

BLUP

Peter von Rohr

09 Oktober 2020

General Principle

- ▶ All methods to predict breeding values follow the same principle
 1. Correct information sources for some population mean
 2. Multiply corrected information source by an appropriate factor
- ▶ Regression Method

$$\hat{u} = b(y - \mu)$$

- ▶ Selection Index
 - ▶ will be presented later
 - ▶ corresponds to multiple regression approach

$$\hat{u} = l = b^T y^*$$

where $b = P^{-1}Gw$ and y^* corrected information sources.

Problem with Correction

- ▶ Population mean is ideal as correction

$$y = \mu + u + e \quad \rightarrow \quad \bar{y} = \bar{\mu} + \bar{u} + \bar{e} = \mu$$

- ▶ Because performances are observed in different
 - ▶ environments and
 - ▶ time points
- ▶ Formation of comparison groups where animals are exposed to the same environments
- ▶ The more groups, the better the correction of environmental effects
- ▶ The more groups, the smaller the single groups

Bias

- ▶ With small comparison groups, it is more likely that mean breeding value of animals in a single group is not 0
- ▶ Average performance of all animals in a comparison group

$$\bar{y}_{CG} = \mu + \bar{u}_{CG} + \bar{e}_{CG}$$

* If \bar{u}_{CG} is not 0, the predicted breeding value \hat{u}_i of animal i is

$$\begin{aligned}\hat{u}_i &= I = b(y_i - (\mu + \bar{u}_{CG})) \\ &= b(y_i - \mu) - b\bar{u}_{CG} \\ &= \hat{u}_i - b\bar{u}_{CG}\end{aligned}$$

where $b\bar{u}_{CG}$ is called bias.

Solution - BLUP

- ▶ Solution to correction problem in selection index: BLUP
- ▶ Estimates environmental effects at the same time as breeding values are predicted
- ▶ Linear mixed effects model
- ▶ Meaning of BLUP
 - ▶ **B** stands for **best** → correlation between true (u) and its prediction (\hat{u}) is maximal or the prediction error variance ($\text{var}(u - \hat{u})$) is minimal.
 - ▶ **L** stands for **linear** → predicted breeding values are linear functions of the observations (y)
 - ▶ **U** stands for **unbiased** → expected values of the predicted breeding values are equal to the true breeding values
 - ▶ **P** stands for **prediction**

Example

	Animal	Sire	Dam	Herd	Weaning Weight
full sibs	12	1	4	1	2.61
	13	1	4	1	2.31
	14	1	5	1	2.44
	15	1	5	1	2.41
	16	1	6	2	2.51
	17	1	6	2	2.55
	18	1	7	2	2.14
	19	1	7	2	2.61
	20	2	8	1	2.34
	21	2	8	1	1.99
	22	2	9	1	3.10
	23	2	9	1	2.81
	24	2	10	2	2.14
	25	2	10	2	2.41
	26	3	11	2	2.54
	27	3	11	2	3.16

known environmental factor

Observation y

Linear Models

In R: use the function `lm()`

► Simple linear model

linear fixed effect model because it only contains fixed effects.

$$y_{ij} = \mu + \text{herd}_j + e_{ij}$$

General mean

fixed effect of herd j

random error term

► Result: Estimate of effect of herd j

► What about breeding value u_i for animal i ?

► Problem: breeding values have a variance σ_u^2

► Cannot be specified in simple linear model

→ Linear Mixed Effects Model (LME)

'Mixed' because they contain fixed and random effects.

$$y_{ijk} = \mu + \beta_j + u_i + e_{ijk}$$

Scalar Notation

In principle, there are also simple random effect models which contain only random effects. Not important for our purpose.

Matrix-Vector Notation

For example dataset with 15 Observations, we have to write 15 equations in scalar notation.

In practice, we often have Millions of observations.

► LME for all animals of a population

→ use matrix-vector notation

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

Same model as in
previous slide

where

y vector of length n of all observations

β vector of length p of all fixed effects

ex: β contains the
effects of herds 1 and 2

X $n \times p$ design matrix linking the fixed effects to the observations

u vector of length n_u of random effects

Z $n \times n_u$ design matrix linking random effect to the observations

e vector of length n of random residual effects.

