Correlation Between Index And Predicted Breeding Value Of A Single Trait

Peter von Rohr

2018-11-26

Disclaimer

This notebook gives a few background information about the correlation between an index and the predicted breeding value of a single trait.

Introduction and Background

Let us assume, we have defined an Index I as a linear function of a vector of predicted breeding values from a set of traits. Hence

$$I = b^T \cdot \hat{u}$$

Furthermore, we assume that the set of traits in the index I is the same as the set of traits in the aggregate genotype H which is defined as

$$H = a^T \cdot u$$

Based on selection index theory, we can derive that the vectors a and b are the same and we can re-write the Index I as

$$I = a^T \cdot \hat{u}$$

where a corresponds to the vector of economic values of the traits in u and in $\hat{u}$.

Evalutate Index I

For any given choice of a and $\hat{u}$, the question is how good is the resulting index I. One possible measure of quality is the correlation (r_{HI}) between the index I and the aggregate genotype H.

$$r_{HI} = \frac{cov(H,I)}{\sqrt{var(H) * var(I)}} = \frac{cov(a^Tu, a^T\hat{u})}{\sqrt{var(a^Tu) * var(a^T\hat{u})}} = \frac{a^Tcov(u, \hat{u}^T)a}{\sqrt{a^Tvar(u)a * a^Tvar(\hat{u})a}} = \frac{a^Tvar(\hat{u})a}{\sqrt{a^Tvar(u)a * a^Tvar(\hat{u})a}}$$

$$= \sqrt{\frac{var(I)}{var(H)}}$$

Alternatively, one could also have a look at the correlations ($r_{I,\hat{u}_k}$) between an index I and the predicted breeding values $\hat{u}_k$ for a single trait k. This correlation is defined as

$$r_{I,\hat{u}_k} = \frac{cov(I, \hat{u}_k)}{\sqrt{var(I) * var(\hat{u}_k)}} = \frac{cov(a^T\hat{u}, \hat{u}_k)}{\sqrt{var(I) * var(\hat{u}_k)}} = \frac{a^T cov(\hat{u}, \hat{u}_k)}{\sqrt{var(I) * var(\hat{u}_k)}}$$

where $cov(\hat{u}, \hat{u}_k)$ corresponds to the k-th column of $var(\hat{u})$ which is the variance-covariance matrix of the predicted breeding values. Defining this variance-covariance matrix to be

$$C = var(\hat{u})$$

and let us define the prediction error variance (PEV) to be

$$PEV = var(u - \hat{u})$$

we can state that

$$C = G - PEV$$

where G is the genetic variance-covariance matrix. In the limit, where accuracy of predicted breeding values are high and PEV is small, C approaches G and hence the covariance $cov(\hat{u}, \hat{u}_k)$ tends towards the weighted mean of the k-th column of G with economic values a as weights.

$$cov(\hat{u}, \hat{u}_k) \approx (G)_k$$

where $(G)_k$ is the k-th column of G. From this, we can say

$$r_{I,\hat{u}_k} \approx \frac{a^T(G)_k}{\sqrt{var(I) * G_{kk})}}$$

where $(G)_{kk}$ is the element on row k and column k of G.

Conclusion

As a consequence of that, the correlation $r_{I,\hat{u}_k}$ does not necessarily have to be large, because it depends in the limit on the genetic variance-covariance matrix for the different traits.