Linear Mixed Effects Models

Kevin O'Brien

September 20, 2015

Contents

0.0.1 Using Linear Mixed Effects Models

It is shown that classical models can give different results to linear mixed effects models , based on the same data. **?** illustrates this with a comparison of simple regression model of prices against sales, with a mixed model, that takes groups of data into account.

0.0.2 Laird Ware Model

Linear mixed effects models (LME) differs from the conventional linear model in that it has both fixed effects and random effects regressors, and coefficients thereof. Further to a paper published by Laird and Ware in 1982, it is conventional to formulate an LME in matrix form as follows:

$$Y_i = X_i\beta + Z_i b_i + \epsilon_i$$

- Y_i is the $n \times 1$ response vector
- X_i is the $n \times p$ Model matrix for fixed effects
- β is the $p \times 1$ vector of fixed effects coefficients
- Z_i is the $n \times q$ Model matrix for random effects
- b_i is the $q \times 1$ vector of random effects coefficients, sometimes denoted as u_i
- ϵ is the $n \times 1$ vector of observation errors

0.0.3 Computation

When tackling linear mixed effects models using the R language, a statistician can call upon the *lme* command found in the *nlme* package.This command fits a LME model to the data set using either Maximum Likelihood (ML) or Restricted Maximum Likelihood (REML).

The first two arguments for *lme* are *fixed* and *data*, which give the model for the expected responses (i.e. the fixed part of the model), and the data that the model should be fitted from. The next argument is *random*, a one-sided formula which describes the random effects, and the grouping structure for the model. The *method* argument can specify whether to use 'REML', the default setting, or 'ML'.

0.1 Linear Mixed Effects Models

Applications

So-called mixed-effect models (or just mixed models) include additional random-effect terms, and are often appropriate for representing clustered, and therefore dependent, data arising, for example, when data are collected hierarchically, when observations are taken on related individuals (such as siblings), or when data are gathered over time on the same individuals.

Mean Centering

It is customary to center data prior to running LMM or HLM. Centering means subtracting the mean, so means become zero. Two main types of centering are group mean centering and grand mean centering.

Restricted Maximum Likelihood

restricted (or residual) maximum likelihood (REML) is a method for fitting linear mixed models. In contrast to conventional maximum likelihood estimation, REML can produce unbiased estimates of variance and covariance parameters.

0.2 Grouped Data Sets

In modern statistical analysis , data sets have very complex structures, such as clustered data, repeated data and hierarchical data (henceforth referred to as grouped data).

Repeated data considers various observations periodically taken from the same subjects. 'Before and after' measurements, as used in paired t tests, are a well known example of repeated measurements. Clustered data is simply the grouping of observations according to common characteristics. For example, an study of pupils of a school would account for the fact that they are grouped according to their classes.

Hierarchical structures organize data into a tree-like structure, i.e. groups within groups. Using the previous example, the pupils would be categorized according to their years (i.e the parent group) and then their classes (i.e the child group). This can be extended again to multiple schools, where each school would be the parent group of each year.

An important feature of such data sets is that there is correlation between observations within each of the groups (**?**). Observations in different groups may be independent, but any assumption that these observations within the same group are independent is inappropriate . Consequently **?** states that there is two sources of variations to be considered, 'within groups' and 'between groups'.

0.2.1 Classical Models

? discusses the inadequacy of 'classical models' in analysing such data types, with particular reference to the simple linear model . The simple linear model is a well known statistical methodology that describes the relationship between dependent variables Y and an independent predictor variable X. Where $Y = y_1, y_2, ..y_k..y_n$ and $X = x_1, x_2, ..x_k..x_n$, an intercept α and slope β are estimated such that the error terms associated with each observation y_i is minimised.

$$y_k = \alpha + \beta x_k + \epsilon_k \tag{1}$$

In classical statistics a typical assumption is that observations are drawn from the same general population, are independent and identically distributed (**?**). Consequently there is no way to account for the grouped nature of data sets described previously, and so there lies the possibility of observations being treated as independent measurements. **?**, pg.3 gives a very informative example wherein a classical approach is compared to an approach that does account for grouping. The conclusion to be drawn from Demidenko's example is that failure to account for grouping leads to an incorrect conclusion about the data. The approach recommended is known as 'mixed models' and shall be introduced presently.

0.3 Fixed Effects and Random Effects Models

Before proceeding to a description of mixed models, an introduction to fixed effects and random effects models is required. This section follows on from the discussion of measurement error models in the last chapter.

0.3.1 Fixed Effects

? gives an example of a study where the observations , occurrences of skin tumours called basal cell epithelioma, were classified according to 'factors', i.e. the gender, age and exposure to sunshine of the patients. Levels are the individual classes of each of these factors (e.g. 'Male' and 'Female' would be the levels of the factor 'Gender'). The scientific interest lies in examining the extent to which different factor levels affect the variable of interest. The effects of a level of a factor are one of two types ; fixed effects and random effects. Fixed effects describe effects due to a finite set of levels for a factor (i.e multichotomous factors). The factors described in the skin tumour example are all fixed effects factors. Fixed effects models are the cases where only fixed effects are present, with the exception of random error terms.

To demonstrate fixed effects model **?** describes a study wherein 24 plants are divided into four groups of six, and each group is subjected to its own treatment regime. Three different fertilizers are used with three of the groups (treatments N,P,K), while no fertilizer is used on the fourth group, (i.e. it is a control group denoted as C). **?** constructs a model to describe the crop yield resultant from the experiment.

$$y_{ij} = \mu + \alpha_i + e_{ij} \tag{2}$$

where y_{ij} is the jth plant (i.e. crop yield) on the ith treatment, with μ as the mean yield, α_i is the effect of each fertilizer treatments (i.e. fixed effects) and e_{ij} is the error term. The fixed effect for each observation is an unknown constant that is to be determined from computing the data.

The μ term would not necessarily be present in all formulations. Some authors, such as **?**, may use a single term as equivalent to the μ and α terms. (It is customary to centre data prior to using mixed effects methodologies. Centering means subtracting the mean of the observations from each observed value, so mean of the resultant values become zero.)

0.3.2 Random Effects

The random effects model describes the case where there is an infinite number of levels in a factor. In other words the factor is a random variable. **?** demonstrates this with

a second example; a study of the maternal ability of mice. In this example 4 female mice , all of the same breed, have 6 litters each over a period of time. The weights w_{ij} of each litter (j) from each mouse (i) were taken to be the proxy for maternal ability and is formulated as follows;

$$w_{ij} = \mu + \delta_i + e_{ij} \tag{3}$$

As in the previous exmaple, α is the mean , e_{ij} is the error term and δ_i is a random effect due to each mouse. Notably these four mice are considered as a sample of the overall population of female mice of that breed, consequently an important characteristic of random effects models is that the δ_i values are a random sample of all δ terms. Therefore these random terms can be used for making inferences about populations. (**?**).

