# Chapter 7
# Introduction to Mixed Models

*Julius van der Werf*

Linear models are commonly used to describe and analyse data in the biological and social sciences. The model needs to represent the sampling nature of the data.
The data vector contains measurements on experimental units. The observations are random variables that follow a multivariate distribution.
The model usually consists of factors. These are variables, either discrete or continuous, which have an effect on the observed data. Different model factors are:

- Discrete factors or class variables such as sex, year, herd
- Continuous factors or covariables such as age

Some factors are of special interest to the researcher but other factors have to be included in the model simply because they explain a significant part of the variation in the data and reduce the residual (unexplained) variation. Such factors are often called 'nuisance variables'.

*Fixed and random effects*

Another distinction that is often used is that between fixed and random effects. A mixed model contains both fixed and random effects (hence 'mixed').

The statistical world is somewhat divided here in more traditional 'frequentists' that make this distinction and Bayesians' that find this distinction artificial and accommodate the properties of different factors in their model specification. However, it is still useful to try to define the difference between fixed and random effects, and acknowledge this dispute. We will discuss this distinction in more detail in a later lecture on Bayesian methods. With the development of mixed models and BLUP methodology, there has always been a clear distinction between random and fixed effects, not in the least due to the influence of C.R. Henderson, the founder of BLUP, and we'll follow that approach here, in first instance.
*Fixed Effects*

- Effects for which the defined classes comprise all the possible levels of interest, eg. sex, age, breed, contemporary group. Effects can be considered as fixed when the number of levels are relatively small and is confined to this number after repeated sampling.

*Random Effects*

- Effects which have levels that are considered to be drawn from an infinite large population of levels. Animal effects are often random. In repeated experiments there maybe other animals drawn from the population.

The distinction is also often determined by the purpose of the experiment. Do we want to know the difference between these specific levels of a factor, or are we interested in how large the differences between levels of a factor might generally be. The effect of management groups could be fixed but arguments for considering them as random could be found just as easily.

*Example* A growth trial for a number of animals from different age groups used several different diets, locations and handlers.

In this case the number of levels for age, diet, location and handler could all conceptually be the same for an infinite number of sampling events. On the other hand different animals would be needed for each repeated sample as the same growth phase could not be repeated in the same animal. Furthermore inferences might be made about diets or locations in general and in this case these effects might be considered random since these could have been sampled from an infinite number of levels. Therefore animals effects would be considered random while all other effects would generally be fixed.

*A checklist that can be used for deciding about fixed or random effects:*

i) What are the number of levels?

| | | |
|---|---|---|
| small | - | fixed |
| large or near infinite | - | possibly random |

ii) Are the levels repeatable?

| | | |
|---|---|---|
| yes | - | fixed |
| no | - | random |

iii) Are there conceptually and infinite number of such levels?

| | | |
|---|---|---|
| yes | - | possibly fixed |
| no | - | possibly random |

iv) Are inferences to be made about levels not included in the sampling?

| | | |
|---|---|---|
| yes | - | possibly random |
| no | - | possibly fixed |

v) Were the levels of the factor determined in a non-random manner?

| | | |
|---|---|---|
| yes | - | possibly random |
| no | - | possibly fixed |

A linear relationship can generally be found to fit most biological data although some transformation may be required. Thus a linear model can generally be used to describe data. All models contain a set of factors composed of three parts which additively affect the observations or records of data:

i) the equation
ii) expectations and variance covariance matrices of random variables
iii) assumptions, limitations and restrictions

**Models:**

Fixed Model $y = X\beta + e$

Random Model $y = Zu + e$

Mixed Model $y = X\beta + Zu + e$

## The Equation

The equation of a model defines the factors that will or could have an effect on an observed trait. The general linear model equation in matrix form is

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e} \qquad \text{...(1)}$$

where

- $\mathbf{y}$ is an $n \times 1$ vector of $n$ observed records
- $\mathbf{b}$ is a $p \times 1$ vector of $p$ levels of fixed effects
- $\mathbf{u}$ is a $q \times 1$ vector of $q$ levels of random effects
- $\mathbf{e}$ is an $n \times 1$ vector of random, residual terms
- $\mathbf{X}$ is a known *design matrix* of order $n \times p$, which relates the records in $\mathbf{y}$ to the fixed effects in $\mathbf{b}$
- $\mathbf{Z}$ is a known *design matrix* of order $n \times q$, which relates the records in $\mathbf{y}$ to the random effects in $\mathbf{u}$

Equation (1) is generally termed a *mixed model* as it contains both fixed and random effects. While not specified directly, interactions between fixed effects are fixed, interactions between random effects are random and interactions between fixed and random effects are random. The mixed model can be reduced to become a fixed effect model by not including **Zu** or a random effects model for which no fixed effects are fitted except the overall mean, i.e. $\mathbf{Xb} = 1m$. In a sense, every model is a mixed mdel as µ is usually fixed and e is usually random.

