CHAPTER 2

VARIANCE COMPONENT AND HIERARCHICAL LINEAR MODELS

2.1 Variance Component Models

2.1.1 Introduction

The variance component models were being used widely for the assessment of school effectiveness and in other areas concerning educational research, before the use of hierarchical and multilevel models. The common feature of all those models is the existence of random effects at each sampling level.

At this point, it is important to make a distinction between the fixed effects and the random effects and between the models that are due to those effects. For that reason, let us mention two examples (Searle, 1971). Consider an agricultural experiment testing the efficacy of three fertilizers on 24 plants (six plants got no fertilizer at all, being considered as control). A very simple model for analyzing the data is the following:

$$y_{ij} = \mu + \alpha_i + e_{ij}, \qquad i = 1, 2, 3, 4, \quad j = 1, 2, 3, 4, 5, 6$$

where y_{ij} is the jth observation on the ith treatment, μ is the mean, α_i is the effect of treatment i and e_{ij} is the error term. In this example our interest is focused on just these four treatments and there is no thought for other fertilizers. Thus, these effects are called fixed effects and consequently the model is the fixed effects model.

On the other hand, consider a laboratory experiment studying the maternal ability of mice, using as a response the weight of six litters from each of four dams. The model used for this kind of data is :

$$y_{ij} = \mu + \delta_i + e_{ij}, \qquad i = 1, 2, 3, 4\ , \quad j = 1, 2, 3, 4, 5, 6$$

where y_{ij} is the weight of j*th* litter from the ith dam, μ is the mean, δ_i is the effect due to i*th* dam and e_{ij} is the error term. In this example, the four female mice used in the experiment are regarded as a sample from a large population of female mice. The experiment is not concerned just for these four mice, but the prime interest is to draw inference about the population. Thus, it can be considered as taking a random sample of mice on each occasion. So, δ_i's are random variables, which are called *random effects* and the corresponding model, random effects model. The error terms are assumed to be a random sample from a population distributed as $(\mathbf{0}, \sigma_e^2 \mathbf{I})$, while δ's are also random sample from a population distributed as $(\mathbf{0}, \sigma_\delta^2 \mathbf{I})$. Thus, the variance of an observation is

$$\sigma_y^2 = \sigma_\delta^2 + \sigma_e^2,$$

and the variances σ_e^2 and σ_δ^2 are called *variance components* and for that reason the corresponding model is also called *variance component model*.

Lastly, there are also models that contain fixed and random effects (apart from the general mean μ and the error term) and these are called *mixed models*.

An example of a mixed linear model is given by Goldstein (1986), using educational data. The data are structured hierarchically and there are three levels: schools, classrooms within schools and children within classrooms. Supposing that there are measurements on a response variable for the jth child in the ith classroom within the kth school, the appropriate model for analyzing that kind of data is:

$$Y_{kij} = \alpha^*_{kij} + \beta^*_{ki} + \gamma^*_k \qquad (1)$$

and at each of the three levels it is possible to set up a linear model that relates the terms of the initial model to a function of explanatory variables. More specifically, for the term representing the level-3 units (schools) it holds that:

$$\gamma^*_k = \gamma_0 + \gamma_1 w_{1,k} + ... + v_k = \sum_{l=0}^{q} \gamma_l w_{l,k} + v_k \qquad (2)$$

with v_k being a random variable with zero mean and variance σ^2_v and γ_l being the school level coefficient for the lth explanatory variable $w_{l,k}$ for school k. The term representing the level-2 units (classrooms) can be written as:

$$\beta^*_{ki} = \beta_0 + \beta_{1,k} z_{1,ki} + ... + u_{ki} = \sum_{l=0}^{p} \beta_{l,k} z_{l,ki} + u_{ki} \qquad (3)$$

with u_{ki} being a random variable with zero mean and variance σ^2_u, while $\beta_{l,k}$ is the classroom level coefficient of the lth explanatory variable $z_{l,ki}$ for classroom ki. Finally, for the term representing the level-1 units (children) we have that:

$$\alpha^*_{kij} = \alpha_0 + \alpha_{1,ki} x_{1,kij} + ... + e_{kij} = \sum_{l=0}^{r} \alpha_{l,ki} x_{l.kij} + e_{kij} \qquad (4)$$

where again as before e_{kij} is a random variable with zero mean and variance σ^2 and $\alpha_{l,ki}$ is the child level coefficient of the lth explanatory variable $x_{l,kij}$ for child kij. Substituting the above three relations to the initial model we finally get:

$$Y_{kij} = \alpha_0 + \beta_0 + \gamma_0 + \sum_{l=1}^{r} \alpha_{l,ki} x_{l,kij} + \sum_{l=1}^{p} \beta_{l,k} z_{l,ki} + \sum_{l=1}^{q} \gamma_l w_{l,k} + (v_k + u_{ki} + e_{kij}) \qquad (5)$$

and the variance will be equal to

$$\mathrm{var}(Y_{kij}) = \sigma_v^2 + \sigma_u^2 + \sigma^2$$

assuming zero covariances between the random variables. A quantity of interest to be estimated is the *intra-class coefficient* given by:

$$\rho_{ki} = \frac{\sigma_v^2 + \sigma_u^2}{\sigma_v^2 + \sigma_u^2 + \sigma^2}$$

and the corresponding quantity for schools is:

$$\rho_k = \frac{\sigma_v^2}{\sigma_v^2 + \sigma_u^2 + \sigma^2}.$$

Random coefficients can be included in model (5). If it were a single level model, the random coefficients would have been defined at that level. But in a multilevel model, as Goldstein (1986) mentioned, '... if a coefficient at any level in (5) is assumed to be a random variable, it can be written in general as the sum of linear functions of explanatory variables at that level and other levels plus error terms, which, in principle, could operate at any level'.

