BLUP

Peter von Rohr

2022-10-21

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
  - uses all available information combined into an index

Selection Index

- will be presented later to estimate aggregate genotype
- Idea: all available information about an animals breeding value is combined into an index of merit (I)
- corresponds to multiple regression approach

$$\hat{u} = I = b_1 * y_1^* + b_2 * y_2^* + ... + b_k * y_k^* = b^T y^*$$

where b the regression coefficients are computed such that the variance ($var(u - \hat{u})$) of the error is minimal.

Index Weights

- Minimization of the variance of the errors means

$$PEV = var(u - \hat{u}) = var(u - I) = var(u - b^T y^*)$$

$$= var(u) + var(b^T y^*) - 2cov(u, (y^*)^T b)$$

$$= \sigma_u^2 + b^T * var(y^*) * b - 2 * b^T * cov(u, (y^*)^T)$$

$$= \sigma_u^2 + b^T * P * b - 2 * b^T * G$$

Solution

- Compute $\frac{\partial PEV}{\partial b} = 0$

$$\frac{\partial PEV}{\partial b} = 2 * P * b - 2 * G = 0$$

$$\rightarrow b = P^{-1} * G$$

Problem with Correction

- Population mean is ideal as correction

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

- 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** $\rightarrow$ correlation between true (u) and its prediction ($\hat{u}$) is maximal or the prediction error variance ($var(u - \hat{u})$) is minimal.
  - **L** stands for **linear** $\rightarrow$ predicted breeding values are linear functions of the observations (y)
  - **U** stands for **unbiased** $\rightarrow$ 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
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

Linear Models

- Simple linear model

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

- Result: Estimate of effect of herd j
- Try with given dataset

Linear Mixed Effects Model

- 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

$\rightarrow$ **Linear Mixed Effects Model** (LME)

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

Matrix-Vector Notation

- LME for all animals of a population

$\rightarrow$ use matrix-vector notation

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

where

y	vector of length n of all observations
β	vector of length p of all fixed effects
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

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

- Variances

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

with $cov(u, e^T) = 0$,

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

Estimates of unknown Parameters

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

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

Mixed Model Equations

- Problem: V^{-1}
- Same solutions obtained with following set of equations

$$\left[\begin{array}{lr} X^T R^{-1} X & X^T R^{-1} Z \\ Z^T R^{-1} X & Z^T R^{-1} Z + G^{-1} \end{array}\right] \left[\begin{array}{c} \hat{\beta} \\ \hat{u} \end{array}\right] = \left[\begin{array}{c} X^T R^{-1} y \\ Z^T R^{-1} y \end{array}\right]$$

Sire Model

- Breeding value of sire as random effect:

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

Example

$$\begin{bmatrix} 2.61 \\ 2.31 \\ 2.44 \\ 2.41 \\ 2.51 \\ 2.55 \\ 2.14 \\ 2.61 \\ 2.34 \\ 1.99 \\ 3.1 \\ 2.81 \\ 2.14 \\ 2.41 \\ 2.54 \\ 3.16 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \\ e_9 \\ e_{10} \\ e_{11} \\ e_{12} \\ e_{13} \\ e_{14} \\ e_{15} \\ e_{16} \end{bmatrix}$$

Animal Model

- Breeding value for all animals as random effects

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