

Livestock Breeding and Genomics - Solution 7

Peter von Rohr

2018-11-09

Problem 1: Predict Breeding Values Using Selection Index

The data set from exercise 6 was extended by an own performance record for our selection candidate i . The new data set is shown in the table below. Predict the breeding value a_i for animal i using all available information using the selection index.

We assume that our selection candidate i has animals 1 and 2 as parents and the observation for our selection candidate is shown last in the above table as **Measurement 10**. The heritability is assumed to be $h^2 = 0.36$ and the phenotypic standard deviation corresponds to $\sigma_y = 31.8$. The population mean μ is assumed to be 250.4.

Table 1: Phenotypic Measurements for Full-Sibs, Half-Sibs and an Own Performance Record

Animal	Sire	Dam	Weight
4	1	2	270.10
5	1	2	263.52
6	1	3	221.49
7	1	3	280.41
8	1	3	215.75
9	1	3	292.45
10	1	2	187.56

Figure 1 shows the structure of the data shown in Table 1.

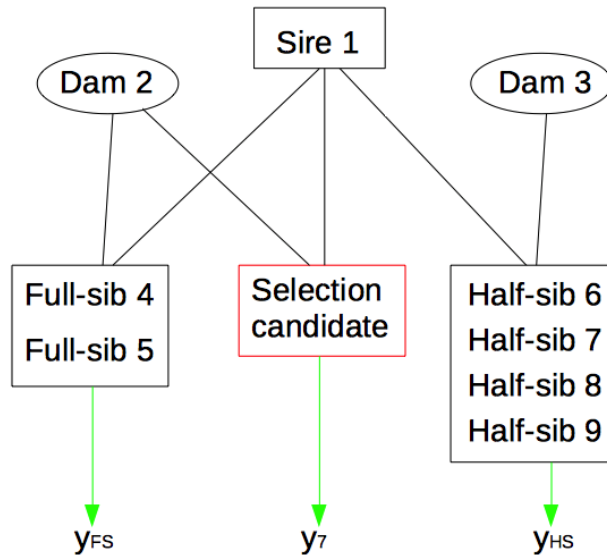


Figure 1: Information Sources To Predict Breeding Values

Solution

First, we determine the matrices P and G . From those matrices, we compute the vector b of index weights.

- When considering the averages of the full-sibs and the half-sibs together with an own performance record, the Matrix P has the following structure:

$$P = \begin{bmatrix} \text{var}(\bar{y}_{FS}) & \text{cov}(\bar{y}_{FS}, \bar{y}_{HS}) & \text{cov}(\bar{y}_{FS}, y_{10}) \\ \text{cov}(\bar{y}_{FS}, \bar{y}_{HS}) & \text{var}(\bar{y}_{HS}) & \text{cov}(\bar{y}_{HS}, y_{10}) \\ \text{cov}(\bar{y}_{FS}, y_{10}) & \text{cov}(\bar{y}_{HS}, y_{10}) & \text{var}(y_{10}) \end{bmatrix}$$

The matrix P shown above is an extended version of the matrix P from Problem 2 in Exercise 6. The new elements here are

- $\text{cov}(\bar{y}_{FS}, y_{10})$

$$\begin{aligned} \text{cov}(\bar{y}_{FS}, y_{10}) &= \text{cov}\left(\frac{1}{2} \sum_{j=4}^5 (\mu + a_j + e_j), (\mu + a_{10} + e_{10})\right) \\ &= \frac{1}{2} \left(\sum_{j=4}^5 \text{cov}(a_j, a_{10}) \right) \\ &= \frac{1}{2} \left(\sum_{j=4}^5 \text{cov}\left(\left(\frac{1}{2}a_1 + \frac{1}{2}a_2 + m_j\right), \left(\frac{1}{2}a_1 + \frac{1}{2}a_2 + m_{10}\right)\right) \right) \\ &= \frac{1}{2} \left(2 * \left(\text{cov}\left(\frac{1}{2}a_1, \frac{1}{2}a_1\right) + \text{cov}\left(\frac{1}{2}a_2, \frac{1}{2}a_2\right) \right) \right) \\ &= \frac{1}{2} \left(\frac{1}{2} (\text{cov}(a_1, a_1) + \text{cov}(a_2, a_2)) \right) \\ &= \frac{1}{2} \left(\frac{1}{2} (\sigma_a^2 + \sigma_a^2) \right) \\ &= \frac{1}{2} \sigma_a^2 = \frac{1}{2} h^2 \sigma_y^2 \end{aligned}$$

- $\text{cov}(\bar{y}_{HS}, y_{10})$

$$\begin{aligned} \text{cov}(\bar{y}_{HS}, y_{10}) &= \text{cov}\left(\frac{1}{4} \sum_{j=6}^9 (\mu + a_j + e_j), (\mu + a_{10} + e_{10})\right) \\ &= \frac{1}{4} \left(\sum_{j=6}^9 \text{cov}(a_j, a_{10}) \right) \\ &= \frac{1}{4} \left(\sum_{j=6}^9 \text{cov}\left(\left(\frac{1}{2}a_1 + \frac{1}{2}a_3 + m_j\right), \left(\frac{1}{2}a_1 + \frac{1}{2}a_2 + m_{10}\right)\right) \right) \\ &= \frac{1}{4} \left(4 * \text{cov}\left(\frac{1}{2}a_1, \frac{1}{2}a_1\right) \right) \\ &= \frac{1}{4} (\text{cov}(a_1, a_1)) \\ &= \frac{1}{4} (\sigma_a^2) = \frac{1}{4} h^2 \sigma_y^2 \end{aligned}$$

