

Livestock Breeding and Genomics - Solution 5

Peter von Rohr

20120-10-23

Prediction of Breeding Values Using the Regression Method

We are using the dataset shown in Table 1 for this exercise. For the animals listed in Table 1, the weaning weight (in 100kg) was observed as phenotypic records. The following parameters are associated with the observed data

- The population mean is assumed to be equal to the average of all observations: $\mu = 2.5$
- The phenotypic variance is assumed to correspond to the empirical variance from the observations and corresponds to $\sigma_p^2 = 0.1014$
- The heritability is assumed to be $h^2 = 0.25$
- The genetic-additive variance can be computed as $\sigma_a^2 = h^2 * \sigma_p^2 = 0.25 * 0.1014 = 0.0254$

Table 1: Example Data Set To Predict Breeding Values

Animal	Sire	Weaning Weight
12	1	2.61
13	1	2.31
14	1	2.44
15	1	2.41
16	1	2.51
17	1	2.55
18	1	2.14
19	1	2.61
20	2	2.34
21	2	1.99
22	2	3.10
23	2	2.81
24	2	2.14
25	2	2.41
26	3	2.54
27	3	3.16

Problem 1: Own performance

Compute the predicted breeding values and the reliabilities for the animals listed in Table 1. Compare the ranking of the animals according to their phenotypic values and according to their predicted breeding values. Compare the reliabilities of the predicted breeding values.

Solution

The predicted breeding values based on own performance are computed as

$$\hat{a}_i = h^2(y_i - \mu)$$

The reliabilities are constant and correspond to

$$B = r_{a,y}^2 = h^2$$

The results are listed in the following table.

Table 2: Predicted Breeding Values Using Own Performance Records

Animal	Weaning Weight	Predicted Breeding Value	Reliability
12	2.61	0.0275	0.25
13	2.31	-0.0475	0.25
14	2.44	-0.0150	0.25
15	2.41	-0.0225	0.25
16	2.51	0.0025	0.25
17	2.55	0.0125	0.25
18	2.14	-0.0900	0.25
19	2.61	0.0275	0.25
20	2.34	-0.0400	0.25
21	1.99	-0.1275	0.25
22	3.10	0.1500	0.25
23	2.81	0.0775	0.25
24	2.14	-0.0900	0.25
25	2.41	-0.0225	0.25
26	2.54	0.0100	0.25
27	3.16	0.1650	0.25

The ranking according to the phenotypic records and according to the predicted breeding values are the same, because each phenotypic record is corrected for the same population mean and is multiplied with the same factor which corresponds to h^2 . The main difference between the phenotypic records and the predicted breeding values is the variability. The predicted breeding values have a much smaller variability compared to the phenotypic records.

Problem 2: Predicted Breeding Values Based on Progeny Records

Compute the predicted breeding values and the reliabilities for the sires based on the progeny records. We are assuming that all progeny for a given sire are half-sibbs. Compare the ranking of the sires according to the average progeny performance values and according to the predicted breeding values.

Solution

The predicted breeding values for the sires based on the average performance of their progeny is computed as

$$\hat{a}_i = \frac{2n_i}{n_i + k}(\bar{y}_i - \mu)$$

where n_i is the number of progeny of sire i , \bar{y}_i is the average of the performance values of the progeny of sire i and $k = \frac{4-h^2}{h^2} = \frac{4-0.25}{0.25} = 15$. The reliabilities are no longer constant, but they depend on the number of progeny

$$B_i = \frac{n_i}{n_i + k}$$

Preparatory Steps

Before we compute the predicted values, we have to prepare a few intermediate quantities that are needed for the computation such as

- the number n_i of progeny records for each sire
- the average progeny performance value \bar{y}_i for each sire

`summarise()` ungrouping output (override with `.groups` argument)

Table 3: Intermediate Results To Predict Breeding Values

Sire	Number of Progeny Records	Average Progeny Performance
1	8	2.4475
2	6	2.4650
3	2	2.8500

Compute Predicted Breeding Values and Reliabilities

We are using the intermediate results to compute the predicted breeding values for each sire.

Table 4: Predicted Breeding Values and Reliabilities for all Sires

Sire	Predicted Breeding Value	Reliability
1	-0.0365217	0.3478261
2	-0.0200000	0.2857143
3	0.0823529	0.1176471

The rankings of the sires according to the progeny averages and according to the predicted breeding values are the same. The reliability for sire 1 is the highest. This is mainly due to the larger number of progeny of sire 1.

Problem 3: Sire Model

We are using the following dataset shown in Table 5 to predict breeding values using a sire model.

Table 5: 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 5.
- 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 $\sigma_u^2 = 0.0254$. Hence the variance-covariance matrix G of all sire effects corresponds to

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

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 5 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 leaning to

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

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

$$\begin{bmatrix} 8 & 0 & 4 & 4 & 0 \\ 0 & 8 & 4 & 2 & 2 \\ 4 & 4 & 19.9685039370079 & 0 & 0 \\ 4 & 2 & 0 & 17.9685039370079 & 0 \\ 0 & 2 & 0 & 0 & 13.9685039370079 \end{bmatrix} \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \end{bmatrix} = \begin{bmatrix} 20.01 \\ 20.06 \\ 19.58 \\ 14.79 \\ 5.7 \end{bmatrix}$$

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

$$\begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{s}_1 \\ \hat{s}_2 \\ \hat{s}_3 \end{bmatrix} = \begin{bmatrix} 8 & 0 & 4 & 4 & 0 \\ 0 & 8 & 4 & 2 & 2 \\ 4 & 4 & 19.9685039370079 & 0 & 0 \\ 4 & 2 & 0 & 17.9685039370079 & 0 \\ 0 & 2 & 0 & 0 & 13.9685039370079 \end{bmatrix}^{-1} \begin{bmatrix} 20.01 \\ 20.06 \\ 19.58 \\ 14.79 \\ 5.7 \end{bmatrix} = \begin{bmatrix} 2.52524956261459 \\ 2.51476201513161 \\ -0.0290480605264477 \\ -0.0189510647027351 \\ 0.047999125229183 \end{bmatrix}$$

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 6: 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