Chapter 6
Linear Mixed Model and Best Linear Unbiased Prediction

Linear mixed model and best linear unbiased prediction (BLUP) were first developed by Henderson (1948) to estimate genetic parameters and genetic merits of animals. In general, BLUP is a method of estimating random effects.

1. Mixed model for a split-plot experiment

Before we introduce mixed model and BLUP, let us look at another split-plot design (xxx lecture notes) to see the advantage of mixed model over the generalized linear model in handling a complicated model. Five varieties of spring wheat were sown in a randomized blocks design in four blocks. The soil was treated with three different levels of nitrogen randomly allocated to equal areas within each plot. The design and yield in tons/ha were shown in **Figure 1**.

Figure 1
Split-plot design of the wheat experiment

	V2	V5	V1	V4	V3
	N1 N3 N2	N2 N3 N1	N1 N2 N3	N1 N3 N2	N2 N1 N3
Block 1	4.6 5.5 5.3	5.0 5.4 4.7	5.5 6.1 6.4	5.0 6.0 5.7	5.5 4.9 5.8
	V1	V3	V2	V5	V4
	N3 N1 N2	N1 N3 N2	N3 N2 N1	N2 N3 N1	N2 N1 N3
Block 2	5.8 5.0 5.5	4.9 5.5 5.4	5.4 5.0 4.7	4.6 5.0 4.2	6.2 5.7 6.5
	V5	V1	V3	V2	V4
	N2 N3 N1	N2 N3 N1	N3 N1 N2	N1 N3 N2	N1 N3 N2
Block 3	4.8 5.0 4.6	5.4 5.9 5.0	5.5 4.8 4.7	5.0 5.8 5.1	5.3 6.7 5.8
	V2	V3	V4	V1	V5
	N3 N1 N2	N3 N2 N1	N2 N1 N3	N1 N2 N3	N3 N2 N1
Block 4	5.9 5.0 5.6	4.8 4.6 4.0	5.1 4.7 5.4	5.2 5.5 5.8	5.2 4.8 4.4

A whole plot is represented by each block×variety combination. The three nitrogen levels are allocated within each whole plot. Therefore, each level of the nitrogen occupies a split-plot (subplot). The model is

$$y_{ijk} = \mu + B_i + V_j + \delta_{ij} + N_k + (VN)_{jk} + e_{ijk} \qquad (1)$$

where

y_{ijk} is the yield of the *k*th nitrogen level for variety *j* in block *i*

μ is the grand mean

B_i is the *i*th block effect (random)

V_j is the effect of the jth variety

δ_{ij} is the whole plot error and is the $(BV)_{ij}$ effect

N_k is the effect of the kth nitrogen level

$(VN)_{jk}$ is the interaction effect between the jth variety and the kth nitrogen level

e_{ijk} is the residual error (split-plot error)

Because there are two errors in the model, a general linear model must be customized to do the appropriate tests. However, the mixed model procedure will do it automatically. The SAS code of PROC GLM is shown below.

```
%let dir=C:\Users\STAT-231B-2016\Text\Chapter 9;
filename aa "&dir\data\wheat.xlsx";
proc import datafile=aa out=wheat dbms=xlsx replace;
run;

ods graphics off;
proc glm data=wheat;
   class block variety nitrogen;
   model yield=block variety block*variety nitrogen variety*nitrogen;
   test h=variety e=block*variety;
run;
```

Table 1
Output of PROC GLM for the wheat data

Source	DF	Type III SS	Mean Square	F Value	Pr > F
block	3	1.00466667	0.33488889	15.18	<.0001
variety	4	6.44933333	1.61233333	73.10	<.0001
block*variety	12	3.83866667	0.31988889	14.50	<.0001
nitrogen	2	6.48700000	3.24350000	147.06	<.0001
variety*nitrogen	8	0.10466667	0.01308333	0.59	0.7756

Tests of Hypotheses Using the Type III MS for block*variety as an Error Term					
Source	DF	Type III SS	Mean Square	F Value	Pr > F
variety	4	6.44933333	1.61233333	5.04	0.0128

The tests for nitrogen and variety*nitrogen are appropriate because they are tested against the split-plot error. However, the test for variety shown in the upper portion of the table is not correct. The correct test for variety is shown in the lower portion of the table.

The MIXED procedure in SAS will test all effects correctly. The SAS code for PROC MIXED is given below.

```
proc mixed data=wheat;
   class block variety nitrogen;
   model yield=variety nitrogen variety*nitrogen;
   random block block*variety;
run;
```

The output of PROC MIXED is given in **Table 2**.

Table 2
Output of PROC MIXED for the wheat data

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
variety	4	12	5.04	0.0128
nitrogen	2	30	147.06	<.0001
variety*nitrogen	8	30	0.59	0.7756

You can see that the tests for the three fixed effects are all correct.

2. Theory of mixed model

2.1. Model and assumption

A generic notation for linear mixed model (in matrix notation) is

$$y = X\beta + Z\gamma + e \tag{2}$$

where

- y is an $n \times 1$ vector for a response variable
- X is an $n \times p$ design matrix for fixed effects
- β is a $p \times 1$ vector for the fixed effects (including the intercept)
- Z is an $n \times q$ design matrix for random effects
- γ is a $q \times 1$ vector for the random effects
- e is an $n \times 1$ vector for the residual error
- γ and e are independent, i.e., $\text{cov}(\gamma, e) = 0$

Because the model contains both fixed (β) effects and random (γ) effects, it is called mixed model. The random effects and the residual errors are assumed to be normally distributed and mutually independent,

$$\gamma \sim N(0, G)$$
$$e \sim N(0, R)$$

where G is a $q \times q$ variance-covariance matrix for the random effects and R is an $n \times n$ variance-covariance matrix for the residuals. When we put the random effects and the residuals together, we have

$$\text{var}\begin{bmatrix}\gamma \\ e\end{bmatrix} = \begin{bmatrix}\text{var}(\gamma) & 0 \\ 0 & \text{var}(e)\end{bmatrix} = \begin{bmatrix}G & 0 \\ 0 & R\end{bmatrix} \tag{3}$$

The expectation of the linear mixed model is $\text{E}(y) = X\beta$ and the variance is

$$\text{var}(y) = V = ZGZ^T + R \tag{4}$$

The data vector *y* is multivariate normal with expectation $X\beta$ and variance *V*. We can estimate parameters $\theta = \{\beta, G, R\}$ and predict the random effects γ. We will learn three methods for estimating parameters, the minimum variance quadratic unbiased estimation (MIVQUE), maximum likelihood (ML) and restricted maximum likelihood (REML).

Let us assume that $G = A\sigma_\gamma^2$ and $R = I\sigma_e^2$ where A is a known $q \times q$ covariance structure, σ_γ^2 is a single variance for the random effects, I is an $n \times n$ identity matrix and σ_e^2 is a single residual variance. We have the following specialized form of the variance matrix,

$$V = ZAZ^T\sigma_\gamma^2 + I\sigma_e^2 \tag{5}$$

In this simplified case, the parameters are $\theta = \beta, \sigma_\gamma^2, \sigma_e^2$.

2.2. MIVQUE

The MIVQUE estimates of the variances are obtained via solving a set of linear equations. Let us define $P = I - X(X^T X)^- X^T$, $V_\gamma = PZA^{1/2}$ and $V_e = PI = P$, where $A^{1/2}$ is the lower half of matrix *A*, i.e., the Cholesky decomposition of *A* with a property of $A^{1/2}(A^{1/2})^T = A$. The two variance components are estimated by solving the following equation system,

$$\begin{bmatrix} H_{\gamma\gamma} & H_{\gamma e} \\ H_{e\gamma} & H_{ee} \end{bmatrix} \begin{bmatrix} \sigma_\gamma^2 \\ \sigma_e^2 \end{bmatrix} = \begin{bmatrix} Q_\gamma \\ Q_e \end{bmatrix} \tag{6}$$

where

$$\begin{aligned} H_{\gamma\gamma} &= \mathrm{tr}(V_\gamma V_\gamma^T V_\gamma V_\gamma^T) \\ H_{\gamma e} &= \mathrm{tr}(V_\gamma V_\gamma^T V_e V_e^T) = \mathrm{tr}(V_\gamma V_\gamma^T P) \\ H_{e\gamma} &= \mathrm{tr}(V_e V_e^T V_\gamma V_\gamma^T) = H_{\gamma e} \\ H_{ee} &= \mathrm{tr}(V_e V_e^T V_e V_e^T) = \mathrm{tr}(P) = n - p \end{aligned} \tag{7}$$

and

$$\begin{aligned} Q_\gamma &= y^T V_\gamma V_\gamma^T y \\ Q_e &= y^T V_e V_e^T y = y^T P y \end{aligned} \tag{8}$$

Therefore, the estimated variance components are given by

$$\begin{bmatrix} \hat{\sigma}_\gamma^2 \\ \hat{\sigma}_e^2 \end{bmatrix} = \begin{bmatrix} H_{\gamma\gamma} & H_{\gamma e} \\ H_{e\gamma} & H_{ee} \end{bmatrix}^{-1} \begin{bmatrix} Q_\gamma \\ Q_e \end{bmatrix} \tag{9}$$

The estimated fixed effects are given by

$$\hat{\beta} = (X^T V^{-1} X)^{-1} X^T V^{-1} y \tag{10}$$

where

$$V = ZAZ^T \hat{\sigma}_\gamma^2 + I\hat{\sigma}_e^2 \tag{11}$$

with variance parameters replaced by the estimated values.

2.3. Maximum likelihood estimation

The maximum likelihood estimates of parameters $\theta = \beta, \sigma_\gamma^2, \sigma_e^2$ are obtained by maximizing the following log likelihood function

$$L(\theta) = -\frac{1}{2}\ln|V| - \frac{1}{2}(y - X\beta)^T V^{-1}(y - X\beta) \tag{12}$$

Unlike MIVQUE, the ML method is an iterative approach requiring some initial values of the parameters to start the iteration process. The variance components cannot be negative and thus the solutions much be restricted within the positive domain. The problem can be simplified into a single parameter problem because β and σ_e^2 can be "absorbed". Such an approach is called profiling. Let $\lambda = \sigma_\gamma^2 / \sigma_e^2$ and rewrite

$$V = ZAZ^T\sigma_\gamma^2 + I\sigma_e^2 = (ZAZ^T\lambda + I)\sigma_e^2 = H\sigma_e^2 \tag{13}$$

where $H = ZAZ^T\lambda + I$. Assuming that λ is known, we can find β and σ_e^2 explicitly by taking the partial derivatives of the log likelihood function with respect to β and σ_e^2, letting them equal to zero and solving for the parameter values. The log likelihood function is rewritten as

$$L(\theta) = -\frac{n}{2}\ln(\sigma_e^2) - \frac{1}{2}\ln|H| - \frac{1}{2\sigma_e^2}(y - X\beta)^T H^{-1}(y - X\beta) \tag{14}$$

We have

$$\begin{aligned} \frac{\partial}{\partial\beta}L(\theta) &= \frac{1}{\sigma_e^2}X^T H^{-1}(y - X\beta) = 0 \\ \frac{\partial}{\partial\sigma_e^2}L(\theta) &= -\frac{n}{2\sigma_e^2} + \frac{1}{2\sigma_e^4}(y - X\beta)^T H^{-1}(y - X\beta) = 0 \end{aligned} \tag{15}$$

The solutions of the above equations are

$$\begin{aligned} \beta &= (X^T H^{-1}X)^{-1}X^T H^{-1}y \\ \sigma_e^2 &= \frac{1}{n}(y - X\beta)^T H^{-1}(y - X\beta) \end{aligned} \tag{16}$$

Substituting β and σ_e^2 in equation (14) by equations (16) yields

$$L(\lambda) = -\frac{n}{2}\ln(r^T H^{-1}r) - \frac{1}{2}\ln|H| + \frac{n}{2}\ln(n) - 1 \tag{17}$$

where

$$r = y - X\beta = y - X(X^T H^{-1}X)^{-1}X^T H^{-1}y \tag{18}$$

and the last term is a constant (irrelevant to the parameter) and thus can be ignored. Therefore, the single parameter log likelihood function is

$$L(\lambda) = -\frac{n}{2}\ln(r^T H^{-1}r) - \frac{1}{2}\ln|H| \tag{19}$$

The MLE of λ is then obtained by maximizing equations (17) or (19), which only involves one parameter. The Newton method works very well for solving a single variable equation.

$$\lambda^{(t+1)} = \lambda^{(t)} - \left[\frac{\partial^2 L(\lambda)}{\partial\lambda^2}\right]^{-1}\frac{\partial L(\lambda)}{\partial\lambda} \tag{20}$$

Once the iteration process converges, the value of λ is the MLE. Substituting the MLE of λ into equations (16), we get the MLE of the fixed effects and the residual error variance. The MLE of the variance for the random effects is $\hat{\sigma}_\gamma^2 = \hat{\lambda}\hat{\sigma}_e^2$ due to the invariance property of the ML method.

