

Chapter 8

Multiple Traits

So far we have seen how to predict breeding values using the BLUP animal model. These breeding values were always only for one trait. From a statistical point of view, experts are calling such analyses **univariate** evaluations. In reality, livestock breeders want to improve their animals in a population with respect to several traits simultaneously. If there are genetic relationships (measured by genetic correlations) between traits, univariate predictions of breeding values do not make optimal use of the available information. This effect is stronger, if certain traits can only be observed in animals of one sex. According to (Mrode, 2005) who cites (Schaeffer, 1984) and (Thompson and Meyer, 1986), the increased accuracy of the evaluations is one of the main advantage of multivariate BLUP analyses. Predictions of breeding values for several traits in a single evaluation is called **multivariate** prediction of breeding values. Such multivariate analyses can be implemented in different ways such as

- combining different corrected information sources in a multivariate selection index as we have seen in chapter 4.
- multivariate prediction of breeding values using BLUP animal model (method of choice).
- combining predicted breeding values from univariate analyses.

Because we have already seen in earlier chapters how to combine information using the selection index, we are directly introducing the multivariate BLUP methods.

8.1 Multivariate Predictions Of Breeding Values Using BLUP

The prediction of breeding values using a multivariate BLUP model is the optimal prediction procedure. It has advantages, if

- some traits have lower heritability than others
- environmental correlations exist between traits measured on the same animal
- some traits are available only a subset of all animals
- some traits were used for a first round of selection

In principle, a multivariate analysis can be thought of as several univariate analyses which are stacked one on top of the other. Let us assume that we have two traits. For the first trait, we define the model

$$y_1 = X_1\beta_1 + Z_1a_1 + e_1$$

Similarly for the second trait, we define the model

$$y_2 = X_2\beta_2 + Z_2a_2 + e_2$$

If we group the data by traits, then we can write the multivariate model as

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

The genetic variance-covariance matrix G_0 for the two traits has the following structure.

$$G_0 = \begin{bmatrix} \sigma_{g_1}^2 & \sigma_{g_1, g_2} \\ \sigma_{g_1, g_2} & \sigma_{g_2}^2 \end{bmatrix} = \begin{bmatrix} g_{11} & g_{12} \\ g_{21} & g_{22} \end{bmatrix}$$

The inverse G_0^{-1} of G_0 can be formulated as

$$G_0^{-1} = \begin{bmatrix} g^{11} & g^{12} \\ g^{21} & g^{22} \end{bmatrix}$$

For the random residual effects, the variance-covariance matrix R_0 for the two traits can be written as

$$R_0 = \begin{bmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{bmatrix}$$

Also the inverse R_0^{-1} can be written as

$$R_0^{-1} = \begin{bmatrix} r^{11} & r^{12} \\ r^{21} & r^{22} \end{bmatrix}$$

The variance-covariance matrix for the complete set of true breeding values and for all random residuals can be written as

$$\text{var}(a) = \text{var} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} g_{11}A & g_{12}A \\ g_{21}A & g_{22}A \end{bmatrix} = G_0 \otimes A = G$$

where A is the numerator relationship matrix and \otimes denotes the so-called **Kroneckerproduct** of two matrices. The variance-covariance matrix R for the random residuals is given by

$$R = \text{var}(e) = \text{var} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} r_{11}I_n & r_{12}I_n \\ r_{21}I_n & r_{22}I_n \end{bmatrix} = R_0 \otimes I_n$$

where n corresponds to the number of animals in the pedigree. The covariances between all elements of a and e are 0. This is denoted by combining both random vectors into a single vector and writing down the variance-covariance matrix of the combined vector as

$$\text{var} \begin{bmatrix} a \\ e \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix} = \text{var} \begin{bmatrix} a_1 \\ a_2 \\ e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} g_{11}A & g_{12}A & 0 & 0 \\ g_{21}A & g_{22}A & 0 & 0 \\ 0 & 0 & r_{11}I_n & r_{12}I_n \\ 0 & 0 & r_{21}I_n & r_{22}I_n \end{bmatrix}$$

The solutions to get estimates of fixed effects and to get predictions for breeding values are obtained from the solutions of mixed model equations. These have exactly the same structure as in the univariate case.

$$\begin{bmatrix} X^T R^{-1} X & Z^T R^{-1} X \\ Z^T R^{-1} X & Z^T R^{-1} Z + G^{-1} \end{bmatrix} \begin{bmatrix} \hat{\beta} \\ \hat{a} \end{bmatrix} = \begin{bmatrix} X^T R^{-1} y \\ Z^T R^{-1} y \end{bmatrix}$$

where

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}, X = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix}, \hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{bmatrix}, Z = \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix}, \hat{a} = \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \end{bmatrix}$$

Based on the specification of the variance-covariance matrices described earlier, we get $R^{-1} = R_0^{-1} \otimes I_n$ and $G^{-1} = G_0^{-1} \otimes A^{-1}$.

