Livestock Breeding and Genomics - Solution 12

Peter von Rohr

2018-12-13

Problem 1 Multivariate BLUP Animal Model

The table below contains data for pre-weaning gain (WWG) and post-weaning gain (PWG) for 5 beef calves.

Animal	Sex	Sire	Dam	WWG	PWG
4	Male	1	NA	4.5	6.8
5	Female	3	2	2.9	5.0
6	Female	1	2	3.9	6.8
7	Male	4	5	3.5	6.0
8	Male	3	6	5.0	7.5

The genetic variance-covariance matrix G_0 between the traits is

$$G_0 = \left[\begin{array}{cc} 20 & 18 \\ 18 & 40 \end{array}\right]$$

The residual variance-covariance matrix R_0 between the traits is

$$R_0 = \left[\begin{array}{cc} 40 & 11 \\ 11 & 30 \end{array}\right]$$

Your Task

Set up the mixed model equations for a multivariate BLUP analysis and compute the estimates for the fixed effects and the predictions for the breeding values.

Solution

The matrices X_1 and X_2 relate records of PWG and WWG to sex effects. For both traits, we have an effect for the male and female sex. Hence the vector β of fixed effects corresponds to

$$\beta = \left[\begin{array}{c} \beta_{M,WWG} \\ \beta_{F,WWG} \\ \beta_{M,PWG} \\ \beta_{F,PWG} \end{array}\right]$$

The matrices X_1 and X_2 are the same and correspond to

$$X_1 = X_2 = \left[\begin{array}{cc} 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 1 & 0 \\ 1 & 0 \end{array}\right]$$

Combining them to the multivariate version leads to

$$X = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Using the matrix X together with matrix $R = I_n \otimes R_0$ to get

$$X^TR^{-1}X = \begin{bmatrix} 0.083 & 0.000 & -0.031 & 0.000 \\ 0.000 & 0.056 & 0.000 & -0.020 \\ -0.031 & 0.000 & 0.111 & 0.000 \\ 0.000 & -0.020 & 0.000 & 0.074 \end{bmatrix}$$

Similarly to the fixed effects, we can put together the vector of breeding values a.

$$a = \begin{bmatrix} a_{1,WWG} \\ a_{2,WWG} \\ a_{3,WWG} \\ a_{4,WWG} \\ a_{5,WWG} \\ a_{6,WWG} \\ a_{7,WWG} \\ a_{8,WWG} \\ a_{1,PWG} \\ a_{2,PWG} \\ a_{3,PWG} \\ a_{4,PWG} \\ a_{5,PWG} \\ a_{6,PWG} \\ a_{7,PWG} \\ a_{8,PWG} \end{bmatrix}$$

The design matrices Z_1 and Z_2 are equal and they link observations to breeding values.

$$Z_1 = Z_2 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$Z = \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix}$$

$$Z = \left[\begin{array}{cccccccccccccccc}
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{array}\right]$$

Together with the numerator relationship matrix A we can get $G = G_0 \otimes A$ and from this $G^{-1} = G_0^{-1} \otimes A^{-1}$

$$A^{-1} = \left[\begin{array}{rrrrrrrr}
1.833 & 0.500 & 0.000 & -0.667 & 0.000 & -1.000 & 0.000 & 0.000 \\
0.500 & 2.000 & 0.500 & 0.000 & -1.000 & -1.000 & 0.000 & 0.000 \\
0.000 & 0.500 & 2.000 & 0.000 & -1.000 & 0.500 & 0.000 & -1.000 \\
-0.667 & 0.000 & 0.000 & 1.833 & 0.500 & 0.000 & -1.000 & 0.000 \\
0.000 & -1.000 & -1.000 & 0.500 & 2.500 & 0.000 & -1.000 & 0.000 \\
-1.000 & -1.000 & 0.500 & 0.000 & 0.000 & 2.500 & 0.000 & -1.000 \\
0.000 & 0.000 & 0.000 & -1.000 & -1.000 & 0.000 & 2.000 & 0.000 \\
0.000 & 0.000 & -1.000 & 0.000 & 0.000 & -1.000 & 0.000 & 2.000
\end{array}\right]$$