- $var(y_{10}) = \sigma_y^2$

Combining all these elements, we can put together the matrix P

$$P = \begin{bmatrix} \frac{1+(n_{FS}-1)h^2/2}{n_{FS}} * \sigma_y^2 & \frac{1}{4}h^2\sigma_y^2 & \frac{1}{2}h^2\sigma_y^2 \\ \frac{1}{4}h^2\sigma_y^2 & \frac{1+(n_{HS}-1)h^2/2}{n_{HS}} * \sigma_y^2 & \frac{1}{4}h^2\sigma_y^2 \\ \frac{1}{2}h^2\sigma_y^2 & \frac{1}{4}h^2\sigma_y^2 & \sigma_y^2 \end{bmatrix}$$

For the matrix G , we can extend the matrix G from Problem 2 in Exercise 6 by the following covariance between the true breeding value a_{10} and the additional source of information corresponding to y_{10} .

$$\begin{aligned} cov(a_{10}, y_{10}) &= cov(a_{10}, (\mu + a_{10} + e_{10})) \\ &= cov(a_{10}, a_{10}) = \sigma_a^2 = h^2\sigma_y^2 \end{aligned}$$

Hence the complete matrix G has the following form

$$G = \begin{bmatrix} \frac{1}{2}h^2\sigma_y^2 \\ \frac{1}{4}h^2\sigma_y^2 \\ h^2\sigma_y^2 \end{bmatrix}$$

The vector b of index weights is now computed using the index normal equations

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

Because all elements of P and G contain σ_y^2 we can divide both sides of the index normal equations by σ_y^2 leading to

$$\begin{aligned} b &= \begin{bmatrix} \frac{1+(n_{FS}-1)h^2/2}{n_{FS}} & \frac{1}{4}h^2 & \frac{1}{2}h^2 \\ \frac{1}{4}h^2 & \frac{1+(n_{HS}-1)h^2/2}{n_{HS}} & \frac{1}{4}h^2 \\ \frac{1}{2}h^2 & \frac{1}{4}h^2 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{2}h^2 \\ \frac{1}{4}h^2 \\ h^2 \end{bmatrix} \\ &= \begin{bmatrix} \frac{1+(2-1)*0.36/2}{2} & \frac{1}{4} * 0.36 & \frac{1}{2} * 0.36 \\ \frac{1}{4} * 0.36 & \frac{1+(4-1)*0.36/2}{4} & \frac{1}{4} * 0.36 \\ \frac{1}{2} * 0.36 & \frac{1}{4} * 0.36 & 1 \end{bmatrix}^{-1} \begin{bmatrix} \frac{1}{2} * 0.36 \\ \frac{1}{4} * 0.36 \\ 0.36 \end{bmatrix} \\ &= \begin{bmatrix} 0.59 & 0.09 & 0.18 \\ 0.09 & 0.385 & 0.09 \\ 0.18 & 0.09 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0.18 \\ 0.09 \\ 0.36 \end{bmatrix} \\ &= \begin{bmatrix} 0.1913 \\ 0.1154 \\ 0.3152 \end{bmatrix} \end{aligned}$$

The predicted breeding value \hat{a}_{10} corresponds to the index I which is computed as

$$\hat{a}_{10} = I = b^T * y^*$$

where y^* corresponds to the information sources (\bar{y}_{FS} , \bar{y}_{HS} and y_{10}) corrected for the population mean μ . The vector y^* can be written as

$$y^* = \begin{bmatrix} \bar{y}_{FS} - \mu \\ \bar{y}_{HS} - \mu \\ y_{10} - \mu \end{bmatrix} = \begin{bmatrix} 266.81 - 250.4 \\ 252.525 - 250.4 \\ 187.56 - 250.4 \end{bmatrix} = \begin{bmatrix} 16.41 \\ 2.125 \\ -62.84 \end{bmatrix}$$

Together with the vector b , we get

$$\hat{a}_i = I = b^T * y^* = \begin{bmatrix} 0.1913313 \\ 0.1153615 \\ 0.3151778 \end{bmatrix}^T * \begin{bmatrix} 16.41 \\ 2.125 \\ -62.84 \end{bmatrix} = -16.42$$

Comparing this result to the result of Problem 2 in Exercise 6, we can see that when we include the own performance of animal 10, the predicted breeding value gets negative. The reason for this is the low value of the own performance record of animal 10 which is included with a high index weight.

Problem 2: Sire Model

We are using the following dataset shown in Table 2 to predict breeding values using a sire model.

