Manuscript version, chapter 2 in J.J. Hox, M. Moerbeek & R. van de Schoot (2018).
Multilevel Analysis. Techniques and Applications. New York, NY: Routledge.

2

The Basic Two-Level Regression Model

Summary. The multilevel regression model has become known in the research literature under a variety of names, such as 'random coefficient model' (Kreft & de Leeuw, 1998), 'variance component model' (Searle, Casella & McCulloch, 1992; Longford, 1993), and 'hierarchical linear model' (Raudenbush & Bryk, 2002; Snijders & Bosker, 2012). Statistically oriented publications generally refer to the model as a 'mixed-effects' or 'mixed linear model (Littell, Milliken, Stroup & Wolfinger, 1996) and sociologists refer to it as 'contextual analysis' (Lazarsfeld & Menzel, 1961). The models described in these publications are not *exactly* the same, but they are highly similar, and we refer to them collectively as 'multilevel regression models'. The multilevel regression model assumes that there is a hierarchical data set, often consisting of subjects nested within groups, with one single outcome or response variable that is measured at the lowest level, and explanatory variables at all existing levels. The multilevel regression model can be extended by adding an extra level for multiple outcome variables (see chapter 10), while multilevel structural equation models are fully multivariate at all levels (see chapters 14 and 15). Conceptually, it is useful to view the multilevel regression model as a hierarchical system of regression equations. In this chapter, we explain the multilevel regression model for two-level data, providing both the equations and an example, and later extend this model with a three-level example.

2.1 **Example**

Assume that we have data from *J* classes, with a different number of pupils n_j in each class. On the pupil level, we have the outcome variable 'popularity' (*Y*), measured by a self-rating scale that ranges from 0 (very unpopular) to 10 (very popular). We have two explanatory variables on the pupil level: *pupil gender* (X_1: 0=boy, 1=girl) and *pupil extraversion* (X_2, measured on a self-rating scale ranging from 1–10), and one class level explanatory variable *teacher experience* (*Z*: in years, ranging from 2–25). There are data on 2000 pupils in 100 classes, so the average class size is 20 pupils. The data are described in Appendix E. The data files and other support materials are also available online.

To analyze these data, we can set up separate regression equations in each class to predict the outcome variable *Y* using the explanatory variables *X* as follows:

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{1ij} + \beta_{2j} X_{2ij} + e_{ij} . \tag{2.1}$$

Using variable labels instead of algebraic symbols, the equation reads:

$$popularity_{ij} = \beta_{0j} + \beta_{1j} gender_{ij} + \beta_{2j} extraversion_{ij} + e_{ij} . \tag{2.2}$$

In this regression equation, β_{0j} is the intercept, β_{1j} is the regression coefficient (regression slope) for the dichotomous explanatory variable gender (i.e., the difference between boys and girls), β_{2j} is the regression coefficient (slope) for the continuous explanatory variable extraversion, and e_{ij} is the usual residual error term. The subscript *j* is for the classes (*j*=1…*J*) and the subscript *i* is for individual pupils (*i*=1…n_j). The difference with the usual regression model is that we assume that each class has a different intercept coefficient β_{0j}, and different slope coefficients β_{1j} and β_{2j}. This is indicated in equations 2.1 and 2.2 by attaching a subscript *j* to the regression coefficients. The residual errors e_{ij} are assumed to have a mean of zero, and a variance to be estimated. Most multilevel software assumes that the variance of the residual errors is the same in all classes. Different authors (cf. Goldstein, 2011; Raudenbush & Bryk, 2002) use different systems of notation. This book uses σ_e^2 to denote the variance of the lowest level residual errors.

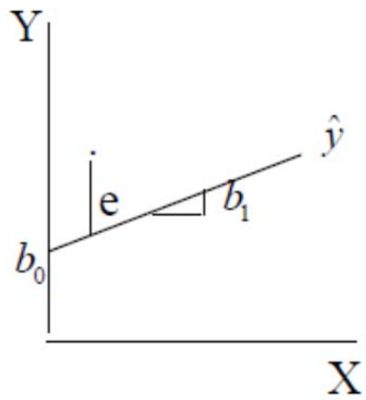
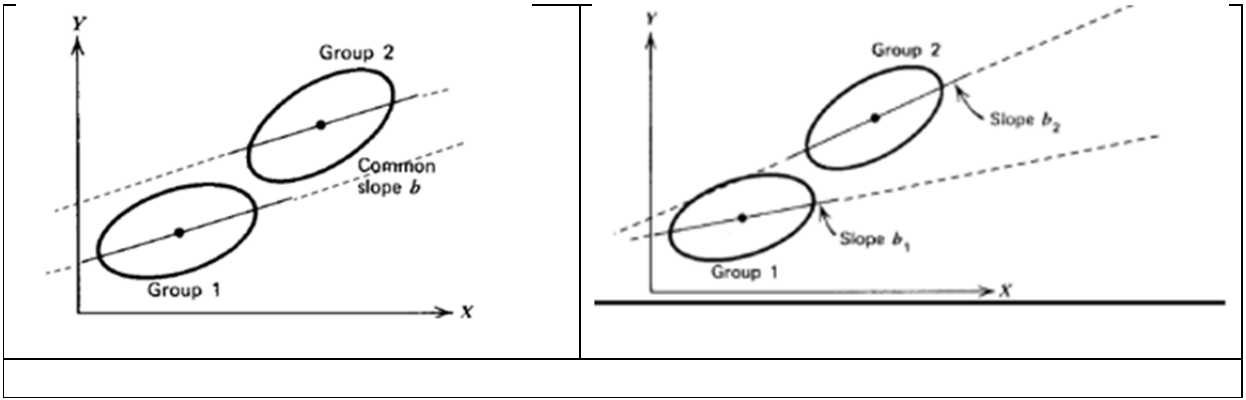


Figure 2.1 shows a single level regression line for a dependent variable Y regressed on a single an explanatory variable X. The regression line represents the predicted values $\hat{y}$ for Y, the regression coefficient b_0 is the intercept, the predicted value for Y if X=0. The regression slope b_1 indicates the predicted increase in Y if X increases by one unit.

