

# Livestock Breeding and Genomics - Solution 10

Peter von Rohr

2020-11-27

## Problem 1 Marker Effect Model

We are given the dataset that is shown in the table below. This dataset contains genotyping results of 10 for 2 SNP loci.

Animal	SNP A	SNP B	Observation
1	0	0	156
2	1	0	168
3	0	1	161
4	1	0	164
5	-1	0	128
6	-1	1	124
7	0	-1	143
8	1	1	178
9	1	0	163
10	0	0	151

### Your Task

- The goal of this problem is to estimate SNP marker effects using a **marker effect model**. Because we have just 2 SNP loci, you can use a fixed effects linear model with the 2 loci as fixed effects. Furthermore you can also include a fixed intercept into the model.
- Specify all the model components including the vector of observations, the design matrix  $X$ , the vector of unknowns and the vector of residuals.
- You can use the R-function `lm()` to get the solutions for estimates of the unknown SNP effects.

### Solution

The fixed effects model to estimate the marker effects can be written as

$$y = X\beta + e$$

where  $y$  is the vector of observations,  $\beta$  is the vector of fixed effects and  $e$  is the vector of residuals. Inserting the data from the dataset into the model components leads to

$$y = \begin{bmatrix} 156 \\ 168 \\ 161 \\ 164 \\ 128 \\ 124 \\ 143 \\ 178 \\ 163 \\ 151 \end{bmatrix} \beta = \begin{bmatrix} \beta_0 \\ \beta_A \\ \beta_B \end{bmatrix} e = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \\ e_9 \\ e_{10} \end{bmatrix}$$

where  $\beta_0$  is the intercept and  $\beta_A$  and  $\beta_B$  correspond to the marker effects (a-values) for both SNPs A and B.

The design matrix  $X$  is taken from the dataset as

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

The solution for the intercept and the marker effects are obtained with

```
##
## Call:
## lm(formula = tbl_all_data$Observation ~ tbl_all_data$`SNP A` +
##     tbl_all_data$`SNP B`, data = tbl_all_data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.40  -4.02   0.52   3.02   7.72
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      148.280      2.172  68.270  3.8e-11 ***
## tbl_all_data$`SNP A`    20.740      2.660   7.797 0.000107 ***
## tbl_all_data$`SNP B`     5.860      3.318   1.766 0.120691
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 6.27 on 7 degrees of freedom
## Multiple R-squared:  0.8985, Adjusted R-squared:  0.8695
## F-statistic: 30.97 on 2 and 7 DF, p-value: 0.0003335
```

## Problem 2 Breeding Value Model

Use the same data as in Problem 1 to estimate genomic breeding values using a **breeding value model**.

### Hints

- The only fixed effect in this model is the mean  $\mu$  which is the same for all observations.
- You can use the following matrix as the genomic relationship matrix

$$G = \begin{bmatrix} 0.141 & -0.124 & -0.123 & -0.124 & 0.288 & 0.083 & 0.287 & -0.329 & -0.124 & 0.082 \\ -0.124 & 0.76 & -0.33 & 0.701 & -0.949 & -1.155 & 0.082 & 0.495 & 0.701 & -0.124 \\ -0.123 & -0.33 & 0.757 & -0.33 & 0.085 & 0.905 & -0.943 & 0.491 & -0.33 & -0.123 \\ -0.124 & 0.701 & -0.33 & 0.76 & -0.949 & -1.155 & 0.082 & 0.495 & 0.701 & -0.124 \\ 0.288 & -0.949 & 0.085 & -0.949 & 1.584 & 1.322 & 0.492 & -1.152 & -0.949 & 0.288 \\ 0.083 & -1.155 & 0.905 & -1.155 & 1.322 & 2.202 & -0.738 & -0.333 & -1.155 & 0.083 \\ 0.287 & 0.082 & -0.943 & 0.082 & 0.492 & -0.738 & 1.576 & -1.148 & 0.082 & 0.287 \\ -0.329 & 0.495 & 0.491 & 0.495 & -1.152 & -0.333 & -1.148 & 1.374 & 0.495 & -0.329 \\ -0.124 & 0.701 & -0.33 & 0.701 & -0.949 & -1.155 & 0.082 & 0.495 & 0.76 & -0.124 \\ 0.082 & -0.124 & -0.123 & -0.124 & 0.288 & 0.083 & 0.287 & -0.329 & -0.124 & 0.141 \end{bmatrix}$$

### Your Tasks

- Specify all model components of the linear mixed model, including the expected values and the variance-covariance matrix of the random effects.

