# CHAPTER 11

# Multilevel and Multivariate Models in Survey Analysis

H. Goldstein and R. Silver

## 11.1 INTRODUCTION

Recent developments in the theory of linear model estimation (Aitkin and Longford, 1986; Goldstein, 1986a) have made possible the specification, efficient estimation and testing of models fitted to data obtained from nested or hierarchical structures. A good example of a hierarchical structure is an educational system where students are 'clustered' or grouped within classes, classes are grouped within schools, schools within local boards or authorities and so forth. Other kinds of data too can be viewed in this manner, most notably repeated measurement longitudinal data, which are an example of two-level data with the first or lower level comprising the measurement occasions 'within' individual subjects, and the second or higher level comprising the subjects themselves (see Goldstein, 1986b). Of course, the individual subjects may themselves be grouped within classes, schools, etc., so giving rise to a three-level or higher-level structure. It is important to emphasize that these hierarchies are intrinsic properties of the systems being studied, and that the use of statistical models to describe these structures is motivated by the structures themselves, independently of any sampling procedure which generates the data. Thus, a sample may be drawn from a school population by a simple random procedure, but it will still generally need to be modelled with due attention paid to the structure of the population itself.

From this viewpoint a multi-stage sampling procedure may be important for providing a valid and efficient analysis, not merely to reduce costs. Thus, if we wish to obtain stable estimates of within-school variation we need sufficiently large numbers of children within individual schools. A simple random sample which produced, say, just one child per school on average would be unsuitable. Further discussion of this point is given in Chapter 12.

The same considerations apply to most social data. Societies tend to have inbuilt hierarchies, for example, individuals within households which are grouped within localities and so on. In this chapter our concern is with a model-based approach with inferences to an assumed superpopulation, and our models are devised explicitly to incorporate the hierarchical structure of the population. In particular, these models can incorporate measurements made, say, at the level of the cluster, and if clusters are geographical areas, we can think of using variables such as the average social composition of the area or its general amenities. Thus the emphasis is on describing the between-unit variability at each level of hierarchical aggregation. This contrasts with the more usual emphasis in survey analysis where this variation is regarded as constituting a 'nuisance' since interest centres on relationships amongst the lowest level units—typically individual subjects.

Because so-called 'multilevel' models have been applied most extensively to educational data, the principal exposition will be in terms of educational variables. Nevertheless, as our second example illustrates, the methods of analysis will apply to data collected from other hierarchical systems, and to social surveys in particular. To introduce the basic idea of multilevel models, consider a simple two-level structure with students grouped within classrooms. Suppose there is a response variable, say a mathematics achievement test score ($y$) measured on each student, which we wish to relate to the gender ($x_1$) and social background ($x_2$) of each student, and to the average social background of all the students in each class ($a_1$) and years of mathematics teaching experience of the classroom teacher ($a_2$). Note that there are two explanatory variables measured at the student level, and two at the classroom level, one of which is a characteristic of the teacher and one is a so-called 'contextual' variable, based on an aggregated characteristic of all the students in the class. Note that the values of the aggregated variables may be available even though the sample itself does not include all the children in the class.

The next section deals with the basic statistical model. The approach is informal, rather than statistically rigorous, and uses a simple notation and terminology. The full statistical details of how to obtain generalized least squares and maximum likelihood estimates can be found in Goldstein (1986a). We deal first with the case of a continuously distributed response variable, and in the following section with discrete response variables.

## 11.2 THE TWO-LEVEL MODEL

The 'fixed' part of the model mentioned above can be written as

$$y_{ct} = \beta_0 x_0 + \beta_1 x_{1ct} + \beta_2 x_{2ct} + \gamma_{10} a_{1c} + \gamma_{20} a_{2c}, \qquad (11.1)$$

where the $c$ subscript indexes classes and the $t$ subscript indexes students within



classes. The coefficients $\beta_0, \beta_1, \beta_2, \gamma_{10}$ and $\gamma_{20}$ are those which we want to estimate. In this example, the variables $x_1$ and $x_2$ are categorical and are defined using 'indicator' variables to denote the levels or categories, and the variables $a_1$ and $a_2$ are basically continuous so that $\gamma_{10}$ and $\gamma_{20}$ can be interpreted in the usual way as regression coefficients. The variable $x_0$ is in fact a constant which is set to 1.0 so that $\beta_0$ becomes the 'overall constant' or 'intercept' term in the model.