2.4. Restricted maximum likelihood estimation

The restricted maximum likelihood estimates of parameters $\theta = \beta, \sigma_\gamma^2, \sigma_e^2$ are obtained by maximizing the following restricted log likelihood function

$$L(\theta) = -\frac{1}{2}\ln|V| - \frac{1}{2}\ln|X^T V^{-1} X| - \frac{1}{2}(y - X\beta)^T V^{-1}(y - X\beta) \tag{21}$$

Again, parameters β and σ_e^2 can be profiled so that only $\lambda = \sigma_\gamma^2 / \sigma_e^2$ is the parameter to be estimated. Let us rewrite the log likelihood function

$$\begin{aligned} L(\theta) = &-\frac{1}{2}\ln|H| - \frac{n}{2}\ln(\sigma_e^2) - \frac{1}{2}\ln|X^T H^{-1} X| + \frac{p}{2}\ln(\sigma_e^2) \\ &- \frac{1}{2\sigma_e^2}(y - X\beta)^T H^{-1}(y - X\beta) \end{aligned} \tag{22}$$

The partial derivatives of the log likelihood function with respect to parameters β and σ_e^2 are

$$\begin{aligned} \frac{\partial}{\partial\beta} L(\theta) &= \frac{1}{\sigma_e^2} X^T H^{-1}(y - X\beta) = 0 \\ \frac{\partial}{\partial\sigma_e^2} L(\theta) &= -\frac{n-p}{2\sigma_e^2} + \frac{1}{2\sigma_e^4}(y - X\beta)^T H^{-1}(y - X\beta) = 0 \end{aligned} \tag{23}$$

The solutions of the above equations are

$$\begin{aligned} \beta &= (X^T H^{-1} X)^{-1} X^T H^{-1} y \\ \sigma_e^2 &= \frac{1}{n-p}(y - X\beta)^T H^{-1}(y - X\beta) \end{aligned} \tag{24}$$

Note that the residual error variance here is different from that of the MLE ($n-p$ replaces *n*). Substituting β and σ_e^2 in equation (22) by (24) yields

$$L(\lambda) = -\frac{1}{2}\ln|H| + \frac{n-p}{2}\ln(r^T H^{-1} r) - \frac{1}{2}\ln|X^T H^{-1} X| + \frac{n-p}{2}\ln(n-p) - 1 \tag{25}$$

where

$$r = y - X\beta = y - X(X^T H^{-1} X)^{-1} X^T H^{-1} y \tag{26}$$

Again, the last term in equation (25) is irrelevant to the parameter (a constant). The Newton method similar to that used in ML can be used to find the REML estimate of λ.

2.5. Proof of the restricted likelihood function

The restricted log likelihood function is

$$L(\theta) = -\frac{1}{2}\ln|V| - \frac{1}{2}\ln|X^T V^{-1} X| - \frac{1}{2}(y - X\beta)^T V^{-1}(y - X\beta) \tag{27}$$

The expectation of the model is $\mathrm{E}(y) = X\beta$ and the variance $\mathrm{var}(y) = V$ is

$$V = ZAZ^T\sigma_\gamma^2 + I\sigma^2 = (ZAZ^T\lambda + I_n)\sigma^2 = V_0\sigma^2 \tag{28}$$

where $V_0 = ZAZ^T\lambda + I_n$ and $\lambda = \sigma_\gamma^2 / \sigma^2$. Note that we have changed notation for the residual error variance. Previously, it was σ_e^2 and now it is σ^2 (the subscript has been deleted).

In the original restricted maximum likelihood method (REML), the fixed effects are removed from the likelihood function prior to the maximization process. For example, we first find an $m \times n$ matrix L so that $LX = 0$. We then use this matrix to convert y into $y^* = Ly$. The linear model for the transformed y is

$$y^* = LX\beta + LZ\gamma + L\varepsilon = Z^*\gamma + \varepsilon^* \tag{29}$$

where $Z^* = LZ$ and $\varepsilon^* = L\varepsilon$. The new model

$$y^* = Z^*\gamma + \varepsilon^* \tag{30}$$

is a pure random model because the fixed effects have been removed. The dimension of the new data becomes $m = n - p$, where $p = \text{rank}(X)$. The expectation and variance for the new model are $\text{E}(y^*) = 0$ and

$$\text{var}(y^*) = Z^*AZ^{*T}\sigma_\gamma^2 + LL^T\sigma^2 = Z^*AZ^{*T}\sigma_\gamma^2 + R\sigma^2 \tag{31}$$

where $R = LL^T$.

How do we find such an L matrix? Try this, $L = I - X(X^TX)^{-1}X^T$. The first $m = n - p$ rows of matrix $I - X(X^TX)^{-1}X^T$ is a good candidate because

$$LX = [I - X(X^TX)^{-1}X^T]X = X - X(X^TX)^{-1}X^TX = X - X = 0 \tag{32}$$

Given the expectation and variance of the transformed variable, we can construct the log likelihood function. Maximizing that likelihood function, we get the maximum likelihood estimates of parameters. Such a method is called the restricted maximum likelihood method (REML) and estimates of the parameters are called the REML estimates.

Let $V^* = Z^*AZ^{*T}\sigma_\gamma^2 + R\sigma^2$. The restricted log likelihood function is

$$L_R(\theta) = -\frac{1}{2}\ln|V^*| - \frac{1}{2}y^{*T}V^{*-1}y^* \tag{33}$$

where $\theta = \left\{\sigma_\gamma^2, \sigma^2\right\}$ are the parameters, which do not include β. Substituting y^* by Ly and V^* by LVL^T, we have

$$L_R(\theta) = -\frac{1}{2}\ln|LVL^T| - \frac{1}{2}y^TL^T(LVL^T)^{-1}Ly \tag{34}$$

Harville (1977) showed that the restricted likelihood function can be written in the following form,

$$L_R(\theta) = -\frac{1}{2}\ln|V| - \frac{1}{2}|X^TV^{-1}X| - \frac{1}{2}y^TPy \tag{35}$$

where

$$V = ZAZ^T\sigma_\gamma^2 + I\sigma^2$$

is the variance matrix of the original data and

$$P = V^{-1} - V^{-1}X(X^{-1}V^{-1}X)^{-1}X^TV^{-1} \tag{36}$$

is a special matrix commonly seen in linear mixed literature. Proof of the restricted log likelihood function in Equation (35) is given later.

Define $V = V_0\sigma^2$, $V_0 = ZAZ^T\lambda + I$ and $\lambda = \sigma_\gamma^2 / \sigma^2$. Define $P = P_0 / \sigma^2$ where

$$P_0 = V_0^{-1} - V_0^{-1}X(X^{-1}V_0^{-1}X)^{-1}X^TV_0^{-1} \tag{37}$$

The restricted log likelihood function with $\theta_0 = \{\lambda, \sigma^2\}$ as the parameters becomes

$$\begin{aligned} L_R(\theta_0) &= -\frac{1}{2}\ln|V_0\sigma^2| - \frac{1}{2}|X^TV_0^{-1}X/\sigma^2| - \frac{1}{2}y^TP_0y/\sigma^2 \\ &= -\frac{1}{2}\ln|V_0| - \frac{1}{2}|X^TV_0^{-1}X| - \frac{n}{2}\ln(\sigma^2) + \frac{p}{2}\ln(\sigma^2) - \frac{1}{2}y^TP_0y/\sigma^2 \\ &= -\frac{1}{2}\ln|V_0| - \frac{1}{2}|X^TV_0^{-1}X| - \frac{1}{2}(n-p)\ln(\sigma^2) - \frac{1}{2}y^TP_0y/\sigma^2 \end{aligned} \tag{38}$$

Taking the partial derivative of this likelihood function with respect to σ^2, we get

$$\frac{\partial L_R(\theta_0)}{\partial\sigma^2} = -\frac{1}{2\sigma^2}(n-p) + \frac{1}{2\sigma^4}y^TP_0y \tag{39}$$

Setting it to zero and solving for σ^2 leads to

$$\hat{\sigma}^2 = \frac{1}{n-p}y^TP_0y \tag{40}$$

Substituting this, equation (40), back to the restricted log likelihood function, equation (38), we get the following profiled likelihood function,

$$L_R(\lambda) = -\frac{1}{2}\ln|V_0| - \frac{1}{2}|X^TV_0^{-1}X| - \frac{1}{2}(n-p)\ln(\hat{\sigma}^2) - \frac{1}{2}(n-p) \tag{41}$$

which is a function of only one parameter λ. Ignoring the constant term, the profiled log likelihood function is

$$L_R(\lambda) = -\frac{1}{2}\ln|V_0| - \frac{1}{2}|X^TV_0^{-1}X| - \frac{1}{2}(n-p)\ln(\hat{\sigma}^2) \tag{42}$$

We now prove that equation (35) is indeed the restricted log likelihood function (with β removed from the likelihood function). There are several different versions of derivation for the restricted log likelihood function (see O'Neill 2013). I choose the derivation from a Bayesian perspective by treating the fixed effects as random variables with a prior normal distribution and then setting the prior variance to infinity. The limit of that likelihood is the restricted likelihood function. Let $\beta \sim N(0, I\sigma_\beta^2)$, where σ_β^2 is the "prior" variance of the fixed effects. When β are treated as random effects, the mixed model becomes a random model, whose expectation is 0 and variance is

$$V_\beta = XX^T\sigma_\beta^2 + ZAZ^T\sigma_\gamma^2 + I\sigma^2 = V + XX^T\sigma_\beta^2 \tag{43}$$

The log likelihood function for this pure random model is

$$L_R(\sigma_\gamma^2, \sigma^2, \sigma_\beta^2) = -\frac{1}{2}\ln|V_\beta| - \frac{1}{2}y^TV_\beta^{-1}y \tag{44}$$

where

$$V_\beta^{-1} = (V + XX^T\sigma_\beta^2)^{-1} = V^{-1} - V^{-1}X(X^{-1}V^{-1}X + I/\sigma_\beta^2)^{-1}X^TV^{-1} \tag{45}$$

and

$$|V_\beta| = |V + XX^T\sigma_\beta^2| = |V||X^TV^{-1}X + I/\sigma_\beta^2||I\sigma_\beta^2| \tag{46}$$

The above inverse and determinant are derived using the Sherman-Morrison-Woodbury matrix identities or simply Woodbury identities (Woodbury 1950). When $\sigma_\beta^2 \to \infty$, the limits of the above two quantities are

$$V_\beta^{-1} = V^{-1} - V^{-1}X(X^{-1}V^{-1}X)^{-1}X^TV^{-1} \tag{47}$$

and

$$\ln|V_\beta| = \ln|V| + \ln|X^TV^{-1}X| + \text{constant} \tag{48}$$

where the $\text{constant} = \ln|I\sigma_\beta^2|$ is irrelevant to parameters σ_γ^2 and σ^2, and thus is considered as a constant. Therefore,

$$\lim_{\sigma_\beta^2\to\infty} V_\beta^{-1} = P = V^{-1} - V^{-1}X(X^{-1}V^{-1}X)^{-1}X^TV^{-1} \tag{49}$$

and

$$\lim_{\sigma_\beta^2\to\infty} \ln|V_\beta| = \ln|V| + \ln|X^TV^{-1}X| \tag{50}$$

The limit of the log likelihood function given in equation (44) becomes

$$\lim_{\sigma_\beta^2\to\infty} L_R(\sigma_\gamma^2, \sigma^2, \sigma_\beta^2) = -\frac{1}{2}\ln|V| - \frac{1}{2}\ln|X^TV^{-1}X| - \frac{1}{2}y^TPy = L_R(\theta) \tag{51}$$

which is exactly the same as the restricted log likelihood function shown defined in equation (35).

