- Schooling systems present an obvious example of a hierarchical structure, with pupils grouped or nested or clustered within schools, which themselves may be clustered within education authorities or boards.
- Figure 1 shows the exam score and intake achievement scores for five students in a school, together with a simple regression line fitted to the data points. The residual variation in the exam scores about this line, is the *level 1 residual variation*, since it relates to level 1 units (students) within a sample level 2 unit (school). In figure 2 the three lines are the simple regression lines for three schools, with the individual student data points removed. These vary in both their slopes and their intercepts (where they would cross the exam axis), and this variation is *level 2 variation*. It is an example of multiple or complex level 2 variation since both the intercept and slope parameters vary.
- While such procedures usually have been regarded as necessary they have not generally merited serious substantive interest. In other words, the population structure, insofar as it is mirrored in the sampling design, is seen as a “nuisance factor”. By contrast, the multilevel modelling approach views the population structure as of potential interest in itself, so that a sample designed to reflect that structure is not merely a matter of saving costs as in traditional survey design, but can be used to collect and analyse data about the higher level units in the population. The subsequent modelling can then incorporate this information and obviate the need to carry out special adjustment procedures, which are built into the analysis model directly.
- Although the direct modelling of clustered data is statistically efficient, it will generally be important to **incorporate weightings in the analysis which reflect the sample design or, for example, patterns of non-response**, so that robust population estimates can be obtained and so that there will be some protection against serious model misspecification. A procedure for introducing external unit weights into a multilevel analysis is discussed in Chapter 3.
- There are particular problems arising when studying event duration data that are encountered when some information is ‘censored’ in the sense that instead of being able to observe the actual duration we only know that it is longer than some particular value, or in some cases less than a particular value. Chapter 9 will discuss ways of dealing with this issue for multilevel event duration models.
- An interesting special case of a 2-level model is the multivariate linear (or generalized linear) model. Suppose we have taken several measurements on an individual, for example their systolic and diastolic blood pressure and their heart rate. If we wish to analyze these together as response variables we can do so by setting up a multivariate,

- An interesting special case of a 2-level model is the multivariate linear (or generalised linear) model. Suppose we have taken several measurements on an individual, for example their systolic and diastolic blood pressure and their heart rate. If we wish to analyse these together as response variables we can do so by setting up a multivariate, in this case 3-variate, model with explanatory variables such as age, gender, social background, smoking exposure, etc.
- The multivariate multilevel model is also the basis for ways of dealing with **missing data** in multilevel models and this is developed in chapter 11.
- It is well known that when variables in statistical models contain relatively large components of such error the resulting statistical inferences can be very misleading unless careful adjustments are made (Fuller, 1987).
- The probability of a child being 'at risk' was estimated by the following (single level) equation logit( ) . . . . p x xx =− + + + 63 59 22 15 1 23
- In the pursuit of causal explanations we require some guiding underlying principles or theories. It is these which will tell us what kinds of things to measure and how to be critical of findings. For example, in studies of the relationship between perinatal mortality and maternal smoking in pregnancy (Goldstein, 1976) we can attempt to adjust for confounding factors, such as poverty, which may be responsible for influencing both smoking habits and mortality. We can also study how the relationship varies across groups and seek measures which explain such variation. we 10 might also, in some circumstances, be able to carry out randomised experiments, assigning for example intensive health education to a randomly selected 'treatment' group and comparing mortality rates with a 'control' group. A multilevel approach could be useful here in two different ways. First, pregnant women will grouped hierarchically, geographically and by medical institution and the between-area and between-institution variation may affect mortality and the relationship between mortality and smoking. Secondly, we will be able often to obtain serial measurements of smoking so allowing the kind of repeated measures 2-level modelling discussed earlier. This will allow us to study how changes in smoking are related to mortality, and permit a more detailed exploration of possible causal mechanisms.
- Finally, many of the concerns addressed by multilevel models are to do with prediction rather than causation.
- Multilevel models are tools to be used with care and understanding..

Consider first a simple model for one school, relating eleven-year-score to eight-year score. We write

$$y_i = \alpha + \beta x_i + e_i \tag{2.1}$$

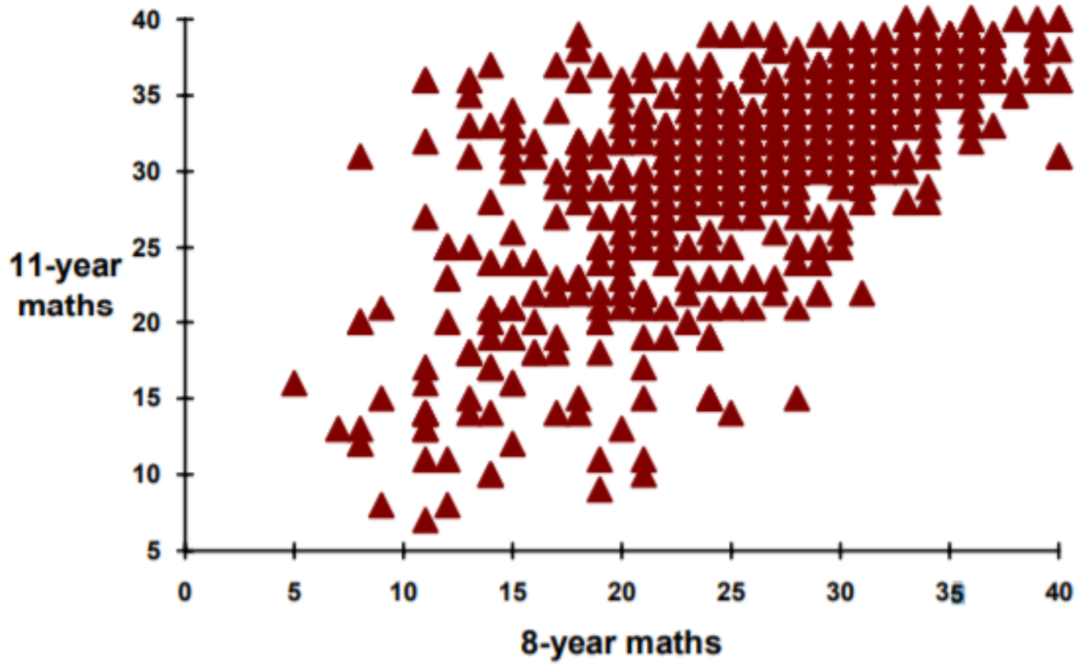

Fig. 2.1 Scatterplot of 11-year by 8-year mathematics test scores. Some points represent more than one child.

