

Chapter 13

The animal model and selection

Julius van der Werf

Introduction

The important genetic properties of the BLUP and the animal model are due to the use of the Numerator Relationships Matrix (NRM), i.e. is the matrix with additive genetic relationships between animals. This matrix was originally used by Henderson to account for covariances between random effects, and therefore to use information from relatives in estimation of breeding value. However, important quantitative genetic properties such as accounting for selection over generations and accounting for inbreeding were revealed later. The essence is that the relationship matrix contains all information about the flow of genes through the population. It also allows an explicit dissection of genetic variation due to having different sires and/or dams, or due to differences from Mendelian sampling

This chapter will discuss some more genetic properties of BLUP EBV's, especially in relation to the NRM.

Usually, statistical models assume random sampling of random effects, as otherwise, predictions and estimates of variance components would be biased by selection. So how can we use BLUP procedures for estimating breeding values, based on data where the whole purpose is to use highly selected individuals? And not only will we collect records on selected animals, also the variance among a group of individuals changes drastically after selection, and genetic variances in populations is reduced as a result of selection. Since these changes are significant and immediate, it is important that the models that we use for genetic evaluation and for prediction of response have the ability to account for such selection. Fortunately they do, provided that the genetic model called 'infinitesimal model' is valid.

This section describes how BLUP and REML can account for selection in mixed models. First we recall the simple BLUP example to show how a breeding value is estimated conditional on its parental mean:

Recalling the BLUP example:

$$\begin{pmatrix} \mathbf{X'X} & \mathbf{X'Z} \\ \mathbf{Z'X} & \mathbf{Z'Z} + \mathbf{A}^{-1}\lambda \end{pmatrix}^{-1} \begin{pmatrix} \hat{\mathbf{b}} \\ \hat{\mathbf{u}} \end{pmatrix} = \begin{pmatrix} \mathbf{X'Y} \\ \mathbf{Z'Y} \end{pmatrix}$$

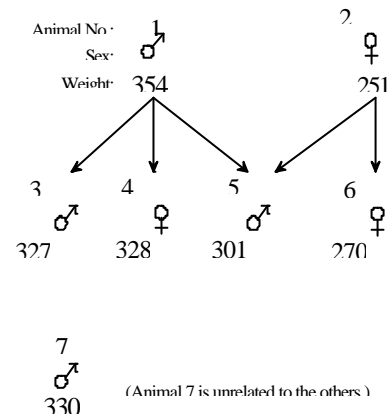
Year of Birth

1990:

1991:

1992:

Pedigree



$$\begin{bmatrix} 7 & 1 & 3 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 3 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & -1 \\ 3 & 1 & 5 & 0 & 0 & 1 & 1 & 1 & 1 & -1 \\ & & & & & & & & & \\ 1 & 1 & 0 & 19/6 & 1/2 & -2/3 & -2/3 & -1 & 0 & 0 \\ 1 & 1 & 0 & 1/2 & 17/6 & 0 & 0 & -1 & -2/3 & 0 \\ 1 & 0 & 1 & -2/3 & 0 & 7/3 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & -2/3 & 0 & 0 & 7/3 & 0 & 0 & 0 \\ 1 & 0 & 1 & -1 & -1 & 0 & 0 & 3 & 0 & 0 \\ 1 & 0 & 1 & 0 & -2/3 & 0 & 0 & 0 & 7/3 & 0 \\ 1 & -1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 2 \end{bmatrix} \begin{pmatrix} \hat{b}_{mean} \\ \hat{b}_{1990} \\ \hat{b}_{1991} \\ \hat{u}_1 \\ \hat{u}_2 \\ \hat{u}_3 \\ \hat{u}_4 \\ \hat{u}_5 \\ \hat{u}_6 \\ \hat{u}_7 \end{pmatrix} = \begin{pmatrix} 2161 \\ 275 \\ 896 \\ \\ 354 \\ 251 \\ 327 \\ 328 \\ 301 \\ 270 \\ 330 \end{pmatrix}$$

and the equation for animal 6:

$$\begin{aligned} \hat{m} + \hat{b}_{1991} - \frac{2}{3}\hat{u}_2 + \frac{7}{3}\hat{u}_6 &= 270 \\ \rightarrow \hat{u}_6 &= \frac{3}{7}(270 - \hat{m} - \hat{b}_{1991}) + \frac{2}{7}\hat{u}_2 \\ \rightarrow \hat{u}_6 &= \frac{3}{7}(270 - \hat{m} - \hat{b}_{1991} - \frac{1}{2}\hat{u}_2) + \frac{1}{2}\hat{u}_2 \end{aligned}$$