Turning to the 'residual' terms, these are random variables with an assumed mean of zero. We start by defining two of them, motivated as follows. Suppose $\beta_0$ varies randomly across individuals and classrooms. That is, for a fixed set of values for $x_1, x_2, a_1, a_2$, assuming we know (or that we have estimates of) $\beta_1, \beta_2, \gamma_{10}, \gamma_{20}$ then we rewrite it with extra subscripts as $\beta_{0ct}$ which varies from student to student with a mean value for the $c$th class of $\beta_{0c}$. We write:

$$\beta_{0c} = \beta_0 + \eta_{0c} \qquad (11.2)$$

and

$$\beta_{0ct} = \beta_0 + \eta_{0c} + \varepsilon_{ct}, \qquad (11.3)$$

where $\eta_{0c}$ is the deviation of the $c$th class from the overall mean and where $\varepsilon_{ct}$ is the deviation of the $t$th student from the mean of the $c$th class, with variances $\sigma_0^2$ and $\sigma^2$ respectively. The $\eta_c$ are mutually independent and so are the $\varepsilon_{ct}$ and they are independent of each other.

Thus, remembering that the variable $x_0 = 1$ we can write:

$$\begin{aligned} y_{ct} &= \beta_{0ct} + \beta_1 x_{1ct} + \beta_2 x_{2ct} + \gamma_{10} a_{1c} + \gamma_{20} a_{2c} \\ &= \beta_0 x_0 + \sum_{k=1}^{2} \beta_k x_{kct} + \sum_{k=1}^{2} \gamma_{k0} a_{kc} + (\eta_{0c} + \varepsilon_{ct}). \end{aligned} \qquad (11.4)$$

The term is brackets is the random part of the model, and we need to estimate the two parameters associated with it, namely $\sigma_0^2, \sigma^2$.

It will be seen that equation (11.4) is the usual form for the linear model, but with the additional random term $\eta_{0c}$. It is the presence of two random terms that requires special estimation procedures and means that ordinary regression techniques cannot be applied (unless of course $\sigma_0^2$ is zero or negligibly small).

We can extend this model to include 'interactions' between level 1 and level 2 variables by allowing, say, $\beta_1$ to be a function of level 2 variables, and writing:

$$\beta_{1ct} = \sum_{k=1}^{2} \gamma_{k1} a_{kc} + \eta_{1c} + \varepsilon_{1ct}. \qquad (11.5)$$

This allows $\beta_1$ to vary at both level 1 and level 2, and in general any coefficient can vary randomly at any level.

As an example, suppose the variable $x_{1ct}$ is coded as (0, 1), so that its coefficient measures the gender difference, and let this coefficient vary randomly across classes. This means that the gender difference is greater in some classes than others, which seems a reasonable assumption. In this case, if the variance of $\beta_{1c}$ is $\sigma_1^2$ with mean $\beta_1$, an extra term is added to the random part of the model and we now have the variance of $y_{ct}$ as

$$\sigma_0^2 + \sigma_1^2 x_{1ct}^2 + 2\sigma_{01} x_{1ct} + \sigma^2, \qquad (11.6)$$

where we allow $\eta_{0c}$ and $\beta_{1c}$ to be correlated with covariance $\sigma_{01}$. Alternatively or in addition, we can allow $\beta_1$ to vary randomly over students within classes, and this leads to a model where the level 1 contribution to the variance for males is

$$\sigma_0^2 + \sigma_m^2$$

and for females

$$\sigma_0^2 + \sigma_f^2.$$

It is clear that we can accumulate a large number of random parameters (variances and covariances) by allowing further coefficients to be random. It is also possible to have a random coefficient whose mean value is 'constrained' to be zero, so that the effect of the explanatory variable is seen only in the random part of the model. A fuller study of such possibilities can be found in Goldstein (1986b, 1987b).