2.1.2 Applications of Variance Component Models in Educational Research

One of the most important contribution to classroom research in the UK was Bennett's study (1976), which was designed to compare formal and informal teaching methods. This study became the subject of a long debate. Bennett's major conclusion was that formal methods of teaching were associated with greater pupil progress.

Aitkin et al. (1981) re-analyzed the educational data that was used by Bennett (1976) and ended up with different conclusions about the teaching styles and their effects on pupil performance.

Primarily, in order that the different teaching styles be defined, a questionnaire that contained 28 items was designed, piloted and administered to a representative sample of teachers. This questionnaire covered six major areas of classroom behaviour, such as:

1. classroom management and organization,
2. teacher control and sanctions,
3. curriculum content and planning,
4. instructional strategies,
5. motivational techniques and
6. assessment procedures.

The 28 items were coded into 38 binary items and the probability model that was adapted in order to examine the existence of distinguishable teaching styles, is a mixture or latent class model. Thus, Aitkin et al. supposed that there are k latent classes or types of teaching style, each one of them characterized by different frequencies of use of different behaviours. If the proportions of each teaching style in the population is λ_1, λ_2, ..., λ_k, where $\sum_{j=1}^{k}\lambda_j = 1$, then the probability that a teacher's vector **X** of responses takes the value **x**, given that he is in the j*th* latent class is:

$$P(\mathbf{X}{=}\mathbf{x}\,|\,j,\theta_j)$$

where θ_j is a vector of parameters, probably different for each latent class, while the unconditional probability of the response **x** is given by:

$$P(\mathbf{X}{=}\mathbf{x}) = \sum_{j=1}^{k} P(\mathbf{X}{=}\mathbf{x}\,|\,j,\theta_j)P(\text{teacher in class j})$$

$$= \sum_{j=1}^{k} \lambda_j P(\mathbf{X}{=}\mathbf{x}\,|\,j,\theta_j)$$

For specifying how the probability $P(\mathbf{X}\,|\,j,\theta_j)$ depends on θ_j the authors of the paper assumed that given the latent class to which a teacher belongs, his responses on the 38 binary items are independent

$$P(\mathbf{X}{=}\mathbf{x}\,|\,j,\theta_j) = \prod_{t=1}^{38} P(X_t = x_t\,|\,j,\theta_{jt}),$$

an assumption widely used in latent class modeling in sociology. Aitkin et al. did not use models with more than three latent classes, as mixture models possess multiple local maxima of the likelihood function according to the initial assignments of teachers to classes, used to start off the iterative algorithm for the maximum likelihood estimates. But with the two-latent class model there exists a unique maximum of the likelihood function.

After having analyzed the data with both two-class and three-class models they concluded that the first latent class in the two-class model is at the formal end of every item, as well as in the three-class model with a few exceptions. The second class in the two-class model is at the informal end of every item, as well as in the three-class model but again with a few exceptions of items. Finally, the third class in the three-class model is intermediate between the first and the second class. For more details about the definition of formal, informal and mixed teaching styles, see Aitkin et al. (1981). Nevertheless, what is of interest to be reported in connection with the teachers with mixed teaching style is the high frequency of disciplinary problems and the low frequency of testing, compared with the formal and informal teachers.

Furthermore, in order to determine any relation of teaching style to pupil progress Aitkin et al. developed '... a 'mixed' or variance component model for 'clustered' or 'nested' sample designs for the one way analysis of covariance for pre-test/test situations'. For this kind of analysis the experimental design is as follows: a random sample of 36 teachers was assigned randomly to classrooms, while a random sample of 921 children was divided into classes of about 25 children. Each teacher was assigned to one of the three teaching methods and taught the one of the classes mentioned above. Finally, the children were tested at the beginning (pre-test) and at the end of the year (achievement test). The analysis was based on the following variance component model

$$Y_{pqr} = \mu + \gamma x_{pqr} + \alpha_p + T_q + E_{pqr}$$

where, Y_{pqr} denote the achievement test score, x_{pqr} the pre-test score of the r*th* child in the q*th* classroom taught by method p. Also, α_p constant and represents the differences between the three methods, T_q is treated as random variable and represents the ability of the teacher in the q*th* classroom. Furthermore, T_q and E_{pqr} are mutually independent random variables assumed normally distributed with zero mean and variance σ_T^2 and σ_E^2 respectively. In the above model the slope of the regression on pre-test score is assumed to be constant for different teaching methods while T_q are treated as random variables. As a consequence of the random teacher effects it holds that:

$$\mathrm{var}(Y_{pqr}) = \mathrm{var}(T_q + E_{pqr}) = \sigma_T^2 + \sigma_E^2$$

$$\mathrm{cov}(Y_{pqr}, Y_{pqr'}) = \mathrm{cov}(T_q + E_{pqr}, T_q + E_{pqr'})$$
$$= \mathrm{var}(T_q) = \sigma_T^2$$

$$\mathrm{cor}(Y_{pqr}, Y_{pqr'}) = \rho = \frac{\sigma_T^2}{\sigma_T^2 + \sigma_E^2}$$

