

# Best linear unbiased prediction (animal model)

Ole F. Christensen

Center for Quantitative Genetics and Genomics, Aarhus University

September 2023

# Simple univariate animal model

Model:  $\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{e}$

$$\mathbf{a} \sim N(\mathbf{0}, \sigma_a^2 \mathbf{A}), \quad \mathbf{e} \sim N(\mathbf{0}, \sigma_e^2 \mathbf{I})$$

- ▶ Phenotype = systematic part + random part
- ▶ Systematic part: sex, treatment, age, herds, season, year of birth, year of recording, etc
- ▶ Random part: variation not explained by systematic component (genetic factor: animal)
- ▶ Genetic effects in  $\mathbf{a}$  correlated due to pedigree-relationships  $\mathbf{A}$

# Best linear unbiased prediction

- ▶ EBV and fixed effects are solution to mixed model equations (MME):

$$\begin{bmatrix} \mathbf{X}^T \mathbf{X} & \mathbf{X}^T \mathbf{Z} \\ \mathbf{Z}^T \mathbf{X} & \mathbf{Z}^T \mathbf{Z} + \mathbf{A}^{-1} \alpha \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T \mathbf{y} \\ \mathbf{Z}^T \mathbf{y} \end{bmatrix},$$

where  $\alpha = \sigma_e^2 / \sigma_a^2$ .

- ▶ Note: Two purposes of matrix  $\mathbf{Z}$ :
  - ▶ EBV for all animals (also those without records).
  - ▶  $\mathbf{A}^{-1}$  is simple to compute for for all animals in pedigree.

# Best linear unbiased prediction

- ▶ **Best**: minimizes mean square error  $E[(a_i - \hat{a}_i)^2]$
- ▶ EBV is **linear** function of data (matrix algebra)
- ▶ **Unbiased**:  $E[\hat{a}_i] = E[a_i]$
- ▶ EBV is a **prediction** of random variable  $\mathbf{a}$  in model for phenotypes  $\mathbf{y}$
- ▶ Note: maximises accuracy  $r_i = \text{Cor}(a_i - \hat{a}_i)$ .
- ▶ Note: prediction error variance (PEV)  $\text{Var}(a_i - \hat{a}_i)$ .

# Accuracy of BLUP

$$\begin{bmatrix} \mathbf{X}^T \mathbf{X} & \mathbf{X}^T \mathbf{Z} \\ \mathbf{Z}^T \mathbf{X} & \mathbf{Z}^T \mathbf{Z} + \mathbf{A}^{-1} \alpha \end{bmatrix} \begin{bmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{a}} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^T \mathbf{y} \\ \mathbf{Z}^T \mathbf{y} \end{bmatrix},$$

- ▶ Invert the LHS matrix, and denote the inverse

$$\begin{bmatrix} \mathbf{C}^{11} & \mathbf{C}^{12} \\ \mathbf{C}^{21} & \mathbf{C}^{22} \end{bmatrix},$$

then  $\mathbf{PEV} = \mathbf{C}^{22} \sigma_e^2$ .

- ▶ Reliability:  $r_i^2 = 1 - PEV_{ii} / (\sigma_a^2 (1 + F_i))$

# Best linear unbiased prediction

- ▶ Infinitesimal model.
- ▶ Multivariate normal distribution.
- ▶ Selection accounted for, as long as data used for selection are in the model.
- ▶ Variance parameters assumed known.
- ▶ Estimation of variance parameters: REML (EM-REML, AI-REML, both relying on MME).

# More general models

- ▶ Non-genetic random effects
- ▶ Multi-trait models
- ▶ Indirect genetic effects
- ▶ Random regression effects (time dependence, .)
- ▶ Routine software for genetic evaluation using BLUP: blupf90, DMU/LMT, etc