

Chapter 33

Selection When Variances are Unequal

C. R. Henderson

1984 - Guelph

The mixed model equations for BLUP are well adapted to deal with variances that differ from one subpopulation to another. These unequal variances can apply to either \mathbf{e} or to \mathbf{u} or a subvector of \mathbf{u} . For example, cows are to be selected from several herds, but the variances differ from one herd to another. Some possibilities are the following.

1. σ_a^2 , additive genetic variance, is the same in all herds but the within herd σ_e^2 differ.
2. σ_e^2 is constant from one herd to another but intra-herd σ_a^2 differ.
3. Both σ_a^2 and σ_e^2 differ from herd to herd, but σ_a^2/σ_e^2 is constant. That is, intra-herd h^2 is the same in all herds, but the phenotypic variance is different.
4. Both σ_a^2 and σ_e^2 differ among herds and so does σ_a^2/σ_e^2 .

1 Sire Evaluation With Unequal Variances

As an example, AI sires are sometimes evaluated across herds using

$$\begin{aligned}y_{ijk} &= s_i + h_j + e_{ijk}. \\ \text{Var}(\mathbf{s}) &= \mathbf{A}\sigma_s^2, \\ \text{Var}(\mathbf{e}) &= \mathbf{I}\sigma_e^2, \\ \text{Cov}(\mathbf{a}, \mathbf{e}') &= \mathbf{0}.\end{aligned}$$

\mathbf{h} is fixed. Suppose, however, that we assume, probably correctly, that within herd σ_e^2 varies from herd to herd, probably related to the level of production. Suppose also that σ_s^2 is influenced by the herd. That is, in the population of sires σ_s^2 is different when sires are used in herd 1 as compared to σ_s^2 when these same sires are used in herd 2. Suppose further that σ_s^2/σ_e^2 is the same for every herd. This may be a somewhat unrealistic assumption, but it may be an adequate approximation. We can treat this as a multiple trait problem, trait 1 being progeny values in herd 1, trait 2 being progeny values in herd 2, etc. For purposes of illustration let us assume that all additive genetic correlations between pairs of traits are 1. In that case if the true rankings of sires for herd 1 were known, then these would be the true rankings in herd 2.

Let us order the progeny data by sire within herd. Then

$$\mathbf{R} = \begin{pmatrix} \mathbf{I}v_1 & 0 & \dots & 0 \\ 0 & \mathbf{I}v_2 & \dots & \\ \vdots & & \ddots & \vdots \\ 0 & \dots & \dots & \mathbf{I}v_t \end{pmatrix},$$

where there are t herds.

$$\mathbf{G} = \begin{pmatrix} \mathbf{A}w_{11} & \mathbf{A}w_{12} & \dots & \mathbf{A}w_{1t} \\ \mathbf{A}w_{12} & \mathbf{A}w_{22} & \dots & \mathbf{A}w_{2t} \\ \vdots & \vdots & & \vdots \\ \mathbf{A}w_{1t} & \mathbf{A}w_{2t} & & \mathbf{A}w_{tt} \end{pmatrix},$$

where v_i/w_{ii} is the same for all $i = 1, \dots, t$. Further $w_{ij} = (w_{ii}w_{jj})^{.5}$. This is, of course, an oversimplified model since it does not take into account season and year of freshening. It would apply to a situation in which all data are from one year and season.

We illustrate this model with a small set of data.

Sires	n_{ij}			y_{ij}		
	1	2	3	1	2	3
1	5	8	0	6	12	-
2	3	4	7	5	8	9
3	0	5	9	-	10	12

$$\mathbf{A} = \begin{pmatrix} 1 & .5 & .5 \\ & 1 & .25 \\ & & 1 \end{pmatrix}.$$

σ_e^2 for the 3 herds is 48, 108, 192, respectively. $Var(\mathbf{s})$ for the 3 herds is

$$\begin{pmatrix} 4\mathbf{A} & 6\mathbf{A} & 8\mathbf{A} \\ & 9\mathbf{A} & 12\mathbf{A} \\ & & 16\mathbf{A} \end{pmatrix}.$$

