Animal Model

1 Introduction

In animal genetics, measurements are taken on individual animals, and thus, the model of analysis should include the animal additive genetic effect. The remaining items in the model are factors that have an effect on the trait of interest. The infinitesimal genetic model is assumed, and the animals that have been measured should represent a random sample (or the entire sample) of progeny of the breeding males and females in the population. That means measurements should not be taken only from the biggest progeny, or the best progeny, or the poorest progeny, but a random sample of progeny.

A simple animal model with one record per animal is

\[ y_i = \mu + a_i + e_i, \]

where \( y_i \) is the phenotypic record or observation, \( \mu \) is the overall mean of the observations, \( a_i \) are the additive genetic values of animals, and \( e_i \) are the residual or environmental effects associated with the \( i^{th} \) observation. Assumptions for this model are

1. The population is large (infinite) in size.
2. The association of alleles from the two parents is assumed to be at random.
3. No selection has been imposed on individual records.
4. Only additive genetic effects exist.
5. The influence of other factors on the observations is absent.

The expectations (for the example data to follow) are

\[ E(a_i) = 0, \quad \text{and} \quad E(e_i) = 0, \]

and the variances are

\[ Var(a_i) = \sigma_a^2, \quad \text{and} \quad Var(e_i) = \sigma_e^2. \]

The variances are assumed known, at least to proportionality. The covariance matrix of the vector of animal additive genetic effects is

\[ Var(a) = A\sigma_a^2. \]

The relationship matrix is assumed to be complete back to the base population (a large, randomly mating population).
The heritability of the trait is

\[
h^2 = \frac{\sigma_a^2}{\sigma_a^2 + \sigma_e^2},
\]

and the ratio of variances is

\[
\alpha = \frac{\sigma_e^2}{\sigma_a^2}.
\]

## 2 Example Problem

Below is the pedigree information and data on 16 animals. The first four animals were base generation animals without records.

### 2.1 Data

<table>
<thead>
<tr>
<th>Animal</th>
<th>Sire</th>
<th>Dam</th>
<th>Record</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
<td>38.5</td>
</tr>
<tr>
<td>6</td>
<td>3</td>
<td>4</td>
<td>48.9</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>2</td>
<td>64.3</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>4</td>
<td>50.5</td>
</tr>
<tr>
<td>9</td>
<td>5</td>
<td>6</td>
<td>36.0</td>
</tr>
<tr>
<td>10</td>
<td>7</td>
<td>8</td>
<td>66.4</td>
</tr>
<tr>
<td>11</td>
<td>5</td>
<td>8</td>
<td>28.9</td>
</tr>
<tr>
<td>12</td>
<td>7</td>
<td>6</td>
<td>73.0</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>6</td>
<td>44.2</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>8</td>
<td>53.4</td>
</tr>
<tr>
<td>15</td>
<td>5</td>
<td>4</td>
<td>33.6</td>
</tr>
<tr>
<td>16</td>
<td>7</td>
<td>8</td>
<td>49.5</td>
</tr>
</tbody>
</table>

The number of observations is \( N = 12 \), and the total sum of squares is 30,811.78. Let \( \sigma_a^2 = 36 \), and \( \sigma_e^2 = 64 \), so that \( h^2 = 0.36 \), and \( \alpha = 1.777778 \).
2.2 Additive Genetic Relationships

The matrix of additive genetic relationships among the sixteen individuals is $A$ (times 16) given below:

$$
\begin{pmatrix}
16 & 0 & 0 & 8 & 0 & 0 & 8 & 0 & 4 & 4 & 8 & 0 & 8 & 4 & 4 & 4 \\
0 & 16 & 0 & 0 & 8 & 0 & 8 & 0 & 4 & 4 & 4 & 0 & 0 & 4 & 4 \\
0 & 0 & 16 & 0 & 0 & 8 & 8 & 0 & 4 & 4 & 0 & 8 & 4 & 8 & 0 & 4 \\
0 & 0 & 0 & 16 & 0 & 8 & 0 & 8 & 4 & 4 & 4 & 4 & 4 & 8 & 4 & 8 \\
8 & 8 & 0 & 0 & 16 & 0 & 4 & 4 & 8 & 4 & 10 & 2 & 4 & 2 & 8 & 4 \\
0 & 0 & 8 & 8 & 0 & 16 & 4 & 4 & 8 & 4 & 2 & 10 & 8 & 6 & 4 & 4 \\
0 & 8 & 8 & 0 & 4 & 4 & 16 & 0 & 4 & 8 & 2 & 10 & 2 & 4 & 2 & 8 \\
8 & 0 & 0 & 8 & 4 & 4 & 0 & 16 & 4 & 8 & 10 & 2 & 6 & 8 & 6 & 8 \\
8 & 4 & 0 & 4 & 10 & 2 & 2 & 10 & 6 & 6 & 18 & 2 & 5 & 5 & 7 & 6 \\
0 & 4 & 8 & 4 & 2 & 10 & 10 & 2 & 6 & 6 & 2 & 18 & 5 & 5 & 3 & 6 \\
8 & 0 & 4 & 4 & 4 & 4 & 8 & 2 & 6 & 6 & 4 & 5 & 5 & 16 & 5 & 4 & 4 \\
4 & 0 & 8 & 4 & 2 & 6 & 4 & 8 & 4 & 6 & 5 & 5 & 5 & 16 & 3 & 6 \\
4 & 4 & 0 & 8 & 8 & 4 & 2 & 6 & 6 & 4 & 7 & 3 & 4 & 3 & 16 & 4 \\
4 & 4 & 4 & 4 & 4 & 8 & 8 & 4 & 8 & 8 & 6 & 6 & 4 & 6 & 4 & 16
\end{pmatrix}
$$

2.3 Application of BLUP with R

The construction of mixed model equations begins by defining $X, Z, G$ and $R$.

$X$ is a vector of ones of order 12 by 1.

$Z$ is a 12 by 16 matrix, in which the first four columns are all zero, and the last 12 columns are an identity matrix.

$G$ is $A\sigma^2_a$ of order 16, and

$R$ is $I\sigma^2_e$ of order 12.

To set up $X$ in R, use the `jd` function below.

```r
# function to form a J matrix - all ones
jd <- function(n,m) matrix(c(1),nrow=n,ncol=m)
X = jd(12,1)
```

There are at least two ways to create $Z$. 


The `desgn` function can be used generally for creating design matrices from a list of the levels of a factor.

The inverse of the relationship matrix is needed. A list of the sires and dams of the 16 animals are needed, with zeros for missing parents, and a list of the $b_i$ values for each animal are needed. In this example, the $b_i$s for animals 1 to 4 are 1, and for animals 5 through 16 are 0.5.
Alpha is the ratio of residual to additive genetic variances. In this example, the ratio is 64 divided by 36, or 1.778. The last line above creates the inverse of $R$ which is an identity matrix. Finally, the mixed model equations are created and solved using the `MME` function given in an earlier section of the notes.

```
Exmp = MME(X,Z,G1,R1,y)
Exmp$LHS
Exmp$RHS
Exmp$C
Exmp$SOLNS
```

2.4 Solutions and SEP

The solutions for the animals are given in the next table. The estimate of the overall mean was 48.8955.

<table>
<thead>
<tr>
<th>Animal</th>
<th>EBV</th>
<th>Diagonals</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-4.7341</td>
<td>.4866</td>
</tr>
<tr>
<td>2</td>
<td>.3728</td>
<td>.4959</td>
</tr>
<tr>
<td>3</td>
<td>5.8579</td>
<td>.4786</td>
</tr>
<tr>
<td>4</td>
<td>-1.4967</td>
<td>.4975</td>
</tr>
<tr>
<td>5</td>
<td>-7.1305</td>
<td>.4198</td>
</tr>
<tr>
<td>6</td>
<td>2.1335</td>
<td>.4260</td>
</tr>
<tr>
<td>7</td>
<td>8.4380</td>
<td>.4137</td>
</tr>
<tr>
<td>8</td>
<td>-2.1543</td>
<td>.4274</td>
</tr>
<tr>
<td>9</td>
<td>-4.7807</td>
<td>.4592</td>
</tr>
<tr>
<td>10</td>
<td>6.2947</td>
<td>.4582</td>
</tr>
<tr>
<td>11</td>
<td>-8.0126</td>
<td>.4795</td>
</tr>
<tr>
<td>12</td>
<td>9.4167</td>
<td>.4751</td>
</tr>
<tr>
<td>13</td>
<td>-2.0456</td>
<td>.4507</td>
</tr>
<tr>
<td>14</td>
<td>2.4341</td>
<td>.4501</td>
</tr>
<tr>
<td>15</td>
<td>-6.7242</td>
<td>.4459</td>
</tr>
<tr>
<td>16</td>
<td>2.5849</td>
<td>.4582</td>
</tr>
</tbody>
</table>