Returning to the basic two-level model of equation (11.4) we see that it implies a positive correlation between the responses of any two students in the same class, but a zero correlation between the responses of any two students chosen from different classes. Thus, the covariance of $y_{ct}$ and $y_{ct'}$ is the covariance between $(\eta_{0c} + \varepsilon_{ct})$ and $(\eta_{0c} + \varepsilon_{ct'})$ and since $\varepsilon_{ct}$ and $\varepsilon_{ct'}$ are assumed to be independent, this covariance then becomes simply $\sigma_0^2$. The variance of $y_{ct}$ or $y_{ct'}$ conditional on the fixed part of the model, is $\sigma_0^2 + \sigma^2$ and so the correlation between the responses is

$$\tau = \sigma_0^2(\sigma_0^2 + \sigma^2)^{-1}. \qquad (11.7)$$

This correlation is the usual 'intraclass' correlation and measures the degree of similarity of students within classrooms, or alternatively how well the response variable $y_{ct}$ is 'clustered' by classrooms. The larger the value of this correlation the greater the clustering and the more important it is to use a fully efficient estimation procedure. In general this correlation will be referred to as the 'intra-unit' correlation since we can determine correlations for each higher level of a model.

One particular feature which gives these models a considerable flexibility is that we can extend the idea of repeated measurements as part of a two-level structure to a specification of multivariate data. This is done by designating level 2 to be that of the student denoted now by $c$, with the variates considered as repeated within-student measurements at level 1, and defined by suitable indicator variables.



Suppose each student $c$ has four measurements $y_{c1}, \ldots, y_{c4}$. These could be different test scores, or alternate forms of the same test. We can write a simple basic two-level model for this as follows:

$$y_{ct} = \beta_{1c} x_{1ct} + \beta_{2c} x_{2ct} + \beta_{3c} x_{3ct} + \beta_{4c} x_{4ct}, \qquad (11.8)$$

where $t$ ranges from 1 to 4 and where $x_1, x_2, x_3, x_4$, are indicator variables (e.g., $x_{1ct} = 1$ if $t = 1$ and 0 otherwise).

We also suppose that the coefficients are random variables at the student level with variances and covariances which are to be estimated. In equation (11.8) this yields ten random parameters. The result of such an analysis will be an estimate of the 4 × 4 covariance matrix together with the means of the four measurements. These are simply the estimates of the coefficient means. Of course, if every student had all four measurements, these estimates would be the same as those obtained by the usual procedure for calculating the means and covariance matrix for a set of variables. The flexibility arises because the two-level model provides efficient estimates where some of the measurements are missing, and also allows nesting of students within higher levels of a hierarchy. In particular, we can see that 'multiple matrix sampling' or 'rotated form' designs, where not every student responds to the same combination of tests of forms, are special cases of data which are 'missing' by design. All such designs can in principle be regarded as special cases of the general multilevel model, and hence analysed by a suitable procedure. Any mixture of rotated designs, longitudinal repeated measurements and multiple levels of institutional organization, can be analysed within the same model structure in an integrated and efficient manner. In Section 11.5 we illustrate some of these possibilities with an example.

## 11.3 ESTIMATION IN THE MULTILEVEL MODEL

There are several methods available for estimating the fixed and random parameters in a multilevel model. The iterative generalized least squares (IGLS) approach used in the examples in this chapter, is outlined in Chapter 12. Under the assumption of multivariate normality the resulting estimates are also maximum likelihood. Goldstein (1986a, 1987a) shows how efficient computer algorithms can be constructed for the general case by making use of the known structure of the data, and how significance tests and confidence intervals can be derived. A trial version of a program is available from the authors.

A further useful extension of the purely hierarchical model is given in Goldstein (1987b, Chapter 7). This allows units to be cross-classified by two or more random factors. Thus the level 2 units could be the cells of a cross-classification of schools by neighbourhoods, with each student belonging to one cell. In sample surveys individuals could be classified by geographical area and workplace. In longitudinal surveys, where migration takes place between measuring occasions, individuals can be described by a cross-classification of their residence neighbourhood at the different occasions.

An important aspect of any analysis is a study of model adequacy. Procedures for the hierarchical fitting of parameters in both the fixed and random parts are available, analogously to those for ordinary linear models. At each level we can also construct residual plots. In the ordinary linear model there is only a single (level 1) residual, but in the models considered here we may have several (correlated) residuals at any level and bivariate residual plots are very useful. In the case of the basic model of equation (11.4) we obtain estimates of the residuals, namely:

$$\hat{\eta}_c, \hat{\varepsilon}_{ct}$$

which are typically referred to as 'shrunken' estimates. Very often these are themselves the focus of interest, as estimates of level 2 (school) 'effects' and from a Bayesian viewpoint would be termed 'posterior' means (Aitkin and Longford, 1986). A detailed discussion of the estimation and interpretation of such residuals is given by Goldstein (1987b, Chapters 2, 3).

