

# Lecture 4

## BLUP Breeding Values

Guilherme J. M. Rosa

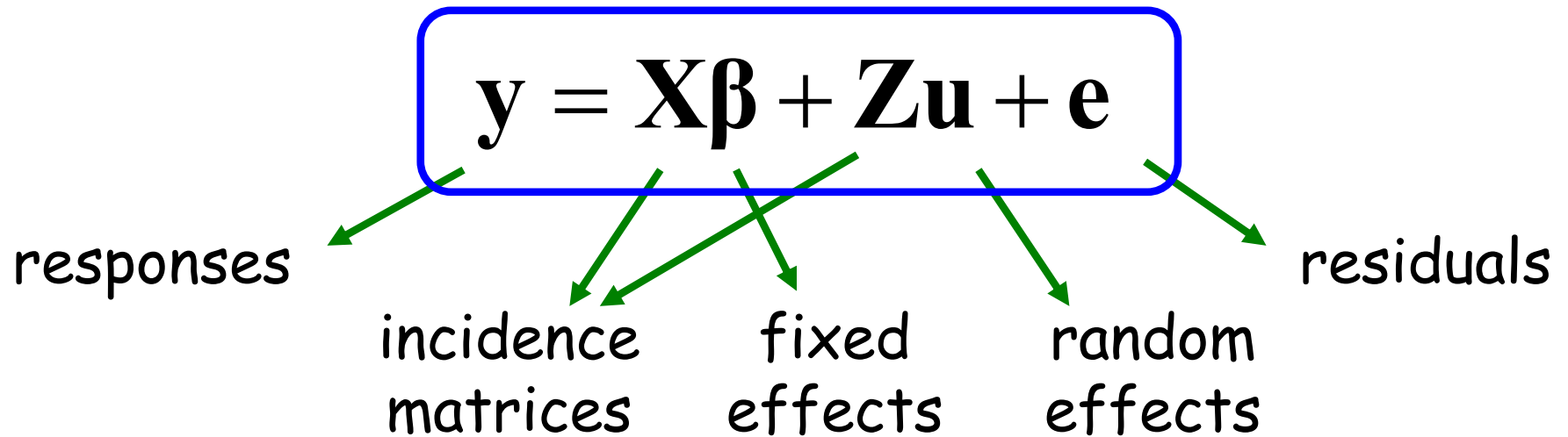
University of Wisconsin-Madison

Mixed Models in Quantitative Genetics

SISG, Seattle

19 - 21 July 2023

# Linear Mixed Effects Model



$$\begin{bmatrix} \mathbf{u} \\ \mathbf{e} \end{bmatrix} \sim \text{MVN} \left( \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{\Sigma} \end{bmatrix} \right)$$

# Estimation of Fixed Effects

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

with  $\boldsymbol{\varepsilon} = \mathbf{Z}\mathbf{u} + \mathbf{e}$ , such that  $\text{Var}[\boldsymbol{\varepsilon}] = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \boldsymbol{\Sigma}$

→ MLE for  $\boldsymbol{\beta}$ :

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1} \mathbf{X}^T \mathbf{V}^{-1} \mathbf{y} \sim \text{MVN}(\boldsymbol{\beta}, (\mathbf{X}^T \mathbf{V}^{-1} \mathbf{X})^{-1})$$

where  $\mathbf{V} = \mathbf{Z}\mathbf{G}\mathbf{Z}^T + \boldsymbol{\Sigma}$

# Prediction of Random Effects

$$\begin{bmatrix} \mathbf{y} \\ \mathbf{u} \end{bmatrix} \sim \text{MVN} \left( \begin{bmatrix} \mathbf{X}\boldsymbol{\beta} \\ \mathbf{0} \end{bmatrix}, \begin{bmatrix} \mathbf{V} & \mathbf{ZG} \\ \mathbf{GZ}^T & \mathbf{G} \end{bmatrix} \right)$$

$$\begin{aligned} \mathbf{E}[\mathbf{u} \mid \mathbf{y}] &= \mathbf{E}[\mathbf{u}] + \text{Cov}[\mathbf{u}, \mathbf{y}^T] \text{Var}^{-1}[\mathbf{y}] (\mathbf{y} - \mathbf{E}[\mathbf{y}]) \\ &= \mathbf{GZ}^T \mathbf{V}^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{GZ}^T (\mathbf{ZGZ}^T + \boldsymbol{\Sigma})^{-1} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) \end{aligned}$$