Since in multilevel regression the intercept and slope coefficients vary across the classes, they are often referred to as *random* coefficients. Of course, we hope that this variation is not totally random, so we can explain at least some of the variation by introducing higher-level variables. Generally, we do not expect to explain all variation, so there will be some unexplained residual variation. In our example, the specific values for the intercept and the slope coefficients are a class characteristic. In general, a class with a high intercept is predicted to have more popular pupils than a class with a low value for the intercept. Since the model contains a dummy variable for gender, the value of the intercept reflects the predicted value for the boys (who are coded as zero). Varying intercepts shift the average value for the entire class, both boys and girls. Differences in the slope coefficient for gender or extraversion indicate that the relationship between the pupils' gender or extraversion and their predicted popularity is not the same in all classes. Some classes may have a high value for the slope coefficient of gender; in these classes, the difference between boys and girls is relatively large. Other classes may have a low value for the slope coefficient of gender; in these classes, gender has a small effect on the popularity, which means that the difference between boys and girls is small. Variance in the slope for pupil extraversion is interpreted in a similar way; in classes with a large coefficient for the extraversion slope, pupil extraversion has a large impact on their popularity, and vice versa.

Figure 2.2 presents an example with two groups. The panel on the left portrays two groups with no slope variation, and as a result the two slopes are parallel. The intercepts for both groups are different. The panel on the right portrays two groups with different slopes, or slope variation. Note that variation in slopes also has an effect on the difference between the intercepts!



Across all classes, the regression coefficients β_{0j} … β_{2j} are assumed to have a multivariate normal distribution. The next step in the hierarchical regression model is to explain the variation of the regression coefficients β_{0j} … β_{2j} by introducing explanatory variables at the class level, for the intercept

$$\beta_{0j} = \gamma_{00} + \gamma_{01}Z_j + u_{0j}, \tag{2.3}$$

and for the slopes

$$\begin{aligned} \beta_{1j} &= \gamma_{10} + \gamma_{11}Z_j + u_{1j} \\ \beta_{2j} &= \gamma_{20} + \gamma_{21}Z_j + u_{2j} \end{aligned}. \tag{2.4}$$

Equation 2.3 predicts the average popularity in a class (the intercept β_{0j}) by the teacher's experience (Z). Thus, if γ_{01} is positive, the average popularity is higher in classes with a more experienced teacher. Conversely, if γ_{01} is negative, the average popularity is lower in classes with a more experienced teacher. The interpretation of the equations under 2.4 is a bit more complicated. The first equation under 2.4 states that the *relationship*, as expressed by the slope coefficient β_{1j}, between the popularity (Y) and the gender (X) of the pupil, depends upon the amount of experience of the teacher (Z). If γ_{11} is positive, the gender effect on popularity is larger with experienced teachers. Conversely, if γ_{11} is negative, the gender effect on popularity is smaller with more experienced teachers. Similarly, the second equation under 2.4 states, if γ_{21} is positive, that the effect of extraversion is larger in classes with an experienced teacher. Thus, the amount of experience of the teacher acts as a *moderator variable* for the relationship between popularity and gender or extraversion; this relationship varies according to the value of the moderator variable.

The u-terms u_{0j}, u_{1j} and u_{2j} in equations 2.3 and 2.4 are (random) residual error terms at the class level. These residual errors u_j are assumed to have a mean of zero, and to be independent from the residual errors e_{ij} at the individual (pupil) level. The variance of the residual errors u_{0j} is specified as $\sigma^2_{u_0}$, and the variance of the residual errors u_{1j} and u_{2j} are specified as $\sigma^2_{u_1}$ and $\sigma^2_{u_2}$. The *covariances* between the residual error terms are denoted by $\sigma_{u_{01}}$, $\sigma_{u_{02}}$ and $\sigma_{u_{12}}$, which are generally *not* assumed to be zero.

Note that in equations 2.3 and 2.4 the regression coefficients γ are not assumed to vary across classes. They therefore have no subscript j to indicate to which class they belong. Because they apply to *all* classes, they are referred to as *fixed* coefficients. All between-class variation left in the β coefficients, after predicting these with the class variable Z_j, is assumed to be residual error variation. This is captured by the residual error terms u_j, which do have subscripts j to indicate to which class they belong.

Our model with two pupil level and one class level explanatory variables can be written as a single complex regression equation by substituting equations 2.3 and 2.4 into equation 2.1. Substitution and rearranging terms gives:

$$\begin{aligned} Y_{ij} = \gamma_{00} + \gamma_{10}X_{1ij} + \gamma_{20}X_{2ij} + \gamma_{01}Z_j + \gamma_{11}X_{1ij}Z_j + \gamma_{21}X_{2ij}Z_j \\ + u_{1j}X_{1ij} + u_{2j}X_{2ij} + u_{0j} + e_{ij} \end{aligned} . \tag{2.5}$$

Using variable labels instead of algebraic symbols, we have

popularity$_{ij}$ = γ_{00}+ γ_{10} *gender*$_{ij}$ + γ_{20} *extraversion*$_{ij}$ + γ_{01} *experience*$_j$
+γ_{11} *gender*$_{ij}$× *experience*$_j$ +γ_{21} *extraversion*$_{ij}$×*experience*$_j$
+ u_{1j} *gender*$_{ij}$ + u_{2j} *extraversion*$_{ij}$ + u_{0j}+ e_{ij} .

The segment [$\gamma_{00} + \gamma_{10} X_{1ij} + \gamma_{20} X_{2ij} + \gamma_{01}Z_j + \gamma_{11} X_{1ij}Z_j + \gamma_{11} X_{2ij}Z_j$] in equation 2.5 contains the fixed coefficients. It is often called the fixed (or deterministic) part of the model. The segment [$u_{1j}X_{1ij} + u_{2j}X_{2ij} + u_{0j} + e_{ij}$] in equation 2.5 contains the random error terms, and it is often called the random (or stochastic) part of the model. The terms $X_{1i}Z_j$ and $X_{2ij}Z_j$ are interaction terms that appear in the model as a consequence of modeling the varying regression slope β_j of a pupil level variable X_{ij} with the class level variable Z_j. Thus, the moderator effect of Z on the relationship between the dependent variable Y and the predictor X, is expressed in the single equation version of the model as a *cross-level interaction*. The interpretation of interaction terms in multiple regression analysis is complex, and this is treated in more detail in chapter 4. In brief, the point made in chapter 4 is that the substantive interpretation of the coefficients in models with interactions is much simpler if the variables making up the interaction are expressed as deviations from their respective means.