An extension of the above model holds if the slope of the regression of achievement test on pre-test is considered to be different for the three teaching methods. Thus,

$$Y_{pqr} = \mu + \gamma_p x_{pqr} + \alpha_p + T_q + E_{pqr}.$$

What is of interest to determine is whether the different teaching methods affect the progress of children in classrooms. If we consider that the null hypothesis is the one of no difference between the teaching methods (α_1=α_2=0), then assuming absence of covariance and equal class size, this hypothesis can be tested with an ANOVA table. In the case of unequal class size efficient estimator can be obtained by maximum likelihood, while the ANOVA table can be replaced by an analysis of deviance table. The results of the analysis showed that the formal classrooms do best in English, the informal classrooms do best in reading, both classrooms are similar in mathematics, while mixed classrooms do worst on all tests.

At this point it was emphasized by the authors that these results are not statistically significant, since they allowed for the random variation among classrooms. Furthermore, the abilities of individual teachers can be estimated treating them as fixed. Another way of estimating teachers' abilities is to use the additional information in the 'prior distribution' of ability. In this case, the ‘posterior distribution’ of T given Y contains all information about teacher ability, given the prior distribution and the data from pupils in each class, while the ‘expected’ ability is the mean of the posterior distribution. The basic conclusion of the reanalysis by Aitkin et al. is condensed as follows: ‘The teaching style differences in achievement which were found in TS (Bennett, (1976)) are not confirmed by the reanalysis. There are two reasons for this. First, the analysis of covariance model which includes the random effect of teachers results in greatly reduced significance of any differences, because of the large variation among teachers. Second, the clustering of teachers by the latent class model changes the nature of the differences among teaching styles...’.

Except from Bennett and Aitkin et al. the educational data was analyzed once again by Prais (1983). Although, Prais did not use a variance component model for the analysis of the data, it is interesting to mention his work and his different conclusions. Prais used straightforward regression techniques and the aim of his study was of course to determine whether formal and informal teaching methods affect pupil performance and in what way. This re-consideration of the data was prompted from the different conclusions of Aitkin and it is in agreement with Cox’s question (Aitkin et al. (1981), discussion): ‘whether some simpler analysis would not be effective’.

Prais instead of using the observations on each of the 921 pupils in the 36 classes, used the 36 class averages. Considering this case, the loss of information is not great since ‘...the variability of class averages is much less than of individual pupils...’. Thus, according to this analysis and for the test scores in mathematics, Prais found out that eleven of the thirteen informal teachers did worse than average, but only two of the twelve formal teachers did worse than average. Although, this difference is significant, according to a test based on binomial distribution, we have to bear in mind the small number of observations and the way the teachers were characterized as formal or informal. As an example of the latter, let us consider one teacher in the informal group whose class showed a gain of 14.1 points above the average (while the best class among the formal classes did 6.1 points above the average). Although this teacher was classified in the informal group of teachers, her methods were not at all informal. Another outlier in the formal group is the teacher whose class did 7.4 points below the average. Before proceeding with the analysis, let us mention that the group consisting of teachers using 'mixed' teaching methods was excluded entirely from Prais's study, as it was very difficult to classify these teachers.

The method that was used by Prais in order to analyze the educational data was the multiple regression analysis with two explanatory variables: (a) the opening score and (b) a dummy variable which takes the value 1 for formal classes and 0 for informal classes. This variable represents the average benefit associated with formal rather than informal instruction. The conclusions of the analysis can be summarized as follows:

1. For the mathematics and English there is a statistically significant difference in favour of formal teaching methods. For the reading scores there is also a difference in favour of formal teaching methods but it is not quite significant.
2. If the three scores are combined in order to obtain a single one, then it is estimated again that formal teaching is better.

3. If the outlying formal teacher is included, then the estimated benefit of formal teaching will lower but the estimated differences will remain important.

4. If the outlying informal teacher is also included, then the obtained results will be consistent with that of Aitkin et al. (1981).

Another study considering the assessment of school effectiveness was performed by Aitkin and Longford (1986). It refers to statistical modelling issues in school effectiveness studies and proposes five models for describing the relationships between examination results and characteristics of the student intake. A major problem, that the authors have to confront with, is when they want to assess the importance of school-level variables, but they have student-level outcomes. That is the so-called *multilevel analysis.* We are going to refer to this more analytically later. It is also possible to use variance component models, mixed models and *hierarchical linear models* because all these are models with random effects at each sampling level.

The data used in the study, were O-level and CSE examination results for a 25% random sample of the 5th year cohort in all the schools of one Local Education Authority in the UK. That is, 907 students in 18 schools. From these 18 schools, two are single-sex Grammar schools and the other 16 are mixed-sex comprehensive schools. It is important to mention that the two single-sex Grammar schools had a substantially higher ability intake than the others. The O-level and CSE passes are available for each student and these are converted to another score (called 'ILEA' score, based on the scoring system of the Inner London Education Authority), in order to have a unique scoring system and be able to fuse the examination results and use them as response variable in the study. The explanatory variables are the VRQ and the gender of the students. The VRQ is the Verbal Reasoning Quotient and is a measure of intellectual ability. It is very important that the VRQ be reliable. If it is not so then it will produce bias towards zero in the estimation of regression coefficients for these variables. A way to overcome this problem is to take two parallel measurements on each student and estimate the reliability of VRQ. These explanatory variables are at the student-level. At the school-level, only the school size will be used. Other variables suggested by the authors are the mean VRQ score for each school, the proportion of girls in each school and the school standard deviations or inter-quartile ranges for VRQ.