Replacing  $\boldsymbol{\beta}$  by its estimate:

$$\hat{\mathbf{u}} = \mathbf{GZ}^T (\mathbf{ZGZ}^T + \boldsymbol{\Sigma})^{-1} (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

# Mixed Model Equations

$$\begin{bmatrix} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} & \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z} \\ \mathbf{Z}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} & \mathbf{Z}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z} + \mathbf{G}^{-1} \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{y} \\ \mathbf{Z}^T \boldsymbol{\Sigma}^{-1} \mathbf{y} \end{bmatrix}$$

BLUP and BLUE:

$$\left\{ \begin{array}{l} \hat{\mathbf{u}} = (\mathbf{Z}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z} + \mathbf{G}^{-1})^{-1} \mathbf{Z}^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{X} \hat{\boldsymbol{\beta}}) \\ \hat{\boldsymbol{\beta}} = \{ \mathbf{X}^T [\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1} \mathbf{Z} (\mathbf{Z}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z} + \mathbf{G}^{-1})^{-1} \mathbf{Z}^T \boldsymbol{\Sigma}^{-1}] \mathbf{X} \}^{-1} \\ \quad \times \mathbf{X}^T [\boldsymbol{\Sigma}^{-1} - \boldsymbol{\Sigma}^{-1} \mathbf{Z} (\mathbf{Z}^T \boldsymbol{\Sigma}^{-1} \mathbf{Z} + \mathbf{G}^{-1})^{-1} \mathbf{Z}^T \boldsymbol{\Sigma}^{-1}] \mathbf{y} \end{array} \right.$$

# Mixed Models in Animal and Plant Breeding

Animal/plant breeding programs are based on the principle that phenotypic observations on related individuals can provide information about their underlying genotypic values

The additive component of genetic variation is the primary determinant of the degree to which offspring resemble their parents, and therefore this is usually the component of interest in artificial selection programs

# Mixed Models in Animal and Plant Breeding

Many statistical methods for analysis of genetic data are specific (or more appropriate) for phenotypic measurements obtained from planned experimental designs and with balanced data sets

While such situations may be possible within laboratory or greenhouse experimental settings, data from natural populations and agricultural species are generally highly unbalanced and fragmented by numerous kinds of relationships

# Animal Model

Culling of data to accommodate conventional statistical techniques (e.g. ANOVA) may introduce bias and/or lead to a substantial loss of information

The mixed model methodology allows efficient estimation of genetic parameters (such as variance components and heritability) and breeding values while accommodating extended pedigrees, unequal family sizes, overlapping generations, sex-limited traits, assortative mating, and natural or artificial selection

To illustrate such application of mixed models in breeding programs, we consider here the so-called Animal Model in situations with a single trait and a single observation (including missing values) per individual



# Animal Model

The animal model can be described as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{e}$$

$\mathbf{y}$  is an  $(n \times 1)$  vector of observations (phenotypic scores)

$\boldsymbol{\beta}$  is a  $(p \times 1)$  vector of fixed effects (e.g. herd-year-season effects)

$\mathbf{u} \sim N(\mathbf{0}, \mathbf{G})$  is a  $(q \times 1)$  vector of breeding values (relative to all individuals with record or in the pedigree file, such that  $q$  is in general bigger than  $n$ )

$\mathbf{e} \sim N(\mathbf{0}, \mathbf{I}_n \sigma_e^2)$  represents residual effects, where  $\sigma_e^2$  is the residual variance

# The Matrix $\mathbf{A}$

The matrix  $\mathbf{G}$  describing the covariances among the random effects (here the breeding values) follows from standard results for the covariances between relatives

It can be shown that the additive genetic covariance between two relatives  $i$  and  $i'$  is given by  $2\theta_{ii'}\sigma_a^2$ , where  $\theta_{ii'}$  is the **coefficient of coancestry** between individuals  $i$  and  $i'$ , and  $\sigma_a^2$  is the additive genetic variance in the base population