- where standard interpretations can be given to the intercept ( ) α , slope ( ) β and residual ( ) ei . We follow the normal convention of using Greek letters for the regression coefficients and place a circumflex over any coefficient (parameter) which is a sample estimate

$$\mathrm{cov}(u_{0j} + e_{0i_1j}, u_{0j} + e_{i_2j}) = \mathrm{cov}(u_{0j}, u_{0j}) = \sigma^2_{u0} \tag{2.5}$$

since the level 1 residuals are assumed to be independent. The correlation between two such students is therefore

$$\rho = \frac{\sigma^2_{u0}}{(\sigma^2_{u0} + \sigma^2_{e0})}$$

which is referred to as the 'intra-level-2-unit correlation'; in this case the intra-school correlation.[1] This correlation measures the proportion of the total variance which is between-schools. In a model with 3 levels, say with schools, classrooms and students, we will have two such correlations; the intra-school correlation measuring the proportion of variance that is between-schools and the intra-classroom correlation measuring that between classrooms.

The existence of a non-zero intra-unit correlation, resulting from the presence of more than one residual term in the model, means that traditional estimation procedures such as 'ordinary least squares' (OLS) which are used for example in multiple regression, are inapplicable. A later section illustrates how the application of OLS techniques leads to incorrect inferences. We now look in more detail at the structure of a 2-level data set, focusing on the covariance structure typified by Figure 2.3.

$$\begin{pmatrix} \sigma^2_{u0} + \sigma^2_{e0} & \sigma^2_{u0} & \sigma^2_{u0} \\ \sigma^2_{u0} & \sigma^2_{u0} + \sigma^2_{e0} & \sigma^2_{u0} \\ \sigma^2_{u0} & \sigma^2_{u0} & \sigma^2_{u0} + \sigma^2_{e0} \end{pmatrix}$$

Figure 2.3 Covariance matrix of three students in a single school for a variance components model.

The matrix in figure 2.3 is the (3 x 3) covariance matrix for the scores of three students in a single school, derived from the above expressions. For two schools, one with three students and one with two, the overall covariance matrix is shown in Figure 2.4. This 'block-diagonal' structure reflects the fact that the covariance between students in different schools is zero, and clearly extends to any number of level 2 units.

$$\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$$

where

- Covariance matrix
- These residuals therefore can have two roles. Their basic interpretation is as random variables with a distribution whose parameter values tell us about the variation among the level 2 units, and which provide efficient estimates for the fixed coefficients. A second interpretation is as individual estimates for each level 2 unit where we use the assumption that they belong to a population of units to predict their values
- The level 1 residuals are generally not of interest in their own right but are used rather for model checking, having first been standardised using the diagnostic standard errors..

Table 2.1 Variance components model applied to JSP data

Parameter	Estimate (s.e.)	OLS Estimate (s.e.)
Fixed:		
Constant	13.9	13.8
8-year score	0.65 (0.025)	0.65 (0.026)
Random:		
σ^2_{u0} (between schools)	3.19 (1.0)	
σ^2_{e0} (between students)	19.8 (1.1)	23.3 (1.2)
Intra-school correlation	0.14	

Comparing the OLS with the multilevel estimates we see that the fixed coefficients are similar, but that there is an intra-school correlation of 0.14. The estimate of the standard error of the between school variance is less than a third of the variance estimate, suggesting a value highly significantly different from zero. This comparison, however, should be treated cautiously, since the variance estimate does not have a Normal distribution and the standard error is only estimated, although the size of the sample here will make the latter caveat less important. It is generally preferable to carry out a likelihood ratio test by estimating the 'deviance' for the current model and the model omitting the level 2 variance (see McCullagh and Nelder, 1989). The next section will deal more generally with inference procedures. The deviances are, respectively, 4294.2 and 4357.3 with a difference of 63.1 which is referred to tables of the chi-squared distribution with one degree of freedom, and is highly significant. Note that if we use the standard error estimate given in Table 2.1 to judge significance we obtain the corresponding value of $(3.19/1.0)^2 = 10.2$ which is very much smaller than the likelihood ratio test statistic.

We elaborate the model first by adding two more explanatory variables, gender and social class. The results are set out in the first column of table 2.2

Table 2.2 Variance components model applied to JSP data with gender and social class

Parameter	Estimate (s.e.)	Estimate (s.e.)
Fixed:		
Constant	14.9	32.9
8-year score	0.64 (0.025)	
Gender (boys - girls)	-0.36 (0.34)	-0.39 (0.47)
Social Class (Non Man. - Manual)	0.72 (0.39)	2.93 (0.51)
Random:		
σ^2_{u0} (between schools)	3.21 (1.0)	4.52 (1.5)
σ^2_{e0} (between students)	19.6 (1.1)	37.2 (2.0)
Intra-school correlation	0.14	0.11