An alternative expression of this restricted log likelihood function is

$$L_R(\theta) = -\frac{1}{2}\ln|V| - \frac{1}{2}\ln|X^TV^{-1}X| - \frac{1}{2}(y - X\hat{\beta})^TV^{-1}(y - X\hat{\beta}) \tag{52}$$

where

$$\hat{\beta} = (X^{-1}V^{-1}X)^{-1}X^TV^{-1}y \tag{53}$$

This can be proved by substituting $\hat{\beta}$ in equation (52) by the $\hat{\beta}$ given in equation (53) and simplifying the expression to achieve the equivalence.

Proof of the equivalence between equation (52) and equation (53) are shown below. The two equations are

$$L_R(\theta) = -\frac{1}{2}\ln|V| - \frac{1}{2}\ln|X^TV^{-1}X| - \frac{1}{2}y^TPy \tag{54}$$

and

$$L_R(\theta) = -\frac{1}{2}\ln|V| - \frac{1}{2}\ln|X^TV^{-1}X| - \frac{1}{2}(y - X\hat{\beta})^TV^{-1}(y - X\hat{\beta}) \tag{55}$$

where

$$P = V^{-1} - V^{-1}X(X^{-1}V^{-1}X)^{-1}X^TV^{-1} \tag{56}$$

and

$$\hat{\beta} = (X^{-1}V^{-1}X)^{-1}X^TV^{-1}y \tag{57}$$

The first two terms of the two equations are the same. Therefore, we only need to prove that

$$y^TPy = (y - X\hat{\beta})^TV^{-1}(y - X\hat{\beta}) \tag{58}$$

This can be proved by substituting $\hat{\beta}$ in equation (56) by the $\hat{\beta}$ given in equation (58) and simplifying the expression to achieve the equivalence, as shown below.

$$
\begin{aligned}
&(y - X\hat{\beta})^T V^{-1}(y - X\hat{\beta}) \\
&\quad = (y^T - \hat{\beta}^T X^T)V^{-1}(y - X\hat{\beta}) \\
&\quad = y^T V^{-1} y - y^T V^{-1} X\hat{\beta} - \hat{\beta}^T X^T V^{-1} y + \hat{\beta}^T X^T V^{-1} X\hat{\beta} \\
&\quad = y^T V^{-1} y - 2y^T V^{-1} X\hat{\beta} + \hat{\beta}^T X^T V^{-1} X\hat{\beta} \\
&\quad = y^T V^{-1} y - 2y^T V^{-1} X(X^{-1}V^{-1}X)^{-1} X^T V^{-1} y + y^T V^{-1} X(X^{-1}V^{-1}X)^{-1} X^T V^{-1} X(X^{-1}V^{-1}X)^{-1} X^T V^{-1} y \\
&\quad = y^T V^{-1} y - 2y^T V^{-1} X(X^{-1}V^{-1}X)^{-1} X^T V^{-1} y + y^T V^{-1} X(X^{-1}V^{-1}X)^{-1} X^T V^{-1} y \\
&\quad = y^T V^{-1} y - y^T V^{-1} X(X^{-1}V^{-1}X)^{-1} X^T V^{-1} y \\
&\quad = y^T \left[V^{-1} - V^{-1} X(X^{-1}V^{-1}X)^{-1} X^T V^{-1} \right] y \\
&\quad = y^T P y
\end{aligned}
$$

Interestingly, we can show that

$$
\begin{aligned}
&(y - X\hat{\beta})^T V^{-1} y \\
&\quad = (y^T - \hat{\beta}^T X^T)V^{-1} y \\
&\quad = y^T V^{-1} y - \hat{\beta}^T X^T V^{-1} y \\
&\quad = y^T V^{-1} y - y^T V^{-1} X\hat{\beta} \\
&\quad = y^T V^{-1} y - y^T V^{-1} X(X^{-1}V^{-1}X)^{-1} X^T V^{-1} y \\
&\quad = y^T \left[V^{-1} - V^{-1} X(X^{-1}V^{-1}X)^{-1} X^T V^{-1} \right] y \\
&\quad = y^T P y
\end{aligned}
\qquad (59)
$$

Therefore, we obtain the following unexpected equivalence

$$
(y - X\hat{\beta})^T V^{-1}(y - X\hat{\beta}) = (y - X\hat{\beta})^T V^{-1} y \qquad (60)
$$

Note again the difference between the likelihood function and the restricted likelihood function,

$$
\begin{aligned}
L_{ML}(\theta) &= -\frac{1}{2}\ln|V| - \frac{1}{2}(y - X\hat{\beta})^T V^{-1}(y - X\hat{\beta}) \\
L_{REML}(\theta) &= -\frac{1}{2}\ln|V| - \frac{1}{2}(y - X\hat{\beta})^T V^{-1}(y - X\hat{\beta}) - \frac{1}{2}\ln|X^T V^{-1} X|
\end{aligned}
\qquad (61)
$$

The two differ by just one term coded in blue.

2.6. Best linear unbiased prediction (BLUP)

The estimated fixed effects are called the best linear unbiased estimates (BLUE) and the estimated random effects are called the best linear unbiased predictions (BLUP). These two types of effects can be obtained via Henderson's mixed model equation,

$$\begin{bmatrix} X^T R^{-1} X & X^T R^{-1} Z \\ Z^T R^{-1} X & Z^T R^{-1} Z + G^{-1} \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} X^T R^{-1} y \\ Z^T R^{-1} y \end{bmatrix} \tag{62}$$

The solutions are

$$\begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} X^T R^{-1} X & X^T R^{-1} Z \\ Z^T R^{-1} X & Z^T R^{-1} Z + G^{-1} \end{bmatrix}^{-1} \begin{bmatrix} X^T R^{-1} y \\ Z^T R^{-1} y \end{bmatrix} \tag{63}$$

Let us define

$$\begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} = \begin{bmatrix} X^T R^{-1} X & X^T R^{-1} Z \\ Z^T R^{-1} X & Z^T R^{-1} Z + G^{-1} \end{bmatrix}^{-1} \tag{64}$$

The variance-covariance matrix of the BLUE and BLUP is

$$\text{var}\begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} \text{var}(\hat{\beta}) & \text{cov}(\hat{\beta}, \hat{\gamma}) \\ \text{cov}(\hat{\gamma}, \hat{\beta}) & \text{var}(\hat{\gamma}) \end{bmatrix} = \begin{bmatrix} C_{11} & C_{12} \\ C_{21} & C_{22} \end{bmatrix} \tag{65}$$

When $G = A\sigma_\gamma^2$ and $R = I\sigma^2$, the estimation can be simplified into

$$\begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} X^T X & X^T Z \\ Z^T X & Z^T Z + A^{-1} / \lambda \end{bmatrix}^{-1} \begin{bmatrix} X^T y \\ Z^T y \end{bmatrix} \tag{66}$$

Let

$$\begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} = \begin{bmatrix} X^T X & X^T Z \\ Z^T X & Z^T Z + A^{-1} / \lambda \end{bmatrix}^{-1} \tag{67}$$

The variance-covariance matrix of the BLUE and BLUP is

$$\text{var}\begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} \text{var}(\hat{\beta}) & \text{cov}(\hat{\beta}, \hat{\gamma}) \\ \text{cov}(\hat{\gamma}, \hat{\beta}) & \text{var}(\hat{\gamma}) \end{bmatrix} = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix} \sigma^2 \tag{68}$$

This means that

$$\text{var}(\hat{\beta}) = D_{11}\sigma^2 \text{ and } \text{var}(\hat{\gamma}) = D_{22}\sigma^2$$

2.7. Derivation of Henderson's mixed model equations

Henderson's BLUP equations are defined as follows

$$\begin{bmatrix} X^T R^{-1} X & X^T R^{-1} Z \\ Z^T R^{-1} X & Z^T R^{-1} Z + G \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} X^T R^{-1} y \\ Z^T R^{-1} y \end{bmatrix}$$

The BLUP equations were derived via Henderson's joint likelihood method. First, we define the density of *y* given γ and all parameters, denoted by $p(y \mid \gamma, \theta)$. We then define the density of γ given θ, denoted by $p(\gamma \mid \theta)$. From these densities, we get the joint density of *y* and γ given θ, denoted by

$$p(y,\gamma \mid \theta) = p(y \mid \gamma,\theta)\, p(\gamma \mid \theta) \tag{69}$$

where $\theta = \left\{\beta, \sigma_\gamma^2, \sigma^2\right\}$ is the parameter set. Recall that $G = A\sigma_\gamma^2$ and $R = I\sigma^2$. The two log densities in the right hand side are

$$\ln p(y \mid \gamma,\theta) = -\frac{1}{2}\ln|R| - \frac{1}{2}(y - X\beta - Z\gamma)^T R^{-1}(y - X\beta - Z\gamma) \tag{70}$$

and

$$\ln p(\gamma \mid \theta) = -\frac{1}{2}\ln|G| - \frac{1}{2}\gamma^T G^{-1}\gamma \tag{71}$$

leading to a log density in the left hand side

$$\begin{aligned}\ln p(y,\gamma \mid \theta) = &-\frac{1}{2}\ln|R| - \frac{1}{2}(y - X\beta - Z\gamma)^T R^{-1}(y - X\beta - Z\gamma) \\ &-\frac{1}{2}\ln|G| - \frac{1}{2}\gamma^T G^{-1}\gamma\end{aligned} \tag{72}$$

Henderson's mixed model equations are obtained by choosing β and γ that maximize the log density of the joint distribution of *y* and γ. In fact, this method of derivation has a logical problem. It is not a typical maximum likelihood problem because the joint density is not a likelihood function. A likelihood function is the density of data given parameter values. In Henderson's problem, the density is the joint density of data (*y*) and parameter (γ). To find the fixed and random effects using Henderson's joint density, we need to find the first partial derivatives of the log joint density with respect to β and γ. The approach mimics the maximum likelihood method. To be consistent with the notation of a maximum likelihood method, let us denote the joint log density by

$$\begin{aligned}L(\beta,\gamma) &= \ln p(y,\gamma \mid \theta) \\ &= -\frac{1}{2}\ln|R| - \frac{1}{2}(y - X\beta - Z\gamma)^T R^{-1}(y - X\beta - Z\gamma) \\ &\quad -\frac{1}{2}\ln|G| - \frac{1}{2}\gamma^T G^{-1}\gamma\end{aligned} \tag{73}$$