**Expectations and Variance Covariance (VCV) Matrices**

In general the expectation of **y** is

$$E\begin{pmatrix} y \\ u \\ e \end{pmatrix} = \begin{pmatrix} Xb \\ 0 \\ 0 \end{pmatrix}$$
...(2)

which is also known as the 1st moment. The 2nd moments describe the variance-covariance structure of y:

$$V\begin{pmatrix} u \\ e \end{pmatrix} = \begin{pmatrix} G & 0 \\ 0 & R \end{pmatrix} \qquad \text{...(3)}$$

where **G** is a dispersion matrix for random effects other than errors and **R** is the dispersion matrix of error terms, for which both are general square matrices assumed to be non-singular and positive definite, with elements that are assumed known.
We usually write

$$\mathbf{V} = \mathbf{ZGZ'} + \mathbf{R}$$

**Assumptions, Limitations and Restrictions**

This part of the model identifies any differences between the operational and ideal models. It may describe the sampling process and to which extend the assumptions that are made can be expected to be true (e.g. about normality, random sampling, uncorrelated error terms, equally distributed error terms, etc).

# Estimation Theory

**Some terminology:**

***Estimation***: We don't know those prior moments (means, variance-covariance, etc), we estimate them. Estimating fixed effect, β;

***Prediction***: We know Variance, covariance, means to calculate other items we want to know. We predict random effects, u;

***Predictand*****:** The quantity to be predicted;

Y = Xβ + Zu + e  u is the quantity we want to predict, it can be said to e a predictand,

***Predictor*****:** the function used to predict the predictand, a linear function of y;

***Best*****:** minimise the mean squared error of the predictor
E$(\hat{u}-u)^2$ to minimised, I use the symbol ~ referring estimated or predicted, to distinguish it from the true value, e.g. $\hat{u}$ is estimated or predicted value of the true value u.

***Linear*****:** set the predictor to be a linear function of y,

***Unbiased*****:** the expectation is the expectation of predictand,  E($\hat{u}$) = E(u).
or a stronger criterion of unbiasedness: û = E(u| û).

We will briefly repeat aspect of the previous lecture on estimation o fixed effects, but now keeping in mind that there are more random effects than just the residual error (which is usually IID)

### Estimating Fixed Effects

Consider a general model

$$y = Xb + \varepsilon$$

$$\text{with } E(y) = Xb \quad \text{and} \quad var(y) = V = var(\varepsilon)$$

We want to estimate fixed effects in b and conduct hypothesis testing about the significance of differences between the different levels of effects. Note that ε is a vector with random effects. They can be caused by several random factors (e.g. animal and residual) and the different levels may be correlated (e.g. due to repeated measurements on the same animals). Hence, var(ε) maybe equal to V = ZGZ'+R.

**Ordinary Least Squares**

The general fixed effects model in matrix form is

$$y = Xb + e \qquad ...(5)$$

For ordinary least squares (OLS), error terms in e are independently and identically distributed random variables with a mean of zero and a variance of $\sigma_e^2$ such that var(y) = var(e) = $I_N\,\sigma_e^2$ where $I_N$ is a dispersion matrix for *n* observations. Given that E(y) = Xb, and the normal equations are X'Xb = X'y providing the least square estimator of $b$:

$$\hat{b} = (X'X)^{-1}\,X'y \qquad ...(6)$$

Thus the OLS approach assumes that all observations are uncorrelated and have a common variance $\sigma_e^2$. If estimates are derived when this is not true then they are no longer 'best'.