- The random parameter estimates are hardly changed, nor is the coefficient of the 8-year maths score. The gender difference is very small and in favour of the girls, but is far from the conventional 5% significance 13 level. The social class difference favours the children of non-manual parents. When we are judging the fixed effects, a simple comparison of the estimate with its standard error is usually adequate. Because the model adjusts for the earlier maths score we can interpret the social class and gender differences is in terms of the relative progress of girls versus boys or non-manual versus manual children. The second column in table 2.2 shows the effects when 8-year maths score is removed from the model and the interpretation is now in terms of the actual differences found at 11 years. Note that the level 1 and level 2 variances are increased, reflecting the importance of the earlier score as a predictor, and the intra-school correlation is slightly reduced. The social class difference is much larger, suggesting that most of the difference is that existing at 8 years with a somewhat greater progress made between 8 and 11 years by those in the non-manual social group. The gender difference remains small. The 8-year score has been used as it stands, without centring it in any way. This is acceptable in the present case, although the strict interpretation of the intercept is the predicted score at an 8 year score of zero, which is outside the range of the observed values. If we were to measure the 8-year-score from its mean, the intercept would be interpreted as the predicted value at the mean 8-year-score. When we introduce random coefficients in chapter 3 we shall see that this becomes an important consideration.

2.9.1 Checking model assumptions

We now check some assumptions of the model by looking at the residuals. Figure 2.7 is a plot of the standardised level 1 residuals against the fixed part predicted value and figure 2.8 is a plot of these residuals against their equivalent Normal scores. Figure 2.7 shows the same pattern as figure 2.1 of a decreasing variance with increasing 8-year score, so that the assumption of a constant level 1 variance is clearly untenable.

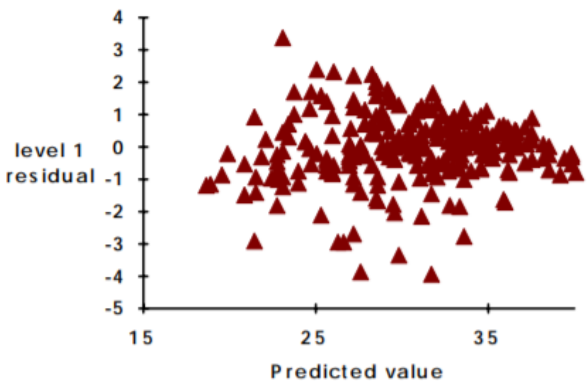

Figure 2.7 Standardised level 1 residuals by predicted values for Table 2.2

In chapter 3 we shall be looking at ways to deal with this. The Normal score plots, on the other hand, are fairly straight, suggesting that the Normal distribution assumption is reasonable for both level 1 and level 2.

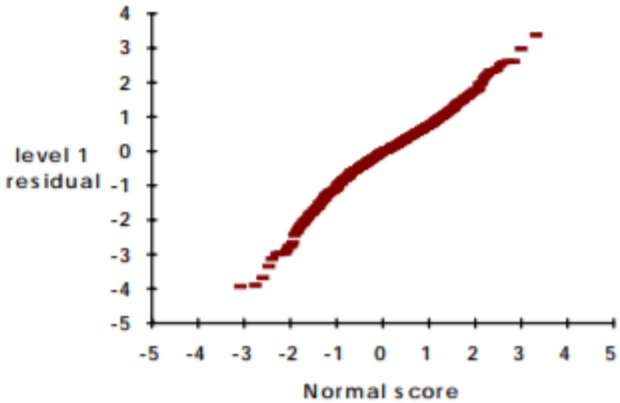

Figure 2.8 Standardised level 1 residuals by Normal equivalent scores for Table 2.2

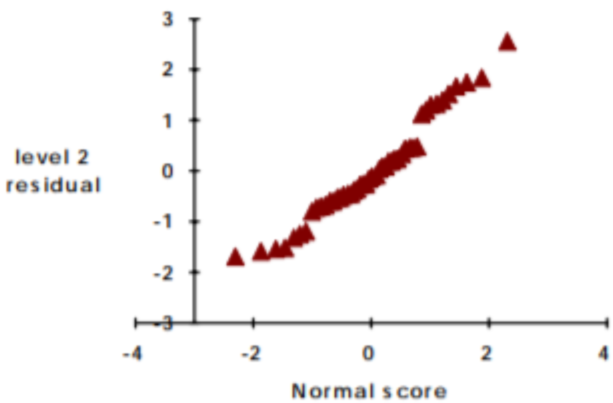

Figure 2.9 Standardised level 2 residuals by Normal equivalent scores for Table 2.2

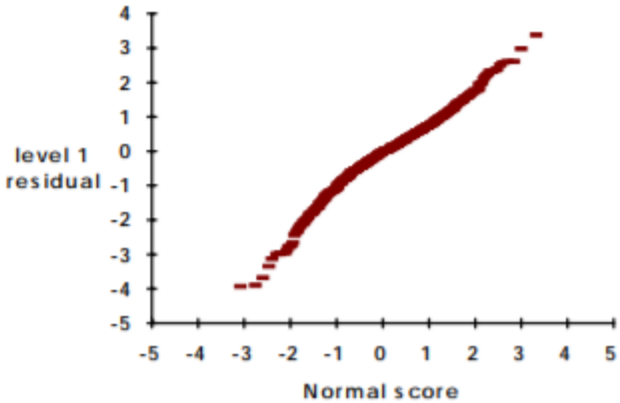

Figure 2.8 Standardised level 1 residuals by Normal equivalent scores for Table 2.2

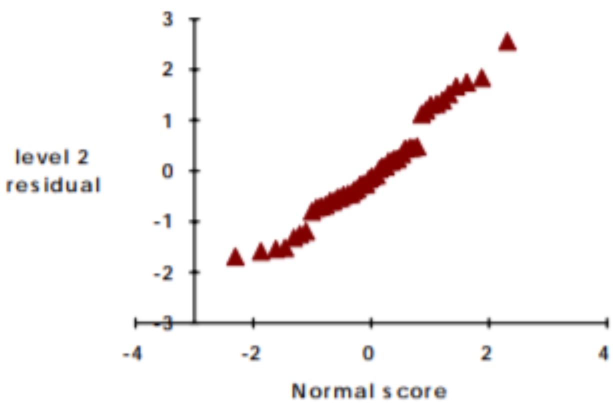

Figure 2.9 Standardised level 2 residuals by Normal equivalent scores for Table 2.2

Table 2.4 Variance components model for JSP data with mean 8-year score measured about sample mean and centring about school mean.