Note that the random error terms u_{1j} are connected to the X_{ij}. Since the explanatory variable X_{ij} and the corresponding error term u_j are multiplied, the resulting error term will be different for different values of the explanatory variable X_{ij}, a situation that in ordinary multiple regression analysis is called 'heteroscedasticity'. The usual multiple regression model assumes 'homo-scedasticity', which means that the variance of the residual errors is independent of the values of the explanatory variables. If this assumption is not true, ordinary multiple regression does not perform very well. This is another reason why analyzing multilevel data with ordinary multiple regression techniques does not perform well.

As explained in the introduction in chapter 1, multilevel models are needed because with grouped data observations from the same group are generally more similar to each other than the observations from different groups, and this violates the assumption of independence of all observations. The amount of dependence can be expressed as a correlation coefficient: the intraclass correlation. The methodological literature contains a number of different formulas to estimate the intraclass correlation ρ. For example, if we use one-way analysis of variance with the grouping variable as independent variable to test the group effect on our outcome variable, the intraclass correlation is given by ρ = [MS(B)-MS(*error*)]/[MS(B)+(*n*-1)×MS(*error*)], where MS(B) is the Between Groups Mean Square and *n* is the common group size. Shrout and Fleiss (1979) give an overview of formulas for the intraclass correlation for a variety of research designs.

The multilevel regression model can also be used to produce an estimate of the intraclass correlation. The model used for this purpose is a model that contains no explanatory variables at all, the so-called *intercept-only* or *empty* model (also referred to as baseline model). The intercept-only model is derived from equations 2.1 and 2.3 as follows. If there are no explanatory variables *X* at the lowest level, equation 2.1 reduces to

$$Y_{ij} = \beta_{0j} + e_{ij} \,. \qquad (2.6)$$

Likewise, if there are no explanatory variables *Z* at the highest level, equation 2.3 reduces to

$$\beta_{0j} = \gamma_{00} + u_{0j} \,. \qquad (2.7)$$

We find the single equation model by substituting 2.7 into 2.6:

$$Y_{ij} = \gamma_{00} + u_{0j} + e_{ij} \,. \qquad (2.8)$$

The intercept-only model of equation 2.8 does not explain any variance in *Y*. It only decomposes the variance into two independent components: σ_e^2, which is the variance of the lowest-level errors e_{ij}, and σ_{u0}^2, which is the variance of the highest-level errors u_{0j}. These two variances sum up to the total variance, hence they are often referred to as variance components. Using this model, we can define the intraclass correlation ρ by the equation

$$\rho = \frac{\sigma_{u_0}^2}{\sigma_{u_0}^2 + \sigma_e^2} \,. \qquad (2.9)$$

The intraclass correlation ρ indicates the proportion of the total variance explained by the grouping structure in the population. Equation 2.9 simply states that the intraclass correlation is the proportion of group level variance compared to the total variance.[1] The intraclass correlation ρ can also be interpreted as the expected correlation between two randomly drawn units that are in the same group.

In the intercept only model we defined variance of the lowest-level errors and variance of the highest-level errors. Both terms can be interpreted as unexplained variance on both levels since there are no predictor in the model specified yet. After adding predictors, just like in ordinary regression analyses, the R^2, which is interpreted as the proportion of variance modeled by the explanatory variables, can be calculated. In the case of multilevel analyses, however, there is variance to be explained at every level (and also for random slope factors). The interpretation of these separate R^2 values are dependent on the ICC-values. For example, if the R^2 at the highest

[1] The intraclass correlation is an estimate of the proportion of group-level variance *in the population*. The proportion of group-level variance in the *sample* is given by the correlation ratio η^2 (eta-squared, cf. Tabachnick & Fidell, 2013, p. 54): η^2=SS(B)/SS(Total).

level appears to be .20 and the ICC is .40, then out of 40% of the total variance 20% is explained. This is further explained in Chapter 4.

2.2 **An extended example**

The intercept-only model is useful as a null-model that serves as a benchmark with which other models are compared. For our pupil popularity example data, the intercept-only model is written as

$popularity_{ij} = \gamma_{00} + u_{0j} + e_{ij}$.

The model that includes pupil gender, pupil extraversion and teacher experience, but not the cross-level interactions, is written as

$$popularity_{ij} = \gamma_{00} + \gamma_{10}\, gender_{ij} + \gamma_{20}\, extraversion_{ij} + \gamma_{01}\, experience_j + u_{1j}\, gender_{ij} + u_{2j}\, extraversion_{ij} + u_{0j} + e_{ij}.$$

Table 2.1. *Intercept-only model and model with explanatory variables*

Model:	Single level model	M_0: Intercept only	M_1: with predictors
Fixed part		Coefficient (s.e.)	Coefficient (s.e.)
Intercept	5.08 (.03)	5.08 (.09)	0.74 (.20)
Pupil gender			1.25 (.04)
Pupil extraversion			0.45 (.03)
Teacher experience			0.09 (.01)
Random part[a]			
σ_e^2	1.91 (.06)	1.22 (.04)	0.55 (.02)
σ_{u0}^2		0.69 (.11)	1.28 (.47)
σ_{u1}^2			0.00 (-)
σ_{u2}^2			0.03 (.008)
Deviance	6970.4	6327.5	4812.8

[a] For simplicity the covariances are not included

Table 2.1 presents the parameter estimates and standard errors for both models.[2] For comparison, the first column presents the parameter estimates of a single level model. The intercept is estimated correctly, but the variance term combines the level-one and level-two variances, and is for that reason not meaningful. M0, the intercept only two-level model, splits this variance term in a variance at the first and a variance at the second level. The intercept-only two-level model estimates the intercept as 5.08, which is simply the average popularity across all classes and pupils. The variance of the pupil level residual errors, symbolized by σ_e^2, is estimated as 1.22.