An estimate of the residual variance is obtained by subtracting the total sum of squares minus the sum of squares due to the model, and dividing that by $N - r(X)$.
\[
\begin{align*}
SST &= 30,811.78 \\
SSR &= 29,618.3158 \\
SST - SSR &= 1193.464 \\
\hat{\sigma}_e^2 &= 108.4967
\end{align*}
\]

The Standard Error of Prediction (SEP) of an Estimated Breeding Value (EBV) is the square root of the product of the diagonal of the inverse of the LHS times the estimate of residual variance. Thus for animal 1 in the table above, the SEP is

\[
SEP = (0.4866 \times 108.4967)^{\frac{1}{2}} = 7.2662.
\]

2.5 Reliability

The reliability (REL) of an EBV is another measure of accuracy of the estimate.

\[
REL_i = \frac{(a_{ii} - c_{ii} \alpha)}{a_{ii}},
\]

where \(a_{ii}\) is the diagonal of the A matrix for animal \(i\), and \(c_{ii}\) is the diagonal of the inverse of the LHS (in the table above). For animal 1,

\[
REL = (1 - 0.4866(1.778))/1 = 0.1349.
\]

The reliability is only 13.49 \%, which is low.

Publication of “official” EBV often requires a minimum reliability, such as 75\%, plus minimums on number of progeny and number of contemporary groups for those progeny.

3 Simulation of Records

The sampling processes involved in generating a set of data are better understood through simulation.

3.1 Generating Breeding Values of Animals

Let the heritability of the trait be 0.25, and the variance of phenotypic records to be 100. Then, \(\sigma_a^2 = 25\), and \(\sigma_e^2 = 75\). Breeding values of base population animals are created by multiplying a
pseudo-random normal deviate, RND, times the genetic standard deviation, $\sigma_a = 5$. Let there be two base animals,

$$a_1 = -1.8014 \times 5 = -9.0070$$
$$a_2 = 0.6556 \times 5 = 3.2780$$

Progeny can be generated from animals previously generated as the average of the parent breeding values plus a Mendelian sampling effect. The Mendelian sampling effect is the product of another random normal deviate times the genetic standard deviation of the Mendelian sampling effect. The variance of the Mendelian sampling effect is $b_i$ times the genetic variance, where $b_i$ is obtained during the calculation of inbreeding coefficients (given in previous notes). That is,

$$b_i = (0.5 - 0.25 \times (F_s + F_d)),$$

where $F_s$ and $F_d$ are the inbreeding coefficients of the sire and dam, respectively.

Let Animal 3 be a progeny of animals 1 and 2,

$$a_3 = 0.5 \times (a_1 + a_2) + m_3$$
$$b_3 = 0.5,$$
$$m_3 = (0.5)^5 \sigma_a \text{ RND}$$
$$= 0.7071 \times 5 \times (-.4045),$$
$$= -1.4301$$

$$a_3 = 0.5(-9.0070 + 3.2780) - 1.4301$$
$$= -4.2946.$$

3.2 Phenotypic Records

To create phenotypic records, the equation of the model must be specified. Assume that

$$y_{ij} = \mu + d_j + a_i + e_{ij},$$

where $\mu$ is an overall mean (assume a value of 50); $d_j$ is a diet effect for one of 3 diets, $a_i$ is the animal breeding value, and $e_{ij}$ is a residual effect.

Animals must be assigned to a diet. One way is to have the experimental design known in advance, such that an animal is created for a specific diet. Another way is to pick one of the three diets randomly and assign it to that animal. The differences among the diets should be set before generating records. Assume that

$$d_1 = -10.0,$$
\[
\begin{align*}
    d_2 &= 0, \\
    d_3 &= 10.0.
\end{align*}
\]

Diet differences would be constant for all animals generated. If the simulation were repeated, then the same differences would probably be used. This would be the traditionalist’s view of a fixed factor.

Residual effects are random and specific to each observation. A residual effect is generated by multiplying a random normal deviate times the residual standard deviation.

