Livestock Breeding and Genomics - Exercise 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.

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.