Parameter	Estimate (s.e.) A	Estimate (s.e.) B	Estimate (s.e.) C
Fixed:			
Constant	31.5	31.5	31.7
8-year score	0.64 (0.025)		0.63 (0.025)
8-year score centred on school mean		0.64 (0.026)	
Gender (boys - girls)	-0.36 (0.34)	-0.36 (0.34)	-0.37 (0.34)
Social Class (Non Man. - Manual)	0.72 (0.38)	0.72 (0.31)	0.79 (0.31)
School mean 8-year score	-0.01 (0.13)	0.63 (0.12)	-0.03 (0.12)
8-year score x school mean 8-year score			-0.02 (0.01)
Random:			
σ^2_{u0} (between schools)	3.21 (1.0)	3.21 (1.0)	3.13 (1.0)
σ^2_{e0} (between students)	19.6 (1.1)	19.6 (1.0)	19.5 (1.1)
Intra-school correlation	0.14	0.14	0.14

In fact, the mean score for students in a school is only one particular summary statistic describing the composition of the students. Another summary would be the spread of scores, measured for example by their standard deviation. We can also consider measures such as the proportions of high or low scoring students and in general any set of such measures. When using the average score we can also consider using the median or modal score rather than the mean. With any of these other measures we may wish to retain the deviation from the school mean as an explanatory variable, and we could even consider introducing a more complex function of this, for example by adding higher order terms. There is here a fruitful area for further study.

Analysis C looks at the possibility of an interaction between student score and school mean and we do find a significant effect which we can interpret as follows. The higher the school mean 8-year score the lower the coefficient of the student's 8-year score. One implication of this is that for two relatively low scoring student's at 8 years, the one in the school with a higher average is predicted to do better at 11 years. To study this further we now need to introduce a model with random coefficients where we explicitly allow each school's coefficient to vary randomly at level 2, as in equation (2.6), see Table 2.5.

The addition of the 8-year score coefficient as a random variable at level 2 somewhat increases the social class difference and somewhat decreases the gender difference, but within their standard errors. The level 1 variance is reduced and we have significant 'slope' variation at level 2; the likelihood ratio test criterion is 52.4 which is referred to chi squared tables with 2 degrees of freedom and is highly significant.

Table 2.5 Random coefficient model for JSP data.

Parameter	Estimate (s.e.)
Fixed:	
Constant	31.7
8-year score	0.62 (0.036)
Gender (boys - girls)	-0.25 (0.32)
Social Class (Non Man. - Manual)	0.96 (0.36)
School mean 8-year score	-0.04 (0.13)
8-year score x school mean 8-year score	-0.02 (0.01)
Random:	
Level 2	
σ^2_{u0} (Intercept)	3.67 (1.03)
σ_{u01} (covariance)	-0.34 (0.09)
σ^2_{u1} (8-year score)	0.03 (0.01)
Level 1	
σ^2_{e0}	17.8 (1.0)

If we calculate the correlation between the intercept and slope at level 2 we obtain a value of -1.03! This sometimes happens as a result of sampling variation and implies that the population correlation is very high. We shall see in chapter 3 we can constrain this correlation to be exactly -1.0 and thus admissible. Alternatively, by suitably elaborating the model or by carrying out certain transformations we can avoid this problem. For now, however, in order to illustrate what this means in the present data we can compute residuals for each school, for the slope and intercept. With these estimates we can then predict the 11-year score for any set of values of the explanatory variables. Figure 2.10 shows the predicted values for manual girls by 8-year score.

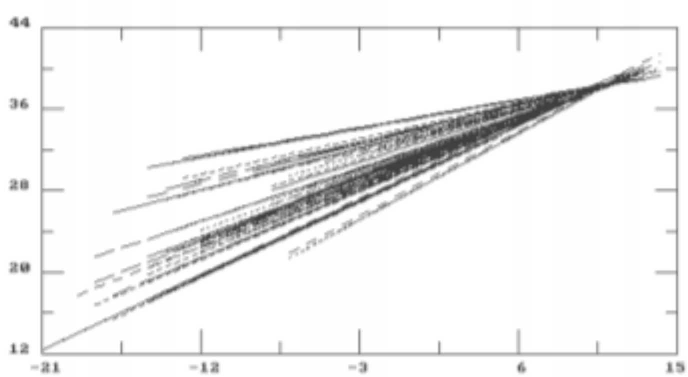

Figure 2.10 Plot of predicted 11-year score by 8-year score for JSP schools

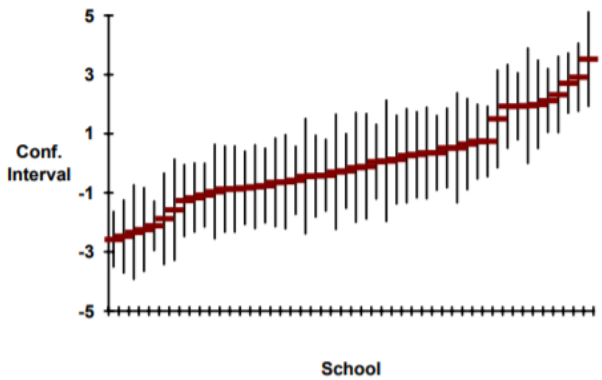

Fig 2.12 Simultaneous confidence intervals for JSP school residuals

Presentations such as that in 2.12 are useful for conveying the inherent uncertainty associated with estimates for individual level 2 (or higher) units, where the number of level 1 units per higher level unit is not large. This uncertainty in turn places inherent limitations upon such comparisons.