## 11.4 LOGLINEAR MULTILEVEL MODELS FOR PROPORTIONS

In this section we describe the analysis of multilevel models where the response variable is a proportion with a logit transformation. We show how this provides a generalization of the usual single-level loglinear model to the multilevel case. Further details are given in Goldstein (1987b).

Consider a two-level model where, within each level 2 unit (cluster), the level 1 units are classified into $I$ response categories. These may themselves have a structure, for example a two-way cross-classification corresponding to a contingency table, which then could be further modelled in terms of constants fitted to the table margins. For simplicity we shall suppose that we have a one-way classification of categories.

Since the response proportions for each cluster add to one we consider only $I-1$ of them. As in Section 11.2, we define a set of $I-1$ dummy variables ($x$) corresponding to the $I-1$ response proportions. Thus we have a multivariate model with a response vector of length $I-1$. As in equation (11.8), the dummy variable coefficients are assumed to vary between clusters, that is, they are random at level 2. For a given cluster the $I-1$ responses themselves have a covariance matrix, constituting the level 1 variation, which if the proportions have a multinomial distribution is given by

$$\mathrm{var}(p_{ci}|c) = \mu_{ci}(1-\mu_{ci})n_c^{-1},$$
$$\mathrm{cov}(p_{ci}, p_{cj}|c) = -\mu_{ci}\mu_{cj}n_c^{-1}, \qquad (11.9)$$

where $c$ refers to cluster, $i$ and $j$ to response categories, $\mu_{ci}$ is the expected proportion and $n_c$ is the cluster size. In general, however, the $\mu_{ci}$ are unknown, and although we could obtain estimates based on the samples for each cluster, these will tend to be unstable unless cluster sizes are very large. Hence we replace $\mu_{ci}$ in effect by an average value by writting:

$$\mathrm{var}(p_{ci}|c) = \sigma_i^2 n_c^{-1},$$
$$\mathrm{cov}(p_{ci}p_{cj}|c) = \sigma_{ij}n_c^{-1}, \qquad (11.10)$$

where the $\sigma_i^2$, and $\sigma_{ij}$ are to be estimated.

An alternative procedure is outlined in Chapter 13. This assumes a multinomial distribution within each cluster and replaces the unknown parameters in equation (11.10) by the corresponding linear or loglinear functions of the (estimated) proportions. As in equation (11.10) the same values are assumed for each cluster. An estimate of the between-cluster (level 2) covariance matrix is then obtained by differencing. We note that where the distribution is not strictly multinomial the variances and covariances will still be proportional to $n_c^{-1}$ and this suggests that equation (11.10) provides a robust procedure.

In the simple model where the $\mu_{ci}$ have the same level 2 variance, we have:

$$\mathrm{var}(p_{ci}) = \sigma^2 + \sigma_i^2 n_c^{-1}. \qquad (11.11)$$

We can specify the covariance structure as follows. For the level 1 variation we define a set of dummy variables taking the values 0 or, $n_c^{-0.5}$ analogously to the dummy variables described above. The coefficients of these dummy variables are random at level 2 so giving the covariance structure (11.10). For the between-cluster variation we have a single constant term with a coefficient random at level 2. In the next section we provide an example of this model.

For the logit and loglinear cases we define a multivariate logit as:

$$\log(p_{ci}/p_{cI}) \qquad (11.12)$$

which reduces to the ordinary logit when $I = 2$. Wong and Mason (1985) give a maximum likelihood procedure for the ordinary logit two-level model. In the general case the choice of denominator as the last proportion is made for convenience.

For the multivariate logit transformation the variances and covariances are also inversely proportional to $n_c$ and so the same procedures for specifying the covariance structures can be used as in the untransformed case. This multivariate logit model is equivalent, in the case of a single-level model, to the corresponding loglinear model based on the underlying frequencies, and so gives us the corresponding loglinear multilevel model for frequency tables.

One difficulty with the use of logit functions occurs when some of the observed proportions are zero or one. In this case we could adopt the common procedure of adding or subtracting 0.25 as appropriate to the relevant frequencies. An iterative version of this can be based upon the EM algorithm (Dempster, *et al.*, 1977), whereby the parameter estimates obtained are used to predict the (0, 1) frequencies and these predicted values treated as if they were observed values for the next cycle of iterations. The process would be repeated until convergence. An alternative is to collapse relevant cells across clusters as described in Chapter 13.