[1] For reasons to be explained later, different options for the details of the Maximum Likelihood estimation procedure may result in slightly different estimates. So, if you re-analyze the example data from this book, the results may differ slightly from the results given here. However, these differences should never be so large that you would draw entirely different conclusions.

The variance of the class level residual errors, symbolized by σ_{u0}^2, is estimated as 0.69. All parameter estimates are much larger than the corresponding standard errors, and calculation of the Z-test shows that they are all significant at p <0.005.[3] The intraclass correlation, calculated by equation 2.9 as $\rho = \sigma_{u0}^2 / \left(\sigma_{u0}^2 + \sigma_e^2\right)$, is 0.69/1.91, which equals 0.36. Thus, 36% of the variance of the popularity scores is at the group level, which is very high for social science data. Since the intercept-only model contains no explanatory variables, the residual variances represent unexplained error variance. The deviance reported in Table 2.1 is a measure of model misfit; when we add explanatory variables to the model, the deviance will go down.

The second model in Table 2.1 includes pupil gender and extraversion and teacher experience as explanatory variables. The regression coefficients for all three variables are significant. The regression coefficient for pupil gender is 1.25. Since pupil gender is coded 0=boy, 1=girl, this means that on average the girls score 1.25 points higher than boys on the popularity measure, when all other variables are kept constant. The regression coefficient for pupil extraversion is 0.45, which means that with each scale point higher on the extraversion measure, the popularity is expected to increase with 0.45 scale points. The regression coefficient for teacher experience is 0.09, which means that for each year of experience of the teacher, the average popularity score of the class goes up with 0.09 points. This does not seem very much, but the teacher experience in our example data ranges from 2 to 25 years, so the predicted difference between the least experienced and the most experienced teacher is (25-2=) 23×0.09=2.07 points on the popularity measure. The value of the intercept is generally not interpreted, it is the expected value of the dependent variable if all explanatory variables have the value zero. We can use the standard errors of the regression coefficients reported in Table 2.1 to construct a 95% confidence interval. For the regression coefficient of pupil gender, the 95% confidence interval runs from 1.17 to 1.33, the confidence interval for pupil extraversion runs from 0.39 to 0.51, and the 95% confidence interval for the regression coefficient of teacher experience runs from 0.07 to 0.11.[4] Note that the interpretation of the regression coefficients in the fixed part is no different than in any other regression model (cf. Aiken & West, 1991).

The model with the explanatory variables includes variance components for the regression coefficients of pupil gender and pupil extraversion, symbolized by σ_{u1}^2 and σ_{u2}^2 in Table 2.1. The variance of the regression coefficients for pupil extraversion across classes is estimated as 0.03, with a standard error of 0.008. The variance of the regression coefficients for pupil gender is estimated as zero and not significant, so the hypothesis that the regression slopes for pupil gender vary across classes is not supported by the data. We should remove the residual variance term for the gender slopes from the model, and estimate the new model again. Table 2.2 presents the estimates for the model with a fixed slope for the effect of pupil gender. Table 2.2 also includes the covariance between the class-level errors for the intercept and the extraversion slope. These covariances are rarely interpreted (for an exception see chapters 5 and 16 where growth models are discussed), and for that reason they are often not included in the reported tables. However, as Table 2.2 demonstrates, they can be quite large and significant, so as a rule they are always included in the model.

[2] Testing variances is preferably done with a test based on the deviance, which is explained in chapter 3.
[4] Chapter 3 treats the interpretation of confidence intervals in more detail.

Table 2.2. *Model with explanatory variables, extraversion slope random*

Model:	M_1: with predictors
Fixed part	Coefficient (s.e.)
Intercept	0.74 (.20)
Pupil gender	1.25 (.04)
Pupil extraversion	0.45 (.02)
Teacher experience	0.09 (.01)
Random part	
σ_e^2	0.55 (.02)
σ_{u0}^2	1.28 (.28)
σ_{u2}^2	0.03 (.008)
$\sigma_{u_{02}}$	-.18 (.05)
Deviance	4812.8

The significant variance of the regression slopes for pupil extraversion implies that we should not interpret the estimated value of 0.45 without considering this variation. In an ordinary regression model, without multilevel structure, the value of 0.45 means that for each point different on the extraversion scale, the pupil popularity goes up with 0.45, for all pupils in all classes. In our multilevel model, the regression coefficient for extraversion varies across the classes, and the value of 0.45 is just the expected value (the mean) across all classes. The varying regression slopes for pupil extraversion are assumed to follow a normal distribution. The variance of this distribution is in our example estimated as 0.034. Interpretation of this variation is easier when we consider the standard deviation, which is the square root of the variance and equal to 0.18 in our example data. A useful characteristic of the standard deviation is that with normally distributed observations about 67% of the observations lie between one standard deviation below and above the mean, and about 95% of the observations lie between two standard deviations below and above the mean. If we apply this to the regression coefficients for pupil gender, we conclude that about 67% of the regression coefficients are expected to lie between (0.45-0.18=) 0.27 and (0.45+0.18=) 0.63, and about 95% are expected to lie between (0.45-0.37=) 0.08 and (0.45+0.37=) 0.82. The more precise value of $Z_{.975}$=1.96 leads to the 95% predictive interval calculated as 0.09–0.81. We can also use the standard normal distribution to estimate the percentage of regression coefficients that are negative. As it turns out, if the mean regression coefficient for pupil extraversion is 0.45, given the estimated slope variance, less than 1% of the classes are expected to have a regression coefficient that is actually negative. Note that the 95% interval computed here is totally different from the 95% confidence interval for the regression coefficient of pupil extraversion, which runs from 0.41 to 0.50. The 95% confidence interval applies to γ_{20}, the mean value of the regression coefficients across all the classes. The 95% interval calculated here is the 95% *predictive interval*, which expresses that 95% of the regression coefficients of the variable 'pupil extraversion' in the classes are predicted to lie between 0.09 and 0.81.