As we have already mentioned, 16 of the schools are mixed-sex comprehensive schools and only 2 are single-sex Grammar schools with high ability intake. Thus, it is obvious that there is heterogeneity in the complete sample and this raises difficulties in interpreting the results. That is why in the analysis a sample containing only the 16 mixed-sex comprehensive schools was used.

Firstly, models with constant slopes for all schools were presented. The first of these models is given below:

$y_{ij} = \alpha + \beta x_{1ij} + \gamma x_{2ij} + \varepsilon_{ij}$, $j = 1,2,\ldots, n_i$; n_i is the number of students from the

i*th* school in the sample

$i = 1,2,\ldots,k$ number of schools (k=18)

y_{ij} : ILEA score of the j student in the i school

x_{1ij}: VRQ intake score for the j student in the i school

x_{2ij} : gender (0 for boys, 1 for girls)

ε_{ij} : random sample from $N(0,\sigma^2)$

The data are treated as a single sample of 907 students and the model does not take into consideration the grouping of students into schools. The response variable is the ILEA score for the j student in the i school, while explanatory variables are the VRQ intake score for j student in i school and the gender of each student. The gender is a dummy variable coded 0 for boys and 1 for girls. Finally, the model includes the random effects ε_{ij}, which are assumed to be random sample from $N(0,\sigma^2)$. In the model, it can also be included the linear and quadratic interactions of VRQ with gender, which are highly significant. If we ignore these interactions and use only the VRQ and gender effect then we will conclude that only the VRQ effect is significant. The school comparisons are based on the school means of residuals. The mean residual for the i*th* school is given by the following formula:

$$e_{1i} = (\bar{y}_i - \bar{y}) + \hat{\beta}_1(\bar{x}_i - \bar{x}) .$$

The second model with constant slopes for each school is given by the formula below:

$y_{ij} = \alpha_i + \beta x_{1ij} + \gamma x_{2ij} + \varepsilon_{ij}$, $j = 1,2,\ldots, n_i$; n_i is the number of students from the i*th* school in the sample

$i = 1,2,\ldots,18$ number of schools

y_{ij} : ILEA score of the j student in the i school

x_{1ij}: VRQ intake score for the j student in the i school

x_{2ij}: gender (0 for boys, 1 for girls)

ε_{ij} : random sample from $N(0,\sigma^2)$

In this model we have the inclusion of specific intercept parameters α_i, i = 1, 2, ..., 18 for each school. This is the only difference between model 1 and model 2. If we consider again the quadratic VRQ and the interaction effects we will find that they are not statistically significant. Also, the gender is not significant for the complete sample and the comprehensive sub-samples. The school comparisons are based on the estimates of α_i in the following formula:

$$e_{2i} = \hat{\alpha} - \bar{\alpha} = (\bar{y}_i - \bar{y}) - \hat{\beta}_2(\bar{x}_i - \bar{x}) .$$

The third model uses aggregated student results, that is, the school means and is as follows:

$\bar{y}_i = \alpha + \beta x_{1i} + \gamma x_{2i} + \eta_{i,}$ i = 1,2,...,18 number of schools

$\bar{y}_i$: mean ILEA score for school i

x_{1i} : mean VRQ intake score for school i

x_{2i} : mean gender score (the proportion of girls) for school i

η_i : random sample from $N(0, \varphi_i^2)$, $\varphi_i^2 = \sigma^2 / n_i$

To fit the model some form of weighted least squares is used. If we take the complete sample with the two single-sex Grammar schools then the gender effect is significant. If we remove the two single-sex Grammar schools then the gender effect will disappear. This happens because in the complete sample the two Grammar schools are of opposite gender and that affects the gender effect. But when we consider only the 16 comprehensive schools then the above statements will not hold because the proportion of girls in the 16 schools is similar. Comparisons are based on individual residuals:

$$e_{3i} = (\bar{y}_i - \bar{y}) - \hat{\beta}_3(\bar{x}_i - \bar{x}) .$$

It is pretty obvious that in the sets of school effects only the slope estimates $\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3$ are different.

The fourth model takes into consideration the fact that the school context affects the student's outcome. For that reason, we include in the first model the school means of individual level variables. Thus, the fourth model has the following form:

$y_{ij} = \alpha + \beta x_{ij} + \delta \bar{x}_i + \varepsilon_{ij}$, j = 1,2,..., n_i ; n_i is the number of students from the *ith* school in the sample

i = 1,2,...,18 number of schools

y_{ij} : ILEA score of the j student in the i school

x_{ij}: VRQ intake score for the j student in the i school

$\overline{x}_i$: mean VRQ for school i

ε_{ij} : random sample from $N(0,\sigma^2)$

The estimate of δ is the difference between the third and second model estimates of β. School comparisons are based on the school means of individual residuals:

$$e_{4i} = (\overline{y}_i - \overline{y}) - \hat{\beta}_3(\overline{x}_i - \overline{x}).$$

As we can see this is identical to the school effects of the third model.