Hence, under the animal model,  $\mathbf{G} = \mathbf{A}\sigma_a^2$ , where  $\mathbf{A}$  is the **additive genetic (or numerator) relationship matrix**, having elements given by  $a_{ii'} = 2\theta_{ii'}$

# The Matrix $A$

For each animal  $i$  in the pedigree ( $i = 1, 2, \dots, n$ ), going from older to younger animals, compute  $a_{ii}$  and  $a_{ij}$  ( $j = 1, 2, \dots, i-1$ ) as follows:

If both parents ( $s$  and  $d$ ) of animal  $i$  are known:

$$a_{ij} = a_{ji} = (a_{js} + a_{jd})/2 \text{ and } a_{ii} = 1 + a_{sd}/2$$

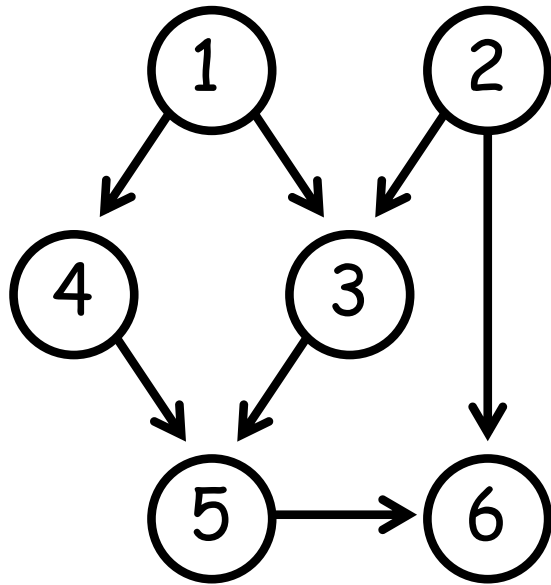
If only one parent (e.g.  $d$ ) of animal  $i$  is known:

$$a_{ij} = a_{ji} = a_{jd}/2 \text{ and } a_{ii} = 1$$

If parents unknown:

$$a_{ij} = a_{ji} = 0 \text{ and } a_{ii} = 1$$

# Example



Animal	Sire	Dam
1	-	-
2	-	-
3	1	2
4	1	-
5	4	3
6	5	2

$$A = \begin{bmatrix} 1 & 0 & .5 & .5 & .5 & .25 \\ 0 & 1 & .5 & 0 & .25 & .625 \\ .5 & .5 & 1 & .25 & .625 & .563 \\ .5 & 0 & .25 & 1 & .625 & .313 \\ .5 & .25 & .625 & .625 & 1.125 & .688 \\ .25 & .625 & .563 & .313 & .688 & 1.125 \end{bmatrix}$$



pedigree matrix A

# Animal Model

In general, in animal/plant breeding interest is on prediction of breeding values (for selection of superior individuals), and on estimation of variance components and functions thereof, such as heritability

The fixed effects are, in some sense, nuisance factors with no central interest in terms of inferences, but which need to be taken into account (i.e., they need to be corrected for when inferring breeding values)

# Animal Model

Since under the animal model  $\mathbf{G}^{-1} = \mathbf{A}^{-1}\sigma_a^{-2}$  and  $\mathbf{R}^{-1} = \mathbf{I}_n\sigma_e^{-2}$ , the mixed model equations can be expressed as:

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

where  $\lambda = \frac{\sigma_e^2}{\sigma_a^2} = \frac{1-h^2}{h^2}$ , such that:

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

Conditional on the variance components ratio  $\lambda$ , the BLUP of the breeding values are given then by:

$$\hat{\mathbf{u}} = (\mathbf{Z}^T \mathbf{Z} + \lambda \mathbf{A}^{-1})^{-1} \mathbf{Z}^T (\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}})$$

These are generally referred to as **Estimated Breeding Values (EBV)**

Alternatively, some breeders associations express their results as Predicted Transmitting Abilities (PTA) (or Estimated Transmitting Abilities (ETA) or Expected Progeny Difference (EPD)), which are equal to half the EBV, representing the portion of an animal's breeding values that is passed to its offspring

The amount of information contained in an animal's genetic evaluation depends on the availability of its own record, as well as how many (and how close) relatives it has with phenotypic information

As a measure of amount of information in livestock genetic evaluations, EBVs are typically reported with its associated accuracies