Note that $6 = [4(9)]^{.5}$, $8 = [4(16)]^{.5}$, and $12 = [9(16)]^{.5}$. Accordingly \mathbf{G} is singular and we need to use the method described in Chapter 5 for singular \mathbf{G} . Now the GLS coefficient matrix for fixed \mathbf{s} is in (33.1) ... (33.3). This corresponds to ordering $(s_{11}, s_{21}, s_{31}, s_{12}, s_{22}, s_{32}, s_{13}, s_{23}, s_{33})$. The first subscript on s refers to sire number and the second to herd number. The right hand side vector is $(.1250, .1042, 0, .1111, .0741, .0926, 0, .0469, .0625, .2292, .2778, .1094)'$.

The upper diagonal element of (33.1) to (33.3) is $5/48$, 5 being the number of progeny of sire 1 in herd 1, and 48 being σ_e^2 for herd 1. The lower diagonal is $16/192$. The first element of the right hand side is $6/48$, and the last is $21/192$.

Upper left 6×6

$$\text{diag}(.10417, .06250, 0, .07407, .03704, .04630). \quad (1)$$

Upper right 6×6 and (lower left 6×6)'

$$\begin{pmatrix} 0 & 0 & 0 & .10417 & 0 & 0 \\ 0 & 0 & 0 & .06250 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & .07407 & 0 \\ 0 & 0 & 0 & 0 & .03704 & 0 \\ 0 & 0 & 0 & 0 & .04630 & 0 \end{pmatrix}. \quad (2)$$

Lower right 6×6

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ .03646 & 0 & 0 & 0 & 0 & .03646 \\ .04687 & 0 & 0 & 0 & .04687 & 0 \\ .16667 & 0 & 0 & 0 & 0 & 0 \\ .15741 & 0 & 0 & 0 & 0 & 0 \\ .08333 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \quad (3)$$

Now we multiply these equations by

$$\begin{pmatrix} 4\mathbf{A} & 6\mathbf{A} & 8\mathbf{A} & 0 \\ & 9\mathbf{A} & 12\mathbf{A} & 0 \\ & & 16\mathbf{A} & 0 \\ & & & \mathbf{I}_3 \end{pmatrix},$$

and add 1 to each of the first 9 diagonal elements. Solving these equations the solution is $(-.0720, .0249, .0111, -.1080, .0373, .0166, -.1439, .0498, .0222, 1.4106, 1.8018, 1.2782)'$. Note that $\hat{s}_{i1}/\hat{s}_{i2} = 2/3$, $\hat{s}_{i1}/\hat{s}_{i3} = 1/2$, $\hat{s}_{i2}/\hat{s}_{i3} = 3/4$. These are in the proportion (2:3:4) which is $(4^5:9^5:16^5)$. Because of this relationship we can reduce the mixed model equations to a set involving s_{i1} and h_j by premultiplying the equations by

$$\begin{pmatrix} 1. & 0 & 0 & 1.5 & 0 & 0 & 2. & 0 & 0 & 0 & 0 & 0 \\ 0 & 1. & 0 & 0 & 1.5 & 0 & 0 & 2. & 0 & 0 & 0 & 0 \\ 0 & 0 & 1. & 0 & 0 & 1.5 & 0 & 0 & 2. & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1. & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1. & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1. \end{pmatrix}. \quad (4)$$

Then the resulting coefficient matrix is post-multiplied by the transpose of matrix (33.4).

This gives equation (33.5).

$$\begin{pmatrix} 15.104 & 4.229 & 4.229 & 3.927 & 5.035 & 2.417 \\ 3.927 & 15.708 & 2.115 & 3.323 & 3.726 & 2.794 \\ 3.927 & 2.115 & 15.708 & 1.964 & 4.028 & 3.247 \\ 0.104 & 0.062 & 0.0 & 0.167 & 0.0 & 0.0 \\ 0.111 & 0.056 & 0.069 & 0.0 & 0.157 & 0.0 \\ 0.0 & 0.073 & 0.094 & 0.0 & 0.0 & 0.083 \end{pmatrix} \begin{pmatrix} \hat{s}_{11} \\ \hat{s}_{21} \\ \hat{s}_{31} \\ \hat{h}_1 \\ \hat{h}_2 \\ \hat{h}_3 \end{pmatrix} = \begin{pmatrix} 16.766 \\ 15.104 \\ 14.122 \\ .229 \\ .278 \\ .109 \end{pmatrix}. \quad (5)$$

The solution is $(-.0720, .0249, .0111, 1.4016, 1.8018, 1.2782)'$. These are the same as before.

