Livestock Breeding and Genomics - Solution 12

Peter von Rohr

2019-12-13

Problem 1 Marker Effect Model

We are given the dataset that is shown in the table below. This dataset contains gentyping 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, β 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 β_0 is the intercept and β_A and β_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 μ which is the same for all observations.
- You can use the following matrix as the genomic relationship matrix

$$G = \left[\begin{array}{rrrrrrrrrr} 0.141 & -0.124 & -0.123 & -0.124 & 0.288 & 0.083 & 0.287 & -0.329 & -0.124 & 0.082 \\ -0.124 & 0.760 & -0.330 & 0.701 & -0.949 & -1.155 & 0.082 & 0.495 & 0.701 & -0.124 \\ -0.123 & -0.330 & 0.757 & -0.330 & 0.085 & 0.905 & -0.943 & 0.491 & -0.330 & -0.123 \\ -0.124 & 0.701 & -0.330 & 0.760 & -0.949 & -1.155 & 0.082 & 0.495 & 0.701 & -0.124 \\ 0.288 & -0.949 & 0.085 & -0.949 & 1.584 & 1.322 & 0.492 & -1.152 & -0.949 & 0.288 \\ 0.083 & -1.155 & 0.905 & -1.155 & 1.322 & 2.202 & -0.738 & -0.333 & -1.155 & 0.083 \\ 0.287 & 0.082 & -0.943 & 0.082 & 0.492 & -0.738 & 1.576 & -1.148 & 0.082 & 0.287 \\ -0.329 & 0.495 & 0.491 & 0.495 & -1.152 & -0.333 & -1.148 & 1.374 & 0.495 & -0.329 \\ -0.124 & 0.701 & -0.330 & 0.701 & -0.949 & -1.155 & 0.082 & 0.495 & 0.760 & -0.124 \\ 0.082 & -0.124 & -0.123 & -0.124 & 0.288 & 0.083 & 0.287 & -0.329 & -0.124 & 0.141 \end{array}\right]$$

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
- β 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

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

The design matrices X and W correspond to

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

The expected values of the random effects are

$$E(u) = 0$$

$$E(e) = 0$$

$$E(y) = X\beta$$

The variance-covariance matrices of the random effects are

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

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

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

where I is the identity matrix and σ_e^2 the residual variance.

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

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

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

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

$$X^TX = \begin{bmatrix} 10 \end{bmatrix}, X^TW = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{bmatrix}, W^TX = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix}$$

$$W^TW + G^{-1} * \lambda = \begin{bmatrix} 50.731 & 1.005 & 1.994 & 1.005 & -3.005 & -0.012 & -3.993 & 3.999 & 1.005 & -1.000 \\ 1.005 & 45.704 & 3.010 & -6.027 & 8.037 & 10.042 & -1.000 & -4.022 & -6.027 & 1.005 \\ 1.994 & 3.010 & 42.740 & 3.010 & 0.977 & -10.008 & 12.978 & -7.974 & 3.010 & 1.994 \\ 1.005 & -6.027 & 3.010 & 45.704 & 8.037 & 10.042 & -1.000 & -4.022 & -6.027 & 1.005 \\ -3.005 & 8.037 & 0.977 & 8.037 & 37.684 & -10.066 & -6.986 & 12.019 & 8.037 & -3.005 \\ -0.012 & 10.042 & -10.008 & 10.042 & -10.066 & 31.669 & 9.985 & 0.046 & 10.042 & -0.012 \\ -3.993 & -1.000 & 12.978 & -1.000 & -6.986 & 9.985 & 30.767 & 15.971 & -1.000 & -3.993 \\ 3.999 & -4.022 & -7.974 & -4.022 & 12.019 & 0.046 & 15.971 & 35.737 & -4.022 & 3.999 \\ 1.005 & -6.027 & 3.010 & -6.027 & 8.037 & 10.042 & -1.000 & -4.022 & 45.704 & 1.005 \\ -1.000 & 1.005 & 1.994 & 1.005 & -3.005 & -0.012 & -3.993 & 3.999 & 1.005 & 50.731 \end{bmatrix},$$

$$\begin{bmatrix} X^Ty \\ W^Ty \end{bmatrix} = \begin{bmatrix} 1536.388 \\ 156.410 \\ 168.379 \\ 161.350 \\ 163.533 \\ 127.857 \\ 124.478 \\ 142.925 \\ 177.661 \\ 162.853 \\ 150.941 \end{bmatrix},$$

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

$$\begin{bmatrix} \hat{\beta} \\ \hat{u} \end{bmatrix} = \begin{bmatrix} 153.639 \\ -3.108 \\ 10.165 \\ -0.248 \\ 10.072 \\ -16.701 \\ -14.002 \\ -6.133 \\ 13.109 \\ 10.059 \\ -3.213 \end{bmatrix},$$