Pedigree BLUP - Sire Model and Animal Model

Peter von Rohr

.

2023-05-08

$$y = Xb + e$$

random residuals

Response

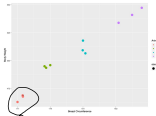Predictors

$$E(e) = 0$$

$$var(e) = I\sigma_e^2$$

$$var(e_i) = \sigma_e^2$$

$$cov(e_i, e_{j'}) = 0$$

Repeated Observations
* Body weight and Breast Circumference measured multiple times for the same animal.

Data Scatterplot

Same characteristics (Body Weight, BC, ...) measured multiple times for the same animal

Animal	Breast Circumference	Body Weight
2	177.0000	463.0000
2	177.3129	468.8940
2	177.3292	467.8753
5	179.0000	496.0000
5	178.6501	495.0033
5	178.7485	493.6563
7	181.0000	518.0000
7	180.9819	509.3221

$cov(e_1, e_2) \neq$

$cov(e_1, e_5)$

$var(e) \neq I\sigma_e^2 \Rightarrow var(e) = V = \begin{bmatrix} \sigma_c^2 & \sigma_1^2 & \sigma_r^2 & 0 & - \\ \sigma_r^2 & \sigma_e^2 & & \\ \sigma^2 & & & \end{bmatrix} \Big\} \quad Var(y) = V$

Ordinary Least Squares where where estimates of fixed effects were determined by the minimization of least squares objective of having a minimum of the sum of the squared residuals.

$$e^T e \to \text{minimal}$$

$$b^0 = (X^T X)^- X^T y$$

$$var(e) = I\sigma_e^2$$

$$q^T b^0$$

Estimable functions providing estimates

With a more general structure in the variance-covariance matrix, e.g. in repeated observations, ordinary least squares cannot be directly

$$var(y) = V = U\sigma_e^2$$

For positive definite matrices like V or U, the cholesky factorization can be applied

$$U = R^T R$$

Transformation of observations with the inverse of R

$$y = Xb + e \longrightarrow \text{Left-multiply both sides with inverse of R}$$

$$R^{-1}y = R^{-1}(Xb + e) \longrightarrow R^{-1}y = R^{-1}Xb + R^{-1}e$$

$$\underbrace{\phantom{R^{-1}y}}_{y^*} \quad \underbrace{\phantom{R^{-1}X}}_{X^*} \quad \underbrace{\phantom{R^{-1}e}}_{e^*}$$

$$y^* = R^{-1}y$$

$$\boxed{y^* = X^*b + e^*} \longrightarrow OLS$$

ordinary least squares

$$var(y^*) = var(R^{-1}y) = R^{-1}var(y)(R^{-1})^T = R^{-1}V(R^{-1})^T = R^{-1}U\sigma_e^2(R^{-1})^T$$

$$V = U\sigma_e^2$$

$$R^{-1}U\sigma_e^2(R^{-1})^T = R^{-1}U(R^{-1})^T\sigma_e^2 = \underbrace{R^{-1}R}_{I}\underbrace{R^T(R^{-1})^T}_{I}\sigma_e^2 = I\sigma_e^2$$

Solutions for the transformed model

$$y^* = X^*b + e^*$$

$$b^0 = ((X^*)^T X^*)^- (X^*)^T y^*$$

$$X^* = R^{-1} X$$

$$y^* = R^{-1} y$$

$$b^0 = ((R^{-1}X)^T (R^{-1}X))^- (R^{-1}X)^T (R^{-1}y)$$

$$b^0 = (X^T \underbrace{(R^{-1})^T R^{-1}}_{\substack{U^{-1} \cdot S_e^2 \\ \underbrace{\quad}_{V^{-1}}}} X)^- X^T \underbrace{(R^{-1})^T R^{-1}}_{\substack{U^{-1} \cdot S_e^2 \\ \underbrace{\quad}_{V^{-1}}}} y$$

$$b^0 = (X^T (V^{-1}) X)^- X^T (V^{-1}) y$$
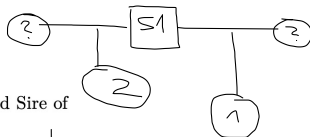
Generalized Least Squares

Independence Assumption

So far

- random residuals: $var(e) = I * \sigma_e^2$ and
- random sire effects: $var(s) = I * \sigma_s^2$