- **EXPECTATION MAXIMIZATION (EM) ALGORITH**
- Markov Chain Monte Carlo estimation
- Variation model
- Bootstrapping for multilevel model

4.2 The basic 2-level multivariate model

To define a multivariate, in the case of our example a 2-variate, model we treat the individual student as a level 2 unit and the 'within-student' measurements as level 1 units. Each level 1 measurement 'record' has a response, which is either the written paper score or the coursework score. The basic explanatory variables are a set of dummy variables that indicate which response variable is present. Further explanatory variables are defined by multiplying these dummy variables

Table 4.1 Data matrix for examination example.

		Intercepts		Gender	
Student	Response	Written	Coursework	Written	Coursework
1 (female)	y_{11}	1	0	1	0
1	y_{21}	0	1	0	1
2 (male)	y_{12}	1	0	0	0
2	y_{22}	0	1	0	0
3 (female)	y_{13}	1	0	1	0

by individual level explanatory variables, for example gender. The data matrix for three individuals, two of who have both measurements and the third of who has only

the written paper score, is displayed in Table 4.1. The first and third students are female (1) and the second is male (0).

The model is written as

$$y_{ij} = \beta_{01}z_{1ij} + \beta_{02}z_{2ij} + \beta_{11}z_{1ij}x_j + \beta_{12}z_{2ij}x_j + u_{1j}z_{1ij} + u_{2j}z_{2ij}$$

$$z_{1ij} = \begin{Bmatrix} 1 \text{ if written} \\ 0 \text{ if coursework} \end{Bmatrix}, \quad z_{2ij} = 1 - z_{1ij}, \quad x_j = \begin{Bmatrix} 1 \text{ if female} \\ 0 \text{ if male} \end{Bmatrix} \tag{4.1}$$

$$\text{var}(u_{1j}) = \sigma_{u1}^2, \quad \text{var}(u_{2j}) = \sigma_{u2}^2, \quad \text{cov}(u_{1j}u_{2j}) = \sigma_{u12}$$

- There are several features of this model. There is no level 1 variation specified because level 1 exists solely to define the multivariate structure. The level 2 variances and covariance are the (residual) between-student variances. In the case where only the intercept dummy variables are fitted, and since every student has both scores, the model estimates of these parameters become the usual between-student estimates of the variances and covariance. The multilevel estimates are statistically efficient even where some responses are missing, and in the case where the measurements have a multivariate Normal distribution they are maximum likelihood. Thus the formulation as a 2-level model allows for the efficient estimation of a covariance matrix with missing responses. In our example the students are grouped within examination centres, so that the centre is the level 3 unit. Table 4.2 presents the results of two models fitted to these data. The first analysis is simply (4.1) with variances and a covariance for the two components added at level 3. In the second analysis additional variance terms for gender have been added. In both analyses the females do worse on the written paper and better on the coursework assessment. There is a greater variability of marks on the coursework element, even though this is marked out of a smaller total, and the intra-centre correlations are approximately the same in the first analysis (0.28 and 0.30). This suggests that the 'moderation' process has been successful in maintaining a similar relative between-centre variation for the coursework marks. The correlation between the two components is 0.50 at the student level and 0.41 at the centre level.

Chapter 5

Nonlinear multilevel models

5.1 Nonlinear models

The models of Chapters 1-4 are linear in the sense that the response is a linear function of the parameters in the fixed part and the elements of are linear functions of the parameters in the random part. In many applications, however, it is appropriate to consider models where the fixed or random parts of the model, or both, contain nonlinear functions. For example, in the study of growth, Jenss and Bayley (1937) proposed the following function to describe the growth in height of young children

$$y_{ij} = \alpha_0 + \alpha_1 t_{ij} + u_{\alpha 0 j} + u_{\alpha 1 j} t_{ij} + e_{\alpha ij} - \exp(\beta_0 + \beta_1 t_{ij} + u_{\beta 0 j} + u_{\beta 1 j} t_{ij} + e_{\beta ij}) \qquad (5.1)$$

where t_{ij} is the age of the j-th child at the i-th measurement occasion. Generalised linear models (McCullagh and Nelder, 1989) are a special case of nonlinear models where the response is a nonlinear function of a fixed part linear predictor. Models for discrete data, such as counts or proportions fall into this category and we shall devote chapter 7 to studying these. For example, a 2-level log linear model can be written

$$E(m_{ij}) = \pi_{ij}, \quad \pi_{ij} = \exp(X_{ij}\beta_j) \qquad (5.2)$$

where m_{ij} is assumed typically to have a Poisson distribution, in this case across level 1 units. Note here, that in the multilevel extension of the standard single level model, the linear predictor contains random variables defined at level 2 or above.

In this chapter we consider a general nonlinear model. Later chapters will use the results for particular applications.