**Deriving Estimates Using Generalised Least Squares**

When the variance among observations is determined by more than uncorrelated residuals with equal variance, we write more generally var(e) = V where can be diagonal:

$$V = \begin{bmatrix} w_1 & & & & \\ & w_2 & & 0 & \\ & & .. & & \\ & 0 & & .. & \\ & & & & w_n \end{bmatrix} \sigma_e^2$$

Leading to *weighted least squares* (*WLS*): $\hat{b} = (X'V^{-1}X)^{-}\,X'V^{-1}Y$

Alternatively V might be non-diagonal and contain variance components such that

$$V = \begin{bmatrix} v_1 & & & & \\ & v_2 & & ij & \\ & & .. & & \\ & ij & & .. & \\ & & & & v_n \end{bmatrix}$$

where $v_i$ is the variance of the $i^{th}$ observation and *ij* are off diagonal elements and are the covariances between them. An example case would be for observations on groups of half sibs such that there would be covariances between measurements due to common sires, or repeated measurements with covariances between repeated observations on the same subject. In most genetic models there is a second random effect (besides error) and there are covariances among the random terms (e.g. due to genetic relationships). Therefore V is generally not diagonal in genetic analysis. This case is conventionally known as *generalised least squares* (*GLS*) where OLS and WLS are merely special cases of GLS. The generalised least squares criterion for simple linear regression is

$$Q_G = Q = (y - X\hat{b})'V^{-1}(y - X\hat{b})$$

Minimising $Q_G$ with respect to $b_0$ and $b_1$ leads to the appropriate normal equations of

$$(X'V^{-1}X)\ \hat{b} = X'V^{-1}Y$$

Determining a generalised inverse for $X'V^{-1}X$ gives the GLS estimates as

$$\hat{b} = (X'V^{-1}X)^{-} X'V^{-1}Y \qquad ...(7)$$

which is a general equation for Best linear Unbiased Estimates of fixed effects model in any linear model.
Note that GLS estimates are better than LS, as the covariance strcture among observations has better been taken into account. Therefore, to estimate fixed effects in a trial where animals maybe genetically related, it would be better to use a mixed model tha a fixed model where individual animal effects including their covariances are ignore. Another good example of possible covariances is among repeated measurements on the same animals. Often, nutritionists measure treatments on animals where the same animal is repeatedly measured. Ignoring covariances among repeated measurements on the same animal would provide a too rosy picture on the accuracy of the estimates obtained (see exercise 1 for an illustration).

**Best Linear Prediction (BLP) of random effects**

Given the model $\qquad y = \mu + u + e$

And the first and second moments (without any assumption of normality).

$$E\begin{bmatrix} u \\ y \end{bmatrix} = \begin{bmatrix} m_u \\ m_y \end{bmatrix} \qquad\qquad Var\begin{bmatrix} u \\ y \end{bmatrix} = \begin{bmatrix} G & C \\ C' & V \end{bmatrix}$$

Then minimising the mean squared error of prediction for $\hat{u}$, , i.e. it minimizes
$S(u-\hat{u})^2$ achieved by

$$BLP(u) = \hat{u} = \mu_u + CV^{-1}(y - \mu_y) \qquad (9)$$

The best predictor of a predictand is the conditional mean of predictor given data , y
It can be written as

$$E(u|y).$$

For example, the milk yield of a daughter j of a bull i is $y_{ij}$, it can expressed as

$$y_{ij} = \mu + s_i + e_{ij}$$

where s is the effect of the bull's breeding value on its progeny (i.e. s = BV/2), we can have first and the second moments (Mean and Variance), that leads to $s_i$ and $y_i$ being jointly distributed with a bivariate normal density having mean and variance,
Without any assumption of normality.

$$E\begin{bmatrix} s_i \\ y_{i.} \end{bmatrix} = \begin{bmatrix} 0 \\ m \end{bmatrix}$$

$$Var\begin{bmatrix} s_i \\ y_{i.} \end{bmatrix} = \begin{bmatrix} s_s^2 & s_s^2 \\ s_s^2 & s_s^2 + s_e^2 \end{bmatrix}$$

Where $s_s^2$ is the variance among sires and $s_e^2$ is the residual variance within a progeny group of a sire.

From the property of the bivariate normal distribution, we have the conditional expectation of $s_i$, given the men of its progeny $\bar{y}_i$

$$E(s_i|\bar{y}_i)=E(s_i)+Cov(s_i, \bar{y}_i)[Var(\bar{y}_i)]^{-1} [\bar{y}_i - E(\bar{y}_i.)]$$

which is

$$E(s_i \mid y_{i.}) = \frac{s_s^2}{s_s^2 + s_e^2/n}[\bar{y}_{i.} - E(\bar{y}_{i.})]$$

Note again that normality is not required here. So our predictor of $s_i$ is

$$\hat{s}_i = \frac{ns_s^2}{ns_s^2 + s_e^2}(\bar{y}_{i.} - m)$$

Where n is the number of daughters of the bull i.