**Accuracy** of predictions is defined as the correlation between true and estimated breeding values, i.e.,  $r_i = \rho(\hat{u}_i, u_i)$

Instead of accuracy, some livestock species genetic evaluations use **reliability**, which is the squared correlation of accuracy ( $r_i^2$ )



# Prediction Accuracy

The calculation of  $\rho(\hat{u}_i, u_i)$  requires the diagonal elements of the inverse of the **MME coefficient matrix**, represented as:

$$\mathbf{C} = \begin{bmatrix} \mathbf{X}^T \mathbf{X} & \mathbf{X}^T \mathbf{Z} \\ \mathbf{Z}^T \mathbf{X} & \mathbf{Z}^T \mathbf{Z} + \lambda \mathbf{A}^{-1} \end{bmatrix}^{-1} = \begin{bmatrix} \mathbf{C}^{\beta\beta} & \mathbf{C}^{\beta u} \\ \mathbf{C}^{u\beta} & \mathbf{C}^{uu} \end{bmatrix}$$

It can be shown that the **prediction error variance** of EBV  $\hat{u}_i$  is given by:

$$\text{PEV} = \text{Var}(\hat{u}_i - u_i) = c_i^{uu} \sigma_e^2$$

where  $c_i^{uu}$  is the  $i$ -th diagonal element of  $\mathbf{C}^{uu}$ , relative to animal  $i$ .

# Prediction Accuracy

The PEV can be interpreted as the fraction of additive genetic variance not accounted for by the prediction

Therefore, PEV can be expressed also as:

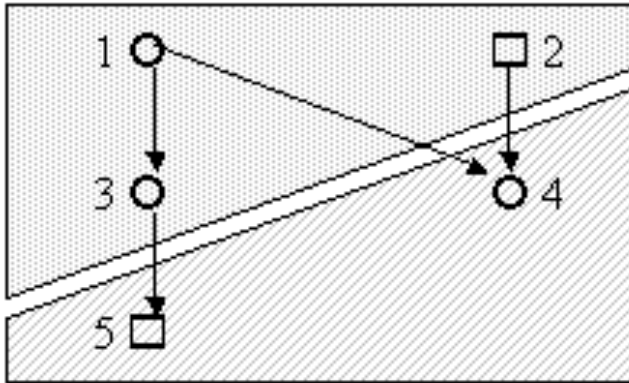
$$\text{PEV} = (1 - r_i^2) \sigma_a^2$$

such that  $c_i^{uu} \sigma_e^2 = (1 - r_i^2) \sigma_a^2$ , from which the reliability is obtained as:

$$r_i^2 = 1 - c_i^{uu} \sigma_e^2 / \sigma_a^2 = 1 - \lambda c_i^{uu}$$

# Animal Model

herd 1



Animal	Sire	Dam	Herd	Observation
1	-	-	h1	310
2	-	-	h1	-
3	-	1	h1	270
4	2	1	h2	350
5	-	3	h2	-

herd 2

$$\begin{bmatrix} 310 \\ 270 \\ 350 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} h_1 \\ h_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_3 \\ e_4 \end{bmatrix}$$


$$\mathbf{y} = \underbrace{\mathbf{X} \boldsymbol{\beta}} + \underbrace{\mathbf{Z} \mathbf{u}} + \mathbf{e}$$

# Animal Model

Breeding values:  $\mathbf{u} \sim \mathbf{N}(\mathbf{0}, \mathbf{A}\sigma_u^2)$  , with

$$\mathbf{A} = \begin{bmatrix} 1 & 0 & 0.5 & 0.5 & 0.25 \\ 0 & 1 & 0 & 0.5 & 0 \\ 0.5 & 0 & 1 & 0.25 & 0.5 \\ 0.5 & 0.5 & 0.25 & 1 & 0.125 \\ 0.25 & 0 & 0.5 & 0.125 & 1 \end{bmatrix}$$