0.3.3 Variance Components

Each random effect has an associated 'variance component' term. This is a model parameter which quantifies random variation due to that effect only. Therefore for every observation there are two sources of variation, random variation and residual variation, and can be expressed as follows $var(y_{ij}) = \sigma^2_p + \sigma^2$. These variations are known as the variance components. This is an important difference with fixed effects models, which is subject to residual variation (i.e σ^2) only (**?**).

In fixed effects models there is no covariance between any pair of observations. **?** shows that, while there is no covariance between observations from different subjects (i.e the mice in Searle's example), there exists correlation between observations from the same subject (i.e. litter weights from the same mouse are correlated). In this case the covariance is the subjects variance component (i.e. σ^2_p)

0.3.4 More complex examples

? offers elaborations on both examples used so far. In the case of the fixed effects model, the model can be amended to take account for different varieties of each plants being studied. (The groups of six plants are subdivided into three variety types.) y_{ijk} is the yield of the kth plant of the jth variety in the ith treatment, and is described as follows;

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} \tag{4}$$

This new formulation includes a fixed effect β_j to account for the variety type, and an 'interaction effect' γ_{ij}. An interaction effect describes the combined effects of two or more variables on the observation. Similarly the random effects example is elaborated

to account for the effects of three different technicians. Again there is a random effect component τ to account for these technicians, and an interaction effect θ to account for the combined effect of the mice and the technicians.

$$w_{ijk} = \mu + \delta_i + \tau_j + \theta_{ij} + e_{ijk} \tag{5}$$

Likelihood Ratio Tests

The problem with REML for model building is that the "likelihoods" obtained for different fixed effects are not comparable. Hence it is not valid to compare models with different fixed effects using a likelihood ratio test or AIC when REML is used to estimate the model. Therefore models derived using ML must be used instead.

0.4 Computation on a mixed effects model

? describes an experiment whereby the productivity of six randomly chosen workers are assessed three times on each of three machines, yielding the 54 observations tabulated below.

Observation	Worker	Machine	score	Observation	Worker	Machine	score
1	1	A	52.00	28	4	B	63.20
2	1	A	52.80	29	4	B	62.80
3	1	A	53.10	30	4	B	62.20
4	2	A	51.80	31	5	B	64.80
5	2	A	52.80	32	5	B	65.00
6	2	A	53.10	33	5	B	65.40
7	3	A	60.00	34	6	B	43.70
8	3	A	60.20	35	6	B	44.20
9	3	A	58.40	36	6	B	43.00
10	4	A	51.10	37	1	C	67.50
11	4	A	52.30	38	1	C	67.20
12	4	A	50.30	39	1	C	66.90
13	5	A	50.90	40	2	C	61.50
14	5	A	51.80	41	2	C	61.70
15	5	A	51.40	42	2	C	62.30
16	6	A	46.40	43	3	C	70.80
17	6	A	44.80	44	3	C	70.60
18	6	A	49.20	45	3	C	71.00
19	1	B	62.10	46	4	C	64.10
20	1	B	62.60	47	4	C	66.20
21	1	B	64.00	48	4	C	64.00
22	2	B	59.70	49	5	C	72.10
23	2	B	60.00	50	5	C	72.00
24	2	B	59.00	51	5	C	71.10
25	3	B	68.60	52	6	C	62.00
26	3	B	65.80	53	6	C	61.40
27	3	B	69.70	54	6	C	60.50

Table 1: Machines Data , Pinheiro Bates

(Overall mean score = 59.65, mean on machine A = 52.35 , mean on machine B = 60.32, mean on machine C = 66.27)

The 'worker' factor is modelled with random effects(u_i), whereas the 'machine' factor is modelled with fixed effects (β_j). Due to the repeated nature of the data, interaction effects between these factors are assumed to be extant, and shall be examined accordingly. The interaction effect in this case (τ_{ij}) describes whether the effect of changing from one machine to another is different for each worker. The productivity score y_{ijk} is the kth observation taken on worker i on machine j, and is formulated as follows;

$$y_{ijk} = \beta_j + u_i + \tau_{ij} + \epsilon_{ijk} \tag{6}$$

$$u_i \sim N(0, \sigma_u^2), \epsilon_{ijk} \sim N(0, \sigma^2), \tau_i \sim N(0, \sigma_\tau^2)$$

The 'nlme' package is incorporated into the R programming to perform linear mixed model calculations. For the 'Machines' data, **?** use the following code, with the hierarchical structure specified in the last argument.

```
lme(score~Machine, data=Machines, random=~1|Worker/Machine)
```

The output of the R computation is given below.

```
Linear mixed-effects model fit by REML
  Data: Machines
  Log-restricted-likelihood: -107.8438
  Fixed: score ~ Machine
(Intercept)    MachineB    MachineC
  52.355556    7.966667   13.916667

Random effects:
 Formula: ~1 | Worker
        (Intercept)
StdDev:     4.78105

 Formula: ~1 | Machine %in% Worker
        (Intercept)  Residual
StdDev:    3.729532 0.9615771

Number of Observations: 54 Number of Groups:
             Worker Machine %in% Worker
                  6                  18
```

The crucial pieces of information given in the programme output are the estimates of the intercepts for each of the three machines. Machine A, which is treated as a control case, is estimated to have an intercept of 52.35. The intercept estimates for machines B and C are found to be 60.32 and 66.27 (by adding the values 7.96 and 13.91 to 52.35 respectively). Estimate for the variance components are also given; $\sigma_u^2 = (4.78)^2$, $\sigma_\tau^2 = (3.73)^2$ and $\sigma_\epsilon^2 = (0.96)^2$.

In simple examples V^{-1} is a straightforward calculation, but with higher dimensions it becomes a very complex calculation. **????** derived the 'Mixed Model Equations (MME)' to provide estimates for β and u without the need to calculate V^{-1}.

$$\begin{pmatrix} X^T R^{-1} X & X^T R^{-1} Z \\ Z^T R^{-1} X & Z^T R^{-1} Z + G^{-1} \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{u} \end{pmatrix} = \begin{pmatrix} X^T R^{-1} y \\ Z^T R^{-1} y \end{pmatrix}. \tag{7}$$

When R and G are diagonal, R^{-1} and G^{-1} are trivial calculations, and therefore the matrices in the equation CC are much simpler to calculate than using V^{-1}.