The first partial derivatives are

$$\begin{aligned}\frac{\partial}{\partial\beta}L(\beta,\gamma) &= X^T R^{-1}(y - X\beta - Z\gamma) \\ &= X^T R^{-1}y - X^T R^{-1}X\beta - X^T R^{-1}Z\gamma\end{aligned}$$

and

$$\begin{aligned}\frac{\partial}{\partial\gamma}L(\beta,\gamma) &= -G^{-1}\gamma + Z^T R^{-1}(y - X\beta - Z\gamma) \\ &= -G^{-1}\gamma + Z^T R^{-1}y - Z^T R^{-1}X\beta - Z^T R^{-1}Z\gamma \\ &= Z^T R^{-1}y - Z^T R^{-1}X\beta - \left(Z^T R^{-1}Z + G^{-1}\right)\gamma\end{aligned}$$

Equating these partial derivatives to zeros, we have

$$\begin{aligned}X^T R^{-1}y - X^T R^{-1}X\beta - X^T R^{-1}Z\gamma &= 0 \\ Z^T R^{-1}y - Z^T R^{-1}X\beta - \left(Z^T R^{-1}Z + G^{-1}\right)\gamma &= 0\end{aligned}$$

Rearrangement of the above equations leads to

$$X^T R^{-1} X\beta + X^T R^{-1} Z\gamma = X^T R^{-1} y$$
$$Z^T R^{-1} X\beta + \left(Z^T R^{-1} Z + G^{-1}\right)\gamma = Z^T R^{-1} y$$

Combining the two equations using compact matrix notation, we have Henderson mixed model equations

$$\begin{bmatrix} X^T R^{-1} X & X^T R^{-1} Z \\ Z^T R^{-1} X & Z^T R^{-1} Z + G \end{bmatrix} \begin{bmatrix} \beta \\ \gamma \end{bmatrix} = \begin{bmatrix} X^T R^{-1} y \\ Z^T R^{-1} y \end{bmatrix}$$

Solving for β and γ, we get the solutions,

$$\begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} X^T R^{-1} X & X^T R^{-1} Z \\ Z^T R^{-1} X & Z^T R^{-1} Z + G \end{bmatrix}^{-1} \begin{bmatrix} X^T R^{-1} y \\ Z^T R^{-1} y \end{bmatrix}$$

2.8. Alternative expression of BLUE and BLUP

The second line of Henderson's mixed model equation (the random effect part) is

$$Z^T R^{-1} X\beta + \left(Z^T R^{-1} Z + G^{-1}\right)\gamma = Z^T R^{-1} y$$

Rearrangement of the equation leads to

$$(Z^T R^{-1} Z + G^{-1})\gamma = Z^T R^{-1} y - Z^T R^{-1} X\beta = Z^T R^{-1}(y - X\beta)$$

Therefore,

$$\gamma = (Z^T R^{-1} Z + G^{-1})^{-1} Z^T R^{-1}(y - X\beta)$$

Substituting γ into the first line of Henderson's mixed model equations, i.e.,

$$X^T R^{-1} X\beta + X^T R^{-1} Z\gamma = X^T R^{-1} y$$

we get

$$X^T R^{-1} X\beta + X^T R^{-1} Z\left(Z^T R^{-1} Z + G^{-1}\right)^{-1}\left(Z^T R^{-1} y - Z^T R^{-1} X\beta\right) = X^T R^{-1} y$$

Further manipulation of this equation yields

$$X^T R^{-1} X\beta - X^T R^{-1} Z\left(Z^T R^{-1} Z + G^{-1}\right)^{-1} Z^T R^{-1} X\beta$$
$$= X^T R^{-1} y - X^T R^{-1} Z\left(Z^T R^{-1} Z + G^{-1}\right)^{-1} Z^T R^{-1} y$$

and

$$X^T\left[R^{-1} - R^{-1} Z\left(Z^T R^{-1} Z + G^{-1}\right)^{-1} Z^T R^{-1}\right] X\beta$$
$$= X^T\left[R^{-1} - R^{-1} Z\left(Z^T R^{-1} Z + G^{-1}\right)^{-1} Z^T R^{-1}\right] y \qquad (74)$$

Therefore,

$$\beta = \left\{X^T\left[R^{-1} - R^{-1} Z\left(Z^T R^{-1} Z + G^{-1}\right)^{-1} Z^T R^{-1}\right] X\right\}^{-1}$$
$$\left\{X^T\left[R^{-1} - R^{-1} Z\left(Z^T R^{-1} Z + G^{-1}\right)^{-1} Z^T R^{-1}\right]\right\} y$$

According to Sherman-Morrision-Woodbury formula (Golub and van Loan 1996, p.50), we show that

$$(ZGZ^T + R)^{-1} = R^{-1} - R^{-1} Z(Z^T R^{-1} Z + G^{-1})^{-1} Z^T R^{-1}$$

However, as defined earlier, the variance of *y* is

$$\text{var}(y) = V = ZGZ^T + R$$

Therefore,

$$R^{-1} - R^{-1}Z(Z^TR^{-1}Z + G^{-1})^{-1}Z^TR^{-1} = (R + ZGZ^T)^{-1} = V^{-1}$$

Substituting all relevant terms in equation (74) by V^{-1}, we have

$$X^TV^{-1}X\beta = X^TV^{-1}y$$

and

$$\beta = (X^TV^{-1}X)^{-1}X^TV^{-1}y$$

Therefore, the BLUE and BLUP can be expressed separately using

$$\hat{\beta} = \left(X^TV^{-1}X\right)^{-1}\left(X^TV^{-1}y\right)$$
$$\hat{\gamma} = \left(Z^TR^{-1}Z + G^{-1}\right)^{-1}Z^TR^{-1}\left(y - X\hat{\beta}\right)$$

An alternative expression of the BLUP can be derived from the conditional expectation of γ given y. The joint distribution between y and γ is multivariate normal of mean

$$\text{E}\begin{bmatrix} y \\ \gamma \end{bmatrix} = \begin{bmatrix} X\beta \\ 0 \end{bmatrix}$$

and variance

$$\text{var}\begin{bmatrix} y \\ \gamma \end{bmatrix} = \begin{bmatrix} V & ZG \\ GZ^T & G \end{bmatrix}$$

The conditional expectation of γ given y is

$$\text{E}(\gamma \mid y) = \text{E}(\gamma) + \text{cov}(\gamma, y)[\text{var}(y)]^{-1}[y - \text{E}(y)]$$
$$= GZ^{-1}V^{-1}(y - X\beta)$$

We can prove that $\hat{\gamma} = \text{E}(\gamma \mid y)$, i.e.,

$$(Z^TR^{-1}Z + G^{-1})^{-1}Z^TR^{-1}(y - X\beta) = GZ^TV^{-1}(y - X\beta)$$

To prove the above identity, we only need to prove

$$(Z^TR^{-1}Z + G^{-1})^{-1}Z^TR^{-1} = GZ^TV^{-1}$$

where

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

Here is the proof,

$$\begin{aligned}
GZ^TV^{-1} &= GZ^T\left[R^{-1} - R^{-1}Z(Z^TR^{-1}Z + G^{-1})^{-1}Z^TR^{-1}\right] \\
&= GZ^TR^{-1} - GZ^TR^{-1}Z(Z^TR^{-1}Z + G^{-1})^{-1}Z^TR^{-1} \\
&= G(Z^TR^{-1}Z + G^{-1})(Z^TR^{-1}Z + G^{-1})^{-1}Z^TR^{-1} - GZ^TR^{-1}Z(Z^TR^{-1}Z + G^{-1})^{-1}Z^TR^{-1} \\
&= \left[G(Z^TR^{-1}Z + G^{-1}) - GZ^TR^{-1}Z\right](Z^TR^{-1}Z + G^{-1})^{-1}Z^TR^{-1} \\
&= \left[GZ^TR^{-1}Z + GG^{-1} - GZ^TR^{-1}Z\right](Z^TR^{-1}Z + G^{-1})^{-1}Z^TR^{-1} \\
&= [I](Z^TR^{-1}Z + G^{-1})^{-1}Z^TR^{-1} \\
&= (Z^TR^{-1}Z + G^{-1})^{-1}Z^TR^{-1}
\end{aligned}$$

We actually used a trick to prove it. In the third step, we inserted an identity matrix between G and $Z^T R^{-1}$ that appears in step two ($GZ^T R^{-1}$), leading to

$$GZ^T R^{-1} = G(Z^T R^{-1} Z + G^{-1})(Z^T R^{-1} Z + G^{-1})^{-1} Z^T R^{-1}$$

This trick allows us to continue the derivation and complete the proof.

3. Examples

3.1. Robinson's dairy cattle example

The data contain nine cows sired by four bulls and distributed among three herds. The trait is milk yield. The data are given in **Table 3.** The data did not generate meaningful estimated variance components for the ML and REML methods. Therefore, a new trait was simulated by adding a variable sampled from a $N(0,\sqrt{500})$ distribution to each observed yield data point. The simulated *y* variable is listed in the last column of **Table 3**. The following analysis was based on this new variable *y*, not the actual yield.

Table 3
Robinson's cattle data

cattle	herd	sire	yield	y
1	1	A	110	139.33
2	1	D	100	85.15
3	2	B	110	122.09
4	2	D	100	145.66
5	2	D	100	132.18
6	3	C	110	130.23
7	3	C	110	122.37
8	3	D	100	92.13
9	3	D	100	82.06

The response variable is *y*, the herd effects are fixed and the sire effects are random. The model in matrix notation is

$$y = X\beta + Z\gamma + e \tag{75}$$

Details of the model are shown below,

$$\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ y_5 \\ y_6 \\ y_7 \\ y_8 \\ y_9 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \\ e_9 \end{bmatrix} \tag{76}$$

Note that the intercept is excluded from the model so that the three herd effects can be uniquely estimated. Let us assume $\gamma \sim N(0, I\sigma_\gamma^2)$ and $e \sim N(0, I\sigma_e^2)$ so that $G = A\sigma_\gamma^2 = I\sigma_\gamma^2$ and $R = I\sigma_e^2$.