It is clear that even where the number of response categories is small, there are, potentially, a large number of random parameters at level 2. In the example in Chapter 13 the full set of variances and covariances are fitted and the structure of the resulting level 2 correlation matrix is examined. The model formulation in this section allows a hierarchically structured design for the level 2 random parameters, since the level 2 and level 1 random parameters are separately specified. Thus, we might first allow the coefficients for each of the (one-way) main effects to vary randomly at level 2, then see whether adding interactions improved the model fit, and so forth. In this way one may be able to arrive at a parsimonious model which can be interpreted and which describes the data adequately.

## 11.5 EXAMPLES

The data for the first example are taken from the second International Educational Achievement (IEA) survey test score results at the eighth grade or year of compulsory schooling. Two populations are used, one from Japan and one from British Columbia in Canada. In each case all the students within one class in each study school were measured. There are 187 and 47 schools respectively. For each student there is a core 'post-test' of geometry items taken towards the end of the school year. In addition each student responded to an 'alternative form' post-test. In fact there were four alternative forms of post-test of which each student took only one. Preliminary analyses revealed no significant differences between the alternative forms and we make no distinction in the analyses presented here. We treat the data as if every student answered the core post-test and the alternative form. Further discussion of this point and an extended model structure to allow for alternative forms of the post-test is given in Goldstein (1987b).

Two classroom-level variables; ($a_1$) the years of experience of the class mathematics teacher and ($a_2$) the cumulative percentage of the post-test items covered in the curriculum were measured. This second variable is often referred to as 'opportunity to learn' (OTL). In addition each student achieved a pretest score taken at the start of the school year.

The fact that only one class of students is measured within each school means that the school and class effects are confounded. We shall refer to this level of clustering as the school effect although the two explanatory variables $a_1$ and $a_2$ actually relate to the particular class sampled.

If we were to analyse only the core post-test scores (putting to one side the alternative form) then a two-level model would suffice where level 2 was the



school and level 1 the individual student. To model overall ability the pretest score would not be used and thus we could write:

$$y_{ct} = \sum_{k=1}^{2} \gamma_{k0} a_{kc} + \beta_{ct},$$

$$\beta_{ct} = \beta + \eta_{0c} + \varepsilon_{ct}, \tag{11.13}$$

where $k = 1, 2$ refers to the two explanatory variables, $c$ refers to school and $t$ to the student. Here $\text{var}(\eta_{0c}) = \sigma_0^2$ and $\text{var}(\varepsilon_{ct}) = \sigma_\varepsilon^2$.

If we introduce into equation (11.13) the individual pretest score $x_{ct}$ with a corresponding regression parameter then the emphasis is altered to modelling the change in performance for each child with the pretest score treated as a student level covariate. The results to be presented will cover both cases of including and excluding the pretest score.

The existence of an alternative form post-test introduces an added complexity which may be incorporated by extending the model to three levels. An additional subscript $j = 1, 2$ identifies the core post-test ($j = 1$) and the alternative form ($j = 2$). The three-level model becomes

$$y_{ctj} = \sum_k \gamma_{k0j} a_{kc} + \beta_{ctj}, \tag{11.14}$$

where

$$\beta_{ctj} = \beta_j + \eta_{0c} + \varepsilon_{ctj}$$

with

$$\text{var}(\eta_{0c}) = \sigma_0^2,$$
$$\text{var}(\varepsilon_{ct1}) = \sigma_{\varepsilon.1}^2,$$
$$\text{var}(\varepsilon_{ct2}) = \sigma_{\varepsilon.2}^2,$$

and

$$\text{cov}(\varepsilon_{ct1}, \varepsilon_{ct2}) = \sigma_{\varepsilon.12}^2$$

As in equation (11.13), the pretest score $x_{ct}$ may be included if the change in attainment is the objective.

The formulation allows for the variances of $\varepsilon_{ctj}$ to be unequal for $j = 1, 2$ and for the two terms to be correlated. Thus the intra-school correlation for the two forms of the post-test may be different since although $\sigma_0^2$ is common, the individual level components of variance $\sigma_{\varepsilon.1}^2$ and $\sigma_{\varepsilon.2}^2$ may differ.

The scores are standardized on a scale which gives the percentage of items correctly answered. The results of the analysis are given in Tables 11.1 and 11.2.

