CHAPTER 3

MULTILEVEL MODELS

3.1 Introduction to Multilevel Models

Multilevel models are concerned with a structure of the data in the population and more specifically, with the hierarchy of the data. A hierarchy consists of units grouped at different levels (Goldstein, (1995)).

There are various kinds of data in which a hierarchical structure can be identified. An example of data with a natural hierarchy is the animals grouped within classes. It is obvious that animals which belong to the same class would be more alike than others which belong to different classes, that is, they would share the same biological characteristics or even similar outward appearance. Furthermore, the data collected in social sciences may also have a hierarchical structure. For example, the students grouped or clustered within schools. In this case, it is observed that students who attend the same school are more alike, on average, than students who attend different schools. Another example of hierarchically structured data in social sciences is the people grouped within different social and economic levels. In this case, people belonging to the same economic level tend to have similar social attitudes and equal mean income. Some data hierarchies can also be created by designed experiments, such as clinical trials, where patients are grouped within hospitals.

In all of the above cases, the data have a hierarchical structure and this cannot be ignored. This structure may be natural or may be created by designed experiments. The provenance of the hierarchy is not important. What is of vital importance is the fact that once hierarchies or groupings are established, the group influences its members and vice versa. For that reason, when one wants to analyze a set of data with hierarchical structure one cannot just ignore this hierarchy and use traditional statistical analysis techniques. If one does so, it would be like overlooking the importance of group effects and on the other hand, the results may be misleading or even invalid.

A typical example of the above situation is Professor Bennett's work on formal and informal teaching styles (Bennett, 1976) and the reanalysis by Aitkin et al (1981), using multilevel analysis. More analytically, Bennett collected data on nearly a thousand students of 36 primary school classes and examined how the progress of the students depended on whether the students were taught by formal methods, by informal methods or a mixture of the two methods. He used multiple regression techniques for analyzing the data and he arrived at the conclusion that the students that were taught by formal methods made greater progress. Several years later, Aitkin analyzed the data set of Bennett, whose analysis ignored the groupings of students according to teachers and into classes, using multilevel analysis. Aitkin took into consideration the hierarchical structure of the data and concluded that the students that were taught by the formal methods did not differ from the others. This work has been the subject of much debate and many papers have been published on this issue. Aitkin's work was the first considerable multilevel analysis of social science data.

Although the issue of the proper analysis of data with hierarchical structure has been long recognized, researchers were not able to confront that problem, because they did not have a powerful research tool in their hands. But now, the development of multilevel modeling solves the problems arising from ignoring hierarchical structures and opens horizons for further research.

3.2 The Linear Multilevel Model

It was explained in the introduction of this chapter that multilevel models take into account the hierarchical structure of the data in the population. A clear example of this structure is the educational system which places students in classes, classes in schools and schools in prefectures.

In a multilevel model, students would be the level-1 units, schools would be the level-2 units and prefectures would be the level-3 units. In this way, multilevel models not only do not ignore the special structure of the population, but also intend to investigate how this structure affects the measurements of interest. Thus, with regard to school education, one may be interested in establishing whether some schools are more effective than others and furthermore in defining those characteristics that are related to the effectiveness of a school.

There are two fundamental principles of multilevel modeling that distinguish multilevel models from the single level ones. The first principle is that multilevel models consider, in the above example, schools as a sample of a wider population of schools. They do not focus just on these schools of the sample, but they regard these schools as source of information for all the schools in the population. Just as a random sample of individuals provides estimates of the characteristics of the population (mean, etc.), so a random sample of schools provides estimates for the characteristics of the schools in the population. To be exact, a random sample of schools can provide estimates of the variation and covariation between schools in the slope and intercept parameters and thus, one will be able to compare schools with different characteristics. The second principle is the existence of different levels of variation. In the example of schools, presented above, there is the level-1 variation, that is the variation between students (because in this example students are the level-1 units). There is also the level-2 variation, that is the variation between schools and the level-3 variation, that is the variation between local authorities. But we are going to refer to variation more analytically later in this chapter.

Below, the 2-level and the 3-level models will be introduced along with the basic notation, the procedures for estimating parameters, forming and testing functions of the parameters and constructing confidence intervals.