Each of the elements of the above matrices are submatrices. $X^T R^{-1} X$ is a $p \times p$ matrix, $Z^T R^{-1} Z + G^{-1}$ is a $q \times q$ matrix. The remaining elements, which are transposes of each other, are of dimensions $p \times q$ and $q \times p$ respectively. Therefore the overall matrix is of dimension $(p+q) \times (p+q)$. These dimensions are notably smaller than $n \times n$, which would have been the case if V^{-1}, and therefore the inversion is easier to compute.

Rearranging the equation CC, the BLUE of β, and the BLUP of u can be shown to be;

$$\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} y \tag{8}$$

$$\hat{u} = G Z^T V^{-1} (y - X\hat{\beta}) \tag{9}$$

0.5 Model Selection

The previous section on estimation assumes the specification of a mixed model in terms of X, Z, G, and R. Even though X and Z have known elements, there is some flexibility is specifying the form and construction is flexible, and for a particular data set, there are numerous possibilities that can be considered. Similarly, various potential covariance structures for G and R may be considered.

First, subject matter considerations and objectives are of great importance when selecting a model; refer to Diggle (1988) and Lindsey (1993).

Second, when the data themselves are looked to for guidance, many of the graphical methods and diagnostics appropriate for the general linear model extend to the mixed model setting as well (Christensen, Pearson, and Johnson 1992).

Likelihood-based approachs to the mixed model allow the comparison of candidate models. The most common of these are the likelihood ratio test and Akaike's and Schwarz's information criteria (Bozdogan 1987; Wolfinger 1993).

Chapter 1

Linear Mixed Effects Models

1.1 Grouped Data Sets

In modern statistical analysis , data sets have very complex structures, such as clustered data, repeated data and hierarchical data (henceforth referred to as grouped data).

Repeated data considers various observations periodically taken from the same subjects. 'Before and after' measurements, as used in paired t tests, are a well known example of repeated measurements. Clustered data is simply the grouping of observations according to common characteristics. For example, an study of pupils of a school would account for the fact that they are grouped according to their classes.

Hierarchical structures organize data into a tree-like structure, i.e. groups within groups. Using the previous example, the pupils would be categorized according to their years (i.e the parent group) and then their classes (i.e the child group). This can be extended again to multiple schools, where each school would be the parent group of each year.

An important feature of such data sets is that there is correlation between observations within each of the groups (**?**). Observations in different groups may be independent, but any assumption that these observations within the same group are independent is inappropriate . Consequently **?** states that there is two sources of variations to be considered, 'within groups' and 'between groups'.

1.1.1 Classical Models

? discusses the inadequacy of 'classical models' in analysing such data types, with particular reference to the simple linear model . The simple linear model is a well known statistical methodology that describes the relationship between dependent variables Y and an independent predictor variable X. Where $Y = y_1, y_2, ..y_k..y_n$ and $X = x_1, x_2, ..x_k..x_n$, an intercept α and slope β are estimated such that the error terms associated with each observation y_i is minimised.

$$y_k = \alpha + \beta x_k + \epsilon_k \tag{1.1}$$

In classical statistics a typical assumption is that observations are drawn from the same general population, are independent and identically distributed (**?**). Consequently there is no way to account for the grouped nature of data sets described previously, and so there lies the possibility of observations being treated as independent measurements. **?**, pg.3 gives a very informative example wherein a classical approach is compared to an approach that does account for grouping. The conclusion to be drawn from Demidenko's example is that failure to account for grouping leads to an incorrect conclusion about the data. The approach recommended is known as 'mixed models' and shall be introduced presently.

1.2 Fixed Effects and Random Effects Models

Before proceeding to a description of mixed models, an introduction to fixed effects and random effects models is required. This section follows on from the discussion of measurement error models in the last chapter.

1.2.1 Fixed Effects

? gives an example of a study where the observations , occurrences of skin tumours called basal cell epithelioma, were classified according to 'factors', i.e. the gender, age and exposure to sunshine of the patients. Levels are the individual classes of each of these factors (e.g. 'Male' and 'Female' would be the levels of the factor 'Gender'). The scientific interest lies in examining the extent to which different factor levels affect the variable of interest. The effects of a level of a factor are one of two types ; fixed effects and random effects. Fixed effects describe effects due to a finite set of levels for a factor (i.e multichotomous factors). The factors described in the skin tumour example are all fixed effects factors. Fixed effects models are the cases where only fixed effects are present, with the exception of random error terms.

To demonstrate fixed effects model **?** describes a study wherein 24 plants are divided into four groups of six, and each group is subjected to its own treatment regime. Three different fertilizers are used with three of the groups (treatments N,P,K), while no fertilizer is used on the fourth group, (i.e. it is a control group denoted as C). **?** constructs a model to describe the crop yield resultant from the experiment.

$$y_{ij} = \mu + \alpha_i + e_{ij} \tag{1.2}$$

where y_{ij} is the jth plant (i.e. crop yield) on the ith treatment, with μ as the mean yield, α_i is the effect of each fertilizer treatments (i.e. fixed effects) and e_{ij} is the error term. The fixed effect for each observation is an unknown constant that is to be determined from computing the data.

The μ term would not necessarily be present in all formulations. Some authors, such as **?**, may use a single term as equivalent to the μ and α terms. (It is customary to centre data prior to using mixed effects methodologies. Centering means subtracting the mean of the observations from each observed value, so mean of the resultant values become zero.)

1.2.2 Random Effects

The random effects model describes the case where there is an infinite number of levels in a factor. In other words the factor is a random variable. **?** demonstrates this with

a second example; a study of the maternal ability of mice. In this example 4 female mice , all of the same breed, have 6 litters each over a period of time. The weights w_{ij} of each litter (j) from each mouse (i) were taken to be the proxy for maternal ability and is formulated as follows;

$$w_{ij} = \mu + \delta_i + e_{ij} \tag{1.3}$$

As in the previous exmaple, α is the mean , e_{ij} is the error term and δ_i is a random effect due to each mouse. Notably these four mice are considered as a sample of the overall population of female mice of that breed, consequently an important characteristic of random effects models is that the δ_i values are a random sample of all δ terms. Therefore these random terms can be used for making inferences about populations. (**?**).

1.2.3 Variance Components

Each random effect has an associated 'variance component' term. This is a model parameter which quantifies random variation due to that effect only. Therefore for every observation there are two sources of variation, random variation and residual variation, and can be expressed as follows $var(y_{ij}) = \sigma_p^2 + \sigma^2$. These variations are known as the variance components. This is an important difference with fixed effects models, which is subject to residual variation (i.e σ^2) only (**?**).