The fifth model, given by the formula below, contains only one explanatory variable, the VRQ:

$y_{ij} = \alpha + \beta x_{ij} + \xi_i + \varepsilon_{ij}$, $j = 1,2,...,n_i$; n_i is the number of students from the *ith* school in the sample

$i = 1,2,...,18$ number of schools

y_{ij} : ILEA score of the j student in the i school

x_{ij}: VRQ intake score for the j student in the i school

ξ_i : school effects, random sample from $N(0, \sigma_I^2)$, independent of e_{ij}

ε_{ij} : random sample from $N(0,\sigma^2)$

The following model can also be used:

$$y_{ij} = \alpha_i + \beta x_{ij} + \varepsilon_{ij},$$

and in this case the α_i, i = 1, 2, ..., 18 are assumed to be a random sample from $N(\alpha,\sigma_I^2)$. According to this model, the α_i are random variables with mean α and variance σ_I^2. If we include the gender in the fifth model we will have:

$y_{ij} = \alpha + \beta x_{1ij} + \gamma x_{2ij} + \xi_i + \varepsilon_{ij},$ j = 1,2,..., n_i ; n_i is the number of students from the *i*th school in the sample

i = 1,2,...,18 (number of schools)

y_{ij} : ILEA score of the j student in the i school

x_{1ij} : VRQ intake score for the j student in the i school

x_{2ij} : gender score for the j student in the i school

ξ_i : school effects, random sample from $N(0, \sigma_I^2)$, independent of ε_{ij}

ε_{ij} : random sample from $N(0,\sigma^2)$

Fitting the above model for the complete sample and for the comprehensive sub-samples, we find that only the VRQ is significant. It is also noticeable that the regression coefficient estimates for the two samples are almost the same. Another important point that should be emphasized is that if the mean of the student-level variables varies over schools then these variables can reduce the school-level variance component but if the mean of the student-level variables is constant within school then the above does not hold.

Comparing the above models we conclude that the first and third model are affected by the presence of the two single-sex Grammar schools and an analysis for school comparisons should not be based on these models. On the other hand, the second and the fifth model are almost unaffected by the presence of the two Grammar schools in the complete sample and give better results. The difference between the two models is that the second uses stratification while the fifth uses clustering or that the second uses fixed effects while the fifth model uses random effects.

In the sequel, models with variable slopes are presented. The first of these models is an extension of the second model. The form of this model is:

$y_{ij} = \alpha_i + \beta_i\, x_{1ij} + \gamma_i\, x_{2ij} + e_{ij}$, $j = 1,2,...,n_i$; n_i is the number of students from the *i*th school in the sample

$i = 1,2,...,18$ (number of schools)

y_{ij}: ILEA score of the j student in the i school
x_{1ij}: VRQ intake score for the j student in the i school
x_{2ij}: gender (0 for boys, 1 for girls)
e_{ij}: random sample from $N(0,\sigma^2)$
In this model we have k regression lines, different for each school. The model can be reduced to the following one where the gender effect is constant over schools:

$y_{ij} = \alpha_i + \beta_i\, x_{1ij} + \gamma\, x_{2ij} + e_{ij}$.

Furthermore, we can totally omit the gender effect, in which case we will have the following model:

$y_{ij} = \alpha_i + \beta_i\, x_{1ij} + e_{ij}$.

There are two ways of comparing the schools. We can either consider the intersections of the regression lines for each pair of schools or we can specify typical values of VRQ and make partial comparisons between schools.

The second model with variable slopes is an extension of the fifth model that was represented before. This extended model has the following form:

$y_{ij} = \alpha + \beta\, x_{ij} + \xi_i + \zeta_i\, x_{ij} + e_{ij}$, $j = 1,2,...,n_i$; n_i is the number of students from the *i*th school in the sample

$i = 1,2,...,18$ (number of schools)

y_{ij}: ILEA score of the j student in the i school
x_{ij}: VRQ intake score for the j student in the i school
ξ_i, ζ_i: random intercept and slope for i*th* school
e_{ij}: random sample from $N(0,\sigma^2)$

The above model can also include a random gender component independent of the random VRQ slope. As we can see from the above, a correct specification of the variables at school and student level is essential for a reliable estimation.

2.2 Hierarchical Linear Models

2.2.1 Introduction

Hierarchical linear models are also of models that can be used in educational research as they contain random effects at each sampling level. The hierarchical structure of educational data inhibits the use of traditional linear models since the latter can produce misleading results.

Raudenbush and Bryk (1986) presented a general statistical model for studying school effects. The data consists of students mathematics achievements and their socioeconomic status (SES). The effectiveness of public and private (Catholic) schools was intended to be examined. A first approach is the use of regression slopes as outcomes. As Raudenbush and Bryk mentioned this is a very appealing method since it extends the kinds of questions that school research can examine and because it permits researchers to explore the effects of school policies on the relationships occurring within-groups. But there are some disadvantages associated with this method. First of all, regression coefficients have considerably greater sampling variability than sample means. Also, the sampling precision of the estimated slopes varies across units depending on the design that has been used for the collection of the data within each unit. Furthermore, the variability in the estimated slopes consists of two components, the parameter variance and the sampling variance. The disadvantage of the use of slopes-as-outcomes is that the corresponding model does not provide separate estimates for the variance components. In order to overcome the above deficiencies, a flexible statistical tool has been developed which provides means for studying how variations in policies and practices influence educational processes and also provides the basis for constructing richer definitions of school effectiveness.