First, we examined the MIVQUE method. The following matrices are required for the method.

$$P = \begin{bmatrix} 0.5 & -0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -0.5 & 0.5 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.667 & -0.333 & -0.333 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.333 & 0.667 & -0.333 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.333 & -0.333 & 0.667 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.75 & -0.25 & -0.25 & -0.25 \\ 0 & 0 & 0 & 0 & 0 & -0.25 & 0.75 & -0.25 & -0.25 \\ 0 & 0 & 0 & 0 & 0 & -0.25 & -0.25 & 0.75 & -0.25 \\ 0 & 0 & 0 & 0 & 0 & -0.25 & -0.25 & -0.25 & 0.75 \end{bmatrix}$$

$$V_\gamma = \begin{bmatrix} 0.5 & 0 & 0 & -0.5 \\ -0.5 & 0 & 0 & 0.5 \\ 0 & 0.667 & 0 & -0.667 \\ 0 & -0.333 & 0 & 0.333 \\ 0 & -0.333 & 0 & 0.333 \\ 0 & 0 & 0.5 & -0.5 \\ 0 & 0 & 0.5 & -0.5 \\ 0 & 0 & -0.5 & 0.5 \\ 0 & 0 & -0.5 & 0.5 \end{bmatrix}$$

Note that $V_e = P$. The MIVQUE of the variance component is

$$\begin{bmatrix} \hat{\sigma}_\gamma^2 \\ \hat{\sigma}_e^2 \end{bmatrix} = \begin{bmatrix} H_{\gamma\gamma} & H_{\gamma e} \\ H_{e\gamma} & H_{ee} \end{bmatrix}^{-1} \begin{bmatrix} Q_\gamma \\ Q_e \end{bmatrix} = \begin{bmatrix} 9.7777778 & 4.3333333 \\ 4.3333333 & 6 \end{bmatrix}^{-1} \begin{bmatrix} 5430.0441 \\ 3366.0483 \end{bmatrix} = \begin{bmatrix} 451.10446 \\ 235.21038 \end{bmatrix}$$

The BLUE and BLUP of the model effects are

$$\begin{bmatrix} \hat{\beta} \\ \hat{\gamma} \end{bmatrix} = \begin{bmatrix} X^T X & X^T Z \\ Z^T X & Z^T Z + A^{-1}/\lambda \end{bmatrix}^{-1} \begin{bmatrix} X^T y \\ Z^T y \end{bmatrix}$$

whose numerical values are

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\gamma}_1 \\ \hat{\gamma}_2 \\ \hat{\gamma}_3 \\ \hat{\gamma}_4 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 3 & 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 4 & 0 & 0 & 2 & 2 \\ 1 & 0 & 0 & 1.52141 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1.52141 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 1.52141 & 0 \\ 1 & 2 & 2 & 0 & 0 & 0 & 1.52141 \end{bmatrix}^{-1} \begin{bmatrix} 224.48 \\ 399.93 \\ 426.79 \\ 139.33 \\ 122.09 \\ 252.60 \\ 537.18 \end{bmatrix} = \begin{bmatrix} 110.76312 \\ 149.96605 \\ 106.92453 \\ 18.776581 \\ -18.32251 \\ 15.368754 \\ -15.82282 \end{bmatrix}$$

The *D* matrix is shown in **Table 4** and the variance-covariance matrix of the BLUE and BLUP is given in **Table 5**.

Table 4
The *D* matrix of the cattle data

	X_1	X_2	X_3	Z_1	Z_2	Z_3	Z_4
X_1	1.1375988	0.4503074	0.4370848	-0.747727	-0.29598	-0.346699	-0.527471
X_2	0.4503074	0.9430378	0.5010258	-0.29598	-0.619845	-0.397417	-0.604634
X_3	0.4370848	0.5010258	0.9006352	-0.287289	-0.329317	-0.71439	-0.58688
Z_1	-0.747727	-0.29598	-0.287289	1.1487545	0.1945434	0.2278799	0.3466987
Z_2	-0.29598	-0.619845	-0.329317	0.1945434	1.0646996	0.2612163	0.3974172
Z_3	-0.346699	-0.397417	-0.71439	0.2278799	0.2612163	0.9632627	0.4655176
Z_4	-0.527471	-0.604634	-0.58688	0.3466987	0.3974172	0.4655176	0.7082431

Multiplying the D matrix by $\hat{\sigma}_e^2 = 235.21038$ yields the covariance matrix of the BLUE and BLUP.

Table 5
The variance-covariance matrix of BLUE and PLUB

	β_1	β_2	β_3	γ_1	γ_2	γ_3	γ_4
β_1	267.57504	105.91696	102.80688	-175.8731	-69.61763	-81.54714	-124.0666
β_2	105.91696	221.81229	117.84647	-69.61763	-145.7939	-93.47664	-142.2163
β_3	102.80688	117.84647	211.83875	-67.57342	-77.45872	-168.032	-138.0403
γ_1	-175.8731	-69.61763	-67.57342	270.19899	45.758627	53.599711	81.547137
γ_2	-69.61763	-145.7939	-77.45872	45.758627	250.4284	61.440795	93.47664
γ_3	-81.54714	-93.47664	-168.032	53.599711	61.440795	226.56939	109.49456
γ_4	-124.0666	-142.2163	-138.0403	81.547137	93.47664	109.49456	166.58612

The ML and REML methods are iterative approaches and therefore it is impossible to write the explicit solutions of the parameters. Using the MIXED procedure in SAS (PROC MIXED will be introduced in next section), we obtained the estimated variance components. For the ML method, the estimated variance components are $\hat{\sigma}_\gamma^2 = 743.14$ and $\hat{\sigma}_e^2 = 35.4968$ with a $\hat{\lambda} = 743.14/35.4968 = 20.93541$. The corresponding estimates for the REML method are $\hat{\sigma}_\gamma^2 = 992.93$ and $\hat{\sigma}_e^2 = 58.5395$ with a ratio given by $\hat{\lambda} = 992.93/58.5395 = 16.96171$. The ML and REML estimates are slightly different but both are very different from the estimates of the MIVQUE. Let us use the REML estimates as an example to demonstrate the BLUE and BLUP of the model effects.

The $H = ZZ^T\lambda + I$ matrix is (Table 6)

Table 6
The H matrix from the REML estimates

	Col1	Col2	Col3	Col4	Col5	Col6	Col7	Col8	Col9
Row1	17.9617	0	0	0	0	0	0	0	0
Row2	0	17.96171	0	16.96171	16.96171	0	0	16.96171	16.96171
Row3	0	0	17.96171	0	0	0	0	0	0
Row4	0	16.96171	0	17.96171	16.96171	0	0	16.96171	16.96171
Row5	0	16.96171	0	16.96171	17.96171	0	0	16.96171	16.96171
Row6	0	0	0	0	0	17.96171	16.96171	0	0
Row7	0	0	0	0	0	16.96171	17.96171	0	0
Row8	0	16.96171	0	16.96171	16.96171	0	0	17.96171	16.96171
Row9	0	16.96171	0	16.96171	16.96171	0	0	16.96171	17.96171

The estimated fixed effects and the residual error variances are

$$\hat{\beta} = (X^T H^{-1} X)^{-1} X^T H^{-1} y = \begin{bmatrix} 105.66011 & 156.59506 & 106.30316 \end{bmatrix}^T$$

$$\hat{\sigma}_e^2 = \frac{1}{n-p}(y - X\beta)^T H^{-1}(y - X\beta) = 58.539621$$

The BLUE and BLUP of the model effects are

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \\ \hat{\gamma}_1 \\ \hat{\gamma}_2 \\ \hat{\gamma}_3 \\ \hat{\gamma}_4 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 3 & 0 & 0 & 1 & 0 & 2 \\ 0 & 0 & 4 & 0 & 0 & 2 & 2 \\ 1 & 0 & 0 & 1.0589 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1.0589 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 & 1.0589 & 0 \\ 1 & 2 & 2 & 0 & 0 & 0 & 1.0589 \end{bmatrix}^{-1} \begin{bmatrix} 224.48 \\ 399.93 \\ 426.79 \\ 139.33 \\ 122.09 \\ 252.60 \\ 537.18 \end{bmatrix} = \begin{bmatrix} 105.66011 \\ 156.59506 \\ 106.30316 \\ 31.795352 \\ -32.58402 \\ 19.424244 \\ -18.63557 \end{bmatrix}$$

3.2. Diallel cross data (hypothetical data)

This is a hypothetical example of hybrid experiment. There are six inbred lines (parents), which cross each other to generate a total of $6\times(6-1)/2=15$ possible crosses. Such a cross experiment is called a diallel cross. The six parents are labeled by 1, 2 …, 6 and the 15 hybrids are labeled by 11, 12, …, 25. The pedigree is illustrated in **Figure 1**. The purpose of the experiment is to use hybrids with known phenotypic values to predict hybrids that are not yet phenotyped. The actual data are given in **Table 7**.

Figure 1
Diallel crosses of six parents

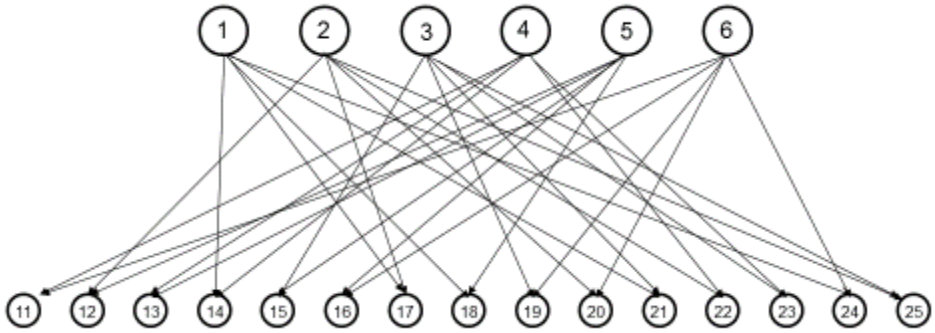

Table 7
Data of the diallel cross experiment

Hybrid	Parent 1	Parent 2	Sample	Phenotype
11	4	6	1	19.7279
12	2	5	1	9.4302
13	4	5	1	5.3604
14	1	4	1	20.7333
15	3	5	1	8.7604
16	5	6	1	11.4847
17	1	2	1	16.5417
18	1	5	1	17.3791
19	3	6	1	12.6979
20	2	6	1	18.0364
21	1	3	2	.
22	2	4	2	.
23	3	4	2	.
24	1	6	2	.
25	2	3	2	.

First, we need to estimate parameters from the sample of ten hybrids. Therefore, the sample size is $n=10$. Since the hybrids are initiated by six inbred parents, we have $q=6$. All the hybrids are evaluated in one single environment and thus $p=1$ and β is only a scalar representing the intercept of the linear mixed model. The parental allelic effects are stored in an $q\times 1=6\times 1$ vector $\gamma=\begin{bmatrix}\gamma_1 & \gamma_2 & \gamma_3 & \gamma_4 & \gamma_5 & \gamma_6\end{bmatrix}^T$. The realized form of the linear mixed model for this special example is

$$\begin{bmatrix} y_{11} \\ y_{12} \\ y_{13} \\ y_{14} \\ y_{15} \\ y_{16} \\ y_{17} \\ y_{18} \\ y_{19} \\ y_{20} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \beta + \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \\ \gamma_5 \\ \gamma_6 \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ e_{13} \\ e_{14} \\ e_{15} \\ e_{16} \\ e_{17} \\ e_{18} \\ e_{19} \\ e_{20} \end{bmatrix} \tag{77}$$

Assume that the six random effects are independent, the REML estimates of the variance components are $\hat{\sigma}_\gamma^2=10.9606$ and $\hat{\sigma}_e^2=11.3919$. The estimated fixed effect is $\hat{\beta}=14.4143$. The predicted random effects are

$$\hat{\gamma}=\begin{bmatrix} \hat{\gamma}_1 \\ \hat{\gamma}_2 \\ \hat{\gamma}_3 \\ \hat{\gamma}_4 \\ \hat{\gamma}_5 \\ \hat{\gamma}_6 \end{bmatrix} = \begin{bmatrix} 3.8223 \\ -0.2883 \\ -1.8023 \\ 0.1606 \\ -3.9006 \\ 2.0083 \end{bmatrix} \tag{78}$$

The predicted phenotypic values for the ten hybrids are

$$\begin{bmatrix} \hat{y}_{11} \\ \hat{y}_{12} \\ \hat{y}_{13} \\ \hat{y}_{14} \\ \hat{y}_{15} \\ \hat{y}_{16} \\ \hat{y}_{17} \\ \hat{y}_{18} \\ \hat{y}_{19} \\ \hat{y}_{20} \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \times 14.4143 + \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 3.8223 \\ -0.2883 \\ -1.8023 \\ 0.1606 \\ -3.9006 \\ 2.0083 \end{bmatrix} = \begin{bmatrix} 16.583219 \\ 10.225359 \\ 10.674256 \\ 18.397211 \\ 8.7112862 \\ 12.521948 \\ 17.948314 \\ 14.335939 \\ 14.62025 \\ 16.134322 \end{bmatrix} \tag{79}$$

To predict the phenotype of the five untested hybrids, we need to find the corresponding *Z* matrix to connect them with the six parents. The *Z* matrix is shown in the following model,

$$\begin{bmatrix}\hat{y}_{21}\\ \hat{y}_{22}\\ \hat{y}_{23}\\ \hat{y}_{24}\\ \hat{y}_{25}\end{bmatrix} = \begin{bmatrix}1\\1\\1\\1\\1\end{bmatrix} \times 14.4143 + \begin{bmatrix}1&0&1&0&0&0\\0&1&0&1&0&0\\0&0&1&1&0&0\\1&0&0&0&0&1\\0&1&1&0&0&0\end{bmatrix} \begin{bmatrix}3.8223\\-0.2883\\-1.8023\\0.1606\\-3.9006\\2.0083\end{bmatrix} = \begin{bmatrix}16.43424\\14.28663\\12.77256\\20.2449\\12.32366\end{bmatrix}$$

The best predicted hybrid is number 24 with a predicted value of 20.2449.