In fixed effects models there is no covariance between any pair of observations. **?** shows that, while there is no covariance between observations from different subjects (i.e the mice in Searle's example), there exists correlation between observations from the same subject (i.e. litter weights from the same mouse are correlated). In this case the covariance is the subjects variance component (i.e. σ_p^2)

1.2.4 More complex examples

? offers elaborations on both examples used so far. In the case of the fixed effects model, the model can be amended to take account for different varieties of each plants being studied. (The groups of six plants are subdivided into three variety types.) y_{ijk} is the yield of the kth plant of the jth variety in the ith treatment, and is described as follows;

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} \tag{1.4}$$

This new formulation includes a fixed effect β_j to account for the variety type, and an 'interaction effect' γ_{ij}. An interaction effect describes the combined effects of two or more variables on the observation. Similarly the random effects example is elaborated

to account for the effects of three different technicians. Again there is a random effect component τ to account for these technicians, and an interaction effect θ to account for the combined effect of the mice and the technicians.

$$w_{ijk} = \mu + \delta_i + \tau_j + \theta_{ij} + e_{ijk} \tag{1.5}$$

1.3 Mixed Models

All models are characterized by the mean α and the error terms. In addition to these terms, any model described so far will have either random effects terms or fixed effects terms and accordingly are referred to as random or fixed models. Models that have both fixed effects terms and random effects terms are known as 'mixed effects models'. Once the theory underlying fixed and random effects models has been fully understood, the progression to understanding mixed models is very simple.

Elaborating on the original mice litter example, the six litters by each mouse were fed according to three different dietary treatments (**?**). Therefore a fixed effect ϕ_j has been added to the model, which is now formulated as follows;

$$y_{ij} = \mu + \delta_i + \phi_j + \gamma_{ij} + \epsilon_{ijk} \tag{1.6}$$

As before, an interaction effect γ_{ij} must also be added to the model. In cases where the interaction term describes the combined effect of fixed and random components, it should be treated as random effect. The variance of the above model is composed of the σ^2_δ, σ^2_γ and σ^2_ϵ .

It may be shown that the interaction factors make no contribution to the outcome, i.e γ_{ij} is consistently calculated as zero. Considering the skin tumour example, a person's age would bear no relation to their gender and hence there would be plausible interaction between the two factors. Indeed , in keeping with the 'Law of Parsimony', factors should be specified such that each would convey separate information. However, interaction terms are extant when the model specifies repeated observations, as there is necessarily a relationship between observations from the same subject. Importantly, interaction effects, being random effects, are attended by variance component terms and therefore also contribute to the overall variance of the model.

? gives a mixed effects model formulation for the Grubbs artillery study. y_{ij} is the muzzle velocity of the ith shell, as measured by the jth chronometer.

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \tag{1.7}$$

In this formulation α_i is the random effect of round i, and the fixed effect component β_j is the bias in chronometer j. (Also, no interaction term is used).

1.3.1 Fixed or Random?

In the examples discussed so far, it is clear when effects are fixed or random. In general, however, the difference is not as clear. **?** discusses the decision whether an effect should be treated as random or fixed, stating that it depends upon the context of the study, and how the data was gathered. *The situation to which the model applies is the deciding factor in determining whether effects are fixed or random.*

Referring to the Grubbs data, the shells fired are a random sample of shells therefore α_i components should be considered random effects. Conversely β_j are fixed effects components because the three measurement devices are the only instruments of interest (**?**).

1.3.2 Advantages of Mixed Models

? discusses the following advantages of using mixed effects models. In the case of repeated measurements , it is appropriate to take account of the correlation of each group of observations. Mixed models lead to more appropriate estimates and standard errors for fixed effects, particularly in the case of repeated measures. Analysis using a mixed model is more appropriate for inference on a hierarchical data. In the case of unbalanced data, mixed models are more appropriate than other methodologies.

? comments that mixed models are the correct approach for dealing with grouped data. The use of linear mixed effects models has advanced greatly with increased usage of statistical software. This author also notes that mixed models are a hybrid of bayesian and frequentist methodologies and that mixed model approaches are more flexible than bayesian.

Unbalanced Data

Unbalanced data refers to situations where these groups are of different sizes. Mixed Effects Models are suitable for studying unbalanced data sets. The variance components of random effects for these set can not be derived using alternative methods such as ANOVA.

1.3.3 Matrix Formulation

There are matrix (i.e multivariate) formulations of both fixed effects models and random effects models. **?** remarks that the matrix notation makes the underlying theory of mixed effects models much easier to work with. The fixed effects models can be specified as follows;

$$\mathbf{Y} = \mathbf{X}\mathbf{b} + \mathbf{e} \tag{1.8}$$

$\mathbf{Y}$ is the vector of n observations, with dimension $n \times 1$. $\mathbf{b}$ is a vector of fixed p effects, and has dimension $p \times 1$. It is composed of coefficients, with the first element being the population mean. For the skin tumour example, with the three specified fixed effects, $p = 4$. $\mathbf{X}$ is known as the design 'matrix', model matrix for fixed effects, and comprises 0s or 1s, depending on whether the relevant fixed effects have any effect on the observation is question. $\mathbf{X}$ has dimension $n \times p$. $\mathbf{e}$ is the vector of residuals with dimension $n \times 1$.

The random effects models can be specified similarly. $\mathbf{Z}$ is known as the 'model matrix for random effects', and also comprises 0s or 1s. It has dimension $n \times q$. $\mathbf{u}$ is a vector of random q effects, and has dimension $q \times 1$.

$$\mathbf{Y} = \mathbf{Z}\mathbf{u} + \mathbf{e} \tag{1.9}$$

Again, once the component fixed effects and random effects components are considered, progression to a mixed model formulation is a simple step. Further to **?**, it is conventional to formulate a mixed effects model in matrix form as follows:

$$\mathbf{Y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e} \tag{1.10}$$

($E(\mathbf{u}) = 0$, $E(\mathbf{e}) = 0$ and $E(\mathbf{y}) = \mathbf{X}\mathbf{b}$)

1.4 Mixed Model Calculations

1.4.1 Formulation of the Variance Matrix V

$\mathbf{V}$, the variance matrix of $\mathbf{Y}$, can be expressed as follows;

$$\mathbf{V} = var(\mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}) \tag{1.11}$$

$$\mathbf{V} = var(\mathbf{X}\mathbf{b}) + var(\mathbf{Z}\mathbf{u}) + var(\mathbf{e})) \tag{1.12}$$