In real livestock datasets, this is not realistic, because sires are related

Table 4.5: Body Weight, Breast Circumference, Breed and Sire of Beef Cattle Animals
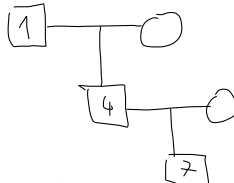
Animal	Body Weight	Breast Circumference	Breed	Sire
1	471	176	Angus	S1
2	463	177	Angus	S1
3	481	178	Simmental	S3
4	470	179	Angus	S2
5	496	179	Simmental	S3
6	491	180	Simmental	S4
7	518	181	Limousin	S5
8	511	182	Limousin	S5
9	510	183	Limousin	S6
10	541	184	Limousin	S6

Half-sibs

no observations for sires

$cov(y_1, y_2) \neq cov(y_1, y_3)$

Example Dataset

Fixed

Animal	Sire	Sex	WWG
4	1	M	4.5
5	3	F	2.9
6	1	F	3.9
7	4	M	3.5
8	3	M	5.0

$$cov(y_4, y_5)$$
$$\neq cov(y_4, y_6)$$

y

The previous two example datasets contain both the sire effect which influences the structure of the variance-covariance matrix of the responses

==> Include these effects into the model such that the variance-covariance structure is appropriately considered.

In contrast to the repeated observations model, we have to include sire as an additional effect into the model. This sire effect will be a "random" effect, because only with random effects, we explicitly specify the variance-covariance structure that these effect should have.

==> The consequence of including a random effect into a linear model makes the resulting model now a so-called Mixed Linear Effects Model (=mixed models) (MLEM).

Fixed linear effects model

$$y = Xb + Zu + e$$

Vector of random effects

$$E(u) = 0 \qquad var(u) = D$$

$$E \begin{bmatrix} \mathbf{y} \\ \mathbf{u} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{Xb} \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix}$$

$$var \begin{bmatrix} \mathbf{y} \\ \mathbf{u} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \overbrace{\mathbf{ZDZ^T + R}}^{var\ y} & \mathbf{ZD} & \mathbf{R} \\ \mathbf{DZ^T} & \mathbf{D} & \mathbf{0} \\ \mathbf{R} & \mathbf{0} & \mathbf{R} \end{bmatrix}$$

cov(y, u^T)

Relationship

E.g. $k = 1$
       $i = 4$

▶ For son i and sire k of i

Probability of sharing two alleles identical by descent at any given location in the genome.

$$cov(s_i, s_k) = 1/2 * \sigma_s^2$$

Sire Relationship Matrix

```
library(pedigreemm)
ped_sire <- pedigree(sire = c(rep(NA,2), 1),
                     dam = rep(NA,3),
                     label = as.character(c(1,3,4)))
mat_A <- getA(ped = ped_sire)
```

$$
A = \begin{array}{c} \\ 1 \\ 3 \\ 4 \end{array}
\begin{array}{ccc} 1 & 3 & 4 \\ \end{array}
\begin{bmatrix} 1 & 0 & 0.5 \\ 0 & 1 & 0 \\ 0.5 & 0 & 1 \end{bmatrix}
$$

Sire Model

▶ pedigreemm

```
lmem_sire <- pedigreemm(
  formula = WWG ~ Sex + (1 | Sire),
  data = tbl_sire_model,
  pedigree = list(Sire = ped_sire)
)
```

```
## boundary (singular) fit: see help('isSingular')
```

```
summary(lmem_sire)
```

```
## Linear mixed model fit by REML ['lmerpedigreemm']
## Formula: WWG ~ Sex + (1 | Sire)
##    Data: tbl_sire_model
##
## REML criterion at convergence: 8.5
##
## Scaled residuals:
```

Mixed model equations

▶ `pedigreemm`: Cannot specify assumed variance components

$$\begin{bmatrix} X^T X & X^T Z \\ Z^T X & Z^T Z + \lambda A_s^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{s} \end{bmatrix} = \begin{bmatrix} X^T y \\ Z^T y \end{bmatrix}$$

Animal Model

- Extension of sire model
- Include dams in pedigree
- Predict breeding values for all animals in the pedigree