We consider first the case when the pretest score is included in the model. As might be expected, in both countries there is a fairly strong relationship with the pretest, as judged by the coefficient of the pretest score. There is a difference between the core and alternate form tests for British Columbia which appears as an interaction with the pretest score. The only other variable showing any relationship with progress between pre- and post-tests is OTL in the case of

*Table 11.1 Summary statistics for British Columbia and Japan: Grade 8 geometry*

| Variable | Pretest | Post-test | Yrs exp. | OTL |
|---|---|---|---|---|
| *British Columbia* | | | | |
| Mean | 88.1 | 54.0 | 8.2 | 50.1 |
| SD | 43.3 | 24.3 | 5.9 | 19.6 |
| *Japan* | | | | |
| Mean | 106.9 | 64.2 | 6.2 | 50.6 |
| SD | 37.2 | 21.6 | 4.6 | 14.1 |

N.B Number of students = 253 (B. Columbia), 671 (Japan),
Number of schools = 47 (B. Columbia), 187 (Japan).

British Columbia. The coefficients for OTL are fairly similar in both countries (0.07 and 0.11) but the standard error for the Japanese coefficient is no smaller despite being based on a larger number of schools. The reason is seen in Table 11.1 where the standard deviation for OTL is larger in British Columbia. Since the OTL values are more dispersed the standard error of the regression coefficient will be smaller and this compensates for fewer schools in the sample. Turning to the fixed-effect terms when the pretest is omitted (i.e. for a purely cross-sectional analysis) the contribution of the alternate form score is less than that of the core in both countries. There is also a suggestion in British Columbia that the years of teaching experience is negatively related to test score.

A persistent problem in cross-cultural studies of educational achievement is to obtain consistent and comparable interpretations of common test scores when the curricula, examination systems and so forth differ. This difficulty applies particularly to the fixed-effect parameters. When we look at the random parameters, however, we are studying how the variation is distributed over the different levels of the educational system and it may be that such comparisons between systems will be more meaningful. This will not remove the difficulty that a common test may be more relevant to some educational systems than to others, so that it seems generally preferable to design and use different assessment measures, each appropriate to its own system. The same argument would apply to studies of a single system subject to curriculum change where comparisons across long time periods were required.

Bearing in mind these caveats, we see that the intra-school correlations for Japan are much smaller than those for British Columbia. After accounting for the pretest, the intra-school correlations in Japan become zero, whereas those for British Columbia are still substantial. Thus Japanese schools appear to have



*Table 11.2 Grade 8 geometry for British Columbia and Japan*

| | British Columbia | Japan |
|---|---|---|
| Post-test related to class-level variables | | |
| *Fixed* | | |
| Constant | 50.8 | 58.2 |
| Alt. minus core | −12.5(1.4) | −5.1(0.7) |
| Years exp. maths | −0.61(0.28) | 0.16(0.17) |
| OTL | 0.17(0.09) | 0.10(0.06) |
| *Random* | | |
| Between schools | 70.0(13.9) | 15.7(8.2) |
| Variance (core) | 550.7(30.3) | 437.7(24.0) |
| Variance (alt.) | 353.9(24.0) | 443.0(20.9) |
| Covariance | 203.9(23.2) | 265.9(17.3) |
| Correlation | 0.46 | 0.60 |
| Intra-school (core) | 0.11 | 0.04 |
| Intra-school (alt.) | 0.17 | 0.03 |
| Post-test related to class-level variables and pretest | | |
| *Fixed* | | |
| Constant | 16.0 | 34.1 |
| Alt. minus core | −1.7(3.1) | −3.8(2.5) |
| Pretest | 0.40(0.03) | 0.24(0.02) |
| Pretest × (alt. minus core) | −0.12(0.03) | −0.01(0.02) |
| Years exp. maths | −0.27(0.17) | 0.19(0.19) |
| OTL | 0.11(0.05) | 0.07(0.06) |
| *Random* | | |
| Between schools | 21.1(10.5) | 0.0 |
| Variance (core) | 319.8(29.8) | 532.3(21.2) |
| Variance (alt.) | 257.1(23.2) | 378.0(20.6) |
| Covariance | 55.9(18.9) | 223.3(16.8) |
| Correlation | 0.20 | 0.50 |
| Intra-school (core) | 0.09 | 0.0 |
| Intra-school (alt.) | 0.08 | 0.0 |