3.2.1 The 2-Level Model

Before the presentation of the 2-level model, let us stress once again that the example of the education system is going to be used throughout this section. First of all, consider the simple-level model

$$Y = \beta_0 + \beta_1 X_i + e_{0i} \qquad (3.1)$$

where Y is a response variable, X an explanatory variable and the intercept (β_0), the slope (β_1) and the residual (e_{0i}) can be interpreted in the standard way. The subscript i refers to the level-1 units, that is the students, in one school. Thus, the model (3.1) describes a single-level relationship. In order to describe simultaneously the relationships of more schools, the following model is used

$$Y_{ij} = \beta_{0j} + \beta_{1j} X_{ij} + e_{0ij} \qquad (3.2)$$

where j refers to the level-2 units, that is the schools and i refers, as above, to the level-1 units. Generally, wherever an item has an ij subscript, it means that this item varies from student to student within a school. When an item has a j subscript, it means that it varies from school to school, but has the same value for the students within the same school. Finally, when an item has neither an ij subscript nor a j subscript, then this item is constant across all students and schools. The model (3.2) is still a single-level model, since it describes separate relationship for each school. For converting the model (3.2) to a 2-level model, we have to consider β_{0j} and β_{1j} as random variables with

$$\beta_{0j} = \beta_0 + u_{0j} \quad \text{and} \quad \beta_{1j} = \beta_1 + u_{1j}$$

where u_{0j} and u_{1j} are also random variables with zero mean and variances and covariance

$$\text{var}(u_{0j}) = \sigma^2_{u0}, \;\; \text{var}(u_{1j}) = \sigma^2_{u1}, \;\; \text{cov}(u_{0j}, u_{1j}) = \sigma_{u01} \qquad (3.3)$$

The model (3.2) now takes the following form

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + (u_{0j} + u_{1j} X_{ij} + e_{0ij}) \qquad (3.4)$$

$$\text{var}(e_{0ij}) = \sigma^2_{e0}$$

In the equation above, β_0 and β_1 are the fixed parameters, while the variances σ^2_{u0}, σ^2_{u1}, σ^2_{e0} and the covariance σ_{u01} are the random parameters. If we considered the simplest 2-level model, then this would include only the random parameters σ^2_{u0} and σ^2_{e0}, and the model would have the following form

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + u_{0j} + e_{0ij} \qquad (3.5)$$

The above model is called *variance components model* and, as mentioned before, the only random parameters are the intercept variances at each level. In these models, the variance of the response about the fixed component is

$$\text{var}(Y_{ij} | \beta_0, \beta_1, X_{ij}) = \text{var}(u_0 + e_{0ij}) = \sigma^2_{u0} + \sigma^2_{e0}$$

where, as mentioned above, σ_{u0}^2 and σ_{e0}^2 are the level-2 and the level-1 variance respectively. Thus, in the variance components models, the variance of the response about the fixed component is the sum of level-1 and level-2 variance. A measure of the extend of clustering of students within schools is the intra-school correlation and is defined as

$$\rho = \frac{\sigma_{u0}^2}{\sigma_{u0}^2 + \sigma_{e0}^2} \tag{3.6}$$

In other words, this correlation measures the proportion of variance that is between schools (Goldstein, 1995).

In order to include further fixed explanatory variables in model (3.4) we extend it and we have

$$y_{ij} = X_{ij}\beta + \sum_{h=0}^{1} u_{hj} z_{hij} + e_{0ij} z_{0ij} \tag{3.7}$$

where **X** is the design matrix for the fixed explanatory variables, X_{ij} is the ij*th* row of X and z_{hij} are the explanatory variables for the random part of the model. In equation (3.7) Z={Z_0 Z_1}, where Z_0 is a vector of ones and Z_1={x_{1ij}}. Any of the explanatory variables can be measured at any of the levels.

Estimation for the Multilevel Model

Let us consider the simple 2-level variance components model (3.5). The variance matrix for the response vector Y for the model (3.5) with two level-2 units (schools) has the following form

$$V = \begin{bmatrix} \sigma_{u0}^2 J_{(n_j)} + \sigma_{e0}^2 I_{(n_j)} & 0 \\ 0 & \sigma_{u0}^2 J_{(n_j)} + \sigma_{e0}^2 I_{(n_j)} \end{bmatrix}$$

where $J_{(n_j)}$ is a $(n_j \times n_j)$ matrix of ones and $I_{(n_j)}$ is a $(n_j \times n_j)$ identity matrix. The above matrix has a block-diagonal structure and this implies that the covariance between students from different schools is zero. If we knew the variances σ^2_{u0} and σ^2_{e0}, then we could construct the covariance matrix V and applying the Generalized Least Squares estimation procedure we could obtain estimates for the fixed coefficients

$$\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} Y \qquad (3.8)$$

When the residuals follow the Normal distribution, then maximum likelihood estimates are obtained from the above equation.