Expected Values and Variances

► Expected values

because they are defined as deviations

$$E(u) = 0 \text{ and } E(e) = 0 \rightarrow E(y) = X\beta$$

► Variances

variance of a vector u and this corresponds to a variance-covariance matrix. For a vector u of length q , then G has dimensions $q \times q$

$$\text{var}(u) = G \text{ and } \text{var}(e) = R$$

with $\text{cov}(u, e^T) = 0$,

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

The Solution For the linear mixed effect model, the vectors u and β are unknown and must be estimated. This estimation is done as such that the resulting estimations or predictions of u and β fulfill the conditions that are imposed by BLUP:

B: Best

L: Linear

U: unbiased

corrected observations

$$\hat{u} = \underline{GZ^T V^{-1}} (\overline{y - X\hat{\beta}})$$

weighting factors

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

Predicted breeding values based on BLUP are consistent with the general principle of prediction of breeding values.

In practice with large data sets, the matrix V has dimensions $n \times n$ with $n = 1-10$ Millions, Computing the inverse of V is computationally too expensive.

Mixed Model Equations

Henderson and his team showed that solving the mixed model equations, results in the same predictions for the breeding values and in the same estimates for the fixed effects.

Mixed Model equations do not contain the matrix V or its inverse. But they contain the matrix R which is the variance-covariance matrix of random residuals. The matrix R has a very simple structure and its inverse is easy to compute.

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

Sire Model

- Breeding value of sire as random effect:

$$y = X\beta + Zs + e$$

**s as random effect instead of u and
s means the random sire effect**

Example: Animal 12 has animal 1 as sire. The observation of animal 12 is decomposed in to the fixed effects (herd) and the effect of sire 1 plus a random residual (e)

Result: the vector s contains entries for each sire. Example: 3 sires (1, 2 and 3) and therefore, the vector s would be of length 3.

The result of the sire model is predicted breeding values for only the sires.

Example

For the unknown fixed herd effects and the unknown random sire effects, the mixed model equations are constructed and solved for the unknowns

$$\begin{array}{c}
 \text{y} \\
 \text{Animal 12}
 \end{array}
 \begin{bmatrix}
 2.61 \\
 2.31 \\
 2.44 \\
 2.41 \\
 2.51 \\
 2.55 \\
 2.14 \\
 2.61 \\
 2.34 \\
 1.99 \\
 3.1 \\
 2.81 \\
 2.14 \\
 2.41 \\
 2.54 \\
 3.16
 \end{bmatrix}
 =
 \begin{array}{c}
 \text{x} \\
 \text{fixed herd effects}
 \end{array}
 \begin{bmatrix}
 1 & 0 \\
 1 & 0 \\
 1 & 0 \\
 1 & 0 \\
 0 & 1 \\
 0 & 1 \\
 0 & 1 \\
 0 & 1 \\
 1 & 0 \\
 1 & 0 \\
 1 & 0 \\
 1 & 0 \\
 0 & 1 \\
 0 & 1 \\
 0 & 1 \\
 0 & 1
 \end{bmatrix}
 \begin{bmatrix}
 \beta_1 \\
 \beta_2
 \end{bmatrix}
 +
 \begin{array}{c}
 \text{Z} \\
 \text{random sire effects}
 \end{array}
 \begin{bmatrix}
 1 & 0 & 0 \\
 1 & 0 & 0 \\
 1 & 0 & 0 \\
 1 & 0 & 0 \\
 1 & 0 & 0 \\
 1 & 0 & 0 \\
 1 & 0 & 0 \\
 1 & 0 & 0 \\
 0 & 1 & 0 \\
 0 & 1 & 0 \\
 0 & 1 & 0 \\
 0 & 1 & 0 \\
 0 & 1 & 0 \\
 0 & 1 & 0 \\
 0 & 0 & 1 \\
 0 & 0 & 1
 \end{bmatrix}
 \begin{bmatrix}
 s_1 \\
 s_2 \\
 s_3
 \end{bmatrix}
 +
 \begin{array}{c}
 \text{e} \\
 e_1 \\
 e_2 \\
 e_3 \\
 e_4 \\
 e_5 \\
 e_6 \\
 e_7 \\
 e_8 \\
 e_9 \\
 e_{10} \\
 e_{11} \\
 e_{12} \\
 e_{13} \\
 e_{14} \\
 e_{15} \\
 e_{16}
 \end{bmatrix}$$

Animal Model

In contrast to the sire model, the animal model contains breeding values of all animals in the pedigree and therefore, we obtain predicted breeding values for all animals.

In our example animals 12 to 27 have observations, but the animal model allows to predict breeding values for all 27 animals.

- Breeding value for all animals as random effects

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