This statistical tool is the hierarchical linear model and the statistical theory behind it includes mixed-model ANOVA, regression with random coefficients, covariance component models and Bayesian estimation for linear models. The paper by Lindley and Smith (1972) constitutes a key contribution to the construction of the hierarchical linear model. Raudenbush and Bryk (1986) described the method followed by Lindley and Smith as follows, 'They explicitly lay out a hierarchical structure in which parameters estimated at one level in the model become the outcome variables at the next level.' The above hierarchical parametric structure enables us to model multilevel phenomena faced in school-effects research. The hierarchical linear model (HLM) is widely applied in social and psychological research.

Let us now define a hierarchical linear model. Firstly, a within-group model is set which determines the relations among various student-level characteristics, X_{ijk} and the achievement scores, y_{ij}. The model has the following form:

$$y_{ij} = \beta_{j0} + \beta_{j1}X_{ij1} + \beta_{j2}X_{ij2} + ... + \beta_{jK-1}X_{ijK-1} + R_{ij} \qquad (1)$$

$i = 1, ..., n_j, \quad j = 0, ..., K-1$

where, y_{ij} is the outcome score for student i in school j, X_{ijk} are the values on the student-level characteristics for student i in school j, β_{jk} are the regression coefficients and R_{ij} is the random error. Furthermore, the within-school regression coefficients, β_{jk}, are allowed to vary across schools, forming the following between-group model:

$$\beta_{jk} = \theta_{0k} + \theta_{1k} Z_{1j} + \theta_{2k} Z_{2j} + ... + \theta_{P-1,k} Z_{P-1,j} + U_{jk} \qquad (2)$$

p = 0, ..., P-1

where, θ_{pk} are regression coefficients that, according to Raudenbush and Bryk (1986), 'capture the effects of school-level variables on the within-school structural relationships (β_{jk})', Z_{pj} are values on the school-level variables for unit j and U_{jk} is the random error. The above between-school model allows a simultaneous investigation of effects on school means and regression coefficients. Now, if we substitute (2) into (1) we get:

$$y_{ij} = \theta_{00} + \sum_{k=1} \theta_{0k} X_{ijk} + \sum_{p=1} \theta_p Z_{pj} + \sum_{k=1} \sum_{p=1} \theta_{pk} X_{ijk} Z_{pj} + U_{j0} + \sum_{k=1} U_{jk} X_{ijk} + R_{ij} \qquad (3)$$

where, the last three terms constitute the error term. If there are no random effects in the between-group model, then the hierarchical linear model is equivalent to an ordinary linear model. We can also set:

$$\beta_{j0} = \gamma_{00} + \gamma_{10} \overline{X}_{1j} + \gamma_{20} \overline{X}_{2j} + ... + \gamma_{K-1,j} + U_{j0}$$

$$\beta_{jk} = \beta_k, \quad \text{for } k = 1, ..., K-1$$

and thus, permitting estimation of both the pooled within-group slopes β_k and the corresponding between-group slopes γ_{k0}.

2.2.2 Applications of Hierarchical Linear Models in Educational Research

Raudenbush and Bryk (1986) used the hierarchical linear model mentioned before, to study the relative effectiveness of public and private (Catholic) schools. Coleman et al. (1982) made a similar study and they concluded that the overall academic achievement was higher in Catholic schools than in public schools and this was more intense for lower-SES students in Catholic schools. The data that was used came from the High School and Beyond (HSB) survey and the analysis sample consisted of 10,231 students from 82 Catholic schools and 94 public schools. The outcome variable is the standardized mathematics achievement score, while as explanatory variables are used the socioeconomic status (SES) of the students and the sector (1 for Catholic and 0 for public schools).

The first model that was applied examines the variability of the SES-achievement among schools. The within-school model is given by:

$$y_{ij} = \beta_{j0} + \beta_{j1}(X_{ij1} - \overline{X}_{.j1}) + R_{ij}$$

where y_{ij} is the mathematics achievement for student i in school j, β_{j0} is the mean mathematics achievement for school j, β_{j1} is the SES-achievement relationship in school j, X_{ij1} is the SES of student i in school j, $\overline{X}_{.j1}$ is the mean SES for school j and R_{ij} is the error term. The between-school model is given by:

$$\beta_{j0} = \theta_{00} + U_{j0},$$
$$\beta_{j1} = \theta_{10} + U_{j1},$$

where θ_{00} is the grand mean for mathematics achievement across all schools, θ_{10} is the mean slope for the SES pooled within all schools, U_{j0} is the effect of school j on the mean mathematics achievement and U_{j1} is the effect of school j on the SES. Having estimated the parameters of the above model the authors of the paper concluded that, ‘there is a positive relationship between SES and mathematics achievement within schools’. Furthermore, the means achievement levels vary across schools and the relationship between mathematics achievement and SES varies also across schools. It is interesting to compare the estimated parameter variance and the total observed variance. Although, of the total observed variability about 90% is parameter variance, only about 35% of the observed slope variability is parameter variance, while the remaining 65% is sampling variance. This is a striking example of the deficiencies of the slopes-as-outcomes model, presented in the previous section.