A record for animal 3, in diet 3, would be

\[
y_{33} = 50 + 10.0 - 4.2946 + .3939(8.6602) = 59.1167.
\]

Usually observations will be to the nearest whole number, so that \( y_{33} = 59 \) would be the final result for animal 3.

If animals are generated in chronological order, then inbreeding coefficients need to be calculated as new animals are generated, so that \( b_i \) values can be obtained for generating Mendelian sampling effects. This is an efficient algorithm for simulating data for large populations. A later set of notes will consider using R to generate large populations of animals for single or multiple traits.

4 Measuring Genetic Change

To measure genetic change, the population needs to be defined precisely. In dairy cattle, one population would be all females born, by year of birth. Then the average EBVs of cows by year of birth would estimate the genetic change in new females coming into the population.

Another population could be all females calving in a particular year. This is different from the first population because cows can have several calvings over their lifetime, and secondly not all females that are born would necessarily calve even once. A cow’s EBV could appear in the calculations in more than one year of calving. This would represent the trend in genetic merit of actively lactating cows.

In beef cattle, the averages for females being bred in a given year could be an interesting trend to monitor versus the average of females that produce a calf. In all cases, EBVs are used in the calculations.

Genetic trends can be overestimated if the heritability used in the calculation of EBV is too large. Similarly, if the heritability is too low, then genetic trend can be underestimated. Thus, having good estimates of the variance parameters is crucial to good estimates of genetic trend.
5 Reduced Animal Models

5.1 Usual Animal Model Approach

Consider an animal model with periods as a fixed factor and one observation per animal, as in the table below.

<table>
<thead>
<tr>
<th>Animal</th>
<th>Sire</th>
<th>Dam</th>
<th>Period</th>
<th>Observation</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>250</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>198</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>245</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>260</td>
</tr>
<tr>
<td>9</td>
<td>2</td>
<td>4</td>
<td>2</td>
<td>235</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>255</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>200</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>1</td>
<td>225</td>
</tr>
</tbody>
</table>

Assume that the ratio of residual to additive genetic variances is 2. The MME for this data would be of order 11 (nine animals and two periods).

\[
\begin{pmatrix}
3 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 5 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 4 & 0 & 2 & 0 & -2 & -2 & 0 & 0 & 0 \\
1 & 0 & 0 & 6 & 0 & 3 & 0 & 0 & -2 & -2 & -2 \\
1 & 0 & 2 & 0 & 5 & 0 & -2 & -2 & 0 & 0 & 0 \\
1 & 0 & 0 & 3 & 0 & 6 & 0 & 0 & -2 & -2 & -2 \\
0 & 1 & -2 & 0 & -2 & 0 & 5 & 0 & 0 & 0 & 0 \\
0 & 1 & -2 & 0 & -2 & 0 & 0 & 5 & 0 & 0 & 0 \\
0 & 1 & 0 & -2 & 0 & -2 & 0 & 0 & 5 & 0 & 0 \\
0 & 1 & 0 & -2 & 0 & -2 & 0 & 0 & 0 & 5 & 0
\end{pmatrix}
\begin{pmatrix}
\hat{b}_1 \\
\hat{b}_2 \\
\hat{a}_1 \\
\hat{a}_2 \\
\hat{a}_3 \\
\hat{a}_4 \\
\hat{a}_5 \\
\hat{a}_6 \\
\hat{a}_7 \\
\hat{a}_8 \\
\hat{a}_9
\end{pmatrix}
= 
\begin{pmatrix}
680 \\
1188 \\
0 \\
225 \\
200 \\
255 \\
250 \\
198 \\
245 \\
260 \\
235
\end{pmatrix}
and the solutions to these equations are

\[
\begin{pmatrix}
\hat{b}_1 \\
\hat{b}_2 \\
\hat{a}_1 \\
\hat{a}_2 \\
\hat{a}_3 \\
\hat{a}_4 \\
\hat{a}_5 \\
\hat{a}_6 \\
\hat{a}_7 \\
\hat{a}_8 \\
\hat{a}_9
\end{pmatrix} =
\begin{pmatrix}
225.8641 \\
236.3366 \\
-2.4078 \\
1.3172 \\
-10.2265 \\
11.3172 \\
-2.3210 \\
-12.7210 \\
6.7864 \\
9.7864 \\
4.7864
\end{pmatrix}.
\]

Some species, such as swine or rainbow trout, have many offspring per mating, and subsequently fewer animals are needed for breeding purposes. The MME would contain many equations (one per animal), even though a large percentage of animals do not have any progeny. Reducing the order of the MME would be an advantage, to save computing time. Quaas and Pollak (1980) proposed the reduced animal model. There is one model for animals that are not parents, and a separate model for animals that are parents. The total number of animal genetic equations in MME is equal to the number of animals that are parents. The solutions for the animals in the reduced animal model are exactly the same as solutions from the regular animal model.