$$G^{-1} = \left[\begin{array}{rrrrrrrrrrrrrrrr}
0.15 & 0.04 & 0.00 & -0.06 & 0.00 & -0.08 & 0.00 & 0.00 & -0.07 & -0.02 & 0.00 & 0.03 & 0.00 & 0.04 & 0.00 & 0.00 \\
0.04 & 0.17 & 0.04 & 0.00 & -0.08 & -0.08 & 0.00 & 0.00 & -0.02 & -0.08 & -0.02 & 0.00 & 0.04 & 0.04 & 0.00 & 0.00 \\
0.00 & 0.04 & 0.17 & 0.00 & -0.08 & 0.04 & 0.00 & -0.08 & 0.00 & -0.02 & -0.08 & 0.00 & 0.04 & -0.02 & 0.00 & 0.04 \\
-0.06 & 0.00 & 0.00 & 0.15 & 0.04 & 0.00 & -0.08 & 0.00 & 0.03 & 0.00 & 0.00 & -0.07 & -0.02 & 0.00 & 0.04 & 0.00 \\
0.00 & -0.08 & -0.08 & 0.04 & 0.21 & 0.00 & -0.08 & 0.00 & 0.00 & 0.04 & 0.04 & -0.02 & -0.09 & 0.00 & 0.04 & 0.00 \\
-0.08 & -0.08 & 0.04 & 0.00 & 0.00 & 0.21 & 0.00 & -0.08 & 0.04 & 0.04 & -0.02 & 0.00 & 0.00 & -0.09 & 0.00 & 0.04 \\
0.00 & 0.00 & 0.00 & -0.08 & -0.08 & 0.00 & 0.17 & 0.00 & 0.00 & 0.00 & 0.00 & 0.04 & 0.04 & 0.00 & -0.08 & 0.00 \\
0.00 & 0.00 & -0.08 & 0.00 & 0.00 & -0.08 & 0.00 & 0.17 & 0.00 & 0.00 & 0.04 & 0.00 & 0.00 & 0.04 & 0.00 & -0.08 \\
-0.07 & -0.02 & 0.00 & 0.03 & 0.00 & 0.04 & 0.00 & 0.00 & 0.08 & 0.02 & 0.00 & -0.03 & 0.00 & -0.04 & 0.00 & 0.00 \\
-0.02 & -0.08 & -0.02 & 0.00 & 0.04 & 0.04 & 0.00 & 0.00 & 0.02 & 0.08 & 0.02 & 0.00 & -0.04 & -0.04 & 0.00 & 0.00 \\
0.00 & -0.02 & -0.08 & 0.00 & 0.04 & -0.02 & 0.00 & 0.04 & 0.00 & 0.02 & 0.08 & 0.00 & -0.04 & 0.02 & 0.00 & -0.04 \\
0.03 & 0.00 & 0.00 & -0.07 & -0.02 & 0.00 & 0.04 & 0.00 & -0.03 & 0.00 & 0.00 & 0.08 & 0.02 & 0.00 & -0.04 & 0.00 \\
0.00 & 0.04 & 0.04 & -0.02 & -0.09 & 0.00 & 0.04 & 0.00 & 0.00 & -0.04 & -0.04 & 0.02 & 0.11 & 0.00 & -0.04 & 0.00 \\
0.04 & 0.04 & -0.02 & 0.00 & 0.00 & -0.09 & 0.00 & 0.04 & -0.04 & -0.04 & 0.02 & 0.00 & 0.00 & 0.11 & 0.00 & -0.04 \\
0.00 & 0.00 & 0.00 & 0.04 & 0.04 & 0.00 & -0.08 & 0.00 & 0.00 & 0.00 & 0.00 & -0.04 & -0.04 & 0.00 & 0.08 & 0.00 \\
0.00 & 0.00 & 0.04 & 0.00 & 0.00 & 0.04 & 0.00 & -0.08 & 0.00 & 0.00 & -0.04 & 0.00 & 0.00 & -0.04 & 0.00 & 0.08
\end{array}\right]$$

Using the matrics X, Z, R^{-1} and G^{-1}, we can compute $Z^T R^{-1} X$ and $Z^T R^{-1} Z + G^{-1}$. These matrices define the right-hand side of the mixed model equations. But they are too be to be shown here.

The vector y of observations contains all observations of both traits

$$y = \left[\begin{array}{c}
4.50 \\
2.90 \\
3.90 \\
3.50 \\
5.00 \\
6.80 \\
5.00 \\
6.80 \\
6.00 \\
7.50
\end{array}\right]$$

