Applied Statistical Methods - Exercise 4

Peter von Rohr

2019-03-18

Problem 1: Traditional Predicted Breeding Values

Given the following data set with observations and a pedigree for a group of animals.

Table 1: Phenotypic Observations

Animal	Observation
1	100.430
2	103.396
3	114.458
4	100.068
5	104.144
6	117.524
7	97.744
8	111.926
9	103.486
10	97.914
11	104.651
12	115.714
13	86.900
14	101.097
15	102.795
16	112.182
17	109.295
18	105.271
19	91.744
20	101.132
21	107.385

The observations in Table 1 can be read from

https://charlotte-ngs.github.io/GELASMSS2019/ex/w05/data_ex04_phe.csv.

The pedigree showing the ancestral relationships is shown below

Table 2: Pedigree

Animal	Sire	Dam
1	NA	NA
2	NA	NA
3	NA	NA
4	NA	NA
5	NA	NA
6	2	3

7	1	3
8	2	5
9	1	5
10	7	8
11	7	8
12	6	9
13	7	8
14	7	9
15	6	8
16	6	9
17	6	8
18	6	8
19	7	8
20	6	9
21	7	8

The pedigree can be read from

https://charlotte-ngs.github.io/GELASMSS2019/ex/w05/data_ex04_ped.csv

Your Task

Predict breeding values for the animals given in the dataset and in the pedigree without using any genotypic information using a BLUP animal model. Set up the mixed model equations for the BLUP animal model and use the package `pedigreemm` to get the inverse of the relationship matrix.

Hints

- Use a mixed linear model with a constant intercept as a fixed effect and the breeding values of all animals as random effects. Hence the following model can be assumed

$$y = Xb + Za + e$$

where y is the vector of all observations, b has just one element and X has one column with all ones. The vector a contains the breeding values for all animals. The matrix Z links the breeding values to the phenotypic observations. The random errors are represented by the vector e.

- Then residual variance σ_e^2 can be assumed to be $\sigma_e^2 = 75$. The genetic additive variance σ_a^2 is $\sigma_a^2 = 25$

Problem 2: Prediction of Genomic Breeding Values Using GBLUP

Use the same phenotypic observations as in Problem 1. In addition to that we use genomic information available in

https://charlotte-ngs.github.io/GELASMSS2019/ex/w05/data_ex04_gen.csv

Your Tasks

Predict the genomic breeding values using the GBLUP approach.

Hints

- Use an analogous mixed linear effect model as was used in Problem 1. Instead of the vector of breeding values use the vector g of genomic breeding values as random effects of the model. Hence the following model can be assumed

$$y = Xb + Zg + e$$

where y is the vector of all observations, b has just one element and X has one column with all ones. The vector g contains the genomic breeding values for all animals. The matrix Z links the breeding values to the phenotypic observations. The random errors are represented by the vector e.

- Use the genomic relationship matrix in the mixed model equations
- The ratio λ of between the variances is assumed to be the same as in Problem 1.
- If the inverse of the genomic relationship matrix cannot be computed, adjust the genomic relationship matrix with the numerator relationship matrix A according to the following formula

$$G^* = 0.95 * G + 0.05 * A$$

where G is the matrix determined based on th given data and the nummerator relationship matrix A can be computed with the function `pedigreemm::getA()` from package `pedigreemm`.