Given the significant variance of the regression coefficient of pupil extraversion across the classes it is attractive to attempt to predict its variation using class level variables. We have one class level variable: teacher experience. The individual level regression equation for this example, using variable labels instead of symbols, is given by:

$$popularity_{ij} = \beta_{0j} + \beta_1 gender_{ij} + \beta_{2j} extraversion_{ij} + e_{ij}. \tag{2.10}$$

The regression coefficient β_1 for pupil gender does not have a subscript *j*, because it is not assumed to vary across classes. The regression equations predicting β_{0j}, the intercept in class *j*, and β_{2j}, the regression slope of pupil extraversion in class *j*, are given by equation 2.3 and 2.4, which are rewritten below using variable labels

$$\begin{aligned} \beta_{0j} &= \gamma_{00} + \gamma_{01} experience_j + u_{0j} \\ \beta_{2j} &= \gamma_{20} + \gamma_{21} experience_j + u_{2j} \end{aligned}. \tag{2.11}$$

By substituting 2.11 into 2.10 we get

$$\begin{aligned} &popularity_{ij} = \gamma_{00} + \gamma_{10} gender_{ij} + \gamma_{20} extraversion_{ij} + \\ &\gamma_{01} experience_j + \gamma_{21} extraversion_{ij} \times experience_j + u_{2j} extraversion_{ij} + u_{0j} + e_{ij} \end{aligned} \tag{2.12}$$

The algebraic manipulations of the equations above make clear that to explain the variance of the regression slopes β_{2j}, we need to introduce an interaction term in the model. This interaction, between the variables pupil extraversion and teacher experience, is a cross-level interaction, because it involves explanatory variables from different levels. Table 2.3 presents the estimates from a model with this cross-level interaction. For comparison, the estimates for the model without this interaction are also included in Table 2.3.

The estimates for the fixed coefficients in Table 2.3 are similar for the effect of pupil gender, but the regression slopes for pupil extraversion and teacher experience are considerably larger in the cross-level model. The interpretation remains the same: extraverted pupils are more popular. The regression coefficient for the cross-level interaction is –0.03, which is small but significant. This interaction is formed by multiplying the scores for the variables ‘pupil extraversion’ and ‘teacher experience,’ and the negative value means that with experienced teachers, the advantage of extraverted is smaller than expected from the direct effects only. Thus, the difference between extraverted and introverted pupils is smaller with more experienced teachers.

Table 2.3. *Model without and with cross-level interaction*

Model:	M_{1A}: main effects	M_2: with interaction
Fixed part	Coefficient (s.e.)	Coefficient (s.e.)
Intercept	0.74 (.20)	-1.21 (.27)
Pupil gender	1.25 (.04)	1.24 (.04)
Pupil extraversion	0.45 (.02)	0.80 (.04)
Teacher experience	0.09 (.01)	0.23 (.02)
Extra*T.experience		-.03 (.003)
Random part		
σ_e^2	0.55 (.02)	0.55 (.02)
σ_{u0}^2	1.28 (.28)	0.45 (.16)
σ_{u2}^2	0.03 (.008)	0.005 (.004)
$\sigma_{u_{02}}$	-.18 (.05)	-.03 (.02)
Deviance	4812.8	4747.6

Comparison of the other results between the two models shows that the variance component for pupil extraversion goes down from 0.03 in the main effects model to 0.005 in the cross-level model. Apparently, the cross-level model explains some of the variation of the slopes for pupil extraversion. The deviance also goes down, which indicates that the model fits better than the previous model. The other differences in the random part are more difficult to interpret. Much of the difficulty in reconciling the estimates in the two models in Table 2.3 stems from adding an interaction effect. This issue is discussed in more detail in Chapter Four.

The coefficients in the tables are all unstandardized regression coefficients. To interpret them properly, we must take the scale of the explanatory variables into account. In multiple regression analysis, and structural equation models (SEM), for that matter, the regression coefficients are often standardized because that facilitates the interpretation when one wants to compare the effects of different variables within one sample. If the goal of the analysis is to compare parameter estimates from different samples to each other, one should always use unstandardized coefficients. To standardize the regression coefficients, as presented in Table 2.1 or Table 2.3, one could standardize all variables before putting them into the multilevel analysis. However, this would in general also change the estimates of the variance components, and their standard errors as well. Therefore, it is better to derive the standardized regression coefficients from the unstandardized coefficients:

$$\begin{matrix} standardized \\ coefficient \end{matrix} = \frac{unstandardized\ coefficient * stand.dev.explanatory\ var.}{stand.dev.outcome\ var.} \qquad (2.13)$$

In our example data, the standard deviations are: 1.38 for popularity, 0.51 for gender, 1.26 for extraversion, and 6.55 for teacher experience. Table 2.4 presents the unstandardized and standardized coefficients for the second model in Table 2.2. It also presents the estimates that we obtain if we first standardize all variables, and then carry out the analysis.

Table 2.4. *Comparing unstandardized and standardized estimates*

Model:	Standardization using 2.13		Standardized variables
Fixed part	Coefficient (s.e.)	standardized	Coefficient (s.e.)
Intercept	0.74 (.20)	-	-.03 (.04)
Pupil gender	1.25 (.04)	0.46	0.45 (.01)
Pupil extraversion	0.45 (.02)	0.41	0.41 (.02)
Teacher experience	0.09 (.01)	0.43	0.43 (.04)
Random part			
σ_e^2	0.55 (.02)		0.28 (.01)
σ_{u0}^2	1.28 (.28)		0.15 (.02)
σ_{u2}^2	0.03 (.008)		0.03 (.01)
$\sigma_{u_{02}}$	-.18 (.01)		-.01 (.01)
Deviance	4812.8		3517.2

Table 2.4 shows that the standardized regression coefficients are almost the same as the regression coefficients estimated for standardized variables. The small differences in Table 2.4 are simply due to rounding errors. However, if we use standardized variables in our analysis, we find very different variance components and a very different value for the deviance. This is not only the effect of scaling the variables differently; the covariance between the slope for pupil extraversion and the intercept is significant for the unstandardized variables, but not significant for the standardized variables. This kind of difference in results is general. The fixed part of the multilevel regression model is invariant for linear transformations, just as the regression coefficients in the ordinary single-level regression model. This means that if we change the scale of our explanatory variables, the regression coefficients and the corresponding standard errors change by the same multiplication factor, and all associated *p*-values remain exactly the same. However, the random part of the multilevel regression model is not invariant for linear transformations. The estimates of the variance components in the random part can and do change, sometimes dramatically. This is discussed in more detail in section 4.2 in Chapter Four. The conclusion to be drawn here is that, if we have a complicated random part, including random components for regression slopes, we should think carefully about the scale of our explanatory variables. If our only goal is to present standardized coefficients in addition to the unstandardized coefficients, applying equation 2.13 is safer than transforming our variables. On the other hand, we may estimate the unstandardized results, including the random part and the deviance, and then re-analyze the data using standardized variables, merely using this analysis as a computational trick to obtain the standardized regression coefficients without having to do hand calculations.