$var(\mathbf{X}\mathbf{b})$ is known to be zero. $var(\mathbf{Z}\mathbf{u})$ can be written as $Zvar(\mathbf{u})Z^T$. Z is a matrix of constants. By letting $var(u) = G$ (i.e $\mathbf{u}$ $N(0, \mathbf{G})$), this becomes ZGZ^T. This specifies the covariance due to random effects. The residual covariance matrix $var(e)$ is denoted as R, ($\mathbf{e}$ $N(0, \mathbf{R})$). Residual are uncorrelated, hence $\mathbf{R}$ is equivalent to $\sigma^2\mathbf{I}$, where $\mathbf{I}$ is the identity matrix. The variance matrix $\mathbf{V}$ can therefore be written as;

$$\mathbf{V} = ZGZ^T + \mathbf{R} \tag{1.13}$$

1.4.2 Estimators and Predictors

The best linear unbiased predictor (BLUP) is used to estimating random effects, i.e to derive **u**. The best linear unbiased estimator (BLUE) is used to estimate the fixed effects, **b**. They were formulated in a paper by **?**, which provides the derivations of both. Inferences about fixed effects have come tobe called 'estimates', whereas inferences about random effects have come be called 'predictions'. hence the naming of BLUP is to reinforce distinction between the two , but it is essentially the same principal involved in both cases, (**?**). The procedures are known as the 'best' in the sense that they minimise the sampling variance and unbiased in the sense that E[BLUE(**b**)]= **b** and E[BLUP(**u**)] = **u**. The BLUE of **b**, and the BLUP of **u** can be shown to be;

$$\hat{b} = (X^T V^{-1} X)^{-1} X^T V^{-1} y \tag{1.14}$$

$$\hat{u} = G Z^T V^{-1}(y - X\hat{b}) \tag{1.15}$$

The practical application of both expressions requires that the variance components be known. Therefore an estimate for the variance components must be derived to analysis by either ANOVA, or REML, a method that shall be introduced shortly. Importantly calculations based on the above formulae require the calculation of the inverse of **V**. In simple examples V^{-1} is a straightforward calculation, but with higher dimensions it becomes a very complex calculation.

1.4.3 Henderson's Mixed Model Equations

???? derived the 'mixed model equations (MME)' to provide estimates for **b** and **u**without the need to calculate the inverse of **V**.

$$\begin{pmatrix} X^T R^{-1} X & X^T R^{-1} Z \\ Z^T R^{-1} X & Z^T R^{-1} Z + G^{-1} \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{u} \end{pmatrix} = \begin{pmatrix} X^T R^{-1} y \\ Z^T R^{-1} y \end{pmatrix}. \tag{1.16}$$

When **R** and **G** are diagonal, determining the inverses thereof are trivial calculations, and therefore the above matrices are much simpler to solve , and overcomes the problem posed by the inverse of **V**.

Each of the elements of the above matrices are submatrices. $X^T R^{-1} X$ is a $p \times p$ matrix, $Z^T R^{-1} Z + G^{-1}$ is a $q \times q$ matrix. The remaining elements, which are transposes of each other, are of dimensions $p \times q$ and $q \times p$ respectively. Therefore the overall matrix is of dimension $(p + q) \times (p + q)$. These dimensions are notably smaller than $n \times n$, which would have been the case if V^{-1}, and therefore the inversion is easier to compute.

1.5 Estimability of Fixed Effects

Potentially it may be impossible to compute unique BLUE estimates for all the fixed factors in a model. This may be due to linear dependence in the model matrix $\mathbf{X}$. Consider the following example;

$$\mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ b_3 \end{pmatrix} \tag{1.17}$$

HERE

1.6 Maximum Likelihood Estimation

Maximum likelihood (ML) estimation is a well known method of obtaining estimates of unknown parameters by optimizing a likelihood function. Models fitted by ML estimation can be compared using the likelihood ratio test. However ML is known to underestimate variance components for finite samples (**?**).

1.6.1 Restricted Likelihood Estimation

A method related to ML is restricted maximum likelihood estimation(REML). REML was developed by **?** and **?** to provide unbiased estimates of variance and covariance parameters. REML obtains estimates of the fixed effects using non-likelihoodlike methods, such as ordinary least squares or generalized least squares, and then using these estimates it maximizes the likelihood of the residuals (subtracting off the fixed effects) to obtain estimates of the variance parameters. In most software packages REML is the default algorithm used to compute coefficients for the predictor variables. REML estimation reduces the bias in the variance component, and also handles high correlations more effectively, and is less sensitive to outliers than ML.

? describes two important outcomes of using REML. Firstly variance components can be estimated without being affected by fixed effects. Secondly in estimating variance components with REML, degrees of freedom for the fixed effects can be taken into account implicitly, whereas with ML they are not. When estimating variance from normally distributed data, the ML estimator for σ^2 is $\frac{S_{yy}}{n}$ whereas the REML estimator is $\frac{S_{yy}}{n-1}$. (S_{yy} is the sum of square identity;

$$S_{yy} = \Sigma_{i=i}^{N}(y - \bar{y})^2 \tag{1.18}$$

Likelihood Ratio Tests

The problem with REML for model building is that the "likelihoods" obtained for different fixed effects are not comparable. Hence it is not valid to compare models with different fixed effects using a likelihood ratio test or AIC when REML is used to estimate the model. Therefore models derived using ML must be used instead.

1.7 Computation of LMEs using R

? advises how to implement LME models in statistical software (ostensibly for S and S PLUS, but R is very similar). When tackling linear mixed effects models using the R language, a statistician can call upon the *lme* command found in the *nlme* package.This command fits a LME model to the data set using either Maximum Likelihood (ML) or Restricted Maximum Likelihood (REML). ML may be referred to as 'full maximum likelihood' estimation.

The first two arguments for *lme* are *fixed* and *data*, which give the model for the expected responses (i.e. the fixed part of the model), and the data that themodel should be fitted from. The next argument is *random*, a one-sided formula which describes the random effects, and the grouping structure for the model. The *method* argument can specify whether to use 'REML', the default setting, or 'ML'.

? describes an experiment whereby the productivity of six randomly chosen workers are assessed three times on each of three machines, yielding the 54 observations in the following table.