The estimation procedure, that is going to be used, is an iterative one. First of all, initial estimates have to be given to the fixed parameters. Thus, from an initial OLS fit (σ^2_{u0}=0), we give the OLS estimates to the fixed coefficients. From these we construct the raw residuals, which are defined as

$$\tilde{y}_{ij} = y_{ij} - \hat{\beta}_0 - \hat{\beta}_1 x_{ij} \qquad (3.9)$$

and the vector is $\tilde{Y}_{ij} = \{\tilde{y}_{ij}\}$. Then, we form the matrix $\tilde{Y}\tilde{Y}^T$, whose expected value is the covariance matrix V. We can construct the vector vec($\tilde{Y}\tilde{Y}^T$) by stacking the columns one on the top of the other. In the same way we construct the vector vec(V), too. Below, the relationship between vec($\tilde{Y}\tilde{Y}^T$) and vec(V) is given, in which R is the residual vector

$$
\begin{bmatrix} \tilde{y}_{11}^2 \\ \tilde{y}_{21}\tilde{y}_{11} \\ . \\ . \\ . \\ \tilde{y}_{22}^2 \end{bmatrix} = \begin{bmatrix} \sigma_{u0}^2 + \sigma_{e0}^2 \\ \sigma_{u0}^2 \\ . \\ . \\ . \\ \sigma_{u0}^2 + \sigma_{e0}^2 \end{bmatrix} + R = \sigma_{u0}^2 \begin{bmatrix} 1 \\ 1 \\ . \\ . \\ . \\ 1 \end{bmatrix} + \sigma_{e0}^2 \begin{bmatrix} 1 \\ 0 \\ . \\ . \\ . \\ 1 \end{bmatrix} + R \qquad (3.10)
$$

The next step is to assume Normality and apply the GLS estimation procedure using the estimated covariance matrix of $\mathrm{vec}(\tilde{Y}\tilde{Y}^T)$. After having obtained estimates by applying GLS to (3.10) we obtain new estimates for the fixed parameters. We continue to apply this procedure until it converges. If the Normality assumption does not hold, the obtained estimates will be consistent but not fully efficient. On the other hand, for models with many random coefficients, the assumption of multivariate Normality is more suitable as it permits parametrization for complex covariance at any level.

The estimation procedure described above is the one that the statistical package MLn uses. This package was developed by H. Goldstein. There are other estimation procedures for the multilevel models, too. One of these is based on a 'Fishing scoring' algorithm which was developed by Longford (1987). There is also a variation of the Iterative Generalized Least Squares (IGLS), the procedure described above, the Expected Generalized Least Squares (EGLS). We can also consider the model (3.2) as a Bayesian linear model (Lindley and Smith, 1972) and assume that β_j is exchangeable with prior distribution which has variance σ_{u0}^2.

It is obvious, that in multilevel model we will have residuals at different levels. Consider again the simple 2-level variance components model (3.5). For this model we obtain residuals at two different levels

$$
\hat{u}_{0j} = \frac{n_j \sigma_u^2}{n_j \sigma_u^2 + \sigma_{e0}^2} \tilde{y}_j
$$

$$\tilde{e}_{0ij} = \tilde{y}_{ij} - \hat{u}_{0j}$$

$$\tilde{y}_j = \frac{\left(\sum_i \tilde{y}_{ij}\right)}{n_j}$$

where n_j is the number of level-1 units in the j*th* level-2 unit. The above residuals are consistent but only conditionally biased. They can be used as the residuals of a single level model for checking the assumptions of the model, that is, the Normality assumption and the assumption of the constant variances in the model. Beyond that, the residuals obtained by a multilevel model can be considered as random variables following a distribution with parameters that give information about the level-2 variation. These residuals can also be viewed as estimates for each level-2 unit.