2.3 **Three- and more-level regression models**

2.3.1 Multiple-level models

In principle, the extension of the two-level regression model to three and more levels is straightforward. There is an outcome variable at the first, the lowest level. In addition, there may be explanatory variables at all available levels. The problem is that three- and more-level models can become complicated very fast. In addition to the usual fixed regression coefficients, we must entertain the possibility that regression coefficients for first-level explanatory variables may vary across units of both the second and the third level. Regression coefficients for second-level explanatory variables may vary across units of the third level. To explain such variation, we must include cross-level interactions in the model. Regression slopes for the cross-level interaction between first-level and second-level variables may themselves vary across third-level units. To explain such variation, we need a three-way interaction involving variables at all three levels.

The equations for such models are complicated, especially when we do not use the more compact summation notation but write out the complete single equation-version of the model in an algebraic format (for a note on notation see section 2.4).

The resulting models are not only difficult to follow from a conceptual point of view; they may also be difficult to estimate in practice. The number of estimated parameters is considerable, and at the same time the highest level sample size tends to become relatively smaller. As DiPrete and Forristal (1994, p. 349) put it, the imagination of the researchers "…can easily outrun the capacity of the data, the computer, and current optimization techniques to provide robust estimates."

Nevertheless, three- and more-level models have their place in multilevel analysis. Intuitively, three-level structures such as pupils in classes in schools, or respondents nested within households, nested within regions, appear to be both conceptually and empirically manageable. If the lowest level is repeated measures over time, having repeated measures on pupils nested within schools again does not appear to be overly complicated. In such cases, the solution for the conceptual and statistical problems mentioned is to keep models reasonably small. Especially specification of the higher-level variances and covariances should be driven by theoretical considerations. A higher-level variance for a specific regression coefficient implies that this regression coefficient is assumed to vary across units at that level. A higher-level covariance between two specific regression coefficients implies that these regression coefficients are assumed to covary across units at that level. Especially when models become large and complicated, it is advisable to avoid higher-order interactions, and to include in the random part only those elements for which there is strong theoretical or empirical justification. This implies that an exhaustive search for second-order and higher-order interactions is not a good idea. In general, we should seek for higher-order interactions only if there is strong theoretical justification for their importance, or if an unusually large variance component for a regression slope calls for explanation. For the random part of the model, there are usually more convincing theoretical reasons for the higher-level variance components than for the covariance components. Especially if the covariances are small and insignificant, analysts sometimes do not include all possible covariances in the model. This is defensible, with some exceptions. First, it is recommended that the covariances between the intercept and the random slopes are always included. Second, it is recommended to include covariances corresponding to slopes of dummy-variables belonging to the same categorical variable, and for variables that are involved in an interaction or belong to the same polynomial expression.

2.3.2 Intraclass correlations in three-level models

In a two-level model, the intraclass correlation is calculated in the intercept-only model using equation 2.9, which is repeated below:

$$\rho = \frac{\sigma_{u_0}^2}{\sigma_{u_0}^2 + \sigma_e^2} . \qquad \text{(2.9, repeated)}$$

The intraclass correlation is an indication of the proportion of variance at the second level, and it can also be interpreted as the expected (population) correlation between two randomly chosen individuals within the same group.

If we have a three-level model, for instance pupils nested within classes, nested within schools, there are two ways to calculate the intraclass correlation. First, we estimate an intercept-only model for the three-level data, for which the single-equation model can be written as follows:

$$Y_{ijk} = \gamma_{000} + v_{0k} + u_{0jk} + e_{ijk} . \qquad (2.15)$$

The variances at the first, second, and third level are respectively σ_e^2, $\sigma_{u_0}^2$, and $\sigma_{v_0}^2$. The first method (cf. Davis & Scott, 1995) defines the intraclass correlations at the class and school level as

$$\rho_{class} = \frac{\sigma^2_{u_0}}{\sigma^2_{v_0} + \sigma^2_{u_0} + \sigma^2_e}, \qquad (2.16)$$

and

$$\rho_{school} = \frac{\sigma^2_{v_0}}{\sigma^2_{v_0} + \sigma^2_{u_0} + \sigma^2_e}. \qquad (2.17)$$

The second method (cf. Siddiqui, Hedeker, Flay & Hu, 1996) defines the intraclass correlations at the class and school level as

$$\rho_{class} = \frac{\sigma^2_{v_0} + \sigma^2_{u_0}}{\sigma^2_{v_0} + \sigma^2_{u_0} + \sigma^2_e}, \qquad (2.18)$$

and

$$\rho_{school} = \frac{\sigma^2_{v_0}}{\sigma^2_{v_0} + \sigma^2_{u_0} + \sigma^2_e}. \qquad (2.19)$$

Actually, both methods are correct (Algina, 2000). The first method identifies the proportion of variance at the class and school level. This should be used if we are interested in a decomposition of the variance across the available levels, or if we are interested in how much variance is located at each level (a topic discussed in section 4.5). The second method represents an estimate of the expected (population) correlation between two randomly chosen elements in the same group. So ρ_{class} as calculated in equation 2.18 is the expected correlation between two pupils within the same class, and it correctly takes into account that two pupils who are in the same class must by definition also be in the same school. For this reason, the variance components for classes and schools must both be in the numerator of equation 2.18. If the two sets of estimates are different, which may happen if the amount of variance at the school level is large, there is no contradiction involved. Both sets of equations express two different aspects of the data, which happen to coincide when there are only two levels. The first method, which identifies the proportion of variance at each level, is the one most often used.