How would one report sire predictions in a problem like this? Probably the logical thing to do is to report them for a herd with average σ_e^2 . Then it should be pointed out that sires are expected to differ more than this in herds with large σ_e^2 and to differ less in herds with small σ_e^2 . A simpler method is to set up equations at once involving only s_{i1} or any other chosen s_{ij} (j fixed). We illustrate with s_{i1} . The $\overline{\mathbf{W}}$ matrix for our example with subclass means ordered sires in herds is

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 \\ 1.5 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1.5 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1.5 & 0 & 1 & 0 \\ 0 & 2 & 0 & 0 & 0 & 1 \\ 0 & 0 & 2 & 0 & 0 & 1 \end{pmatrix}.$$

This corresponds to $\hat{s}_{i2} = 1.5 \hat{s}_{i1}$, and $\hat{s}_{i3} = 2 \hat{s}_{i1}$. Now compute the diagonal matrix

$$\text{diag}(5, 3, 8, 4, 5, 7, 9) [\text{dg}(48, 48, 108, 108, 108, 192, 192)]^{-1} \equiv \mathbf{D}.$$

Then the GLS coefficient matrix is $\overline{\mathbf{W}}' \mathbf{D} \overline{\mathbf{W}}$ and the right hand side is $\overline{\mathbf{W}}' \mathbf{D} \overline{\mathbf{y}}$, where $\overline{\mathbf{y}}$ is the subclass mean vector. This gives

$$\begin{pmatrix} .2708 & 0 & 0 & .1042 & .1111 & 0 \\ & .2917 & 0 & .0625 & .0556 & .0729 \\ & & .2917 & 0 & .0694 & .0937 \\ & & & .1667 & 0 & 0 \\ & & & & .1574 & 0 \\ & & & & & .0833 \end{pmatrix} \begin{pmatrix} \hat{s}_{11} \\ \hat{s}_{21} \\ \hat{s}_{31} \\ \hat{h}_1 \\ \hat{h}_2 \\ \hat{h}_3 \end{pmatrix} = \begin{pmatrix} .2917 \\ .3090 \\ .2639 \\ .2292 \\ .2778 \\ .1094 \end{pmatrix}. \quad (6)$$

Then add $(4\mathbf{A})^{-1}$ to the upper 3 x 3 submatrix of (33.6) to obtain mixed model equations. Remember $4\mathbf{A}$ is the variance of the sires in herd 1. The solution to these equation is as before, $(-.0720, .0249, .0111, 1.4106, 1.8018, 1.2782)'$.

2 Cow Evaluation With Unequal Variances

Next we illustrate inter-herd joint cow and sire when herd variances are unequal. We assume a simple model

$$y_{ij} = h_i + a_j + e_{ij}.$$

\mathbf{h} is fixed, \mathbf{a} is additive genetic merit with

$$Var(\mathbf{a}) = \begin{pmatrix} \mathbf{A}g_{11} & \mathbf{A}g_{12} & \cdots & \mathbf{A}g_{1t} \\ \mathbf{A}g_{12} & \mathbf{A}g_{22} & \cdots & \mathbf{A}g_{2t} \\ \vdots & \vdots & & \vdots \\ \mathbf{A}g_{1t} & \mathbf{A}g_{2t} & & \mathbf{A}g_{tt} \end{pmatrix}.$$

\mathbf{A} is the numerator relationship for all animals. There are t herds, and we treat production as a different trait in each herd. We assume genetic correlations of 1. Therefore $g_{ij} = (g_{ii}g_{jj})^{.5}$.

$$Var(\mathbf{e}) = \begin{pmatrix} \mathbf{I}v_1 & & & \mathbf{0} \\ & \mathbf{I}v_2 & & \\ & & \ddots & \\ \mathbf{0} & & & \mathbf{I}v_t \end{pmatrix}.$$

First we assume σ_a^2/σ_e^2 is the same for all herds. Therefore g_{ii}/v_i is the same for all herds.