-
- <u>Generalised linear models (McCullagh and Nelder, 1989) are a special case of nonlinear models where the response is a nonlinear function of a fixed part linear predictor.</u>

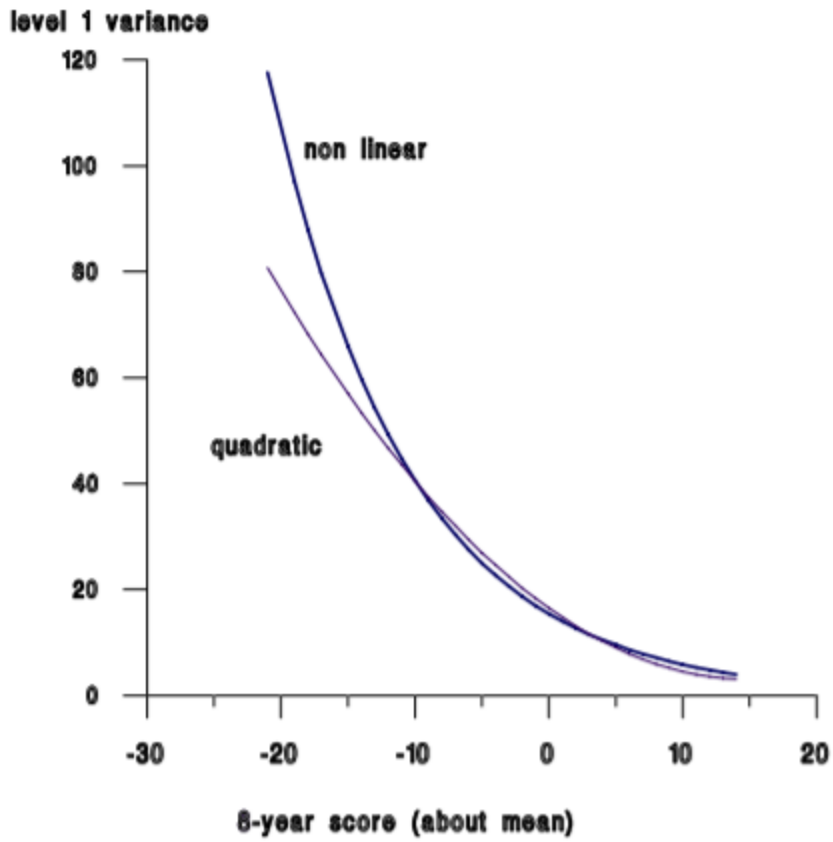

Fig 5.1 Level 1 variance as a function of 8-year Maths score

Table 5.2 Nonlinear model estimates with first order fixed part prediction. Age is measured about 8.0 years.

Fixed coefficient	**Estimate (s.e.)**
Intercept (linear)	90.3
Intercept (nonlinear)	3.58
Age	0.15 (0.10)
Age squared	-0.016 (0.02)
Age cubed	0.002 (0.004)

Nonlinear model level 2 covariance matrix (s.e.)

	Intercept	Age
Intercept	0.025 (0.003)	
Age	-0.0027 (0.0003)	0.00036 (0.00005)

Level 1 variance = 0.25

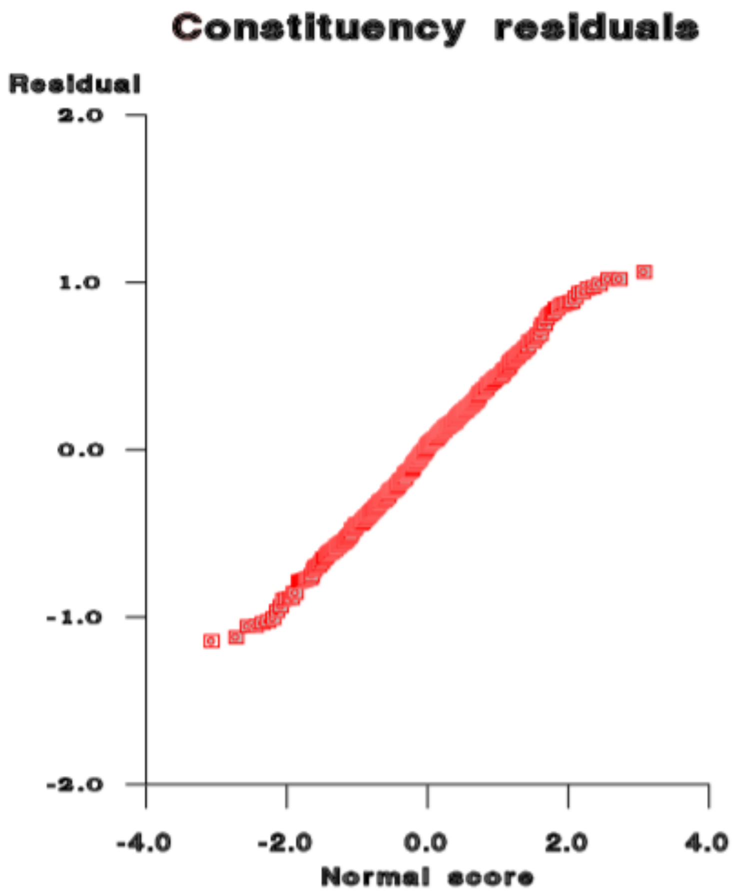

Figure 7.1 Normal residual plot for constituencies: Analysis A in Table 7.3

- <u>Instead of using a set of proportions as the response we can consider the underlying event counts as the set of responses.</u>

Appendix 7.1

Differentials for some discrete response models

The Logit - Binomial model

$$f = [1 + \exp(-X\beta)]^{-1}$$
$$f' = f[1 + \exp(X\beta)]^{-1}$$
$$f'' = f'[1 - \exp(X\beta)][1 + \exp(X\beta)]^{-1}$$

The Logit - Multinomial (Multivariate Logit) model

$$f^{(s)} = \exp(X\beta^{(s)})[1 + \sum_{h=1}^{t-1} \exp(X\beta^{(h)})]^{-1}, \quad s = 1,\dots,t-1$$
$$f'^{(s)} = f^{(s)}[1 + \sum_{h \neq s} \exp(X\beta^{(h)})][1 + \sum_{h=1}^{t-1} \exp(X\beta^{(h)})]^{-1}$$
$$f''^{(s)} = f'^{(s)}[1 - 2f^{(s)}]$$