The second model that Raudenbush and Bryk (1986) applied for the analysis of the data identifies the variability among schools as a function of sector. Thus, the within-school model is the same as before, but now the between-school model includes the sector and we have:

$$\beta_{j0} = \theta_{00} + \theta_{01} Z_{j1} + U_{j0},$$
$$\beta_{j1} = \theta_{10} + \theta_{11} Z_{j1} + U_{j1},$$

where Z_{j1} is the sector (1 for Catholic and 0 for public schools), θ_{00} is the mean mathematics achievement in the public sector, θ_{10} is the mean slope for the SES in the public sector, θ_{01} is the Catholic school effect on means mathematics achievement, θ_{11} is the Catholic school effect on the SES slope, U_{j0} is the effect of school j on the mean mathematics achievement after accounting for the sector effects and U_{j1} is the effect of school j on the SES after accounting for the sector effects. From this second model the authors concluded that, ‘the relationship of SES to mathematics achievement is smaller in Catholic schools than in public schools’. Besides, the parameter variance estimates are now conditional variances and they measure the amount of variability remaining among the schools means and slopes after sector is taken into account. A natural estimator of the proportion of parameter variance explained by sector is given, for each β_{jk} (k=0, 1), by the following formula:

$$R^{2*} = 1 - \frac{\mathrm{var}(\beta_{jk}) - \mathrm{var}(\beta_{jk} | Z_1)}{\mathrm{var}(\beta_{jk})}.$$

The sector accounts for 11.3% of the parameter variance among school means and 71.6% of the variance in SES achievement slopes. Finally, although the parameter variance among school means has been reduced by taking sector into account, more variability remains to be explained.

The third model is a hierarchical linear model and accounts for the sector effects. There are two methods for adjusting the sector effects; the first is by controlling the student -level differences and the second is by controlling for student-level differences. It is assumed that the amount of homework is a confounding variable in the within-school model and that the SES composition of the school and its interaction with sector are confounding variables in the between-school model. The within-school model is now given by:

$$y_{ij} = \beta_{j0} + \beta_{j1}(X_{ij1} - \overline{X}_{.j1}) + \beta_{j2}(X_{ij2} - \overline{X}_{.j2}) + R_{ij}$$

where y_{ij} is the mathematics achievement for student i in school j, β_{j0} is the mean mathematics achievement in school j, β_{j1} and β_{j2} are the SES achievement and homework achievement relationships in school j, X_{ij1} is the SES of student i in school j, X_{ij2} is the amount of time spent on homework by student i in school j, $\overline{X}_{.j1}$ and $\overline{X}_{.j2}$ are the mean SES and mean hours of homework in school j and R_{ij} is the error term. The between-school model is given by:

$$\beta_{j0} = \theta_{00} + \theta_{01}Z_{j1} + \theta_{02}(Z_{j2} - \overline{Z}_{.2}) + \theta_{03}Z_{j1}(Z_{j2} - \overline{Z}_{.2}) + U_{j0},$$
$$\beta_{j1} = \theta_{01} + \theta_{11}Z_{j1} + \theta_{12}(Z_{j2} - \overline{Z}_{.2}) + \theta_{13}Z_{j1}(Z_{j2} - \overline{Z}_{.2}) + U_{j1},$$
$$\beta_{j2} = \theta_{20} + \theta_{21}Z_{j1} + \theta_{22}(Z_{j2} - \overline{Z}_{.2}) + \theta_{23}Z_{j1}(Z_{j2} - \overline{Z}_{.2}) + U_{j2},$$

where Z_{j1} is the sector, Z_{j2} is the school SES, θ_{00} is the mean mathematics achievement in the public sector, θ_{10} and θ_{20} are the mean slopes in the public sector for the SES achievement and homework achievement relationships, θ_{01} is the Catholic school effect on mean mathematics achievement slopes, θ_{02} is the effect of school SES on mean achievement, θ_{12} and θ_{22} are the effects of school SES on the SES achievement and homework achievement slopes, θ_{03}, θ_{13} and θ_{23} are the effects of the sector-by-school-SES interactions on mean achievement, the SES achievement slope and the homework achievement slope, and finally, U_{j0}, U_{j1} and U_{j2} are the remaining random effects associated with school j. From the analysis of this third model, it is concluded by Raudenbush and Bryk (1986) that, 'school SES is related to mean achievement', while, 'differences in social class composition between public and Catholic schools account for the observed mean mathematics achievement differences between the two sectors'. Furthermore, school SES affects the strength of the SES achievement relationship within schools. Also, according to Raudenbush and Bryk (1986), 'a student's social class has a stronger effect on individual achievement in higher social class schools than in schools with a larger proportion of less advantaged students'. Another important result is that if homework, school SES and the sector-by-school SES interaction are taken into account, then the effect of the Catholic school on the SES achievement relationship persists. On the other hand, the interactions of sector with school SES have no effect either on mean achievement or on the within-school relationships and thus these terms will be excluded from the model. Finally, with the inclusion of the school SES effect, no significant variation among slopes remains to be explained.

Another application of the hierarchical linear models on educational research was made by W. Raudenbush and J. Willms (1995). The authors of this paper specify and estimate the school effects. They also define two types of school effect. The first type refers to the effect of a particular policy on a student outcome, while the second one describes how the attendance of a particular school modifies a student's outcome.

In other words, the first type of school effects, the type A effect, is the difference between a given student's performance and the performance that would have been expected if that student had attended another school. The second type of school effects, the type B effect, is the difference between a student's performance in a particular school and the performance that would have been expected if that student had attended a school with the same context and with practice of average effectiveness.