The right-hand side is computed as

$$\left[\begin{array}{c}
X^T R^{-1} y \\
Z^T R^{-1} y
\end{array}\right]$$

The solutions are

$$
\left[\begin{array}{c}
\widehat{\beta_{M,WWG}} \\
\widehat{\beta_{F,WWG}} \\
\widehat{\beta_{M,PWG}} \\
\widehat{\beta_{F,PWG}} \\
\widehat{a_{1,WWG}} \\
\widehat{a_{2,WWG}} \\
\widehat{a_{3,WWG}} \\
\widehat{a_{4,WWG}} \\
\widehat{a_{5,WWG}} \\
\widehat{a_{6,WWG}} \\
\widehat{a_{7,WWG}} \\
\widehat{a_{8,WWG}} \\
\widehat{a_{1,PWG}} \\
\widehat{a_{2,PWG}} \\
\widehat{a_{3,PWG}} \\
\widehat{a_{4,PWG}} \\
\widehat{a_{5,PWG}} \\
\widehat{a_{6,PWG}} \\
\widehat{a_{7,PWG}} \\
\widehat{a_{8,PWG}}
\end{array}\right]
=
\left[\begin{array}{r}
4.3609 \\
3.3973 \\
6.7999 \\
5.8803 \\
0.1509 \\
-0.0154 \\
-0.0784 \\
-0.0102 \\
-0.2703 \\
0.2758 \\
-0.3161 \\
0.2438 \\
0.2796 \\
-0.0076 \\
-0.1703 \\
-0.0127 \\
-0.4778 \\
0.5172 \\
-0.4790 \\
0.3920
\end{array}\right]
$$

Problem 2 Variance Components Estimation

The simplest forM of variance components estimation is based on the residuals of a fitted linear model and is shown in the summary results of the R-function `lm()`. Let us assume that we are given the dataset in the table shown below to which we fit a simple sire model.

Table 2: Example Dataset for Variance Components Estimation Based on Residuals Using a Sire Model

Animal	Sire	WWG
4	2	4.5
5	1	2.9
6	1	3.9
7	2	3.5
8	1	5.0

The sire model is simplified to have a common mean μ. For a moment we are setting the sire effects to be fixed effects. This leads to the following model with $var(e) = I * \sigma_e^2$

$$y = X\mu + Z_s s + e$$

Using the above shown dataset we can use the R-function `lm()` to fit this simple linear model. Because, we want to have the sires as fixed effects, we have to convert them into factors before calling `lm()`.

```
tbl_data_sol12p02$Sire <- as.factor(tbl_data_sol12p02$Sire)
lm_data_sol12p02 <- lm( WWG ~ 1 + Sire, data = tbl_data_sol12p02 )
summary(lm_data_sol12p02)
```

```
##
## Call:
## lm(formula = WWG ~ 1 + Sire, data = tbl_data_sol12p02)
##
## Residuals:
##        1        2        3        4        5
##  0.50000 -1.03333 -0.03333 -0.50000  1.06667
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.93333    0.54840   7.172  0.00558 **
## Sire2        0.06667    0.86709   0.077  0.94356
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.9499 on 3 degrees of freedom
## Multiple R-squared:  0.001967,   Adjusted R-squared:  -0.3307
## F-statistic: 0.005911 on 1 and 3 DF,  p-value: 0.9436
```

From the output of `summary()` we are given the residual standard error to be 0.9499. This residual standard error is an estimate of σ_e. The question is where does it come from. The least-squares procedure does not yield this estimate for σ_e. The answer is that this estimate comes from the residuals r of the model. For our model the vector r of residuals is defined as

$$r = y - X\widehat{\mu} - Z_s\widehat{s}$$

where $\widehat{\mu}$ and $\widehat{s}$ can be taken from the ouput of the `summary()` function. They correspond to

$$\widehat{\mu} = 3.933333$$

$$\widehat{s} = \left[\begin{array}{c} 0.0000 \\ 0.0667 \end{array}\right]$$

The estimate $\widehat{\sigma_e^2}$ for σ_e^2 is obtained by

$$\widehat{\sigma_e^2} = \frac{1}{n-p}\sum_{i=1}^{n} r_i^2$$

where n is the total number of observations and p is the number of parameters that are estimated by `lm` which is 2 for our sire model. The term $n-p$ is also called **degrees of freedom** (df). What is given as residual standard error by the output of `summary()` is the square root of $\widehat{\sigma_e^2}$.

Your Task

Verify for the above given dataset and the proposed sire model the residual standard error given by `summary()` by using the computation based on the residuals shown above.

Solution

The vector r of residuals can be obtained using the function `residuals()`

```
(vec_res <- residuals(lm_data_sol12p02))
```

```
##           1           2           3           4           5
##  0.50000000 -1.03333333 -0.03333333 -0.50000000  1.06666667
```

The degrees of freedom for the residuals $(n - p)$ are obtained by the function `df.residual()`

```
(n_df_e <- df.residual(lm_data_sol12p02))
```

```
## [1] 3
```

From this the residual standard error is computed as

```
(n_res_sd <- sqrt(sum(vec_res^2) / n_df_e))
```

```
## [1] 0.9498538
```

The same result can be obtained using the function `sigma()`

```
sigma(lm_data_sol12p02)
```

```
## [1] 0.9498538
```