greater homogeneity of achievement than those in British Columbia. It is also interesting to note that the correlation between core and alternate forms is only moderate and appears to be stronger in Japan. The variances $\sigma^2_{\varepsilon.2}$ are generally smaller than the corresponding terms $\sigma^2_{\varepsilon.1}$ for the core pretest which suggests that the alternative form may be preferred in terms of reliability. It should be remembered, however, that we have made no adjustment for measurement error in the pretest score, and this might alter the inferences we make if measurement error were taken into account (see Goldstein, 1986a). In general, a major aim of these kinds of analysis will be to see how far the between-school variation

*Table 11.3 Support for law on racial discrimination*

| | | Numbers of individuals responding (row percentages in parentheses) | | |
|---|---|---|---|---|
| | | 1984 | | |
| | | Yes | No | Total |
| 1983 | Yes | 253 (79.1) | 67 (20.9) | 320 (100) |
| | No | 71 (47.3) | 79 (52.7) | 150 (100) |
| | Total | 324 | 146 | 470 |

can be reduced by the further fitting of explanatory variables, at either the school or student level.

The data for the second example are taken from a survey of social attitudes (Jowell and Airey, 1985). The response is the proportion of individuals in 1984 who said that they supported the law in the United Kingdom which outlawed racial discrimination in housing, employment, etc. There is one dichotomous explanatory variable, namely the (two-category) response to the same question asked of the same individuals one year previously in 1983.

Table 11.3 shows the overall numbers responding in each category. The data is based on a sample of 49 randomly selected clusters, and in one sense this example, too, could be regarded as a three-level structure. There are two responses for each individual who are in turn nested within the 49 clusters. However, for data of this kind we may consider the observed proportions within each cluster as the random variables to be analysed rather than the individual-level responses. We formulate a two-level model involving clusters (level 2) and the categories for the contingency table within each cluster are level 1. Thus for each cluster there are four categories defined by the responses in 1983 and 1984. In each cluster we observe the proportions falling into each cell of the table. We analyse three of these proportions directly or calculate a set of three logits by using one cell as a baseline response. In either case we refer to this variable as $y_{ct}$ for the $t$th proportion (or logit) in the $c$th cluster.

Table 11.4 contains a set of dummy variables which are used in various model formulations for the fixed and random parts of models for these data. We first consider the model reported under (A) in Table 11.5, which is an analysis of



*Table 11.4 Explanatory design variables for the* c*th level 2 unit*

| | 1983 = yes 1984 = yes | 1983 = yes 1984 = no | 1983 = no 1984 = yes |
|---|---|---|---|
| $x_1$ | 1 | 0 | 0 |
| $x_2$ | 0 | 1 | 0 |
| $x_3$ | 0 | 0 | 1 |
| $x_4$ | 1 | 1 | 0 |
| $x_5$ | 1 | 0 | 1 |
| $x_6$ | $n_c^{-0.5}$ | 0 | 0 |
| $x_7$ | 0 | $n_c^{-0.5}$ | 0 |
| $x_8$ | 0 | 0 | $n_c^{-0.5}$ |
| $x_9$ | 1 | 1 | 1 |

*Table 11.5 A two-level multivariate analysis of proportions. Social attitudes data*

| Explanatory variable | Analysis A | B | C | D |
|---|---|---|---|---|
| *Fixed* | | | | |
| $x_1$ | — | 1.77 (0.23) | — | 0.54 (0.03) |
| $x_2$ | — | — | — | 0.14 (0.02) |
| $x_3$ | — | — | — | 0.15 (0.02) |
| $x_4$ | 0.63 (0.12) | −0.27 (0.17) | 0.20 (0.01) | — |
| $x_5$ | 0.67 (0.12) | −0.23 (0.17) | 0.21 (0.01) | — |
| *Random* | | | | |
| *Level 1* | | | | |
| $\sigma_6^2$ | 2.15 | 3.09 | 0.51 | 0.34 |
| $\sigma_7^2$ | 13.1 | 6.49 | 0.16 | 0.13 |
| $\sigma_8^2$ | 13.2 | 6.94 | 0.19 | 0.16 |
| $\sigma_{67}$ | −2.57 | −1.69 | −0.19 | −0.13 |
| $\sigma_{68}$ | −2.18 | 1.23 | −0.20 | −0.13 |
| $\sigma_{78}$ | −4.41 | −1.99 | 0.008 | −0.02 |
| *Level 2* | | | | |
| $\sigma_9^2$ | 0.77 (0.56) | 0.66 (0.47) | 0.0001 (0.007) | 0.0001 (0.007) |