4. The MIXED procedure in SAS

The MIXED procedure fits a variety of mixed linear models to data and enables you to use these fitted models to make statistical inferences about the data. A mixed linear model is a generalization of the standard linear model used in the GLM procedure (general linear model), the generalization being that the data are permitted to exhibit correlation and nonconstant variability. The linear mixed model, therefore, provides you with the flexibility of modeling not only the means of your data (as in the standard linear model) but their variances and covariances as well.

4.1. Cattle data

We now use the two examples introduced above to demonstrate the features of PROC MIXED. Take the dairy cattle data as the first example. The SAS codes to read the data and perform mixed model analysis are shown below,

```
data cattle;
   input id herd sire$ yield y;
cards;
1 1 A 110 139.33
2 1 D 100  85.15
3 2 B 110 122.09
4 2 D 100 145.66
5 2 D 100 132.18
6 3 C 110 130.23
7 3 C 110 122.37
8 3 D 100  92.13
9 3 D 100  82.06
;
proc mixed data=cattle method=mivque0 nobound noprofile;
  class herd sire;
  model y=herd/noint solution;
  random sire/solution;
run;
```

The `method` option in the PROC MIXED statement is set to MIVQUE0, which performs variance component analysis using the MIVQUE method. Both `herd` and `sire` are class variables. The fixed effects `herd` appears in the model statement. The `noint` option in the model statement is to exclude the intercept from the model. The `solution` option in the `model` statement will display the estimated fixed effects. The random effects `sire` appears in the `random` statement. The `solution` option in the random statement will display the estimated random effects. By default, PROC MIXED forces on all variance components to be non-negative.

The `parms` statement is optional that allows you to specify the initial values of the variance components. For the MIVQUE0 method, initial values are not permitted because it is a non-iterative approach. However, for some unexplained reason, even though we set `method = mivque0`, the program still iterates. Therefore, we need to add an option `noiter` in the `parms` statement to prevent the program from iteration. The output of the above code is shown below,

Table 8
Output of PROC MIXED with the MIVQUE0 method

Covariance Parameter Estimates	
Cov Parm	Estimate
Sire	451.10
Residual	235.21

Solution for Fixed Effects						
Effect	herd	Estimate	Standard Error	DF	t Value	Pr > |t|
herd	1	110.76	16.3577	3	6.77	0.0066
herd	2	149.97	14.8934	3	10.07	0.0021
herd	3	106.92	14.5547	3	7.35	0.0052

Solution for Random Effects						
Effect	sire	Estimate	Std Err Pred	DF	t Value	Pr > |t|
sire	A	18.7766	16.4377	3	1.14	0.3362
sire	B	-18.3225	15.8249	3	-1.16	0.3307
sire	C	15.3688	15.0522	3	1.02	0.3824
sire	D	-15.8228	12.9068	3	-1.23	0.3077

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
herd	3	3	37.00	0.0072

The results are identical to what are shown in the previous section using the analytic method of MIVQUE.

The SAS code for PROC MIXED using the ML method is

```
proc mixed data=cattle method=ml;
  class herd sire;
  model y=herd/noint solution;
  random sire/solution;
run;
```

The output is shown in **Table 9** below.

Table 9
Output of PROC MIXED with the ML method

Covariance Parameter Estimates	
Cov Parm	Estimate
Sire	743.14
Residual	35.4968

Solution for Fixed Effects						
Effect	herd	Estimate	Standard Error	DF	t Value	Pr > |t|
herd	1	105.42	14.5882	3	7.23	0.0055
herd	2	156.85	14.2926	3	10.97	0.0016
herd	3	106.29	14.2855	3	7.44	0.0050

Solution for Random Effects						
Effect	sire	Estimate	Std Err Pred	DF	t Value	Pr > |t|
sire	A	32.3617	15.0908	3	2.14	0.1213
sire	B	-33.1795	14.8310	3	-2.24	0.1113
sire	C	19.5446	14.5603	3	1.34	0.2720
sire	D	-18.7268	13.9838	3	-1.34	0.2729

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
herd	3	3	56.48	0.0039

The estimate variance components are $\hat{\sigma}^2_{\gamma} = 743.14$ and $\hat{\sigma}^2_{e} = 35.4968$, which are different from the estimates of MIVQUE. Do not take the differences seriously because the sample size is too small to generate any meaningful estimates.
The SAS code for the REML method is

```
proc mixed data=cattle method=reml mmeq;
  class herd sire;
  model y=herd/noint solution;
  random sire/solution;
run;
```

By default, PROC MIXED uses the REML method. Therefore, you may delete the `method = reml` option in the PROC MIXED statement. The The output is shown in **Table 10** below.

Table 10
Output of PROC MIXED with the REML method

Covariance Parameter Estimates	
Cov Parm	Estimate
Sire	992.93
Residual	58.5395

Solution for Fixed Effects						
Effect	herd	Estimate	Standard Error	DF	t Value	Pr > |t|
herd	1	105.66	17.1045	3	6.18	0.0085
herd	2	156.60	16.6914	3	9.38	0.0026
herd	3	106.30	16.6792	3	6.37	0.0078

Solution for Random Effects						
Effect	sire	Estimate	Std Err Pred	DF	t Value	Pr > |t|
sire	A	31.7954	17.7813	3	1.79	0.1717
sire	B	-32.5840	17.4278	3	-1.87	0.1583
sire	C	19.4242	17.0565	3	1.14	0.3375
sire	D	-18.6356	16.2557	3	-1.15	0.3348

Type 3 Tests of Fixed Effects				
Effect	Num DF	Den DF	F Value	Pr > F
herd	3	3	37.72	0.0070

The estimate variance components are $\hat{\sigma}^2_{\gamma} = 992.93$ and $\hat{\sigma}^2_{e} = 58.5395$, which are very close to the ML estimates. Again, do not take the differences seriously because of the small sample size.

4.2. Diallele data

The diallel cross data (**Table 7**) are atypical and the design matrix for the random effects cannot be generated in an automatic manner. Therefore, we have to generate the design matrix using PROC IML. The PROC IML code is shown below,

```
proc iml;
     use hybrid;
        read all var{id p1 p2};
     close;
     n=nrow(id);
     q=max(p1||p2);
     z=j(n,q,0);
     do i=1 to n;
        z[i,p1[i]]=1;
        z[i,p2[i]]=1;
     end;
     print z;
quit;
```

The generated *Z* matrix along with the phenotypic values (*y*) are listed in Table 11 Next page). We then read the *Z* matrix to a SAS dataset and then call the mixed procedure to use the *Z* matrix for analysis. The following code read the Z matrix as six numerical variables, named Z_1, Z_2, …, Z_6.

```
data z;
  input ID Z1 Z2 Z3 Z4 Z5 Z6 y;
cards;
11    0     0     0     1     0     1     19.7279
12    0     1     0     0     1     0     9.4302
13    0     0     0     1     1     0     5.3604
14    1     0     0     1     0     0     20.7333
15    0     0     1     0     1     0     8.7604
16    0     0     0     0     1     1     11.4847
17    1     1     0     0     0     0     16.5417
18    1     0     0     0     1     0     17.3791
19    0     0     1     0     0     1     12.6979
20    0     1     0     0     0     1     18.0364
21    1     0     1     0     0     0     .
22    0     1     0     1     0     0     .
23    0     0     1     1     0     0     .
24    1     0     0     0     0     1     .
25    0     1     1     0     0     0     .
;
```

Table 11

The design matrix for the random effects of the diallel cross experiment

ID	Z1	Z2	Z3	Z4	Z5	Z6	y
11	0	0	0	1	0	1	19.7279
12	0	1	0	0	1	0	9.4302
13	0	0	0	1	1	0	5.3604
14	1	0	0	1	0	0	20.7333
15	0	0	1	0	1	0	8.7604
16	0	0	0	0	1	1	11.4847
17	1	1	0	0	0	0	16.5417
18	1	0	0	0	1	0	17.3791
19	0	0	1	0	0	1	12.6979
20	0	1	0	0	0	1	18.0364
21	1	0	1	0	0	0	.
22	0	1	0	1	0	0	.
23	0	0	1	1	0	0	.
24	1	0	0	0	0	1	.
25	0	1	1	0	0	0	.

Unfortunately, the MIXED procedure in SAS cannot directly use a numerical design matrix. You may think that the following code would work,

```
proc mixed data=z method=reml mmeq;
   model y = /solution;
   random z1-z6/solution;
run;
```

You would receive the following unexpected result (**Table 12**). The program would report seven variance components, one for each of the *Z* variables and a residual variance. This is not what we want. We want to have a single variance from the six *Z* variables. The random statement of PROX MIXED treats the six numerical variables as six separate classes of variable and estimates a variance for each *Z*. There are two ways we can fool the program to do what we need. One is to use the `type = lin(m)` option in the random `statement` to define the type of covariance structure to overwrite the default `type = vc` structure. The other way is to use a high performance MIXTED procedure with a new feature of EFFECT statement. This SAS procedure is called PROC HPMIXED.

Table 12
Estimated variance components from the above SAS code

Covariance Parameter Estimates	
Cov Parm	Estimate
Z1	51.1178
Z2	0
Z3	0
Z4	0
Z5	16.3091
Z6	18.3755
Residual	8.2660

Let us examine the first approach using PROC MIXED. In the random statement, we can add the `type = lin(m)` option to define a *G* structure with the following type,

$$G = \sum_{k=1}^{m} V_k \sigma_k^2 \tag{80}$$

where *G* is the covariance structure of the random effects. The dimension of G is $q \times q$ where q is the number of columns of the Z matrix. In this special example,

$$\text{var}(y) = V = ZGZ^T + R = ZZ^T \sigma_\gamma^2 + I\sigma_e^2 \tag{81}$$

Therefore, $G = I\sigma_\gamma^2 = V_1 \sigma_1^2$ and the structure of V_1 is an identity matrix. The following SAS code will take this special covariance structure and estimate a single variance for the random effects.

```
data v1;
  input parm row col1 col2 col3 col4 col5 col6;
cards;
1 1 1 0 0 0 0 0
1 2 0 1 0 0 0 0
1 3 0 0 1 0 0 0
1 4 0 0 0 1 0 0
1 5 0 0 0 0 1 0
1 6 0 0 0 0 0 1
;
proc mixed data=z method=reml mmeq;
   model y = /solution;
   random z1-z6/type=lin(1) ldata=v1 solution;
run;
```

Data V1 defines a SAS dataset containing the identity matrix and two additional variables named `parm` and `row`, where `parm` takes a value of 1 (the first covariance structure) and `row` identifies the row number of the structural matrix. The remaining variables must with names `col1, col2, …, colq`. The output of the above code is (**Table 13**)

Table 13
Output of PROC MIXED with a special structure of the *G* matrix

Covariance Parameter Estimates	
Cov Parm	Estimate
LIN(1)	10.9601
Residual	11.3922

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	14.4143	2.9459	4	4.89	0.0081

Solution for Random Effects					
Effect	Estimate	Std Err Pred	DF	t Value	Pr > |t|
Z1	3.8222	2.2071	4	1.73	0.1584
Z2	-0.2882	2.1704	4	-0.13	0.9008
Z3	-1.8023	2.2963	4	-0.78	0.4764
Z4	0.1607	2.1704	4	0.07	0.9445
Z5	-3.9005	2.0768	4	-1.88	0.1336
Z6	2.0082	2.1609	4	0.93	0.4053

The second approach is to use PROC HPMIXED by invoking the EFFECT statement. The SAS code is

```
proc hpmixed data=z method=reml;
   effect z=collection(z1-z6);
   model y = /solution;
   random z /solution;
run;
```

The effect statement can define new variables from current variables. The new variable is called z, which is a collection of all six variables Z1, Z2, …, Z6. The new variable z is

the design matrix and is treated as a single class variable. The default type = vc applies to this random statement and generates the correct result.