Observation	Worker	Machine	score	Observation	Worker	Machine	score
1	1	A	52.00	28	4	B	63.20
2	1	A	52.80	29	4	B	62.80
3	1	A	53.10	30	4	B	62.20
4	2	A	51.80	31	5	B	64.80
5	2	A	52.80	32	5	B	65.00
6	2	A	53.10	33	5	B	65.40
7	3	A	60.00	34	6	B	43.70
8	3	A	60.20	35	6	B	44.20
9	3	A	58.40	36	6	B	43.00
10	4	A	51.10	37	1	C	67.50
11	4	A	52.30	38	1	C	67.20
12	4	A	50.30	39	1	C	66.90
13	5	A	50.90	40	2	C	61.50
14	5	A	51.80	41	2	C	61.70
15	5	A	51.40	42	2	C	62.30
16	6	A	46.40	43	3	C	70.80
17	6	A	44.80	44	3	C	70.60
18	6	A	49.20	45	3	C	71.00
19	1	B	62.10	46	4	C	64.10
20	1	B	62.60	47	4	C	66.20
21	1	B	64.00	48	4	C	64.00
22	2	B	59.70	49	5	C	72.10
23	2	B	60.00	50	5	C	72.00
24	2	B	59.00	51	5	C	71.10
25	3	B	68.60	52	6	C	62.00
26	3	B	65.80	53	6	C	61.40
27	3	B	69.70	54	6	C	60.50

Table 1.1: Machines Data , Pinheiro Bates

(Overall mean score = 59.65, mean on machine A = 52.35 , mean on machine B = 60.32, mean on machine C = 66.27)

The 'worker' factor is modelled with random effects(u_i), whereas the 'machine' factor is modelled with fixed effects (β_j). Due to the repeated nature of the data, interaction effects between these factors are assumed to be extant, and shall be examined accordingly. The interaction effect in this case (τ_{ij}) describes whether the effect of changing from one machine to another is different for each worker. The productivity score y_{ijk} is the kth observation taken on worker i on machine j, and is formulated as follows;

$$y_{ijk} = \beta_j + u_i + \tau_{ij} + \epsilon_{ijk} \tag{1.19}$$
$$u_i \sim N(0, \sigma^2_u), \epsilon_{ijk} \sim N(0, \sigma^2), \tau_i \sim N(0, \sigma^2_\tau)$$

The 'nlme' package is incorporated into the R programming to perform linear mixed model calculations. For the 'Machines' data, **?** use the following code, with the hierarchical structure specified in the last argument.

```
lme(score~Machine, data=Machines, random=~1|Worker/Machine)
```

The output of the R computation is given below.

```
Linear mixed-effects model fit by REML
  Data: Machines
  Log-restricted-likelihood: -107.8438
  Fixed: score ~ Machine
(Intercept)    MachineB    MachineC
  52.355556    7.966667   13.916667

Random effects:
 Formula: ~1 | Worker
        (Intercept)
StdDev:     4.78105

 Formula: ~1 | Machine %in% Worker
        (Intercept)  Residual
StdDev:    3.729532 0.9615771

Number of Observations: 54 Number of Groups:
             Worker Machine %in% Worker
                  6                  18
```

The crucial pieces of information given in the programme output are the estimates of the intercepts for each of the three machines. Machine A, which is treated as a control case, is estimated to have an intercept of 52.35. The intercept estimates for machines B and C are found to be 60.32 and 66.27 (by adding the values 7.96 and 13.91 to 52.35 respectively). Estimate for the variance components are also given; $\sigma^2_u = (4.78)^2$, $\sigma^2_\tau = (3.73)^2$ and $\sigma^2_\epsilon = (0.96)^2$.

1.8 Model Selection

1.8.1 Akaike Information Critierion

This is a model selection method, assessing how the goodness of fit of a model. It is computed as follows:

$$AIC = -2l_{max} + 2k$$

with l_{max} as the log-likelihood maximum and k as the number of parameters. The candidate model with the lowest AIC value is considered the best fitting of the candidate models.

?, p.13 reports that some researchers have noted that there is a bias present in AIC estimation, and have proposed alternative formulations to rectify it. It is also reported that AIC doesn't address the issue of multicollinearity sufficiently. **?** formulate an adaption; the Healthy AIC. It is constructed to overcome the issue of ill posed models. The HAIC will choose the candidate model with the shortest parameter vector length.

Chapter 2

Errata

2.1 Distinction between Classical Models and Mixed Models

? discusses the inadequacy of 'classical models' in analysing such data types, with particular reference to the simple linear model. The simple linear model is a well known statistical methodology that describes the relationship between dependent variables Y and an independent predictor variable X. Where $Y = y_1, y_2, ..y_k..y_n$ and $X = x_1, x_2, ..x_k..x_n$, an intercept α and slope β are estimated such that the error terms associated with each observation y_i is minimized.

$$y_k = \alpha + \beta x_k + \epsilon_k \tag{2.1}$$

In classical statistics a typical assumption is that observations are drawn from the same general population, are independent and identically distributed (**?**). Consequently there is no way to account for the grouped nature of data sets described previously, and so there lies the possibility of observations being treated as independent measurements. **?**, pg.3 gives a very informative example wherein a classical approach is compared to an approach that does account for grouping. The conclusion to be drawn from Demidenko's example is that failure to account for grouping leads to an incorrect conclusion about the data. The approach recommended is known as 'mixed models' and shall be introduced presently.

2.2 Maximum Likelihood Estimation

Maximum likelihood (ML) estimation is a well known method of obtaining estimates of unknown parameters by optimizing a likelihood function. Models fitted by ML estimation can be compared using the likelihood ratio test. However ML is known to underestimate variance components for finite samples (**?**).

2.2.1 Restricted Likelihood Estimation

A method related to ML is restricted maximum likelihood estimation(REML). REML was developed by **?** and **?** to provide unbiased estimates of variance and covariance parameters. REML obtains estimates of the fixed effects using non-likelihoodlike methods, such as ordinary least squares or generalized least squares, and then using these estimates it maximizes the likelihood of the residuals (subtracting off the fixed effects) to obtain estimates of the variance parameters. In most software packages REML is the default algorithm used to compute coefficients for the predictor variables. REML estimation reduces the bias in the variance component, and also handles high correlations more effectively, and is less sensitive to outliers than ML.

? describes two important outcomes of using REML. Firstly variance components can be estimated without being affected by fixed effects. Secondly in estimating variance components with REML, degrees of freedom for the fixed effects can be taken into account implicitly, whereas with ML they are not. When estimating variance from normally distributed data, the ML estimator for σ^2 is $\frac{S_{yy}}{n}$ whereas the REML estimator is $\frac{S_{yy}}{n-1}$. (S_{yy} is the sum of square identity;

$$S_{yy} = \Sigma_{i=i}^{N}(y - \bar{y})^2 \tag{2.2}$$

2.3 Fixed Effects and Random Effects Models

Before proceeding to a description of mixed models, an introduction to fixed effects and random effects models is required. This section follows on from the discussion of measurement error models in the last chapter.