The Log - Poisson model

$$f = \exp(X\beta)$$
$$f' = \exp(X\beta)$$
$$f'' = \exp(X\beta)$$

The log log - Binomial model

$$f = 1 - \exp[-\exp(X\beta)]$$
$$f' = (1 - f)\exp(X\beta)$$
$$f'' = f'[1 - \exp(X\beta)]$$

•

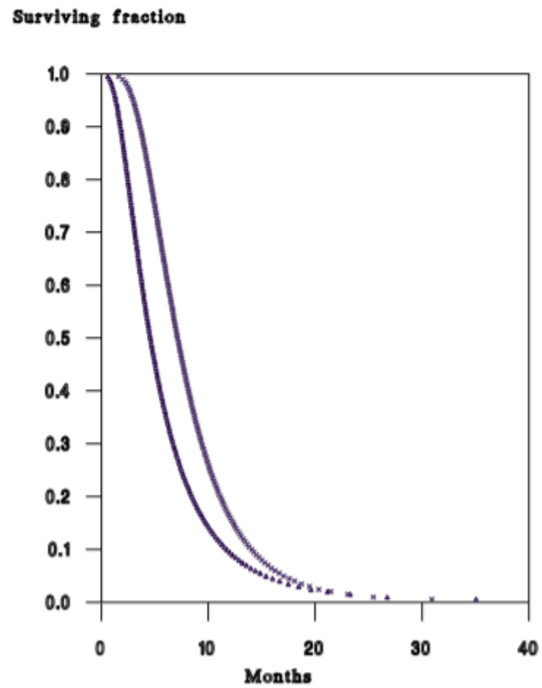

Figure 9.3 Estimated survival functions for women with previous live births (upper) and a previous death; born in 1900, age 20, 12 months marriage.

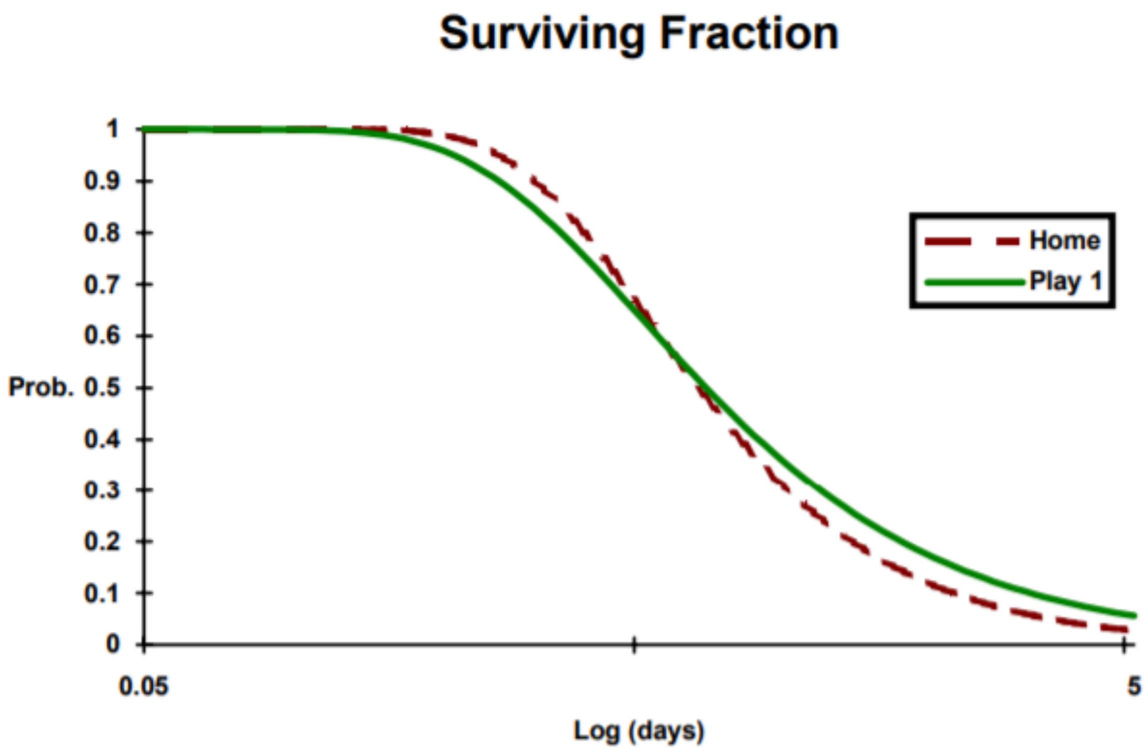

figure 9.4 Estimated surviving probability of play episodes.

- 11.2 Design issues When designing a study where the multilevel nested structure of a population is to be modelled, the allocation of level 1 units among level 2 units and the allocation of these among level 3 units etc. will clearly affect the precision of the resulting estimates of both the fixed and random parameters. The situation becomes more complex when there are random cross classifications and where there are several random coefficients. There are generally differential costs associated with sampling more level 1 units within an existing level 2 unit as opposed to selecting further level 1 units in a new level 2 unit. At the present time there appears to be little empirical or theoretical work on issues of optimum design for multilevel models.. Some approximations for studying the standard errors of the fixed coefficients have been derived by Snijders and Bosker (1993) in the case of a simple 2-level variance components model. They are concerned with students sampled within schools and assume that the cost of selecting a student in a new school is a fixed constant times the cost of selecting a student in an already selected school. They also assume that there is a minimum of 11 students per school. They tend to find that, where this constant is greater than 1 and the total number of students to be sampled is fixed, the sample of schools should be as large as possible, although this will not necessarily be true for all the coefficients of interest. Where cost information is available, together with some idea of parameter values, perhaps from a pilot study, then a guide to design efficiency can be obtained by simulating the effect of different design strategies and studying the resulting characteristics of the parameter estimates, such as their mean squared errors. This will be time consuming however, since for each design a number of simulated samples will be required. On the other hand, in certain areas, such as that of school effectiveness or animal and human growth studies, where information about costs and parameter values is often available, it would be possible to derive some generally useful results.