This equation can rewritten as

$$\hat{s}_i = \frac{n}{n + s_e^2/s_s^2}(\bar{y}_{i.} - m) = \frac{n}{n+a}(\bar{y}_{i.} - m)$$

Where $\alpha = s_e^2/s_s^2 = (1-h^2/4) / (h^2/4) = (4-h^2)/h^2$

Note that this is the same regression of breeding value on information source as is applied in selection index theory.

## Mixed Model Estimation

As presented previously the mixed linear model in matrix form is

$$y = Xb + Zu + e$$

Recall that G is the VCV matrix of *u* and R is the VCV matrix of *e* such that

$$V = V(y) = V(Zu + e) = ZGZ' + R \qquad ...(8)$$

Note that if R was reduced to its simplest form, namely $I\sigma^2_e$ and $u$ was ignored, the mixed model equation would reduce to the standard linear model (5).

If G and R are known, estimates of $b$ and the predicted value of $u$ are

$$\hat{b} = (X'V^{-1}X)^{-} X'V^{-1}y$$

$$\hat{u} = GZ'V^{-1}(y - Xb)$$

which as a result of V given in (8) means that these effects have been estimated simultaneously and thus

i) $\hat{b}$ is the GLS solution for $b$ as well as its best linear unbiased estimator (BLUE)

ii) $\hat{u}$ is the best linear unbiased predictor (BLUP) of $u$

Henderson(1959) developed a set of equation that simultaneously generate BLUE($X\beta$) and BLUP(u), these equation being called mixed model equations: MME.

$$\begin{bmatrix} X'R^{-1}X & X'R^{-1}Z \\ Z'R^{-1}X & Z'R^{-1}Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X'R^{-1}y \\ Z'R^{-1}y \end{bmatrix}$$

**Proof of MME to give BLUE and BLUP:**

There are two subsets of equations:

$$X'R^{-1}X\hat{b} + X'R^{-1}Z\hat{u} = X'R^{-1}y \qquad (9)$$

$$Z'R^{-1}X\hat{b} + (Z'R^{-1}Z + G^{-1})\hat{u} = Z'R^{-1}y \qquad (10)$$

Substituting for $\hat{u}$ gives

$$X'R^{-1}X\hat{b} + X'R^{-1}Z(Z'R^{-1}Z+G^{-1})^{-1}Z'R^{-1}(y - X\hat{b}) = X'R^{-1}y$$

$$X'R^{-1}X\hat{b} - X'R^{-1}Z(Z'R^{-1}Z+G^{-1})^{-1}Z'R^{-1}X\hat{b}$$

$$= X'R^{-1}y - X'R^{-1}Z(Z'R^{-1}Z+G^{-1})^{-1}Z'R^{-1}y$$

$$\rightarrow X'(R^{-1} - X'R^{-1}Z(Z'R^{-1}Z+G^{-1})^{-1}Z'R^{-1})X\hat{b}$$

$$= X'(R^{-1} - X'R^{-1}Z(Z'R^{-1}Z+G^{-1})^{-1}Z'R^{-1})y$$

$$\rightarrow \qquad X'V^{-1}X\hat{b} = X'V^{-1}y$$

where $V^{-1} = R^{-1} - R^{-1}Z(Z'R^{-1}Z+G^{-1})^{-1}Z'R^{-1}$

This can be shown by proving that $VV^{-1} = I$ (Mrode, Appendix C) writing $W = (Z'R^{-1}Z+G^{-1})^{-1}$

$V^{-1} = R^{-1} - R^{-1}ZWZ'R^{-1}$

and $VV^{-1} =$

$ZGZ'(R^{-1} - R^{-1}ZWZ'R^{-1}) + R(R^{-1} - R^{-1}ZWZ'R^{-1}) =$

$ZGZ'R^{-1} + RR^{-1} - ZGZ'R^{-1}ZWZ'R^{-1} + RR^{-1}ZWZ'R^{-1} =$

$ZGZ'R^{-1} + I - Z(GZ'R^{-1}Z + I)WZ'R^{-1} =$

$ZGZ'R^{-1} + I - ZG(Z'R^{-1}Z + G^{-1})WZ'R^{-1} =$

$ZGZ'R^{-1} + I - ZGW^{-1}WZ'R^{-1} =$

$ZGZ'R^{-1} + I - ZGZ'R^{-1} = I$

Thus showing that the mixed model solution for b is a GLS estimate.