Hypothesis Testing and Confidence Intervals for Fixed and Random Parameters

When hypothesis testing is needed for combinations of fixed parameters, then it is necessary to construct linearly independent functions of the p fixed parameters of the model. These functions will have the form $f = C\beta$, where C is a $(r \times p)$ contrast matrix. More precisely, if we want to test the null hypothesis that two explanatory variables are zero (H_0: f = k, where k = {0} in this case), then we have to define the contrast matrix C and construct the functions f. The next step in the hypothesis testing is to form the following equation

$$R = (\hat{f} - k)^T [C(X^T \hat{V}^{-1} X)^{-1} C^T](\hat{f} - k) \quad \text{where} \quad \hat{f} = C\hat{\beta} \qquad (3.11)$$

Under the null hypothesis $R \sim \chi^2_r$.

Now, if we want to construct an α% confidence region for the fixed parameters first of all we form the following equation

$$\hat{R} = (f - \hat{f})^T [C(X^T \hat{V}^{-1} X)^{-1} C^T]^{-1} (f - \hat{f}) \qquad (3.12)$$

and then set it equal to the α% tail region of χ_r^2. The above expression of $\hat{R}$ gives an r-dimensional ellipsoid region, which is the confidence region. Furthermore, if we want to construct a (1-α)% confidence interval separately for each linearly independent function of the parameters or for linear functions of a subset of the parameters, then the following formula will be used

$$(C_i \hat{\beta} - d_i, C_i \hat{\beta} + d_i) \qquad (3.13)$$

where C_i is the i*th* row of the matrix C and d_i is

$$d_i = [C_i (X^T \hat{V}^{-1} X)^{-1} C_i^T \chi^2_{q,(\alpha)}]^{0.5}$$

where q is the number of parameters involved in the subset.

The above methods for hypothesis testing and confidence intervals (or regions) are applied to the fixed parameters. For the random parameters, and when the sample is large, we can use the same procedures as those described for the fixed parameters. But we generally use procedures based on the likelihood statistic. So we form the deviance statistic

$$D_{01} = -2\log\left(\frac{\lambda_0}{\lambda_1}\right) \qquad (3.14)$$

where λ_0 and λ_1 are the likelihoods for the null and the alternative hypotheses, respectively. The deviance statistic $D_{01} \sim \chi_q^2$, where q is the difference in the number of parameters in the two models. Based on the same method we can also construct a (1-α)% confidence region. For doing this, we set D_{01} equal to the α% point of χ^2_q, and a region is constructed to satisfy the equation of deviance statistic.

Finally, consider the level-2 residuals. According to the example that is used throughout the section, these residuals would be the school-level residuals. In this particular case, an interesting thing to do is to compare a subset of schools and identify those schools whose residuals are different (larger or smaller) from others. For doing this, first of all we have to order the residuals from smallest to largest. Then, we construct a (1-α)% confidence interval for each residual and judge if the residuals are significantly different by whether their confidence intervals overlap. For example, under the Normality assumption and for two residuals with equal standard errors, the width of the confidence intervals for judging whether the residuals are significantly different is

$$\pm 1.39\sigma$$

at a 5% significant level. If we want to obtain confidence intervals for each of the level-2 residuals, we have to assume first of all that, on average, each pair of level-2 units will be compared the same numbers of times. The procedure for constructing the confidence intervals for each residual is

$$\hat{u}_j \pm z_b s_j \qquad (3.15)$$

and is widely discussed by Goldstein and Healy (1995). In the equation above z_β is the positive normal deviate with a two-tailed probability β.

Complex Variation

It was mentioned in the beginning of the chapter, that one fundamental principle that distinguishes multilevel models from the single level models is the existence of variation at each level. Furthermore, in multilevel models we have the potential to model the variation, at the various levels, as a function of explanatory variables. More precisely, let consider u_j and e_{ij} as the level-2 and level-1 variation, respectively. These can be expressed as

$$u_j = \sum_{h=0}^{r_2} u_{hj} z_{hj} \quad \text{and} \quad e_{ij} = \sum_{h=0}^{r_1} e_{hij} z_{hij} \qquad (3.16)$$

where z's are explanatory variables of the random part of the model and in most cases z_{0j} and z_{0ij} define the intercept variance term at each level. The simplest case of the complex variation is to model variance as a linear function of one explanatory variable. Thus, assume that we have the following model