5.2 Theoretical Development

The vector of additive genetic values of animals in the animal model is \( \mathbf{a} \). Denote animals with progeny as \( \mathbf{a}_p \), and those without progeny as \( \mathbf{a}_o \), so that

\[
\mathbf{a}' = \begin{pmatrix} \mathbf{a}_p' & \mathbf{a}_o' \end{pmatrix}.
\]

In terms of the example data,

\[
\mathbf{a}_p' = \begin{pmatrix} a_1 & a_2 & a_3 & a_4 \end{pmatrix}, \\
\mathbf{a}_o' = \begin{pmatrix} a_5 & a_6 & a_7 & a_8 & a_9 \end{pmatrix}.
\]

The additive genetic value of an animal may be written as the average of the additive genetic values of the parents plus a Mendelian sampling effect, which is the animal’s specific deviation from the parent average, i.e.

\[
a_i = .5(a_s + a_d) + m_i.
\]

Therefore,

\[
\mathbf{a}_o = \mathbf{P}\mathbf{a}_p + \mathbf{m},
\]
where $\mathbf{P}$ is a matrix that indicates the parents of each animal in $\mathbf{a}_o$, with elements equal to 0.5, and $\mathbf{m}$ is the vector of Mendelian sampling effects. Then

$$
\mathbf{a} = \begin{pmatrix}
\mathbf{a}_p \\
\mathbf{a}_o
\end{pmatrix} = \begin{pmatrix}
\mathbf{I} \\
\mathbf{P}
\end{pmatrix} \mathbf{a}_p + \begin{pmatrix}
0 \\
\mathbf{m}
\end{pmatrix}.
$$

and

$$
\text{Var}(\mathbf{a}) = \mathbf{A} \sigma_a^2
= \begin{pmatrix}
\mathbf{I} \\
\mathbf{P}
\end{pmatrix} \mathbf{A}_{pp} \begin{pmatrix}
\mathbf{I} & \mathbf{P}'
\end{pmatrix} \mathbf{A}_o \mathbf{a}_p + \begin{pmatrix}
0 \\
0 \\
\mathbf{D}
\end{pmatrix} \sigma_a^2
$$

where $\mathbf{D}$ is a diagonal matrix with diagonal elements equal to $b_i$, and $b_i = 0.5 - 0.25 \star (F_s + F_d)$ for the $i^{th}$ animal, and

$$
\text{Var}(\mathbf{a}_p) = \mathbf{A}_{pp} \sigma_a^2.
$$

The animal model can now be written as

$$
\begin{pmatrix}
\mathbf{y}_p \\
\mathbf{y}_o
\end{pmatrix} = \begin{pmatrix}
\mathbf{X}_p \\
\mathbf{X}_o
\end{pmatrix} \mathbf{b} + \begin{pmatrix}
\mathbf{Z}_p & 0 \\
0 & \mathbf{Z}_o
\end{pmatrix} \begin{pmatrix}
\mathbf{I} \\
\mathbf{P}
\end{pmatrix} \mathbf{a}_p + \begin{pmatrix}
\mathbf{e}_p \\
\mathbf{e}_o + \mathbf{Z}_o \mathbf{m}
\end{pmatrix}.
$$

The residual vector has two types of residuals and the additive genetic values of animals without progeny have been replaced with $\mathbf{P} \mathbf{a}_p$. Because every individual has only one record, then $\mathbf{Z}_o = \mathbf{I}$, but $\mathbf{Z}_p$ may have fewer rows than there are elements of $\mathbf{a}_p$ because not all parents may have observations themselves. In the example data, animal 1 does not have an observation, therefore,

$$
\mathbf{Z}_p = \begin{pmatrix}
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1
\end{pmatrix}.
$$