### Solution

The breeding value model is a linear mixed effects model which can be written as

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

where

- $y$  is the vector of observations
- $\beta$  is the vector of fixed effects
- $u$  is the vector of random genomic breeding values
- $e$  is the vector of random residuals
- $X$  and  $W$  are design matrices linking fixed effects and genomic breeding values to observations.

Inserting the information from the dataset into the model leads to

$$y = \begin{bmatrix} 156 \\ 168 \\ 161 \\ 164 \\ 128 \\ 124 \\ 143 \\ 178 \\ 163 \\ 151 \end{bmatrix} \quad \beta = [\mu] \quad u = \begin{bmatrix} u_1 \\ u_2 \\ u_3 \\ u_4 \\ u_5 \\ u_6 \\ u_7 \\ u_8 \\ u_9 \\ u_{10} \end{bmatrix} \quad e = \begin{bmatrix} e_1 \\ e_2 \\ e_3 \\ e_4 \\ e_5 \\ e_6 \\ e_7 \\ e_8 \\ e_9 \\ e_{10} \end{bmatrix}$$

The design matrices  $X$  and  $W$  correspond to

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

The expected values of the random effects are

$$\begin{aligned} E(u) &= 0 \\ E(e) &= 0 \\ E(y) &= X\beta \end{aligned}$$

The variance-covariance matrices of the random effects are

$$\text{var}(u) = G * \sigma_u^2$$

where  $G$  is the genomic relationship matrix and  $\sigma_u^2$  the genetic additive variance explained by the SNPs

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

where  $I$  is the identity matrix and  $\sigma_e^2$  the residual variance.

$$\text{var}(y) = WGW^T * \sigma_u^2 + R$$

The solutions for the fixed effects are obtained from mixed model equations.

$$\begin{bmatrix} X^T X & X^T W \\ W^T X & W^T W + G^{-1} * \lambda \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} X^T y \\ W^T y \end{bmatrix}$$

The parameter  $\lambda = \sigma_e^2 / \sigma_u^2$  is the ratio between residual variance and genetic variance. We assume that this value corresponds to  $\lambda = 3$ .

The single components of the mixed model equations are

$$X^T X = [10], X^T W = [1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1 \ 1], W^T X = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$W^T W + G^{-1} = \begin{bmatrix} 50.731 & 1.005 & 1.994 & 1.005 & -3.005 & -0.012 & -3.993 & 3.999 & 1.005 & -1 \\ 1.005 & 45.704 & 3.01 & -6.027 & 8.037 & 10.042 & -1 & -4.022 & -6.027 & 1.005 \\ 1.994 & 3.01 & 42.74 & 3.01 & 0.977 & -10.008 & 12.978 & -7.974 & 3.01 & 1.994 \\ 1.005 & -6.027 & 3.01 & 45.704 & 8.037 & 10.042 & -1 & -4.022 & -6.027 & 1.005 \\ -3.005 & 8.037 & 0.977 & 8.037 & 37.684 & -10.066 & -6.986 & 12.019 & 8.037 & -3.005 \\ -0.012 & 10.042 & -10.008 & 10.042 & -10.066 & 31.669 & 9.985 & 0.046 & 10.042 & -0.012 \\ -3.993 & -1 & 12.978 & -1 & -6.986 & 9.985 & 30.767 & 15.971 & -1 & -3.993 \\ 3.999 & -4.022 & -7.974 & -4.022 & 12.019 & 0.046 & 15.971 & 35.737 & -4.022 & 3.999 \\ 1.005 & -6.027 & 3.01 & -6.027 & 8.037 & 10.042 & -1 & -4.022 & 45.704 & 1.005 \\ -1 & 1.005 & 1.994 & 1.005 & -3.005 & -0.012 & -3.993 & 3.999 & 1.005 & 50.731 \end{bmatrix}$$

with

$$rhs = \begin{bmatrix} X^T y \\ W^T y \end{bmatrix} = \begin{bmatrix} 1536.388 \\ 156.41 \\ 168.379 \\ 161.35 \\ 163.533 \\ 127.857 \\ 124.478 \\ 142.925 \\ 177.661 \\ 162.853 \\ 150.941 \end{bmatrix}$$

The solution vector for the estimate of the fixed effect  $\mu$  and the genomic breeding values for all animals are given by

$$sol = \begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} 153.6388 \\ -3.1077 \\ 10.1654 \\ -0.2477 \\ 10.0717 \\ -16.7013 \\ -14.0022 \\ -6.1328 \\ 13.1093 \\ 10.0586 \\ -3.2134 \end{bmatrix}$$