The proof that the MME provide a BLUP solution for u:

From (9) and (10), the MME give as solution:

$(Z'R^{-1}Z+G^{-1})\ \hat{u} = ZR^{-1}(y - X\hat{b})$

$\rightarrow \quad \hat{u} = (Z'R^{-1}Z+G^{-1})^{-1} ZR^{-1}(y - X\hat{b})$

$\rightarrow \quad = W ZR^{-1}(y - X\hat{b})$

Henderson (1963) proved that this is equal to the BLUP estimate $GZ'V^{-1}(y - X\hat{b})$

as $\quad GZ'V^{-1}(y - X\hat{b})$

$$\begin{aligned}
&= GZ'(R^{-1} - R^{-1}ZWZ'R^{-1})(y - X\hat{b}) \\
&= G(Z'R^{-1} - ZR^{-1}ZWZ'R^{-1})(y - X\hat{b}) \\
&= G(I - Z'R^{-1}ZW)Z'R^{-1}(y - X\hat{b}) \\
&= G(W^{-1} - Z'R^{-1}Z)WZ'R^{-1}(y - X\hat{b}) \\
&= G(Z'R^{-1}Z+G^{-1}) - Z'R^{-1}Z)WZ'R^{-1}(y - X\hat{b}) \\
&= (GZ'R^{-1}Z + I - GZ'R^{-1}Z)WZ'R^{-1}(y - X\hat{b}) \\
&= (I)WZ'R^{-1}(y - X\hat{b}) \\
&= WZ'R^{-1}(y - X\hat{b})
\end{aligned}$$

Hence, in the MME we estimate

BLUE (b) = $\hat{b} = (X'V^{-1}X)^{-1}X'V^{-1}y$ is a GLS estimate, and

BLUP (u)= $\hat{u} = (Z'R^{-1}Z+G^{-1})^{-1}Z'R^{-1}(y-X\hat{b})$

which is equal to

$GZ'(ZGZ' + R)^{-1}(y-X\hat{b}) \qquad = GZ'V^{-1}(y-X\hat{b})$

The solution for û is identical to the selection index equation, except that *b* is replaced by its estimate $\hat{b}$. Note the difference between estimating a fixed effect (GLS) and a random effect (BLUP). The first is estimated at its (corrected) mean whereas the second is regressed towards zero, depending on the amount of information available (or better, depending n the ratio of covariance and variance of the information used).

# Appendix:

**Useful Identities for Variances and covariances**

(1) $\sigma(\mathbf{x},\mathbf{x}) = \sigma^2{}_{\mathbf{x}}$

(2) $\sigma(\mathbf{x},\mathbf{a}) = 0$

(3) $\sigma(\mathbf{ax},\mathbf{y}) = \mathbf{a}\sigma_{\mathbf{xy}}$

(4) $\sigma(\mathbf{ax},\mathbf{by}) = \mathbf{ab}\sigma_{\mathbf{xy}}$

(5) $\sigma[(\mathbf{a}+\mathbf{x}),\mathbf{y}] = \sigma_{\mathbf{xy}}$

(6) $\sigma^2{}_{\mathbf{ax}} = \mathbf{a}^2\sigma^2{}_{\mathbf{x}}$

(7) $\sigma^2(\mathbf{x}-\mathbf{y}) = \sigma^2{}_{\mathbf{x}} + \sigma^2{}_{\mathbf{y}} - 2\sigma_{\mathbf{xy}}$

(8) $\sigma^2(\mathbf{x}+\mathbf{y}) = \sigma^2{}_{\mathbf{x}} + \sigma^2{}_{\mathbf{y}} + 2\sigma_{\mathbf{xy}}$

(9) $\sigma(\mathbf{x},\mathbf{x}-\mathbf{y}) = \sigma^2{}_{\mathbf{x}} - \sigma_{\mathbf{xy}}$

(10) $\sigma[(\mathbf{x}_1+\mathbf{x}_2),(\mathbf{y}_1+\mathbf{y}_2)] = \sigma_{\mathbf{x}_1\mathbf{y}_1} + \sigma_{\mathbf{x}_1\mathbf{y}_2} + \sigma_{\mathbf{x}_2\mathbf{y}_1} + \sigma_{\mathbf{x}_2\mathbf{y}_2}$

(11) $\sigma[(\mathbf{x}_1-\mathbf{x}_2),(\mathbf{y}_1-\mathbf{y}_2)] = \sigma_{\mathbf{x}_1\mathbf{y}_1} - \sigma_{\mathbf{x}_1\mathbf{y}_2} - \sigma_{\mathbf{x}_2\mathbf{y}_1} + \sigma_{\mathbf{x}_2\mathbf{y}_2}$

Where x and y are variables and a is constant