2.3.1 Fixed Effects

? gives an example of a study where the observations , occurrences of skin tumours called basal cell epithelioma, were classified according to 'factors', i.e. the gender, age and exposure to sunshine of the patients. Levels are the individual classes of each of these factors (e.g. 'Male' and 'Female' would be the levels of the factor 'Gender'). The scientific interest lies in examining the extent to which different factor levels affect the variable of interest. The effects of a level of a factor are one of two types ; fixed effects and random effects. Fixed effects describe effects due to a finite set of levels for a factor (i.e multichotomous factors). The factors described in the skin tumour example are all fixed effects factors. Fixed effects models are the cases where only fixed effects are present, with the exception of random error terms.

To demonstrate fixed effects model **?** describes a study wherein 24 plants are divided into four groups of six, and each group is subjected to its own treatment regime. Three different fertilizers are used with three of the groups (treatments N,P,K), while no fertilizer is used on the fourth group, (i.e. it is a control group denoted as C). **?** constructs a model to describe the crop yield resultant from the experiment.

$$y_{ij} = \mu + \alpha_i + e_{ij} \tag{2.3}$$

where y_{ij} is the jth plant (i.e. crop yield) on the ith treatment, with μ as the mean yield, α_i is the effect of each fertilizer treatments (i.e. fixed effects) and e_{ij} is the error term. The fixed effect for each observation is an unknown constant that is to be determined from computing the data.

The μ term would not necessarily be present in all formulations. Some authors, such as **?**, may use a single term as equivalent to the μ and α terms. (It is customary to centre data prior to using mixed effects methodologies. Centering means subtracting the mean of the observations from each observed value, so mean of the resultant values become zero.)

2.3.2 Random Effects

The random effects model describes the case where there is an infinite number of levels in a factor. In other words the factor is a random variable. **?** demonstrates this with

a second example; a study of the maternal ability of mice. In this example 4 female mice , all of the same breed, have 6 litters each over a period of time. The weights w_{ij} of each litter (j) from each mouse (i) were taken to be the proxy for maternal ability and is formulated as follows;

$$w_{ij} = \mu + \delta_i + e_{ij} \tag{2.4}$$

As in the previous exmaple, α is the mean , e_{ij} is the error term and δ_i is a random effect due to each mouse. Notably these four mice are considered as a sample of the overall population of female mice of that breed, consequently an important characteristic of random effects models is that the δ_i values are a random sample of all δ terms. Therefore these random terms can be used for making inferences about populations. (**?**).

2.3.3 Variance Components

Each random effect has an associated 'variance component' term. This is a model parameter which quantifies random variation due to that effect only. Therefore for every observation there are two sources of variation, random variation and residual variation, and can be expressed as follows $var(y_{ij}) = \sigma_p^2 + \sigma^2$. These variations are known as the variance components. This is an important difference with fixed effects models, which is subject to residual variation (i.e σ^2) only (**?**).

In fixed effects models there is no covariance between any pair of observations. **?** shows that, while there is no covariance between observations from different subjects (i.e the mice in Searle's example), there exists correlation between observations from the same subject (i.e. litter weights from the same mouse are correlated). In this case the covariance is the subjects variance component (i.e. σ_p^2)

2.3.4 More complex examples

? offers elaborations on both examples used so far. In the case of the fixed effects model, the model can be amended to take account for different varieties of each plants being studied. (The groups of six plants are subdivided into three variety types.) y_{ijk} is the yield of the kth plant of the jth variety in the ith treatment, and is described as follows;

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + e_{ijk} \tag{2.5}$$

This new formulation includes a fixed effect β_j to account for the variety type, and an 'interaction effect' γ_{ij}. An interaction effect describes the combined effects of two or more variables on the observation. Similarly the random effects example is elaborated

to account for the effects of three different technicians. Again there is a random effect component τ to account for these technicians, and an interaction effect θ to account for the combined effect of the mice and the technicians.

$$w_{ijk} = \mu + \delta_i + \tau_j + \theta_{ij} + e_{ijk} \tag{2.6}$$

2.4 Mixed Models

All models are characterized by the mean α and the error terms. In addition to these terms, any model described so far will have either random effects terms or fixed effects terms and accordingly are referred to as random or fixed models. Models that have both fixed effects terms and random effects terms are known as 'mixed effects models'. Once the theory underlying fixed and random effects models has been fully understood, the progression to understanding mixed models is very simple.

Elaborating on the original mice litter example, the six litters by each mouse were fed according to three different dietary treatments (**?**). Therefore a fixed effect ϕ_j has been added to the model, which is now formulated as follows;

$$y_{ij} = \mu + \delta_i + \phi_j + \gamma_{ij} + \epsilon_{ijk} \tag{2.7}$$

As before, an interaction effect γ_{ij} must also be added to the model. In cases where the interaction term describes the combined effect of fixed and random components, it should be treated as random effect. The variance of the above model is composed of the σ^2_δ, σ^2_γ and σ^2_ϵ .

It may be shown that the interaction factors make no contribution to the outcome, i.e γ_{ij} is consistently calculated as zero. Considering the skin tumour example, a person's age would bear no relation to their gender and hence there would be plausible interaction between the two factors. Indeed , in keeping with the 'Law of Parsimony', factors should be specified such that each would convey separate information. However, interaction terms are extant when the model specifies repeated observations, as there is necessarily a relationship between observations from the same subject. Importantly, interaction effects, being random effects, are attended by variance component terms and therefore also contribute to the overall variance of the model.

? gives a mixed effects model formulation for the Grubbs artillery study. y_{ij} is the muzzle velocity of the ith shell, as measured by the jth chronometer.

$$y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \tag{2.8}$$

In this formulation α_i is the random effect of round i, and the fixed effect component β_j is the bias in chronometer j. (Also, no interaction term is used).

2.4.1 Fixed or Random?

In the examples discussed so far, it is clear when effects are fixed or random. In general, however, the difference is not as clear. **?** discusses the decision whether an effect should be treated as random or fixed, stating that it depends upon the context of the study, and how the data was gathered. *The situation to which the model applies is the deciding factor in determining whether effects are fixed or random.*

Referring to the Grubbs data, the shells fired are a random sample of shells therefore α_i components should be considered random effects. Conversely β_j are fixed effects components because the three measurement devices are the only instruments of interest (**?**).

2.4.2 Advantages of Mixed Models

? discusses the following advantages of using mixed effects models. In the case of repeated measurements , it is appropriate to take account of the correlation of each group of observations. Mixed models lead to more appropriate estimates and standard errors for fixed effects, particularly in the case of repeated measures. Analysis using a mixed model is more appropriate for inference on a hierarchical data. In the case of unbalanced data, mixed models are more appropriate than other methodologies.

