for, say, results from a convenience sample of five cities in Utah to be generalized to all U. S. cities. The statement would be true only if the cities were randomly selected from all U. S. cities, and the number of these level 2 groups (cities) was sufficient (>20 is a common rule of thumb).

What are split plot experiments in LMM?

Split plot experiments are a common type of linear mixed model. Arising out of agricultural applications, there would be one factor such as levels of irrigation applied randomly to whole plots of land within farms, then another factor such as seed types applied randomly to subplots (the "split plots") nested within the plots. The whole plots reflect diffent randomly-selected levels of irrigation and contain a set of subplots with randomly selected seed types. There may also be control variables, called "blocking variables," such as the farm on which the plots are located. There is, of course, also a dependent variable such as plant height, called the "response variable." In this example, irrigation level, seed type, and farm would all be "classification variables" for plant height as dependent. A variety of mixed models could be constructed using these factors as fixed and random effects. For instance, one could specify irrigation, seed type, and irrigation\*seed type as fixed effects and farm and irrigation\*farm as random effects. Output of LMM would include covariance parameter estimates for Farm, Farm\*Irrigation interaction, and Residual, along with AIC and other fit statistics for the model, interpreted as discussed above.

What are latent growth models?

Latent growth models (LGM) are a type of multilevel model suitable for clustered data with repeated observations of variables at multiple levels. While multiple group SEM is a conventional approach discussed in the SEM section of Statnotes, multilevel SEM (MSEM) is a newer methodological development which handles large numbers of groups (ex, 100 - 200). Strictly speaking, an LGM model adapts SEM to analyze change over time when both individual and group-level variables are in the model. However, LGM has been extended and generalized to other types of multi-level models and may refer simply to multilevel models based on SEM software as opposed to regression-based multilevel models.. As such, LGM is a more versatile alternative to repeated measures ANOVA (see Tomarken & Waller, 2005: 36 for a list of nine ways LGM is more versatile). For an introduction to LGM, see Duncan, Duncan, Strycker, et al. (1999) and Hox (2002). See also du Toit & du Toit (2005).

What software is available for multilevel modeling?

Several software packages for multi-level modeling have emerged in the last decade. Leading packages are listed below.

SPSS's "Linear Mixed Models" module, part of its SPSS Advanced Models extension, handles hierarchical linear models (HLM) as well as related models for random or mixed ANOVA and ANCOVA, repeated measures ANOVA and MANOVA, and variance component estimation (VARCOMP).

AMOS and LISREL. It is possible to implement multi-level models in structural equation modeling programs like AMOS and LISREL. LISREL's MLM module is called MULTILEV.

SAS's PROC MIXED procedure can implement several models: simple random-effect only, simple mixed with a single fixed and random effect, split-plot, multilocation, repeated measures, analysis of covariance, random coefficients, and spatial correlation See Littell et al. (1999).

HLM, authored Steve Raudenbush and Tony Bryk. Raudenbush headed the longitudinal and multi-level methods project at Michigan State University. HLM can read data from a variety of statistical packages, including SPSS, SAS, SYSTAT, and STATA, and it covers nonlinear as well as linear models. A free student version is available. This was perhaps the leading package during the development of multi-level modeling in the 1990s. HLM does not have a bulit-in data editor: data preparation must be done in SPSS (which HLM imports) or another program. HLM does not read ordinal variables, which must be converted to a series of dummy variables in the data preparation stage. Cross-level interaction terms are created automatically by HLM, and there is an option for automatic centering of variables (group mean centered or grand mean centered).

MLWin, a Windows program produced by the UK/Canada Multilevel Models Project, for models with any number of levels. It is the Windows version of the earlier MLn multi-level modeling software package.

MPlus supports multi-level modeling with latent variables.

Bibliography

Booth, J.G. & Hobert, J.P. (1998). Standard errors of prediction in generalized linear mixed models. Journal of the American Statistical Association 93, 262 -272.

Breslow, N. E., & Clayton, D. G. (1993). Approximate inference in generalized linear mixed models. Journal of the American Statistical Association 88, 9-25.

Bryk, A.S., and Raudenbush,S.W. (1992). Hierarchical linear models. Newbury Park, Sage Publications.

Bryk, A. S., S. W. Raudenbush, and R. Congdon (1996). Hierarchical Linear and nonlinear modeling with the HLM/2L and HLM/3L Programs. Chicago: Scientific Software International.