- Chapter 11 Software, missing data and structural equation models 11.1 Software for multilevel analysis **Traditionally, statistical analysis packages for the analysis of linear or generalised linear models have assumed a single level model with a single random variable. For the models described in this book such software packages are clearly inadequate, and this led, in the mid 1980's, to the development of four special purpose packages for fitting multilevel models.** One of these, GENMOD (Mason et al., 1988), is no longer generally available. The other three are HLM (Bryk et al., 1988), MLwiN (Rasbash et al., 1999) and VARCL (Longford, 1988). A detailed review of these four packages (including ML3 which subsequently became MLn and then MLwiN) has been carried out by Kreft et al (1994). In their original form HLM, ML3 and VARCL were designed for continuous Normally distributed response variables and all three produced maximum likelihood (ML) or restricted maximum likelihood (REML) estimates. All three were soon able to fit 3-level models and VARCL and ML3 developed procedures for fitting Binomial and Poisson response models using the first order marginal approximation described in chapter 5. In addition VARCL is able to fit a variance components model with up to nine levels. Subsequently, the major statistical packages, notably BMDP, SAS and GENSTAT, have begun to incorporate procedures for ML and REML estimation for Normal response models. The packages EGRET and SABRE will obtain ML estimates for a 2-level logit response model. A Bayesian package using Markov Chain Monte Carlo (MCMC) estimation, BUGS, is also available and MLwiN allows MCMC estimation for a range of models. Appendix 11.1 contains details of where these and other programs can be obtained. The two packages, MLwiN and BUGS, are able to fit nearly all the models described in this book, although not currently structural equation models. These latter models can be fitted by the program BIRAM, listed in Appendix 11.1. The programs Mln and MLwiN allow an effectively unlimited number of levels to be fitted, together with case weights, measurement errors and robust estimates of standard errors. They also have a high level MACRO language which will allow a wide range of special purpose facilities to be incorporated. A number of the papers referenced in earlier chapters have carried out their estimation procedures using special purpose software written in statistical programming languages such as S-Plus or Gauss. For the most part, however, this approach is computationally inefficient for the analysis of large and complex data sets, and the use of one of the special purpose packages is then essential, even when powerful mainframe computers are used. The general purpose packages, SAS, GENSTAT and also MLwiN allow a wide variety of data manipulations to be carried out within the software whereas the others tend to demand a somewhat rigid data format with limited possibilities for data transformations etc. It is reasonable to expect that the standard multilevel models will soon be available within most of the major general purpose statistical packages. For the more complex models, such as those with multivariate outcomes, nonlinear relationships and complex variation at all levels, it will be important to have a user interface which assists understanding the complexity of structure when specifying models and when interpreting output. Because the level of complexity of multilevel models is greater than that associated with single level linear or generalised linear models, the importance of helpful user interfaces cannot be overemphasised if the best use is to be made of these models. The ability to work interactively in a graphical environment will also be important and it will be necessary for programs to optimise computations so that very large and complex datasets can be handled within a reasonable time (Goldstein and Rasbash, 1992).

- 11.2 Design issues **When designing a study where the multilevel nested structure of a population is to be modelled, the allocation of level 1 units among level 2 units and the allocation of these among level 3 units etc. will clearly affect the precision of the resulting estimates of both the fixed and random parameters.** The situation becomes more complex when there are random cross classifications and where there are several random coefficients. There are generally differential costs associated with sampling more level 1 units within an existing level 2 unit as opposed to selecting further level 1 units in a new level 2 unit. At the present time there appears to be little empirical or theoretical work on issues of optimum design for multilevel models.. Some approximations for studying the standard errors of the fixed coefficients have been derived by Snijders and Bosker (1993) in the case of a simple 2-level variance components model. They are concerned with students sampled within schools and assume that the cost of selecting a student in a new school is a fixed constant times the cost of selecting a student in an already selected school. They also assume that there is a minimum of 11 students per school. They tend to find that, where this constant is greater than 1 and the total number of students to be sampled is fixed, the sample of schools should be as large as possible, although this will not necessarily be true for all the coefficients of interest. Where cost information is available, together with some idea of parameter values, perhaps from a pilot study, then a guide to design efficiency can be obtained by simulating the effect of different design strategies and studying the resulting characteristics of the parameter estimates, such as their mean squared errors. This will be time consuming however, since for each design a number of simulated samples will be required. On the other hand, in certain areas, such as that of school effectiveness or animal and human growth studies, where information about costs and parameter values is often available, it would be possible to derive some generally useful results.

- 11.3 Missing data A characteristic of most large scale studies is that some of the intended measurements are unavailable. In surveys, for example, this may occur through chance or because certain questions are unanswered by particular groups of respondents. We are concerned with missing values of explanatory variables in a multilevel model. **An important distinction is made between situations where the existence of a missing data item can be considered a random event and where it is informative and the result of a non random mechanism.** Randomly missing data may be missing 'completely at random' or 'at random' conditionally on the values of other measurements. The following exposition will be concerned with these two types of random event. Where data cannot be assumed to be missing at random, one approach is to attempt to model the missingness mechanism, and then to predict values from this model. Such predictions can be treated in similar ways to those described below. We consider the problem of missing data in two parts. First we develop a procedure for predicting data values which are missing and then we study ways of obtaining model parameter estimates from the resulting 'filled-in' or 'completed' data set. The prediction will use those measurements which are available, so that data values which are missing at random conditional on these measurements can be incorporated. Detailed discussions of missing data procedures are given by Rubin (1987) and Little (1992). The basic exposition will be in terms of a single level model for simplicity, pointing out the extensions for multilevel models.