2.3.3. An example of a three-level model

The data in this example are from a hypothetical study on stress in hospitals. The data are from nurses working in wards nested within hospitals. In each of 25 hospitals, four wards are selected and randomly assigned to an experimental and control condition. In the experimental condition, a training program is offered to all nurses to cope with job-related stress. After the program is completed, a sample of about 10 nurses from each ward is given a test that measures job-related stress. Additional variables are: nurse age (years), nurse experience (years), nurse gender (0=male, 1=female), type of ward (0=general care, 1=special care), and hospital size (0=small, 1=medium, 2=large).

This is an example of an experiment where the experimental intervention is carried out on a higher level, in this example the ward level. In biomedical research this design is known as a multisite cluster randomized trial. They are quite common, also in educational and organizational research, where entire classes or schools are assigned to experimental and control conditions. Since the design variable Experimental versus Control group (ExpCon) is manipulated at the second (ward) level, we can study whether the experimental effect is different in different hospitals, by defining the regression coefficient for the ExpCon variable as random at the hospital level.

In this example, the variable ExpCon is of main interest, and the other variables are covariates. Their function is to control for differences between the groups, which can occur even is randomization is used, especially with small samples, and to explain variance in the outcome variable stress. To the extent that these variables successfully explain variance, the power of the test for the effect of ExpCon will be increased. Therefore, although logically we can test if explanatory variables at the first level have random coefficients at the second of third level, and if explanatory variables at the second level have random coefficients at the third level, these possibilities are not pursued. We do test a model with a random coefficient for ExpCon at the third level, where there turns out to be significant slope variation. This varying slope can be predicted by adding a cross-level interaction between the variables *expcon* and *hospsize*. In view of this interaction, the variables *expcon* and *hospsize* have been centered on their overall mean.[5] Table 2.5 presents the results for a series of models.

Table 2.5. *Models for stress in hospitals and wards*

Model:	M_0: Intercept only	M_1: with predictors	M_2: with random slope ExpCon	M_3: with cross-level interaction
Fixed part	Coef. (s.e.)	Coef. (s.e.)	Coef. (s.e.)	Coef. (s.e.)
Intercept	5.00 (0.11)	5.50 (.12)	5.46 (.12)	5.50 (.11)
ExpCon[a]		-.70 (.12)	-.70 (.18)	-0.50 (.11)
Age		0.02 (.002)	0.02 (.002)	0.02 (.002)
Gender		-.45 (.03)	-.45 (.03)	-.45 (.03)
Experience		-0.06 (.004)	-.06 (.004)	-.06 (.004)
Ward type		0.05 (.12)	0.05 (.07)	0.05 (.07)
Hosp. Size[a]		0.46 (.12)	0.29 (.12)	.46 (.12)
Exp*HSize				1.00 (.16)
Random part				
$\sigma^2_{e\ ijk}$	0.30 (.01)	0.22 (.01)	0.22 (.01)	0.22 (.01)
$\sigma^2_{u0\,jk}$	0.49 (.09)	0.33 (.06)	0.11 (.03)	0.11 (.03)
σ^2_{v0k}	0.16 (.09)	0.10 (0.05)	0.166 (.06)	0.15 (.05)
σ^2_{u1k}			0.66 (.22)	0.18 (.09)
Deviance	1942.4	1604.4	1574.2	1550.8

[a] Centered on grand mean

The equation for the first model, the intercept-only model is

$$stress_{ijk} = \gamma_{000} + v_{0k} + u_{0jk} + e_{ijk}\,. \tag{2.20}$$

[5] Chapter 4 discusses the interpretation of interactions and centering.

This produces the variance estimates in the M0 column of Table 2.5. The proportion of variance (ICC) is 0.52 at the ward level, and 0.17 at the hospital level, calculated following equations 2.18 and 2.19. The nurse level and the ward level variances are evidently significant. The test statistic for the hospital level variance is Z=0.162/0.0852=1.901, which produces a one-sided *p*-value of 0.029. The hospital level variance is significant at the 5% level. The sequence of models in Table 2.5 shows that all predictor variables have a significant effect, except the ward type, and that the experimental intervention significantly lowers stress. The experimental effect varies across hospitals, and a large part of this variation can be explained by hospital size; in large hospitals the experimental effect is smaller.

2.4 **Notation and software**

2.4.1 Notation

In general, there will be more than one explanatory variable at the lowest level and more than one explanatory variable at the highest level. Assume that we have *P* explanatory variables *X* at the lowest level, indicated by the subscript *p* (*p*=1…*P*). Likewise, we have *Q* explanatory variables *Z* at the highest level, indicated by the subscript *q* (*q*=1…*Q*). Then, equation 2.5 becomes the more general equation:

$$Y_{ij} = \gamma_{00} + \gamma_{p0} X_{pij} + \gamma_{0q} Z_{qj} + \gamma_{pq} Z_{qj} X_{pij} + u_{pj} X_{pij} + u_{0j} + e_{ij} \,. \qquad (2.21)$$

Using summation notation, we can express the same equation as

$$Y_{ij} = \gamma_{00} + \sum_p \gamma_{p0} X_{pij} + \sum_q \gamma_{0q} Z_{qj} + \sum_p \sum_q \gamma_{pq} X_{pij} Z_{qj} + \sum_p u_{pj} X_{pij} + u_{0j} + e_{ij} \,. \qquad (2.22)$$

The errors at the lowest level e_{ij} are assumed to have a normal distribution with a mean of zero and a common variance σ_e^2 in all groups. The u-terms u_{0j} and u_{pj} are the residual error terms at the highest level. They are assumed to be independent from the errors e_{ij} at the individual level, and to have a multivariate normal distribution with means of zero. The variance of the residual errors u_{0j} is the variance of the intercepts between the groups, symbolized by $\sigma_{u_0}^2$. The variances of the residual errors u_{pj} are the variances of the slopes between the groups, symbolized by $\sigma_{u_p}^2$. The *covariances* between the residual error terms $\sigma_{u_{pp'}}$ are generally not assumed to be zero; they are collected in the higher level variance/covariance matrix Ω.[6]