Burton, B. (1993). Some observations on the effect of centering on the results obtained from hierarchical linear modeling. Washington, DC: National Center for Education Statistics, U. S. Department of Education.

Cheung, M. W.-L. & Au, K. (2005). Applications of multilevel structural equation modeling to cross-cultural research. Structural Equation Modeling 12(4), 598-619.

Curran, P. J. (2003). Have multilevel models been structural equation models all along? Multivariate Behavioral Research 38(4): 529–569.

de Leeuw, J. and I. G. G. Kreft (1986). Random coefficient models for multi-level analysis. Journal of Educational Statistics 11: 57-86.

Diggle, P. J. (1988). An approach to the analysis of repeated measurements. Biometrics 44, 959-971.

Duncan, T. E., Duncan, S. C., Strycker, L. A., Li. F, & Alpert, A. (1999). An introduction to latent variable growth curve modeling: Concepts, issues, and applications. Mahwah, NJ: Erlbaum

du Toit, S. H. C. & du Toit, M. (2005). Multilevel structural equation modeling. In de Jeeuw, J. & Kreft, I. G. G., eds., .Handbook of quantitative multilevel analysis. Boston: Kluwer Acad.

Ehlers, Margaret (2004). Assessment of covariance selection strategies with repeated measures data. 34th Annual SCASA Meeting, American Statistical Association. Retrieved 4/223/07 from http://www.stat.sc.edu/scasa/Apr04t.html.

Goldstein, H. (1991). Nonlinear multi-level models with an application to discrete response data. Biometrika 78: 45-51.

Goldstein, H. (1995). Multilevel statistical models. London, Edward Arnold; New York, Halstead Press.

Hoffman, D. A. & Gavin, M. B. (1998). Centering decisions in hierarchical linear models: Implications for research organizations. Journal of Management 24(5), 623-641.

Hox, J. J. (1995). Applied multi-level analysis, 2nd Ed.. Amsterdam: TT-Publikaties. A basic, non-technical introductory text.

Hox J. J. (2002). Multilevel analysis: Techniques and applications. Mahwah, NJ: Erlbaum.

Hox, J. J. & Maas, C. J. M. (2001). The accuracy of mulitlevel structural equation modeling with pseudobalanced groups and small samples. Structural Equation Modeling 8, 157-174.

Jiang, Jiming (2007). Linear and generalized linear mixed models and their applications Springer Series in Statistics. New York: Springer.

Julian, M. W. (2001). The consequences of ignoring multilevel data structures in nonhierarchical covariance modeling. Structural Equation Modeling 8, 325-352.

Kreft, Ita G. G. (1996). Are multi-level techniques necessary? An overview, including simulation studies. Obtained online on Nov. 22, 1999, at http://www.calstatela.edu/faculty/ikreft/quarterly/quarterly.html.

Kreft, Ita G. G., Jan de Leeuw, J., and Aiken, Leona S. (1995). The effect of different rorms of centering in hierarchical linear models. Multivariate Behavioral Research 30(1), 1-21.

Kreft, Ita G. G., Jan de Leeuw, J., and R. van der Leeden (1994). Review of five multi-level analysis programs: BMDP-5V, GENMOD, HLM, ML3, VARCL. American Statistician 48(4): 324-335.

Kreft, Ita G. G. and Leeuw, Jan de (1991). Model-based ranking of schools. International Journal of Education 15: 45-59.

Kreft, Ita G. G. and Leeuw, Jan de (1998). Introducing multi-level modeling. Thousand Oaks, CA: Sage Publications. Introducing Statistical Methods Series. A user-oriented introductory text with a minimum of formal mathematics. Includes worked examples using real datasets using MLn, the predecessor of MLWin.

Littell, Ramon C., George A. Milliken, Walter W. Stroup, and Russell D. Wolfinger (1999). SAS System for mixed models. Cary, NC: Sas Institute.

Longford, N.T. and Muthen,B.O. (1992). Factor analysis for clustered populations. Psychometrika, 57, 581-97.

Longford, N.T. (1993). Random coefficient models. Oxford, Clarendon Press.

Luke, Douglas (2004). Multilevel modeling. Thousand Oaks, CA: Sage Publications. Vol. 142 in Quantitative Applications in the Social Sciences Series.