As an example suppose that we have 2 herds with cows 2, 3 making records in herd 1 and cows 4, 5 making records in herd 2. These animals are out of unrelated dams, and the sire of 2 and 4 is 1. The records are 3, 2, 5, 6.

$$\mathbf{A} = \begin{pmatrix} 1 & .5 & 0 & .5 & 0 \\ & 1 & 0 & .25 & 0 \\ & & 1 & 0 & 0 \\ & & & 1 & 0 \\ & & & & 1 \end{pmatrix}.$$

Ordering the data by cow number and the unknowns by h_1, h_2, \mathbf{a} in herd 1, \mathbf{a} in herd 2 the incidence matrix is

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

Suppose that

$$\mathbf{G} = \begin{pmatrix} 4\mathbf{A} & 8\mathbf{A} \\ 8\mathbf{A} & 16\mathbf{A} \end{pmatrix}, \quad \mathbf{R} = \begin{pmatrix} 12\mathbf{I} & \mathbf{0} \\ \mathbf{0} & 48\mathbf{I} \end{pmatrix}.$$

Then $\sigma_e^2/\sigma_a^2 = 3$ in each herd, implying that $h^2 = .25$. Note that \mathbf{G} is singular so the method for singular \mathbf{G} is used. With these parameters the mixed model solution is

$$\begin{aligned}\hat{\mathbf{h}} &= (2.508, 5.468). \\ \hat{\mathbf{a}} \text{ in herd 1} &= (.030, .110, -.127, -.035, .066). \\ \hat{\mathbf{a}} \text{ in herd 2} &= (.061, .221, -.254, -.069, .133).\end{aligned}$$

Note that \hat{a}_i in herd 2 = 2 \hat{a}_i in herd 1 corresponding to $(16/4)^{.5} = 2$.

A simpler method is to use an incidence matrix as follows.

$$\begin{pmatrix} 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 2 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 2 \end{pmatrix}$$

This corresponds to unknowns h_1, h_2, \mathbf{a} in herd 1. Now $\mathbf{G} = 4\mathbf{A}$ and \mathbf{R} is the same as before. The resulting solution is the same as before for $\hat{\mathbf{h}}$ and $\hat{\mathbf{a}}$ in herd 1. Then $\hat{\mathbf{a}}$ in herd 2 is 2 times $\hat{\mathbf{a}}$ in herd 1.

Now suppose that \mathbf{G} is the same as before but $\sigma_e^2 = 12,24$ respectively. Then h^2 is higher in herd 2 than in herd 1. This leads again to the \hat{a} in herd 2 being twice \hat{a} in herd 1, but the \hat{a} for cows making records in herd 2 are relatively more variable, and if we were selecting a single cow, say for planned mating, the chance that she would come from herd 2 is increased. The actual solution in this example is

$$\begin{aligned}\hat{\mathbf{h}} &= (2.513, 5.468). \\ \hat{\mathbf{a}} \text{ in herd 1} &= (.011, .102, -.128, -.074, .106). \\ \hat{\mathbf{a}} \text{ in herd 2} &= \text{twice those in herd 1.}\end{aligned}$$

The only reason we can compare cows in different herds is the use of sires across herds.

A problem with the methods of this chapter is that the individual intra-herd variances must be estimated with limited data. It would seem, therefore, that it might be advisable to take as the estimate for an individual herd, the estimate coming from that herd regressed toward the mean of variances of all herds, the amount of regression depending upon the number of observations. This would imply, perhaps properly, that intra-herd variances are a sample of some population of variances. I have not derived a method comparable to BLUP for this case.