Consequently,

$$
\mathbf{R} = \text{Var} \begin{pmatrix}
\mathbf{e}_p \\
\mathbf{e}_o + \mathbf{m}
\end{pmatrix}
= \begin{pmatrix}
\mathbf{I} \sigma_e^2 & 0 \\
0 & \mathbf{I} \sigma_e^2 + \mathbf{D} \sigma_a^2
\end{pmatrix}
= \begin{pmatrix}
\mathbf{I} & 0 \\
0 & \mathbf{R}_o
\end{pmatrix} \sigma_e^2
$$

The mixed model equations for the reduced animal model are

$$
\begin{pmatrix}
\mathbf{X}_p' \mathbf{X}_p + \mathbf{X}_o' \mathbf{R}_o^{-1} \mathbf{X}_o & \mathbf{X}_p' \mathbf{Z}_p + \mathbf{X}_o' \mathbf{R}_o^{-1} \mathbf{P} \\
\mathbf{Z}_p' \mathbf{X}_p + \mathbf{P}' \mathbf{R}_o^{-1} \mathbf{X}_o & \mathbf{Z}_p' \mathbf{Z}_p + \mathbf{P}' \mathbf{R}_o^{-1} \mathbf{P} + \mathbf{A}_o^{-1} \mathbf{a}_o
\end{pmatrix}
\begin{pmatrix}
\hat{\mathbf{b}} \\
\hat{\mathbf{a}}_p
\end{pmatrix}
= \begin{pmatrix}
\mathbf{X}_p' \mathbf{y}_p + \mathbf{X}_o' \mathbf{R}_o^{-1} \mathbf{y}_o \\
\mathbf{Z}_p' \mathbf{y}_p + \mathbf{P}' \mathbf{R}_o^{-1} \mathbf{y}_o
\end{pmatrix}.
Solutions for $\hat{a}_o$ are derived from the following formulas:

$$\hat{a}_o = P\hat{a}_p + \hat{m},$$

where

$$\hat{m} = (Z_o'Z_o + D^{-1}\alpha)^{-1}(y_o - X_o\hat{b} - P\hat{a}_p).$$

5.3 Analysis of Example Data

Using the example data,

$$P = \begin{pmatrix} .5 & 0 & .5 & 0 \\ .5 & 0 & .5 & 0 \\ 0 & .5 & 0 & .5 \\ 0 & .5 & 0 & .5 \end{pmatrix},$$

and

$$D = \text{diag} \left( .5, .5, .5, .5 \right),$$

then the MME with $\alpha = 2$ are

$$\begin{pmatrix} 3 & 0 & 0 & 1 & 1 & 1 \\ 0 & 4 & .8 & 1.2 & .8 & 1.2 \\ 0 & .8 & 2.4 & 0 & .4 & 0 \\ 1 & 1.2 & 0 & 3.6 & 0 & .6 \\ 1 & .8 & 4 & 0 & 3.4 & 0 \\ 1 & 1.2 & 0 & .6 & 0 & 3.6 \end{pmatrix} \begin{pmatrix} \hat{b}_1 \\ \hat{b}_2 \\ \hat{a}_1 \\ \hat{a}_2 \\ \hat{a}_3 \\ \hat{a}_4 \end{pmatrix} = \begin{pmatrix} 680. \\ 950.4 \\ 179.2 \\ 521. \\ 379.2 \\ 551. \end{pmatrix}$$

The solutions are as before, i.e.

$$\hat{b}_1 = 225.8641, \quad \hat{a}_1 = -2.4078, \quad \hat{a}_3 = -10.2265,$$

$$\hat{b}_2 = 236.3366, \quad \hat{a}_2 = 1.3172, \quad \hat{a}_4 = 11.3172.$$

To compute $\hat{a}_o$, first calculate $\hat{m}$ as:

$$\begin{pmatrix} 5 & 0 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 & 0 \\ 0 & 0 & 5 & 0 & 0 \\ 0 & 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 0 & 5 \end{pmatrix} = \begin{pmatrix} 250 \\ 198 \\ 245 \\ 260 \\ 235 \end{pmatrix}$$
\begin{align*}
\mathbf{X}_o \hat{\mathbf{b}} &= \begin{pmatrix}
0 & 1 \\
0 & 1 \\
0 & 1 \\
0 & 1 \\
0 & 1 \\
\end{pmatrix} \begin{pmatrix}
225.8641 \\
236.3366 \\
\end{pmatrix} \\
\mathbf{P} \hat{\mathbf{a}}_p &= \begin{pmatrix}
-6.3172 \\
-6.3172 \\
6.3172 \\
6.3172 \\
\end{pmatrix} \\
\hat{\mathbf{m}} &= (\mathbf{I} + \mathbf{D}^{-1} \alpha)^{-1} (\mathbf{y}_o - \mathbf{X}_o \hat{\mathbf{b}} - \mathbf{P} \hat{\mathbf{a}}_p) \\
&= \begin{pmatrix}
3.9961 \\
-6.4039 \\
.4692 \\
3.4692 \\
-1.5308 \\
\end{pmatrix} \\
\mathbf{P} \hat{\mathbf{a}}_p + \hat{\mathbf{m}} &= \begin{pmatrix}
-2.3211 \\
-12.7211 \\
6.7864 \\
9.7864 \\
4.7864 \\
\end{pmatrix}.
\end{align*}

Generally, with today’s computer power, there is little need to use reduced animal models. However, routine genetic evaluation schemes for species with high reproductive rates may benefit from using a reduced animal model.

6 Sire and Dam Models

Another type of reduced animal model is called a sire and dam model. The assumptions to use this model are

1. Animals have only one record each,
2. Animals that have records are not parents of other animals, and
3. None of the parents have records of their own.
4. Sires and dams are mated randomly.
5. Sires and dams are random samples of parents, and not the result of intense selection.
6. Progeny have only one record each.
Parents have only progeny, and do not have records themselves. The animal model equation is re-written from

\[ y = Xb + Ia + e \]

to

\[ y = Xb + 0.5(Z_s a_s + Z_d a_d) + m + e \]

where the subscripts \( s \) and \( d \) refer to sire and dam, respectively. The analysis is conducted by combining \((m + e)\) into one residual term, say \( \epsilon \). Also,

\[
\text{Var} \begin{pmatrix} a_s \\ a_d \\ m + e \end{pmatrix} = \begin{pmatrix} \sigma_s^2 & 0 & 0 \\ 0 & \sigma_d^2 & 0 \\ 0 & 0 & \sigma_m^2 + \sigma_e^2 \end{pmatrix},
\]

which implies that sires are unrelated to each other or to any of the dams, and dams are unrelated. The sire variance is usually assumed to be equal to one quarter of the additive genetic variance. The dam variance generally equals one quarter of the additive genetic variance plus any common environmental effects, dominance genetic effects, and maternal genetic effects. Heritability is usually estimated from the sire variance for that reason.

In a sire and dam model, there is no interest in obtaining predictions for the progeny. However, the Mendelian sampling effects can be obtained by backsolving.

Sires and dams have usually been highly selected, especially sires, and so are not random samples of males or females amongst all males and females that are born in the population. They should not be random factors in the model, but more likely should be fixed factors.

### 7 Sire-Maternal Grandsire Models

The sire and dam model may be further simplified, when appropriate, to a sire-maternal grandsire model. The assumptions for this model are the same as those for the sire-dam model plus:

1. Each dam has only one progeny, and the dams of the dams (i.e. maternal granddams) have only one daughter, and
2. The daughters of a maternal grandsire (MGS) represented in the data are a random sample of all daughters of that MGS.

The last assumption is the most critical, and probably not valid. In dairy cattle, usually dams of cows are selected, and only good cows are allowed to have progeny. Poor cows are either culled or not allowed to produce replacement cows. For this reason, Sire-Maternal Grandsire models are not recommended in dairy cattle, and possibly other species too.

The \( a_d \) vector in the sire and dam model may be further partitioned as

\[ a_d = 0.5(a_{mgs} + a_{mgd}) + m_d \]
so that if \( Z_{mgd} = I \), and \( Z_d = I \), then

\[
y = Xb + .5Z_s a_s + .5(.5Z_{mgs} a_{mgs} + .5a_{mgd} + m_d) + \epsilon
\]

\[
= Xb + Z_s(.5a_s) + .5Z_s(.5a_{mgs}) + (.25a_{mgd} + m_d + \epsilon).
\]

The vectors \( .5a_s \) and \( .5a_{mgs} \) contain many of the same sires, and so the two can be combined into one vector, say \( s \), and the two incidence matrices can also be combined into one matrix, say \( Z_{smgs} \), which contains a 1 in each row for the sire of an animal, and a .5 for the MGS of that animal. A computational requirement at this point is that the sire and MGS are not the same individual for any one animal. The combined model is

\[
y = Xb + Z_{smgs} s + \xi.
\]

The solutions for \( s \) are in terms of estimated transmitting abilities and must be doubled to give estimated breeding values.