Maas, CoraJ. M. & Hox, Joop J. (2005) Sufficient sample sizes for multilevel modeling. Methodology 1(3), 86-92.

McCulloch, Charles E. (1997). Maximum likelihood algorithms for generalized linear mixed models. Journal of the American Statistical Association 92, 162-170.

McCulloch, Charles E. & and Searle, Shayle R.(2001). Generalized, linear, and mixed models. NY: Wiley-Interscience.

McCulloch, Charles E. (2003). Generalized linear mixed models. Regional Conference Series. Bethesda, MD: Institute of Mathematical Statistics

McDonald, R.P. (1994). The bilevel reticular action model for path analysis with latent variables. Sociological Methods and Research, 22, 399-413.

McDonald, R.P. and Goldstein, H. (1988). Balanced versus unbalanced designs for linear structural relations in two level data. British Journal of Mathematical and Statistical Psychology, 42, 215-32.

Moerbeek, Mirjam; van Breukelen, Gerard J. P.; & Berger, Martijn P. F. (2003). A comparison between traditional methods and multilevel regression for the analysis of multicenter intervention studies. Journal of Clinical Epidemiology 56, 341-350.

Morris, C. (1995). Hierarchical models for educational data - an overview. Journal of Educational and Behavioral Statistics 20(2).

Muthén, B.O. (1989). Latent variable modelling in heterogeneous populations. Psychometrika, 54, 557-85.

Muthén, B.O. (1994). Multilevel covariance structure analysis. Sociological Methods and Research 22: 376-399.

Peugh, James L. & Enders. Craig K. (2005). Using the SPSS Mixed procedure to fit cross-sectional and longitudinal multilevel models. Educational and Pcyhological Measurement 65(5), 717-741. Replicates Singer (1998) using SPSS Mixed instead of SAS. Enders has made supplementary instructional material available at http://www.asu.edu/clas/psych/people/documents/FittingmultilevelmodelsusingSPSSpull-downs.pdf.

Raudenbush, Stephen W. and Anthony S. Bryk (2002). Hierarchical linear models: Applications and data analysis methods, Second edition. Thousand Oaks, CA: Sage Publications. Advanced Quantative Techniques in the Social Sciences Series No. 1.

Singer, J. (1998). Using SAS Proc Mixed to fit multilevel models, hierarchical models, and individual growth curves. Journal of Educational and Behavioral Statistics 24(4), 323-355.

Singer, J. & Willett, J. (2003). Applied logitudinal data analysis. Oxford, UK: Oxford University Press.

Snijders, T. & Bosker, R. (2000). Multilevel analysis: An introduction to basic and advanced multilevel modeling. London: Sage.

SPSS, Inc. (2005). Linear mixed-effects modeling in SPSS: An introduction to the MIXED procedure. SPSS Technical Report. Chicago, IL: SPSS, Inc.

Sullivan, Lisa M.; Dukes, Kimberly A.; & Losina, Elena (1999). Tutorial in biostatistics: An introduction to hierarchical linear modelling. Statistics in Medicine/i> 18: 855-888. Illustrated using HLM and SAS.

Thompson, R. (1985). A note on restricted maximum likelihood estimation with an alternative outlier model, Journal of the Royal Statistical Society. Series B (Methodological). 47(1), 53-55.

Tomarken, A. J. & Waller, N. G. (2005). Structural equation modeling: Strengths, limitations, and misconceptions. Annual Review of Clinical Psychology 1, 31-65.

Verbeke, Geert & Molenberghs, Geert (1997). Linear mixed models in practice: A SAS-oriented approach. Lecture Notes in Statistics. New York: Springer.

Verbeke, G., & Molenberghs, G. (2000). Linear mixed models for longitudinal data. New York: Springer.

West, Brady; Welch, Kathleen B.; & Galecki, Andrzej T. (2006). Linear mixed models: A practical guide using statistical software. Boca Raton, FL: Chapman & Hall/CRC.

Wong, G. and W. Mason (1985). The hierarchical logistic regression model for mulitlevel analysis. Journal of the American Statistical Association 80: 513-524.

--------------------------------------------------------------------------------

@c 2006, 2008, 2009 G. David Garson

Last update: 12/16/2009.

--------------------------------------------------------------------------------

--------------------------------------------------------------------------------