Below, a statistical model is presented for school effects. According to this model, the outcome of a student depends on the effect of school practice, on the contribution of school context and on the student's background. An interaction effect between the school practice and the context is also possible. But for simplicity let us consider the following additive model:

$$Y_{ij} = \mu + P_{ij} + C_{ij} + S_{ij} + e_{ij}$$

Y_{ij}: the outcome for student i in school j

μ: grand mean of Y

P_{ij}: the effect of school practice (school resources, organizational structure, instructional leadership) on student i in school j

C_{ij}: the contribution of school context (mean socioeconomic level of school's students, unemployment rate of the community)

S_{ij}: the influence of measured student background variables (pre-entry aptitude, socioeconomic status)

e_{ij}: random error, independent of P, C, S, with zero mean and homogeneous variance $\sigma^2(e)$.

Furthermore, it is pretty obvious that a school has not a uniform effect on all who attend it. That is why we can include in the model the main effects of school-level variables as well as interactions between school and student-level variables. According to the above, we have that:

$P_{ij} = P_j + (PS)_{ij}$
$C_{ij} = C_j + (CS)_{ij}$
$S_{ij} = S_j + (S_{ij} - S_j)$

where $(PS)_{ij}$ and $(CS)_{ij}$ are the interactions of school practice and school context with the student background with zero means. In the above model the type A effect is this term:

$A_{ij} = P_{ij} + C_{ij}$

and with interactions

$A_{ij} = A_j + (AS)_{ij}$

where $A_j = Pj + Cj$ and $(AS)_{ij} = (PS)_{ij} + (CS)_{ij}$.

The type B effect is this term:

$B_{ij} = P_{ij}$

and with interactions

$B_{ij} = B_j + (BS)_{ij}$

where $B_j = Pj$ and $(BS)_{ij} = (PS)_{ij}$.

Another important aspect is the variation in school effects. For example, if the variance of type A effect were zero, then the choice of the school would not be important. On the other hand, if the variance of type A was large, then choosing among a set of schools would make a great difference. The same holds for the variance of the type B effect. The between-schools variances are given by the formula:

$$E(\mu_j - \mu)^2 = \tau(A) + \tau(S) + 2\mathrm{cov}(A,S)$$

and the within-schools variances are given by this formula:

$$E(\mu_j - \mu)^2 = \sigma^2(S) + \sigma^2(AS) + 2\mathrm{cov}(S,AS) + \sigma^2(e).$$

According to the above formulas, the school practice, the school context and the student background contribute to the between and within-school variation. It is also clear that both types of variation include a covariation term between school effects and student background.

The authors use two ways of estimating the type A effect. The first is by addition $A_{ij} = P_{ij} + C_{ij}$ and the second is by subtraction $A_{ij} = Y_{ij} - \mu - S_{ij}$. The first approach requires that all the variables have been measured and included in the additive model so that P_{ij} and C_{ij} are unbiased. The second one requires that S_{ij} is estimated without bias. In the same way, the type B effect can be estimated by addition and by subtraction. For estimating the type B effect by subtraction we use this formula, $B_{ij} = P_{ij} = y_{ij} - \mu - C_{ij} - S_{ij}$, which requires that C_{ij} and S_{ij} must be estimated without bias.

The example presented in the above mentioned paper, analyses data from 5,054 students from 20 secondary schools. The response variable is the fourth-year English outcome for each student and the explanatory variable is the primary 7 reading test score for each student. The following statistical models were used:

<u>Uniform Effects Model:</u> $Y_{ij} = \alpha + \beta_\omega X_{ij} + \gamma_c X_j + u_j + e_{ij}$

Y_{ij}: fourth year English outcome for student i in school j

X_{ij}: primary 7 reading test score

X_j: school sample mean reading test score for school j

u_j: random effect of school j, ~ N(0, τ)

e_{ij}: student-level random error, iid, ~ N(0, σ^2)

After an orthogonal reparamaterization the model takes the form:

$$Y_{ij} = \alpha + \beta_\omega (X_{ij} - X_j) + \beta_b X_j + u_j + e_{ij},$$

where Y_{ij}, u_j and e_{ij} are defined as above and X_{ij} - X_j is the student reading ability deviated around the school mean. An unbiased estimate of type A effect for each school is given by the formula

$$Aij = Y_j - \alpha - \beta_\omega X_j$$

If we assume that X and u are orthogonal then a consistent estimate of type B effect is given by

$$B_j = u_j = Y_j - \alpha - \beta_b X_j.$$

<u>Nonuniform Effects Model:</u> If we include in the model a randomly varying coefficient for students' reading test scores then the form of the model will be:

$$Y_{ij} = \alpha + \beta_\omega X_{ij} + \gamma_c X_j + u_{0j} + u_{1j} X_{ij} + e_{ij}.$$

In the above model the student background, the school context and the school practice are defined in the following way

$$S_{ij} = \beta_\omega X_{ij}$$

$C_j = \gamma_c X_j$

$P_{ij} = u_{0j} + u_{1j} X_{ij}$.

As we can see in this model P_{ij} varies from student to student within schools. The estimates for type A and type B effects are given by the formulas

$A_{ij} = Y_j - \alpha - \beta_\omega X_{ij}$ and $B_{ij} = A_{ij} - \gamma_c X_j$.

As biased estimators of school effects the means of each school of the pupil-level residuals from the first model of Aitkin and Longford (1986) can be used. Also, estimates can be produced based on residuals from means-on-means regression, on banding or via Empirical Bayes.