Note that in equation 2.15, γ_{00}, the regression coefficient for the intercept, is not associated with an explanatory variable. We can expand the equation by providing an explanatory variable that is a constant equal to one for all observed units. This yields the equation

$$Y_{ij} = \gamma_{p0} X_{pij} + \gamma_{pq} Z_{qj} X_{pij} + u_{pj} X_{pij} + e_{ij} \qquad (2.23)$$

where X_{0ij}=1, and *p=0…P*. Equation 2.23 makes clear that the intercept is a regression coefficient, just like the other regression coefficients in the equation. Some multilevel software, for instance HLM (Raudenbush, Bryk, Cheongh, Congdon & Du Toit, 2011) puts the intercept

[6] We may attach a subscript to Ω to indicate to which level it belongs. As long as there is no risk of confusion, the simpler notation without the subscript is used.

variable X_0=1 in the regression equation by default. Other multilevel software, for instance MLwiN (Rasbash, Steele, Browne & Goldstein, 2015), requires that the analyst includes a variable in the data set that equals one in all cases, which must be added explicitly to the regression equation.

Equation 2.23 can be made very general if we let X be the matrix of all explanatory variables in the fixed part, symbolize the residual errors at all levels by $u^{(l)}$ with *l* denoting the level, and associate all error components with predictor variables Z, which may or may not be equal to the X. This produces the very general matrix formula $Y=X\beta+Z^{(l)}u^{(l)}$ (cf. Goldstein, 2011, appendix 2.1). Since this book is more about applications than about mathematical statistics, it generally uses the algebraic notation, except when multivariate procedures such as structural equation modeling are discussed.

The notation used in this book is close to the notation used by Goldstein (2011) and Kreft and de Leeuw (1998). The most important difference is that these authors indicate the higher-level variance by σ_{00} instead of our $\sigma_{u_0}^2$. The logic is that, if σ_{01} indicates the covariance between variables *0* and *1*, then σ_{00} is the covariance of variable *0* with itself, which is its variance. Raudenbush and Bryk (2002), and Snijders and Bosker (2012) use a different notation; they denote the lowest level error terms by r_{ij}, and the higher-level error terms by u_j. The lowest level variance is σ^2 in their notation. The higher-level variances and covariances are indicated by the Greek letter *tau*; for instance, the intercept variance is given by τ_{00}. The τ_{pp} are collected in the matrix TAU, symbolized as T. The HLM program and manual in part use a different notation, for instance when discussing longitudinal and three-level models.

In models with more than two levels, two different notational systems are used. One approach is to use different Greek characters for the regression coefficients at different levels, and different (Greek or Latin) characters for the variance terms at different levels. With many levels, this becomes cumbersome, and it is simpler to use the same character, say β for the regression slopes and *u* for the residual variance terms, and let the number of subscripts indicate to which level these belong.

2.4.2 Software

Multilevel models can be formulated in two ways: (1) by presenting separate equations for each of the levels, and (2) by combining all equations by substitution into a single model-equation. The softwares HLM (Raudenbush et al., 2011) and Mplus (Muthén & Muthén, 1998-2015) require specification of the separate equations at each available level. Most other software, e.g., MLwiN (Rasbash et al., 2015), SAS Proc Mixed (Littell et al., 1996), SPSS command Mixed (Norusis, 2012), and the R package LME4 (Bates et al., 2015) use the single equation representation. Both representations have their advantages and disadvantages. The separate-equation representation has the advantage that it is always clear how the model is built up. The disadvantage is that it hides from view that modeling regression slopes by other variables is equivalent to adding a cross-level interaction to the model. As will be explained in Chapter Four, estimating and interpreting interactions correctly requires careful thinking. On the other hand, while the single-equation representation makes the existence of interactions obvious, it conceals the role of the complicated error components that are created by modeling varying slopes. In practice, to keep track of the model, it is recommended to start by writing the separate equations for the separate levels, and to use substitution to arrive at the single-equation representation.

To take a quote from Singer's excellent introduction to using SAS Proc Mixed for multilevel modeling (Singer, 1998, p. 350): "Statistical software does not a statistician make. That said, without software, few statisticians and even fewer empirical researchers would fit the kinds of sophisticated models being promulgated today." Indeed, software does not make a statistician, but the advent of powerful and user-friendly software for multilevel modeling has had a large impact in research fields as diverse as education, organizational research, demography, epidemiology, and medicine. This book focuses on the conceptual and statistical issues that arise in multilevel modeling of complex data structures. It assumes that researchers who apply these techniques have access to and familiarity with *some* software that can estimate these models. Specific software is mentioned in some places, but only if a technique is discussed that requires specific software features or is only available in a specific program.

Since statistical software evolves rapidly, with new versions of the software coming out much faster than new editions of general handbooks such as this, we do not discuss software setups or output in detail. As a result, this book is more about the possibilities offered by the various techniques than about how these things can be done in a specific software package. The techniques are explained using analyses on small but realistic data sets, with examples of how the results could be presented and discussed. At the same time, if the analysis requires that the software used have some specific capacities, these are pointed out. This should enable interested readers to determine whether their software meets these requirements, and assist them in working out the software setups for their favorite package.

In addition to the relevant program manuals, several software programs have been discussed in introductory articles. Using SAS Proc Mixed for multilevel and longitudinal data is discussed by Singer (1998). Peugh and Enders (2005) discuss SPSS Mixed using Singer's examples. Both Arnold (1992), and Heck, Thomas and Tabata (2012, 2014) discuss multilevel modeling using SPSS. Sullivan, Dukes and Losina (1999) discuss HLM and SAS Proc Mixed. West, Welch and Galecki (2007) present a series of multilevel analyses using SAS, SPSS, R, Stata and HLM. Finally, the multilevel modeling program at the University of Bristol maintains a multilevel homepage that contains a series of software reviews. The homepage for this book contains links to these and other multilevel resources.

The data sets used in the examples are described in appendix E, and are all available through the Internet.