Table 2: Example Data Set for Weaning Weight in Beef Cattle

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

Your Tasks

- Specify the sire model for the dataset given in Table 2.
- Besides the model indicate also the expected values and the variances for all the random components in the model.
- Set up the mixed model equations for the sire model and compute the estimates for the fixed effects (Herd) and the predicted breeding values for the sires.

Assumptions

We assume that the sires are unrelated and that the genetic additive variance $\sigma_a^2 = 0.0254$. Hence the variance-covariance matrix G of all sire effects corresponds to

$$\text{var}(s) = G = I * \sigma_s^2 = I * \frac{\sigma_a^2}{4}$$

Furthermore, the residuals e are not correlated which means that the variance-covariance matrix R is

$$\text{var}(e) = R = I * \sigma_e^2$$

with $\sigma_e^2 = 0.076$.

Solution

The model

The sire model for the data set given in Table 2 has the following structure

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

where

- y vector of length 16 of phenotypic observations
- β vector of length 2 of unknown fixed herd effects
- X 16×2 design matrix linking observations to fixed effects
- s vector of length 3 of unknown random sire effects
- Z 16×3 design matrix linking observations to sire effects
- e vector of length 16 of unknown random residual effects

As in the lecture notes, we can put the information from the dataset into the model leading to

$$\begin{bmatrix} 2.61 \\ 2.31 \\ 2.44 \\ 2.41 \\ 2.51 \\ 2.55 \\ 2.14 \\ 2.61 \\ 2.34 \\ 1.99 \\ 3.10 \\ 2.81 \\ 2.14 \\ 2.41 \\ 2.54 \\ 3.16 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} s_1 \\ s_2 \\ s_3 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \\ e_9 \\ e_{10} \\ e_{11} \\ e_{12} \\ e_{13} \\ e_{14} \\ e_{15} \\ e_{16} \end{bmatrix}$$

Expected Values and Variances

We fix the expected values $E(s)$ and $E(e)$ of the random components s and e to be

$$E(s) = 0 \quad \text{and} \quad E(e) = 0$$

From this we can compute $E(y) = X\beta$.

The variances $var(s)$ and $var(e)$ of the random components are defined as

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

The variance $var(y)$ can be computed as $var(y) = V = Z^T G Z + R$.

Using the assumptions, we can further specify

$$var(s) = G = I * \sigma_s^2 = I * \frac{\sigma_a^2}{4} \quad \text{and} \quad var(e) = R = I * \sigma_e^2$$

Mixed Model Equations

The general form of the mixed model equations was presented in the lecture notes. We now use that form for the sire model which results in

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

Using the above specified assumptions regarding the variance-covariance matrices G and R , the mixed model equations can be simplified to

$$\begin{bmatrix} X^T X & X^T Z \\ Z^T X & Z^T Z + I * 4 * \lambda \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{s} \end{bmatrix} = \begin{bmatrix} X^T y \\ Z^T y \end{bmatrix}$$

where $\lambda = \frac{\sigma_e^2}{\sigma_a^2}$.

Inserting the number leads to

$$\begin{bmatrix} 8.0000 & 0.0000 & 4.0000 & 4.0000 & 0.0000 \\ 0.0000 & 8.0000 & 4.0000 & 2.0000 & 2.0000 \\ 4.0000 & 4.0000 & 19.9685 & 0.0000 & 0.0000 \\ 4.0000 & 2.0000 & 0.0000 & 17.9685 & 0.0000 \\ 0.0000 & 2.0000 & 0.0000 & 0.0000 & 13.9685 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \end{bmatrix} = \begin{bmatrix} 20.01 \\ 20.06 \\ 19.58 \\ 14.79 \\ 5.70 \end{bmatrix}$$

The Solution

The solution consists of the estimates for the fixed effects and the predictions of the sire breeding values. Because, we are working with a very small dataset, we can obtain the solutions by pre-multiplying both sides of the mixed model equations with the inverse of the coefficient matrix. Hence

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \end{bmatrix} = \begin{bmatrix} 8.0000 & 0.0000 & 4.0000 & 4.0000 & 0.0000 \\ 0.0000 & 8.0000 & 4.0000 & 2.0000 & 2.0000 \\ 4.0000 & 4.0000 & 19.9685 & 0.0000 & 0.0000 \\ 4.0000 & 2.0000 & 0.0000 & 17.9685 & 0.0000 \\ 0.0000 & 2.0000 & 0.0000 & 0.0000 & 13.9685 \end{bmatrix}^{-1} \begin{bmatrix} 20.01 \\ 20.06 \\ 19.58 \\ 14.79 \\ 5.70 \end{bmatrix} = \begin{bmatrix} 2.5252 \\ 2.5148 \\ -0.0290 \\ -0.0190 \\ 0.0480 \end{bmatrix}$$

Ranking the sires

Now that we have predicted breeding values for all sire, we can rank the accordingly. For our example the ranking of the sires is

Table 3: Ranking of Sires According To Predicted Breeding Values

Sire	Predicted Breeding Value	Rank
1	-0.0290481	3
2	-0.0189511	2
3	0.0479991	1