Therefore, the breeding value of animal 6 is estimated as a deviation of her phenotypic record from the expected mean and from her dams breeding value. The expected mean is the sum of all fixed effects plus the family mean. Since animal 6 has only one parent known, we take the deviation from one parent only (half sib family mean) and the weighting factor is the regression of the within half sib family deviation on the within half sib family breeding value.

The variance of the within half sib breeding value is 0.75 times the genetic variance. The weight is therefore equal to the following regression coefficient

$$\frac{\text{cov}(u_{whsf}, (y - y_{hsfmean}))}{\text{var}(y - y_{hsfmean})} = \frac{\frac{3}{4}s_g^2}{s_e^2 + \frac{3}{4}s_g^2} = \frac{\frac{3}{4}h^2}{1 - h^2 + \frac{3}{4}h^2} = \frac{3}{7}$$

We can do the same thing for **animal 5**:

$$\begin{aligned} \mathbf{m} + b_{1991} - \hat{u}_1 - \hat{u}_2 + 3\hat{u}_5 &= 301 \\ \rightarrow \hat{u}_5 &= \frac{1}{3}(301 - \mathbf{m} - b_{1991}) + \frac{1}{3}(\hat{u}_1 + \hat{u}_2) \end{aligned}$$

$$\rightarrow \hat{u}_5 = \frac{1}{3}(301 - \mathbf{m} - b_{1991} - \frac{1}{2}(\hat{u}_1 + \hat{u}_2)) + \frac{1}{2}(\hat{u}_1 + \hat{u}_2)$$

The breeding value of animal 5 is also estimated as a deviation of her phenotypic record from the expected mean and from the mean of his parents' breeding values. The expected mean is again the sum of all fixed effects plus the family mean. Since animal 5 has both parents known, we take the deviation from the full sib family mean and the weighting factor is the regression of the within full sib family deviation on the within full sib family breeding value.

The variance of the within full sib breeding value is 0.5 times the genetic variance. The weight is therefore equal to the following regression coefficient

$$\frac{\text{cov}(u_{\text{whsf}}, (y - y_{\text{fjmean}}))}{\text{var}(y - y_{\text{fjmean}})} = \frac{\frac{1}{2}\mathbf{s}_g^2}{\mathbf{s}_e^2 + \frac{1}{2}\mathbf{s}_g^2} = \frac{\frac{1}{2}h^2}{1 - h^2 + \frac{1}{2}h^2} = \frac{1}{3}$$

Writing out the equation for **animal 2** who is a parents with progeny:

$$\mathbf{m} + b_{1990} + \frac{1}{2}\hat{u}_1 + \frac{17}{6}\hat{u}_2 - \hat{u}_5 - \frac{2}{3}\hat{u}_6 = 251$$

$$\rightarrow \hat{u}_2 = \frac{6}{17}(251 - \mathbf{m} - b_{1990}) - \frac{3}{17}\hat{u}_1 + \frac{6}{17}\hat{u}_5 + \frac{4}{17}\hat{u}_6$$

$$\rightarrow \hat{u}_2 = \frac{6}{17}(251 - \mathbf{m} - b_{1990}) + \frac{6}{17}(\hat{u}_5 - \frac{1}{2}\hat{u}_1) + \frac{4}{17}\hat{u}_6$$

Hence we see that the breeding value for animal 2 is estimated from her own record as deviation from the fixed effects (we have no family mean since she has no parents known), and from the estimated breeding values of her progeny.

Notice that the breeding values are corrected for the other parent (i.e. there's a correction for the mate), if the mate is known. In this case, the EBV of animal 5 is corrected for the contribution of his sire.