? comments that mixed models are the correct approach for dealing with grouped data. The use of linear mixed effects models has advanced greatly with increased usage of statistical software. This author also notes that mixed models are a hybrid of bayesian and frequentist methodologies and that mixed model approaches are more flexible than bayesian.

Unbalanced Data

Unbalanced data refers to situations where these groups are of different sizes. Mixed Effects Models are suitable for studying unbalanced data sets. The variance components of random effects for these set can not be derived using alternative methods such as ANOVA.

2.4.3 Matrix Formulation

There are matrix (i.e multivariate) formulations of both fixed effects models and random effects models. **?** remarks that the matrix notation makes the underlying theory of mixed effects models much easier to work with. The fixed effects models can be specified as follows;

$$\mathbf{Y} = \mathbf{X}\mathbf{b} + \mathbf{e} \tag{2.9}$$

$\mathbf{Y}$ is the vector of n observations, with dimension $n \times 1$. $\mathbf{b}$ is a vector of fixed p effects, and has dimension $p \times 1$. It is composed of coefficients, with the first element being the population mean. For the skin tumour example, with the three specified fixed effects, $p = 4$. $\mathbf{X}$ is known as the design 'matrix', model matrix for fixed effects, and comprises 0s or 1s, depending on whether the relevant fixed effects have any effect on the observation is question. $\mathbf{X}$ has dimension $n \times p$. $\mathbf{e}$ is the vector of residuals with dimension $n \times 1$.

The random effects models can be specified similarly. $\mathbf{Z}$ is known as the 'model matrix for random effects', and also comprises 0s or 1s. It has dimension $n \times q$. $\mathbf{u}$ is a vector of random q effects, and has dimension $q \times 1$.

$$\mathbf{Y} = \mathbf{Z}\mathbf{u} + \mathbf{e} \tag{2.10}$$

Again, once the component fixed effects and random effects components are considered, progression to a mixed model formulation is a simple step. Further to **?**, it is conventional to formulate a mixed effects model in matrix form as follows:

$$\mathbf{Y} = \mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e} \tag{2.11}$$

($E(\mathbf{u}) = 0$, $E(\mathbf{e}) = 0$ and $E(\mathbf{y}) = \mathbf{X}\mathbf{b}$)

2.5 Mixed Model Calculations

2.5.1 Formulation of the Variance Matrix V

$\mathbf{V}$, the variance matrix of $\mathbf{Y}$, can be expressed as follows;

$$\mathbf{V} = var(\mathbf{X}\mathbf{b} + \mathbf{Z}\mathbf{u} + \mathbf{e}) \tag{2.12}$$

$$\mathbf{V} = var(\mathbf{X}\mathbf{b}) + var(\mathbf{Z}\mathbf{u}) + var(\mathbf{e})) \tag{2.13}$$

$var(\mathbf{X}\mathbf{b})$ is known to be zero. $var(\mathbf{Z}\mathbf{u})$ can be written as $Zvar(\mathbf{u})Z^T$. Z is a matrix of constants. By letting $var(u) = G$ (i.e $\mathbf{u}$ $N(0, \mathbf{G})$), this becomes ZGZ^T. This specifies the covariance due to random effects. The residual covariance matrix $var(e)$ is denoted as R, ($\mathbf{e}$ $N(0, \mathbf{R})$). Residual are uncorrelated, hence $\mathbf{R}$ is equivalent to $\sigma^2\mathbf{I}$, where $\mathbf{I}$ is the identity matrix. The variance matrix $\mathbf{V}$ can therefore be written as;

$$\mathbf{V} = ZGZ^T + \mathbf{R} \tag{2.14}$$

2.5.2 Estimators and Predictors

The best linear unbiased predictor (BLUP) is used to estimating random effects, i.e to derive **u**. The best linear unbiased estimator (BLUE) is used to estimate the fixed effects, **b**. They were formulated in a paper by **?**, which provides the derivations of both. Inferences about fixed effects have come tobe called 'estimates', whereas inferences about random effects have come be called 'predictions'. hence the naming of BLUP is to reinforce distinction between the two , but it is essentially the same principal involved in both cases, (**?**). The procedures are known as the 'best' in the sense that they minimise the sampling variance and unbiased in the sense that E[BLUE(**b**)]= **b** and E[BLUP(**u**)] = **u**. The BLUE of **b**, and the BLUP of **u** can be shown to be;

$$\hat{b} = (X^T V^{-1} X)^{-1} X^T V^{-1} y \tag{2.15}$$

$$\hat{u} = G Z^T V^{-1} (y - X\hat{b}) \tag{2.16}$$

The practical application of both expressions requires that the variance components be known. Therefore an estimate for the variance components must be derived to analysis by either ANOVA, or REML, a method that shall be introduced shortly. Importantly calculations based on the above formulae require the calculation of the inverse of **V**. In simple examples V^{-1} is a straightforward calculation, but with higher dimensions it becomes a very complex calculation.

2.5.3 Henderson's Mixed Model Equations

???? derived the 'mixed model equations (MME)' to provide estimates for **b** and **u**without the need to calculate the inverse of **V**.

$$\begin{pmatrix} X^T R^{-1} X & X^T R^{-1} Z \\ Z^T R^{-1} X & Z^T R^{-1} Z + G^{-1} \end{pmatrix} \begin{pmatrix} \hat{\beta} \\ \hat{u} \end{pmatrix} = \begin{pmatrix} X^T R^{-1} y \\ Z^T R^{-1} y \end{pmatrix}. \tag{2.17}$$

When **R** and **G** are diagonal, determining the inverses thereof are trivial calculations, and therefore the above matrices are much simpler to solve , and overcomes the problem posed by the inverse of **V**.

Each of the elements of the above matrices are submatrices. $X^T R^{-1} X$ is a $p \times p$ matrix, $Z^T R^{-1} Z + G^{-1}$ is a $q \times q$ matrix. The remaining elements, which are transposes of each other, are of dimensions $p \times q$ and $q \times p$ respectively. Therefore the overall matrix is of dimension $(p + q) \times (p + q)$. These dimensions are notably smaller than $n \times n$, which would have been the case if V^{-1}, and therefore the inversion is easier to compute.

Likelihood Ratio Tests

The problem with REML for model building is that the "likelihoods" obtained for different fixed effects are not comparable. Hence it is not valid to compare models with different fixed effects using a likelihood ratio test or AIC when REML is used to estimate the model. Therefore models derived using ML must be used instead.