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


$$\lambda = \frac{\sigma_e^2}{\sigma_u^2} = \frac{1-h^2}{h^2}$$

# R Code



animal model  
toy example

```
y<-matrix(c(310,270,350),nrow=3)
X<-matrix(c(1,1,0,0,0,1),nrow=3)
Z<-matrix(c(1,0,0,0,0,0,0,1,0,0,0,0,0,1,0),nrow=3, byrow = TRUE)
A<-matrix(c(1,0,0.5,0.5,0.25,
            0,1,0,0.5,0,
            0.5,0,1,0.25,0.5,
            0.5,0.5,0.25,1,0.125,
            0.25,0,0.5,0.125,1),nrow=5)
```

```
h2<-1/3 # heritability
a=(1-h2)/h2

# crossproducts
XX<-crossprod(X,X)
XZ<-t(X) %*% Z
ZX<-t(Z) %*% X
ZZ<-crossprod(Z,Z)+a*solve(A)

# mixed model equations
# coefficient matrix and right hand side
C<-rbind(cbind(XX,XZ),cbind(ZX,ZZ))
rhs<-rbind(t(X) %*% y,t(Z) %*% y)

#solution
theta.hat <- solve(C) %*% rhs
```

$$h^2 = \frac{1}{3} \rightarrow \alpha = 2 \Rightarrow \begin{cases} \hat{h}_1 = 290 \\ \hat{h}_2 = 348 \\ \hat{u}_1 = 4.0 \\ \hat{u}_2 = 0.0 \\ \hat{u}_3 = -4.0 \\ \hat{u}_4 = 2.0 \\ \hat{u}_5 = -2.0 \end{cases}$$

# Animal Model

The animal model can be extended to model multiple (correlated) traits, multiple random effects (such as maternal effects and common environmental effects), repeated records (e.g. test day models), and so on

Example (Mrode 1996, pp74-76): Weaning weight (kg) of piglets, progeny of three sows mated to two boars:

Piglet	Sire	Dam	Sex	Weight
6	1	2	1	90
7	1	2	2	70
8	1	2	2	65
9	3	4	2	98
10	3	4	1	106
11	3	4	2	60
12	3	4	2	80
13	1	5	1	100
14	1	5	2	85
15	1	5	1	68

A linear model with the (fixed) effect of sex, and the (random) effects of common environment (related to each litter) and breeding values can be expressed as  $\mathbf{X}$ :

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u} + \mathbf{W}\mathbf{c} + \mathbf{e}$$

Weight →  $\mathbf{y}$ 
Sex →  $\mathbf{X}\boldsymbol{\beta}$ 
Breeding values →  $\mathbf{Z}\mathbf{u}$ 
Common environment →  $\mathbf{W}\mathbf{c}$ 
Residual →  $\mathbf{e}$

Assuming that  $\sigma_u^2 = 20$ ,  $\sigma_c^2 = 15$  and  $\sigma_e^2 = 65$ , the MME are as follows:

$$\begin{bmatrix} \mathbf{X}^T\mathbf{X} & \mathbf{X}^T\mathbf{Z} & \mathbf{X}^T\mathbf{W} \\ \mathbf{Z}^T\mathbf{X} & \mathbf{Z}^T\mathbf{Z} + \mathbf{A}^{-1}\lambda_1 & \mathbf{Z}^T\mathbf{W} \\ \mathbf{W}^T\mathbf{X} & \mathbf{W}^T\mathbf{Z} & \mathbf{W}^T\mathbf{W} + \mathbf{I}\lambda_2 \end{bmatrix} \begin{bmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\mathbf{u}} \\ \hat{\mathbf{c}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T\mathbf{y} \\ \mathbf{Z}^T\mathbf{y} \\ \mathbf{W}^T\mathbf{y} \end{bmatrix}$$

where  $\lambda_1 = \frac{\sigma_e^2}{\sigma_u^2} = 3.25$  and  $\lambda_2 = \frac{\sigma_e^2}{\sigma_c^2} = 4.3$

The BLUEs and BLUPs  
(inverting the numerator  
relationship matrix) are:



Mrode example

Effects	Solutions
<i>Sex</i>	
1	91.493
2	75.764
-----	
<i>Animals</i>	
1	-1.441
2	-1.175
3	1.441
4	1.441
5	-0.266
6	-1.098
7	-1.667
8	-2.334
9	3.925
10	2.895
11	-1.141
12	1.525
13	0.448
14	0.545
15	-3.819
-----	
<i>Environ.</i>	
2	-1.762
4	2.161
5	-0.399