Dataset

Animal	Sire	Dam	Sex	WWG
4	1	NA	M	4.5
5	3	2	F	2.9
6	1	2	F	3.9
7	4	5	M	3.5
8	3	6	M	5.0

where `NA` stands for unknown

The Model

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zu} + \mathbf{e}$$

► Random effects $\mathbf{u}$ and $\mathbf{e}$

$$E(\mathbf{e}) = \mathbf{0}$$
$$var(\mathbf{e}) = \mathbf{I} * \sigma_e^2$$
$$E(\mathbf{u}) = \mathbf{0}$$
$$var(\mathbf{u}) = \mathbf{A} * \sigma_u^2$$

with numerator relationship matrix $\mathbf{A}$

Numerator Relationship Matrix $\mathbf{A}$

▶ Diagonal elements

$$(\mathbf{A})_{ii} = 1 + F_i$$

where F_i is the inbreeding coefficient

$$F_i = \frac{1}{2} * (\mathbf{A})_{sd}$$

▶ Off-diagonal elements

$$cov(u_i, u_j) = (\mathbf{A})_{ij} * \sigma_u^2$$

Example

```
library(pedigreemm)
ped_ani <- pedigree(sire = c(rep(NA, n_nr_founder),
                             1,3,1,4,3),
                    dam  = c(rep(NA, n_nr_founder),
                             NA,2,2,5,6),
                    label = as.character(1:n_nr_animal))
mat_A <- as.matrix(getA(ped = ped_ani))
```

$$
A = \begin{bmatrix}
1 & 0 & 0 & 0.5 & 0 & 0.5 & 0.25 & 0.25 \\
0 & 1 & 0 & 0 & 0.5 & 0.5 & 0.25 & 0.25 \\
0 & 0 & 1 & 0 & 0.5 & 0 & 0.25 & 0.5 \\
0.5 & 0 & 0 & 1 & 0 & 0.25 & 0.5 & 0.125 \\
0 & 0.5 & 0.5 & 0 & 1 & 0.25 & 0.5 & 0.375 \\
0.5 & 0.5 & 0 & 0.25 & 0.25 & 1 & 0.25 & 0.5 \\
0.25 & 0.25 & 0.25 & 0.5 & 0.5 & 0.25 & 1 & 0.25 \\
0.25 & 0.25 & 0.5 & 0.125 & 0.375 & 0.5 & 0.25 & 1
\end{bmatrix}
$$

Solution

- ▶ `pedigreemm` cannot handle such small datasets with only one observation per animal
- ▶ Mixed model equations

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

with $\lambda = \sigma_e^2 / \sigma_u^2$

Genomic BIUP

1. Marker effect models (MEM): Linear mixed effects models with marker effects as random effects
2. Breeding-value based models (BVM): Genomic breeding values as random effects

Marker Effect Models

▶ Model
$$y = 1_n\mu + Wq + e$$

▶ Solution

$$\begin{bmatrix} 1_n^T 1_n & 1_n^T W \\ W^T 1_n & W^T W + \lambda_q * I \end{bmatrix} \begin{bmatrix} \hat{\mu} \\ \hat{q} \end{bmatrix} = \begin{bmatrix} 1_n^T y \\ W^T y \end{bmatrix}$$

with $\lambda_q = \sigma_e^2/\sigma_q^2$.

Breeding Value Models

▶ Model

$$y = Xb + Zg + e$$

▶ Solution

$$\begin{bmatrix} X^T X & X^T Z \\ Z^T X & Z^T Z + \lambda_g * G^{-1} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} X^T y \\ Z^T y \end{bmatrix}$$

with $\lambda_g = \sigma_e^2 / \sigma_g^2$.

Genomic Relationship Matrix

$$g = U \cdot q$$

with $U = W - P$ and P has columns $2p_j - 1$ with p_j being the frequency of the positive allele at locus j.

$$var(g) = G * \sigma_g^2$$

$$var(g) = UU^T * \sigma_q^2$$

$$\sigma_g^2 = 2 \sum_{j=1}^{m} p_j(1 - p_j)\sigma_q^2$$

Genomic Relationship Matrix II

$$var(g) = G * \sigma_g^2 = UU^T \sigma_q^2$$

$$G = \frac{UU^T}{2 \sum_{j=1}^m p_j (1 - p_j)}$$