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + (u_{0j} + u_{1j} z_{ij} + e_{0ij} + e_{1ij} z_{ij}) \qquad (3.17)$$

where $\mathrm{var}(e_{0ij}) = \sigma^2_{e0}$, $\mathrm{var}(e_{1ij}) = 0$, $\mathrm{cov}(e_{0ij}, e_{1ij}) = \sigma_{e01}$, $\mathrm{var}(u_{0j}) = \sigma^2_{u0}$, $\mathrm{var}(u_{1j}) = \sigma^2_{u1}$, $\mathrm{cov}(u_{0j}, u_{1j}) = \sigma_{u01}$. Therefore, the level-1 variation has the following form

$$\mathrm{var}(e_{0ij} + e_{1ij} z_{ij}) = \sigma^2_{e0} + 2\sigma_{e01} z_{ij} + \sigma^2_{e1} z^2_{ij} = \sigma^2_{e0} + 2\sigma_{e01} z_{ij}$$

while the level-2 variation can be written as

$$\mathrm{var}(u_{0j} + u_{1j} z_{ij}) = \sigma^2_{u0} + 2\sigma_{u01} z_{ij} + \sigma^2_{u1} z^2_{ij}$$

There is also the potential to model the variation at several levels as a nonlinear function of explanatory variables, or even to take the predicted value $\hat{y}_{ij}$, form the random term $e_{1ij}\sqrt{\hat{y}_{ij}}$ and then the level-1 variation will become $\sigma^2_{e1}\hat{y}_{ij}$, known as constant coefficient of variation model. Furthermore, we can model the level-1 variation using as explanatory variables dummy variables defined by two or more subgroups. It is possible to include in the model of the level-1 variation the covariances between two dummy variables coefficients, a dummy variable coefficients and a continuous variable coefficient, but it is not possible to include covariances between the dummy variables categories of the same explanatory variable. Everything that has been mentioned about the complex variation holds for all levels of a multilevel model.

3.2.2 The 3-Level Model

What has been said about the 2-level model can be extended in order to cover the case of the 3-level model. Based again on the example of the education system, a 3-level structure could consist of the students being the level-1 units, the schools being the level-2 units and the prefectures being the level-3 units. In this case, a simple linear 3-level model would have the following form

$$Y_{ijk} = \beta_0 + \beta_1 X_{ijk} + (v_k + u_{jk} + e_{ijk})$$

where β_0 and β_1 are the fixed parameters of the model and v_k, u_{jk} and e_{ijk} are considered as random variables with zero mean and variances

$$\text{var}(v_k) = \sigma^2_v, \quad \text{var}(u_{jk}) = \sigma^2_u, \quad \text{and} \quad \text{var}(e_{ijk}) = \sigma^2_e$$

These variances are the random parameters of the model. Let us now consider the 3-level model of this form

$$Y_{ijk} = \beta_0 + \beta_1 X_{ijk} + (v_{ok} + v_{1k} z_{ijk} + u_{0jk} + e_{0ijk})$$

In the above case, an explanatory variable has been introduced in the random part of the model and thus it is possible to have complex variation at level-3. This can be done by modeling the level-3 variance as a linear function of the explanatory variable z_{ijk}. The same extensions can be made to any of the other two levels, while a more general form of the 3-level model is

$$Y_{ijk} = X_{ijk}\beta + \sum_{h=0}^{r_3} v_{hk} z_{hk} + \sum_{h=0}^{r_2} u_{hjk} z_{hjk} + \sum_{h=0}^{r_1} e_{hijk} z_{hijk}$$

where **X** is the design matrix of the fixed explanatory variables and z's are the explanatory variables of the random part of the model.

3.2.3 Applications of Multilevel Models in Educational Research

Multilevel models are widely used in educational research. To be more precise, the need for assessment of school effectiveness gave rise to the development of multilevel modeling. As Aitkin and Longford (1986) argued '... the minimal requirement for valid institutional comparisons was an analysis based upon individual level data which adjusted for intake differences...'. Goldstein et al. (1993) analyzed data on examination results from inner London schools in relation to intake achievement, pupil gender and school type. More specifically, the data concern examination results from 5748 students in 66 schools in six Inner London Education Authorities. The examination results includes students' General Certificate of Secondary Examination (GCSE) grades in Mathematics and English, a total score for all subjects taken in that examination, scores on a common reading test at the age of 11 (LRT) and grades into three categories on the basis of a verbal reasoning (VR) test also at 11 years. If students did not have both intake measures or they obtained an ungraded result, they were omitted from the analysis. The response variables have been transformed to normality, while two kinds of models have been fitted to the data. The first model analyzes the total examination score and the second model constitutes a bivariate analysis of the English and mathematics scores.