Table 14
Output of PROC HPMIXED

Covariance Parameter Estimates	
Cov Parm	Estimate
z	10.9606
Residual	11.3919

We now show the predicted phenotypic values using the MIXED procedure by adding the `outp = pred` option in the model statement.

```
proc mixed data=z method=reml mmeq;
   model y = /solution outp=pred;
   random z1-z6/type=lin(1) ldata=v1 solution;
run;
proc print data=pred;
run;
```

The predicted phenotypes and the errors of the prediction along with the original data will be included in a new data dataset call `pred`, as shown in **Table 15** (next page). The five untested hybrids, hybrid ID 21, … 25, also have predicted values. The StdErrPred column gives the standard error of prediction for each predicted phenotypic value. For hybrids with ID 11 to 20 who have observed phenotypic values in the data (training data), the standard error of prediction ranges from 2.17 to 2.37. However, for hybrids with ID 21 to 25 who have not yet been tested in the field (test sample), the standard error of prediction ranges from 2.642 to 3.989, which is much higher than the error for the hybrids in the training sample.

PROC MIXED can handle missing data more efficiently than most other software package. Even If you do not have missing values in the data, you can still evaluate the predictability of your model via cross validation. You can set the response variable of the test sample by a missing value and then perform regular mixed model analysis. When every individual is predicted this way, you can calculate the correlation between the predicted and observed trait values. The squared correlation coefficient is a measure of the predictability.

Table 15
Predicted values of the diallel cross data

ID	Z1	Z2	Z3	Z4	Z5	Z6	y	Pred	StdErrPred
11	0	0	0	1	0	1	19.7279	16.5831	2.17363
12	0	1	0	0	1	0	9.4302	10.2255	2.18550
13	0	0	0	1	1	0	5.3604	10.6744	2.18550
14	1	0	0	1	0	0	20.7333	18.3971	2.28768
15	0	0	1	0	1	0	8.7604	8.7114	2.31236
16	0	0	0	0	1	1	11.4847	12.5219	2.17399
17	1	1	0	0	0	0	16.5417	17.9482	2.28768
18	1	0	0	0	1	0	17.3791	14.3359	2.22128
19	0	0	1	0	0	1	12.6979	14.6202	2.36547
20	0	1	0	0	0	1	18.0364	16.1342	2.17363
21	1	0	1	0	0	0	.	16.4342	2.83030
22	0	1	0	1	0	0	.	14.2867	2.82650
23	0	0	1	1	0	0	.	12.7726	2.98918
24	1	0	0	0	0	1	.	20.2447	2.64221
25	0	1	1	0	0	0	.	12.3237	2.98918

You can explicitly predict the phenotypes of the test sample from the training sample using the multivariate normal theorem. Let us partition *y* into y_1 and y_2, where y_1 is the phenotypic values for n_1 individuals in the training sample and y_2 is the trait values of n_2 individuals in the test sample and $n_1 + n_2 = n$. Other terms are also partitioned accordingly. The mixed model can be rewritten as

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1\beta \\ X_2\beta \end{bmatrix} + \begin{bmatrix} Z_1\gamma \\ Z_2\gamma \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \tag{82}$$

The variance-covariance matrix is also partitioned similarly,

$$\text{var}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} = \begin{bmatrix} Z_1Z_1^T\sigma_\gamma^2 & Z_1Z_2^T\sigma_\gamma^2 \\ Z_2Z_1^T\sigma_\gamma^2 & Z_2Z_2^T\sigma_\gamma^2 \end{bmatrix} + \begin{bmatrix} I\sigma_E^2 & 0 \\ 0 & I\sigma_E^2 \end{bmatrix} \tag{83}$$

Let $K_{ij} = Z_iZ_j^T$ be the relationship between individuals in sample *i* and individuals in sample *j*, for $i, j = 1, 2$. We have

$$\text{var}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} = \begin{bmatrix} K_{11}\sigma_\gamma^2 & K_{12}\sigma_\gamma^2 \\ K_{21}\sigma_\gamma^2 & K_{22}\sigma_\gamma^2 \end{bmatrix} + \begin{bmatrix} I\sigma_e^2 & 0 \\ 0 & I\sigma_e^2 \end{bmatrix} \tag{84}$$

Let $G_{ij} = K_{ij}\sigma_\gamma^2$ and $R_{ij} = I\sigma_e^2$, and note that $V_{ij} = K_{ij}\sigma_\gamma^2 + I\sigma_e^2 = G_{ij} + R_{ij}$. We have

$$\text{var}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{bmatrix} = \begin{bmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{bmatrix} + \begin{bmatrix} R_{11} & 0 \\ 0 & R_{22} \end{bmatrix} \tag{85}$$

To predict the trait values in the testing sample, we use the conditional expectation of y_2 given y_1, which is denoted by $\text{E}(y_2 \mid y_1)$ and expressed as

$$\begin{aligned} \hat{y}_2 &= \text{E}(y_2 \mid y_1) \\ &= X_2\hat{\beta} + G_{21}V_{11}^{-1}(y_1 - X_1\hat{\beta}) \\ &= X_2\hat{\beta} + G_{21}\left(G_{11} + R_{11}\right)^{-1}(y_1 - X_1\hat{\beta}) \\ &= X_2\hat{\beta} + \hat{\sigma}_\gamma^2 K_{21}\left(K_{11}\hat{\sigma}_\gamma^2 + I\hat{\sigma}_e^2\right)^{-1}(y_1 - X_1\hat{\beta}) \end{aligned} \tag{86}$$

For the diallel cross data, $n_1 = 10$ and $n_2 = 5$. Note that $K_{21} = K_{12}^T$ and $K_{12} = Z_1 Z_2^T$ is a 10×5 matrix representing the relationship between the 10 hybrids in the training sample and the 5 hybrids in the test sample and $K_{11} = Z_1 Z_1^T$ is a 10×10 matrix representing the internal relationship of the 10 hybrids in the training sample. These two matrices (K_{12} and K_{11}) are given in **Table 16** and **Table 17**, respectively (next page). You can verify that the predicted phenotypic values of the five hybrids obtained from equation (86) is the same as those given in **Table 15**.

Table 16

The relationship matrix between the ten tested hybrids and the five untested hybrids

K_{12}	H_{21}	H_{22}	H_{23}	H_{24}	H_{25}
H_{11}	0	0.5	0.5	0.5	0
H_{12}	0	0.5	0	0	0.5
H_{13}	0	0.5	0.5	0	0
H_{14}	0.5	0.5	0.5	0.5	0
H_{15}	0.5	0	0.5	0	0.5
H_{16}	0	0	0	0.5	0
H_{17}	0.5	0.5	0	0.5	0.5
H_{18}	0.5	0	0	0.5	0
H_{19}	0.5	0	0.5	0.5	0.5
H_{20}	0	0.5	0	0.5	0.5

Table 17
The relationship matrix for the ten hybrids tested in the field

K_{11}	H_{11}	H_{12}	H_{13}	H_{14}	H_{15}	H_{16}	H_{17}	H_{18}	H_{19}	H_{20}
H_{11}	1	0	0.5	0.5	0	0.5	0	0	0.5	0.5
H_{12}	0	1	0.5	0	0.5	0.5	0.5	0.5	0	0.5
H_{13}	0.5	0.5	1	0.5	0.5	0.5	0	0.5	0	0
H_{14}	0.5	0	0.5	1	0	0	0.5	0.5	0	0
H_{15}	0	0.5	0.5	0	1	0.5	0	0.5	0.5	0
H_{16}	0.5	0.5	0.5	0	0.5	1	0	0.5	0.5	0.5
H_{17}	0	0.5	0	0.5	0	0	1	0.5	0	0.5
H_{18}	0	0.5	0.5	0.5	0.5	0.5	0.5	1	0	0
H_{19}	0.5	0	0	0	0.5	0.5	0	0	1	0.5
H_{20}	0.5	0.5	0	0	0	0.5	0.5	0	0.5	1

4.3. Diallel cross with correlated parents

In the diallel cross experiment analyzed before, we assumed that the six parents are independent. We now relax this assumption and assume that they are related with the following covariance matrix. In quantitative genetics, this matrix is called the coancestry matrix. The model needs to be revised by taking into the coancestry matrix. The expectation remains to be $\mathrm{E}(y) = X\beta$ and the variance matrix is $\mathrm{var}(y) = ZAZ^T\sigma_\gamma^2 + I\sigma_e^2$, where A is a 6×6 covariance matrix given in **Table 18**.

Table 18
Coancestry coefficient matrix of the six inbred parents

A	P_1	P_2	P_3	P_4	P_5	P_6
P_1	1	0.012156	0.215946	0.329827	0.330796	0.300371
P_2	0.012156	1	0.268336	0.20332	0.338522	0.25434
P_3	0.215946	0.268336	1	0.028454	0.169486	0.041214
P_4	0.329827	0.20332	0.028454	1	0.052249	0.131461
P_5	0.330796	0.338522	0.169486	0.052249	1	0.10023
P_6	0.300371	0.25434	0.041214	0.131461	0.10023	1

The SAS code to read the A matrix and perform mixed model analysis is shown below.

```
data A;
  input parm row col1 col2 col3 col4 col5 col6;
cards;
1 1 1 0.012156 0.215946 0.329827 0.330796 0.300371
1 2 0.012156 1 0.268336 0.203320 0.338522 0.254340
1 3 0.215946 0.268336 1 0.028454 0.169486 0.041214
1 4 0.329827 0.203302 0.028454 1 0.052249 0.131461
1 5 0.330796 0.338522 0.169486 0.052249 1 0.100230
1 6 0.300371 0.254340 0.041214 0.131461 0.100230 1
;
ods graphics off;
proc mixed data=z method=reml;
   model y = /solution outp=pred;
   random z1-z6/type=lin(1) ldata=A solution;
run;
proc print data=pred;
run;
```

Table 19.
The output of PROC MIXED with correlated parents

Covariance Parameter Estimates	
Cov Parm	Estimate
LIN(1)	13.9595
Residual	12.2890

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	13.5935	4.3708	4	3.11	0.0359