### 8 Sire Models

The next progression away from the animal model is a sire model. More assumptions are needed than the previous models. The sire model was one of the first models used in animal breeding for the evaluation of dairy bulls for the milk producing ability of their daughters. The additional assumptions of the sire model are

1. Sires are mated randomly to dams, and
2. Dams are mated to only one sire and have just one progeny.

The sire model is relatively simple to apply. Relationships among sires are usually formed based on the sire and MGS of each bull, rather than sire and dam.

Because of the many assumptions that are implied with a sire model (when compared to an animal model), the use of a sire model in the genetic evaluation of animals should be a last resort. If at all possible, an animal model should be employed.
9 Exercises

1. Below are pedigrees and data on 20 animals in three contemporary groups.

<table>
<thead>
<tr>
<th>Animal</th>
<th>Sire</th>
<th>Dam</th>
<th>Group</th>
<th>Record</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>2</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>3</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>4</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>40</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>28</td>
</tr>
<tr>
<td>7</td>
<td>3</td>
<td>4</td>
<td>1</td>
<td>34</td>
</tr>
<tr>
<td>8</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>35</td>
</tr>
<tr>
<td>9</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>17</td>
</tr>
<tr>
<td>10</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>41</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>25</td>
</tr>
<tr>
<td>12</td>
<td>3</td>
<td>2</td>
<td>2</td>
<td>38</td>
</tr>
<tr>
<td>13</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>29</td>
</tr>
<tr>
<td>14</td>
<td>3</td>
<td>4</td>
<td>2</td>
<td>27</td>
</tr>
<tr>
<td>15</td>
<td>5</td>
<td>7</td>
<td>3</td>
<td>37</td>
</tr>
<tr>
<td>16</td>
<td>6</td>
<td>14</td>
<td>3</td>
<td>30</td>
</tr>
<tr>
<td>17</td>
<td>8</td>
<td>7</td>
<td>3</td>
<td>42</td>
</tr>
<tr>
<td>18</td>
<td>1</td>
<td>10</td>
<td>3</td>
<td>46</td>
</tr>
<tr>
<td>19</td>
<td>3</td>
<td>13</td>
<td>3</td>
<td>24</td>
</tr>
<tr>
<td>20</td>
<td>5</td>
<td>9</td>
<td>3</td>
<td>26</td>
</tr>
</tbody>
</table>

(a) Construct $A^{-1}$ and set up the MME.

(b) Apply the following model,

$$y_{ij} = \mu + g_i + a_j + e_{ij},$$

where group, animal additive genetic, and residual effects are random. Let

$$\frac{\sigma_e^2}{\sigma_a^2} = 1.5, \quad \frac{\sigma_e^2}{\sigma_g^2} = 5.0.$$

(c) Compute SEP and reliabilities for all animals.

(d) Estimate the average EBVs for each contemporary group.

(e) Apply a reduced animal model to the same data. Compare solutions.

2. Generate phenotypes for animals 7 to 16 in the table below. Contemporary groups ($g_i$) are random effects with variance, $\sigma_g^2 = 120$. Age effects are fixed with differences $A_1 = 0$, $A_2 = 15$, $A_3 = 20$, and $A_4 = 22$. Let the overall mean ($\mu$) be 300, and the additive genetic variance be 2000, and the residual variance be 6500. The model equation is

$$y_{ijk} = \mu + g_i + A_j + a_k + e_{ijk},$$

where $a_k$ are the animal additive genetic values, and $e_{ijk}$ is a residual effect.
(a) Analyze the data you have created with the assumed model to obtain EBV on the animals.

(b) Correlate the EBVs with the true breeding values (which are known in a simulation exercise).

(c) Repeat the generation of new data sets, in the same manner using a different set of random numbers, a total of 10 times, and average the correlation of EBV with true breeding values over the 10 replicates. What is the variability of the correlation?

(d) Analyze the data with an appropriate Sire-Dam model and compare the sire and dam solutions with the solutions from the Animal model.