As far as the first model is concerned, the explanatory variables that Goldstein et al. (1993) used are: the standardized London reading test (LRT), the verbal reasoning category (VR), the student's gender, the school gender (mixed, girls, boys) and the school religions denomination (State, Church of England, Roman Catholic, other). Thus, the first model is written as:

$$y_{ij} = \sum_{h=0}^{5} \beta_h x_{hij} + \sum_{h=6}^{10} \beta_h x_{hj} + \sum_{h=0}^{2} u_{hj} x_{hij} + \sum_{h=0}^{1} e_{hij} x_{hij}$$

where the subscript 0 refers to the constant term, 1 to LRT, 2 to the dummy variable for VR group 1, 3-5 refer to the square of LRT, the dummy variable for VR group 2, and the dummy variable for gender. Finally, the subscripts 6-10 refer to the five school level defined variables. Furthermore, the first summation refers to the explanatory variables defined at the student level, the second to those defined at the school level, the third summation refers to the random part of the model at the school level and the fourth summation defines the random variation at the student level. At level-2 it holds that

$$\text{var}(u_{hj}) = \sigma_h^2 \quad \text{and} \quad \text{var}(e_{hij}) = \sigma_{eh}^2$$

while the level-1 contribution to the variance is

$$\sigma_{e0}^2 + 2\sigma_{e01}x + \sigma_{e1}^2 x^2 .$$

After having fitted the first model Goldstein et al. concluded that the effect of school gender is small while there is a small advantage for the students attended Roman Catholic schools. Furthermore, girls do better than boys with large differences among the different verbal reasoning categories. For the between school variation it is concluded that the relationship between examination score and LRT varies and also the difference between verbal reasoning categories 1 and 3 varies with high positive correlation. Finally, at level-1 the variance increases with the increase of LRT score. For each school it is possible to form the school residuals as follows:

$$u_{0j} - 2u_{1j}, \quad u_{0j} + 2u_{1j} + u_{2j},$$

and because the above residuals are estimated, confidence intervals for them can be constructed. With the construction of these intervals it is concluded that it is not possible to distinguish which school is more effective, since there is a considerable overlap between these intervals.

The second model includes the student-level variables as explanatory variables, while at the between-school-level the model includes only the intercept and the LRT coefficient. A multivariate model is specified by treating the multiple variates within each student as the level-1 classification. Thus, in the fixed part of the model it is concluded that there is a small difference in favour of the boys for mathematics, while the difference is in favour of the girls for English. On the other hand, there is little difference between the relationships for mathematics and English for LRT and VR categories. In the random part of the model the LRT for mathematics and English does not vary. Finally, the between-student variation decreases from the first verbal reasoning category to the third. According to the school residuals, for students with average LRT scores, holds that the school with the greatest English residual is average for mathematics and one of the schools with high mathematics residual has low value for English.

Another model was introduced by H. Goldstein and S. Thomas (1996). Their aim was to investigate the properties of league tables and compare institutions after adjusting for the intake achievements of the students. First, it is necessary to mention that the league tables are tables with the average General Certificate of Secondary Education examination results of each school. Every secondary school in England and Wales has to publish those results in order that the tables are used by parents in choosing schools and colleges.

In the analysis agreed to participate 436 institutions but finally, only 325 had all the institution level information available. The number of students for those 325 institutions was 21,654. For each one of them, there were selected information about the A-level and AS-level results, the GCSE results together with the number of GCSE examinations taken. It must be mention here that the A-level exams are the exams of advanced level General Certificate of Education (GCE), the AS-level exams are the exams of advanced supplementary level GCE and the GCSE exams are the exams of General Certificate of Secondary Education. They also collected data on the gender of the student, the gender composition of the institution and the age of the student.

What is of interest in comparing institutions, is the progress that the students make from the time they enter the specific institution to the time they graduate. That is why the existing achievements or performances of the students before entering the institutions are also important. For that reason, in school effectiveness research, the final examination result is used as response variable and the existing achievements as explanatory variable. In this case, a simple model of this form is the following:

$$y_{ij} = \beta_0 + \beta_1 x_{ij} + u_j + e_{ij},$$

y_{ij}: A-AS examination score for the *i*th student in the *j*th institution

x_{ij}: GCSE score for the *i*th student in the *j*th institution
u_j: effect for the *j*th institution
e_{ij}: student residual

We regard u_j as random effect and hence the above model is a 'multilevel' model with two-level hierarchy of students nested within schools. For comparing the institutions the estimates of $u_{j,}$ derived either as Bayesian posterior estimates or as regression predictors of the unknown u_j were used, given the responses and the model parameter estimates.