Solution for Random Effects					
Effect	Estimate	Std Err Pred	DF	t Value	Pr > |t|
Z1	3.7604	2.8375	4	1.33	0.2557
Z2	-0.5498	2.7689	4	-0.20	0.8523
Z3	-1.5413	2.7950	4	-0.55	0.6107
Z4	0.9571	2.6494	4	0.36	0.7362
Z5	-3.2093	2.6350	4	-1.22	0.2902
Z6	2.7109	2.7337	4	0.99	0.3775

Table 20
Predicted values of the diallel cross data for correlated parents

Obs	ID	Z1	Z2	Z3	Z4	Z5	Z6	y	Pred	StdErrPred
1	11	0	0	0	1	0	1	19.7279	17.2614	2.24626
2	12	0	1	0	0	1	0	9.4302	9.8343	2.29561
3	13	0	0	0	1	1	0	5.3604	11.3412	2.18524
4	14	1	0	0	1	0	0	20.7333	18.3109	2.39052
5	15	0	0	1	0	1	0	8.7604	8.8428	2.41195
6	16	0	0	0	0	1	1	11.4847	13.0950	2.21990
7	17	1	1	0	0	0	0	16.5417	16.8041	2.11789
8	18	1	0	0	0	1	0	17.3791	14.1445	2.31112
9	19	0	0	1	0	0	1	12.6979	14.7631	2.44059
10	20	0	1	0	0	0	1	18.0364	15.7546	2.25143
11	21	1	0	1	0	0	0	.	15.8125	2.91053
12	22	0	1	0	1	0	0	.	14.0007	2.95280
13	23	0	0	1	1	0	0	.	13.0092	3.10081
14	24	1	0	0	0	0	1	.	20.0648	2.75052
15	25	0	1	1	0	0	0	.	11.5023	3.14534

The results from correlated parents are different from those of independent parents, but the differences are not very significant. The variance matrix can be rewritten as

$$\text{var}(y) = V = ZAZ^T\sigma_\gamma^2 + I\sigma_e^2 = K\sigma_\gamma^2 + I\sigma_e^2 \tag{87}$$

where $K = ZAZ^T$ is called the kinship matrix for the hybrids. PROC MIXED can directly handle this *K* matrix as the input covariance structure. There are several different ways to fit the data to PROC MIXED. We will show two of them in this section using the diallel hybrid data for correlated parents. The model is rewritten as

$$y = X\beta + \xi + e \tag{88}$$

where $\xi = Z\gamma$ and the variance of ξ is

$$\text{var}(\xi) = ZAZ^T\sigma_\gamma^2 = K\sigma_\gamma^2 \tag{89}$$

Assume that the *K* matrix is stored in a SAS dataset with two extra columns named `parm (all 1's)` and `row` (with values of 1, 2,…,15). The 15 variables corresponding to the 15 columns of the *K* matrix are named `col1, col2, …, col15`. The dataset is named K. The SAS code to handle this K matrix is

```
proc mixed data=z method=reml;
   class ID;
   model y = /solution outp=pred;
   random id/subject=intercept type=lin(1) ldata=K solution;
   parms /lowerb=1e-8,1e-8;
run;
proc print data=pred;
run;
```

The results are identical to those shown before except that the predicted random effects are no longer Z but ξ, as shown below (**Table 21**),

Table 21
Predicted random effects $\hat{\xi}$

Solution for Random Effects						
Effect	ID	Estimate	Std Err Pred	DF	t Value	Pr > |t|
ID	11	3.6680	4.5515	0	0.81	.
ID	12	-3.7591	4.7703	0	-0.79	.
ID	13	-2.2523	4.5634	0	-0.49	.
ID	14	4.7175	4.6727	0	1.01	.
ID	15	-4.7507	4.5643	0	-1.04	.
ID	16	-0.4984	4.8076	0	-0.10	.
ID	17	3.2106	4.6949	0	0.68	.
ID	18	0.5511	4.8965	0	0.11	.
ID	19	1.1696	4.5379	0	0.26	.
ID	20	2.1611	4.7091	0	0.46	.
ID	21	2.2191	4.8535	0	0.46	.
ID	22	0.4072	4.8673	0	0.08	.
ID	23	-0.5843	4.7028	0	-0.12	.
ID	24	6.4713	5.0814	0	1.27	.
ID	25	-2.0911	4.8817	0	-0.43	.

The second way to handle this data is to fit a structure R matrix with no random effects. The model can be rewritten as

$$y = X\beta + \varepsilon \tag{90}$$

where $\varepsilon = \xi + e$ is the residual error vector with a complicated error structure

$$R = \text{var}(\varepsilon) = \text{var}(\xi) + \text{var}(e) = K\sigma_\gamma^2 + I\sigma_e^2 \tag{91}$$

To fit this complicated R structure, we need the `type = lin(2)` option in a `repeated` statement of PROC MIXED. The covariance structure is a $(2n)\times n$ matrix defined as $W = K//I$ or

$$W = \begin{bmatrix} K \\ I \end{bmatrix}$$

where the top part of the *W* matrix is the covariance structure for ξ and the bottom part is the structure for *e*. Adding two extra columns to the *W* matrix, we generate a *W* dataset, as shown below for part of the *W* dataset,

Table 22
Sketch of the *W* matrix for the error structure

parm	row	col1	col2	.	.	.	col15
1	1	2.2629	0.610	.	.	.	0.52731
1	2	0.6101	2.677	.	.	.	1.77634
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
1	15	0.5273	1.776	.	.	.	2.53667
2	1	1	0	.	.	.	0
2	2	0	1	.	.	.	0
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.
2	15	0	0	.	.	.	1

The SAS code to use the above W matrix is shown below,

```
proc mixed data=z method=reml;
   model y = /solution;
   repeated /subject=intercept type=lin(2) ldata=w;
run;
```

Here, we did not use the random statement but replaced it by the repeated statement to define the covariance structure of R. This way of handling complicated error structure is much faster than fitting random effects to the model. The result is demonstrated in **Table 23**.

Table 23
Output of PROC MIXED by fitting complicated structure of *R*

Covariance Parameter Estimates		
Cov Parm	Subject	Estimate
LIN(1)	Intercept	13.9580
LIN(2)	Intercept	12.2870

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	13.5934	4.3705	0	3.11	

The estimated fixed effect and the variance components are identical to those estimated from all previous models. Unfortunately, because we did not use the random statement, we were not able to predict ξ and the phenotypes.

5. Genomic prediction in rice (real application)

This is a real data in rice. The population consists of 210 recombinant inbred lines (RIL) with two traits, yield (yd) and thousand grain weight (kgw). The covariance structure is stored in a SAS dataset called kk and the trait values are stored in a dataset called phe. Both the kk and phe data are available on the course website. The model is

$$y = X\beta + \xi + e \tag{92}$$

where $X = 1_n$ is a unity vector, β is simply the intercept, ξ is called the polygenic effect with $\mathrm{E}(\xi) = 0$ and $\mathrm{var}(\xi) = K\sigma_\xi^2$, and $e \sim N(0, I\sigma_e^2)$ is the residual error. The parameters are $\theta = \left\{\beta, \sigma_\xi^2, \sigma_e^2\right\}$. The SAS code to read the data and perform mixed model analysis is

```
filename kk "&dir\kk.csv" lrecl=20000;
filename phe "&dir\phe.csv";
proc import datafile=kk out=kk dbms=csv replace;
proc import datafile=phe out=phe dbms=csv replace;
run;
%let y=kgw;
proc mixed data=phe method=reml;
   class RIL;
   model &y = /outp=pred solution;
   random ril /type=lin(1) ldata=kk solution;
   parms /lowerb=1e-8,1e-8;
run;
```

where the trait analyzed in this code is kgw. The estimated variance components are given in **Table 24**.

Table 24
Estimated parameters for the rice data

Covariance Parameter Estimates	
Cov Parm	Estimate
LIN(1)	10.6314
Residual	0.5544

Solution for Fixed Effects					
Effect	Estimate	Standard Error	DF	t Value	Pr > |t|
Intercept	24.1397	0.5490	209	43.97	<.0001

Therefore, the estimate parameters are

$$\hat{\theta} = \begin{bmatrix} \hat{\beta} \\ \hat{\sigma}^2_{\xi} \\ \hat{\sigma}^2_{e} \end{bmatrix} = \begin{bmatrix} 24.1397 \\ 10.6314 \\ 0.5544 \end{bmatrix}$$

We partitioned the data into 10 parts of equal size (21 lines) and then used the first 9 parts (189 lines) to predict the last part (21 lines). The way to do this is to set the phenotypic values of the last part to a missing value.

```
data phe1;
  set phe;
  if foldid=10 then kgw=.;
run;
proc mixed data=phe1 method=reml;
   class RIL;
   model kgw = /outp=pred solution;
   random ril /type=lin(1) ldata=kk solution;
   parms /lowerb=1e-8,1e-8;
run;
proc print data=pred;
   where foldid=1 | foldid=10;
run;
```

We printed the predicted values for `foldid = 1` and `foldid = 10`, which are shown in **Table 25** (next two pages).

Table 25
Predicted KGW for the last fold (fold 10) using the first nine folds (folds 1 – 9) of RILs of the rice data

Obs	RIL	foldid	kgw	Pred	StdErrPred
5	R005	1	26.1275	26.0224	0.56319
7	R007	1	25.44	25.0325	0.52055
9	R009	1	25.6825	25.7999	0.60180
18	R019	1	25.9475	26.2213	0.53005
20	R021	1	26.0975	26.0380	0.56216
23	R025	1	19.0225	19.6185	0.58693
24	R027	10	.	23.0677	0.96494
26	R029	10	.	23.3125	1.05078
38	R044	1	25.345	24.5623	0.56245
39	R045	10	.	26.6711	1.01460
42	R049	1	20.3375	20.6093	0.56983
43	R050	1	24.2975	24.5279	0.56586
47	R055	1	20.005	19.8681	0.59097
51	R060	10	.	24.1092	0.99124
56	R065	10	.	27.2296	0.90022
61	R070	1	21.3675	21.4318	0.60238
67	R076	10	.	24.5044	1.08206
70	R079	10	.	22.6257	1.03372
72	R081	1	24.0175	24.0350	0.53865
82	R094	1	23.005	22.6556	0.57335
87	R101	1	22.905	22.8226	0.56712
89	R105	10	.	26.6674	0.99274
91	R107	10	.	22.8552	1.07220
96	R113	10	.	24.4979	0.93989
97	R115	10	.	25.3988	1.16317
100	R119	10	.	27.3223	0.95280
106	R126	1	21.895	21.9162	0.56799
108	R128	10	.	22.3844	0.94351
110	R130	10	.	23.2819	0.96825
111	R131	10	.	24.8594	0.95653
114	R135	1	22.2975	22.7832	0.58705
116	R138	1	22.0425	21.7962	0.55459

Obs	RIL	foldid	kgw	Pred	StdErrPred
124	R146	10	.	27.4823	1.19694
130	R152	1	24.1025	24.0929	0.57698
132	R154	10	.	22.9277	0.97817
133	R155	1	20.045	20.2834	0.56427
134	R156	1	25.645	25.7709	0.58058
136	R158	1	24.7925	25.1080	0.57176
137	R159	1	24.505	24.3602	0.55413
153	R177	10	.	28.2500	1.04994
156	R180	1	28.015	27.4795	0.52219
176	R202	10	.	20.9278	0.99167
183	R210	1	26.6125	25.6195	0.57205
190	R220	1	22.29	21.6874	0.55212
192	R222	1	29.09	28.6163	0.54385
200	R230	10	.	24.3856	1.04276

The predicted kgw for foldid = 1 is also presented in the table to compare the standard error of prediction with `foldid = 10`. You can see that the prediction errors for lines with `foldid = 10` are almost doubled compared with the errors for lines with `foldid = 1`.