*Note*: Analyses A and B use the logits as responses: analyses C and D use the proportions.

the logits of the proportions.

$$\begin{aligned} y_{ct} &= \beta_{0ct} + \beta_4 x_{4ct} + \beta_5 x_{5ct}, \\ \beta_{0ct} &= \beta_0 + \eta_9 x_{9ct} + \eta_6 x_{6ct} + \eta_7 x_{7ct} + \eta_8 x_{8ct}, \\ \text{var}(\eta_i) &= \sigma_i^2 \qquad i = 6\ldots 9, \\ \text{cov}(\eta_i, \eta_{i'}) &= \sigma_{ii'} \qquad i, i' = 6, 7, 8, \\ \text{cov}(\eta_i, \eta_9) &= 0 \qquad i = 6, 7, 8. \end{aligned} \tag{11.15}$$

In the fixed part of the model the dummy variables $x_4$ and $x_5$ represent the two main effects for responses in 1983 and 1984. In the random part of the model the dummy variable $x_9$ is the explanatory variable defining the between-cluster variance and $x_6$, $x_7$ and $x_8$ define the level 1 variances and covariances. Note that the variables $x_6$, $x_7$ and $x_8$ include the factor $n_c^{-0.5}$ to reflect the fact that observed proportions are based on a sample of size $n_c$ in the $c$th cluster.

In this description we have taken the four cell proportions together and used dummy variables to define the model structure of main effects and interactions. An alternative approach would have been to analyse the conditional proportions of responses in 1983. A comparison between this approach and that described above is analagous to the use of logistic or loglinear models for contingency tables. The conditional approach could be adopted using the same sorts of models as described above and a discussion of this is given in Goldstein (1987b).

In column B of Table 11.4 an extra dummy variables $x_1$, is included in the fixed part of the model to allow for interaction between the responses in 1983 and 1984.

A comparison of columns A and B shows, as we would expect, that the interaction term is very important with standard error much smaller than the coefficient. In the random part of the model we observe that the components of variance reduce with the introduction of the extra explanatory variable. This is often the case.

Columns C and D of Table 11.4 contain the corresponding analyses for the cell proportions rather than the logits. In column D, for illustrative purposes a different parametrization of the three fixed terms has been used but still accounts for all three available degrees of freedom. The results in columns C and D give the same general picture as those in columns A and B, although because of the change in scale from the logistic to the ordinary proportions the actual numerical values are different. As before, the interaction term is needed in the fixed part of the model and the random components are generally smaller in column D.

The variances and covariances between $x_6, x_7$ and $x_8$ generally take the signs and approximate values that we would expect under the within-cluster multinomial sampling assumption. This approach is described in Chapter 13. Ths cluster component $\sigma_9^2$ is very small. In this analysis the level 1 random parameter estimates have little substantive interest attached to them, apart from acting as a check on the adequacy of a multinomial distribution. They may be regarded as nuisance parameters. Also, their standard errors are complicated and have not been calculated.



For more complex models we can obtain an approximate chi-squared goodness-of-fit test based on the observed and predicted frequencies, the latter being derived from the predicted proportions. For degrees of freedom we would use the total number of frequencies less the number of fitted model parameters.

## 11.6 DISCUSSION

In our examples we have shown how multilevel modelling can explore the contribution of measurements made at different levels of a hierarchical system to both the fixed and random parts of the model. In many cases, primary interest will centre on the random variation rather than the fixed coefficients. This offers researchers a new range of techniques for exploring and understanding the effects of hierarchical systems in terms of the relative heterogeneity of units. The introduction of variables measured on higher level units allows us to attempt to 'explain' such variation. In social surveys we can introduce 'ecological' variables at the cluster level, and this opens up many new analysis possibilities.

## 11.7 ACKNOWLEDGEMENTS

Our grateful thanks are due to the following for their help and comments during the preparation of this paper. Leigh Burstein, Skip Kifer, Rod McDonald, Les McLean and Richard Wolfe. We are particularly grateful to Richard Wolfe for his help with the analysis of the mathematics attainment data and for discussions of models.