8.2 Multitrait Selection

Now that we have predicted breeding values for a collection of traits available, the question is how to integrate these predicted breeding values into a consistent selection criterion. As we have seen earlier, selection index theory provides a tool to optimally combine different sources of information in order to approximate the aggregate genotype H . In section 8.2.3, we will return to this topic once again. Although the selection index provides an ideal framework to combine estimated breeding values into an optimal selection criterion, we are going to describe to other selection procedures that are commonly used in practical livestock breeding. The two procedures are

1. Tandem-Selection and
2. Selection based on independent selection thresholds.

The description of these procedures aims at showing the negative consequences that results from the use of these sub-optimal selection strategies.

8.2.1 Tandem-Selection

The term **Tandem-Selection** stands for the strategy of always improving the population with respect to just one trait. Once the breeding goal for this trait is reached the population is improved with respect to a different trait. This sequence of single-trait improvements is continued until the breeding goal for all traits is reached.

The problem with Tandem-Selection is that while improving the population for a given trait, the population can only realize correlated selection responses for all other traits. These correlated selection responses might be very small or can even be negative which causes the time that it takes to reach the breeding goal for all relevant traits to be very long.

8.2.2 Selection Based On Independent Selection Thresholds

This method was used before the selection index was discovered. This selection procedure is very easy to apply. First selection thresholds are defined for all traits. In the next steps, all animals which are above the thresholds for all traits are selected as parents of the next generation. With this method, selection responses for all traits can be obtained in the early generations after the implementation of this selection strategy.

We are using the following example to show how selection based on independent selection thresholds is used. For reasons of simplicity, we restrict ourselves on two traits. But the results can be generalized without any problems.

8.2.2.1 Example: Selection On Independent Thresholds

In a dairy cattle population, breeders want to improve milk yield and protein content. We assume the following selection thresholds for the two traits of interest during first lactation

- milk yield: 6900 kg
- protein content: 3.5 %

Figure 8.1 shows the performance data for a herd of dairy cows. The selection thresholds (red line for milk yield and blue line for protein content) divide the diagram into four quadrants. None of the cows in the lower right quadrant does meet any of the selection criterion imposed by the thresholds. The cows in the upper left quadrant fulfill the requirements for protein content and the cows in the lower right quadrant fulfill the requirements for milk yield. Only the cows in the upper right quadrant fulfill the requirements for both traits.

The disadvantage of this selection strategy becomes apparent with the cows in the upper left and in the lower right quadrant. They are culled and thereby not considered as parents of the next generation even though, they have good performances in one of the traits. From a statistical genetics point of view there are three problems associated with a selection strategy that is based on independent selection thresholds

1. livestock breeders tend to put the thresholds for all traits in the range of positive predicted breeding values. This leads to an exclusion of very many animals and a dramatic reduction in genetic variability
2. genetic relationships between traits are completely ignored. These relationships must be considered when defining selection thresholds. Otherwise the expected genetic gain will not be as expected.
3. differences in the economic relevance of the different traits are completely ignored. Putting the threshold in all traits into the range of positive predicted breeding values leads to a high emphasis on traits with a high heritability. Traits with lower heritability will have only very small selection responses.

8.2.3 Selection Index

In section 4.3, we have already briefly described how we can use selection index theory to approximate the aggregate genotype H in an optimal way. Just as a reminder, the aggregate genotype H combines all economically relevant traits into a single value using a linear function of the true breeding values a and taking the economic values w as weighting factors. Given that H corresponds to the linear function

$$H = w^T a$$

and we want to approximate H by an Index I which is a linear function of all predicted breeding values \hat{a} , we can write

$$I = b^T \hat{a}$$

where b is a vector of unknown index weights. The vector b is determined using the optimality condition of minimum prediction error variance which results in

$$b = P^{-1} G w \tag{8.1}$$

where P is the variance-covariance matrix between all information sources and G is the covariance matrix between the information sources and the traits in the aggregate genotype. In case where the traits in the aggregate genotype H and in the index I are the same, the matrices P and G are defined as

$$P = \text{var}(\hat{a})$$

and

$$G = \text{cov}(a, \hat{a})$$

For predicted breeding values using BLUP, it can be shown that $\text{cov}(a, \hat{a}) = \text{var}(\hat{a})$ and therefore $P = G$. Using that equality in equation (8.1), we get

$$b = w$$

which means that the vector b of index weights corresponds to the vector of economic values w .

The use of the selection index theory to combine predicted breeding values as information sources to approximate the aggregate genotype has the following advantages

- genetic relationships (correlations) between traits in the aggregate genotype are considered correctly
- relationships between information sources in the index are considered correctly
- information from auxiliary traits can be used
- differences in economic relevance of different traits are considered correctly
- expected selection responses can be estimated and thereby quantified

Despite all these advantages, index selection alone is very rarely used in practical livestock breeding. The reason for this is that every population has a few traits that are difficult to associate with an economic value or for some traits it is difficult to come up with genetic parameters. As a consequence of that, in practical livestock breeding we will always find a mix of index selection and a variety of independent selection thresholds.