Also, normal scores for the whole sample can be used in order to determine the following model which describes the relationship between A-AS-level score and GCSE score:

$y_{ij} = \beta_0 + \Sigma\, \beta_h\, x_{ij}^h + \beta_5\, z_{ij} + u_j + e_{ij,}$

$u_j \sim N(0, \sigma_u^2)$,
$e_{ij} \sim N(0, \sigma_e^2)$,
$\rho = \sigma_u^2\, (\sigma_u^2 + \sigma_e^2)^{-1}$
$z_{ij} = 1$, for girls
0, for boys
ρ: intra-institution correlation

The model that has just be presented takes into account the GCSE score for every student. It is therefore adjusted for the GCSE score and if we do so, the between-school and between-student variance will be reduced. The results would be completely different if these necessary adjustments were not made or if the model was adjusted for the mean institutional A-AS-score only for one year. In those cases the unadjusted analyses would produce biased estimates of institutional effects. Furthermore, this may be a reason why the league tables are misleading and partial.

Goldstein and Thomas (1996) considered the gender effect and defined the following model:

$$y_{ij} = \beta_{0j} + \Sigma\, \beta_h\, x_{ij}^{h} + \beta_{5j}\, z_{ij} + e_{ij,}$$

$$\beta_{0j} \sim N(\beta_0, \sigma_{u0}^{2})$$
$$\beta_{5j} \sim N(\beta_5, \sigma_{u5}^{2})$$
$$\text{cov}(\beta_{0j}, \beta_{5j}) = \sigma_{u05}.$$

The above model introduces the complex variation by modeling this variance as a function of gender:

$$e_{ij} = e_{0ij} + e_{5ij}\, x_{5ij},$$

$$e_{ij} \sim N(0, \sigma^{2}_{e0} + 2\, \sigma_{e05}\, x_{5ij})$$
$$\text{var}(e_{0ij}) = \sigma^{2}_{e0}$$
$$\text{cov}(e_{0ij}, e_{5ij}) = \sigma_{e05}.$$

The variable x_{5ij} is a dummy variable which takes the value 1 for females and 0 for males. Furthermore, the quantity $2\sigma_{e05}$ is the difference between the variances for males and females.

At this point, the dependence of the variation at both levels on the GCSE score can be studied. For that reason, the GCSE score will be grouped in three groups in order to have homogeneity within but not between groups. This can be done by extending the above model into the following one:

$$y_{ij} = \beta_0 + \Sigma\, \beta_h\, x_{ij}^{h} + \beta_{5j}\, z_{ij} + \beta_{6j}\, x_{ij}\, zij + \beta_{7j}\, x_{ij}^{2}\, zij + u_{1j}\, w_{1ij} + u_{2j}\, w_{2ij} + u_{3j}\, w_{3ij} + u_{5j}\, w_{5ij} + e_{ij}.$$

In this model the w_{hij}, for h = 1, 2, 3, are dummy (0, 1) variables for the three GCSE groups and the level 1 variation is given by:

$$var(e_{ij}) = \sigma^2_{e0} + 2\ \sigma_{e05}\ z_{ij} + 2\ \sigma_{e02}\ w_{2ij} + 2\ \sigma_{e03}\ w_{3ij}.$$

It is obvious that the level 1 variation is an additive function of parameters for GCSE score and gender. Also, it can be observed that in each factor one category is missing. It could be considered an interactive function with parameters for each of the six combinations of GCSE score by gender, too.

The number of GCSE examinations taken by the students can be used as one more explanatory variable. The age of the student and the type and the status of the institution can also be included. For comparing between institutions, it is important to have in mind the fact that the residual estimates, with which the comparisons are going to be made, have standard errors which have to be estimated. Furthermore, by assuming that these standard errors are independent and also with the assumption of normality, normal confidence intervals about each estimate can be constructed. Thus, one can present the estimates of interest graphically accompanied by error bars corresponding to confidence intervals at a level β. A separation will be significant if the intervals do not overlap.

It is obvious from the above that the use of unadjusted results are misleading as well as the league tables, which do not take account the previous achievements of the students. Thus, if we want to get statistically significant estimates, we will have to make adjustments of the previous performance of the students. If we use only the average examination results as measure of comparison between institutions, then we will arrive at a wrong choice of institutions.