The weights for animal 2 are not very easy to recognize, but they are the same as selection index weights. We can check this by simplifying a bit the example, and ignore animal 5 as a progeny.

If animal 2 had only one progeny (animal 6), than her BLUP equation would look like

$$\hat{u}_2 = \frac{6}{14}(251 - \mathbf{m} - b_{1990}) + \frac{4}{14}\hat{u}_6$$

but we saw earlier how the breeding value of animal 6 is estimated, therefore:

$$\rightarrow \hat{u}_2 = \frac{6}{14}(251 - \mathbf{m} - b_{1990}) + \frac{4}{14}[\frac{3}{7}(270 - \mathbf{m} - b_{1991} - \frac{1}{2}\hat{u}_2) + \frac{1}{2}\hat{u}_2]$$

$$\rightarrow \hat{u}_2 = \frac{6}{14}(251 - \mathbf{m} - b_{1990}) + \frac{6}{49}(270 - \mathbf{m} - b_{1991}) + \frac{4}{49}\hat{u}_2$$

$$\rightarrow \hat{u}_2 = \frac{7}{15}(251 - m - b_{1990}) + \frac{2}{15}(270 - m - b_{1991})$$

and the two weights can be found exactly by selection index to estimate the breeding value of an animal based on her own and her progeny's phenotypic record using a heritability of 0.5. This again shows that BLUP uses the same weights as selection index. The difference is that the means (fixed effects) are more properly corrected for.

BLUP accounts also for possible genetic differences. It takes deviations from expected genetic means, and the EBV is regressed towards the expected genetic mean. This expected genetic mean is in most cases the family mean. This is an illustration of a important properties of BLUP:

- It corrects for selection.
Since the better parents have usually (more) offspring, we expect that the average breeding value goes up in later generations (in case of selection).
- The genetic trend is estimated from the average EBVs over time, i.e. the EBVs are plotted against the birth year of the animals.

Breeding values are estimated as a deviation of the phenotypic record from the expected mean and from the mean of his parents' breeding values. Hence, they are estimated conditional on the parental breeding values. The regression coefficient is based on the within family variance (genetic over total). For one parent known this is $\frac{3}{4}V_A/(V_E + \frac{3}{4}V_A)$, and for two parents known this is $\frac{1}{2}V_A/(V_E + \frac{1}{2}V_A)$. Offspring within families will be estimated unbiasedly against each other, as long as all, and not a selected group of offspring gets evaluated. The family mean will be estimated from parents EBV, and this will be unbiased in its turn based on the same argument. The parents are likely better than their family mates (otherwise they wouldn't be selected) and this will be expressed in their EBV as they will be positively deviating within their family. Hence, by using pedigree information and all information that has been used to base selection decisions on, all animals will obtain unbiased estimates of BV as long as parents are unbiased. This argument can be carried back to the base animals; those with no parents known. Since there is no pedigree information from base animals, we will have to assume that they are sampled from the same homogeneous base population (as we will regress their phenotype back to this mean). If base animals are really from different subpopulations, possibly with different genetic means, we should regress their performances back to the appropriate mean. This can be achieved by assigning them to their appropriate genetic group.

Genetic trend

Presumably, in effective breeding programs, the average genetic merit of the traits selected for changes in a desired direction. BLUP accounts also for possible genetic differences as it takes deviations from expected genetic means, and the EBV is regressed towards the expected genetic mean. This expected genetic mean is in most cases the family mean.

This is an illustration of a important properties of BLUP:

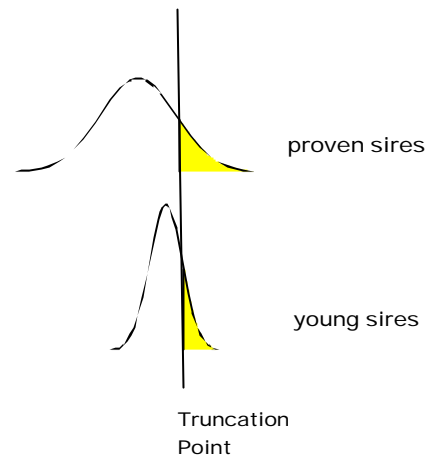
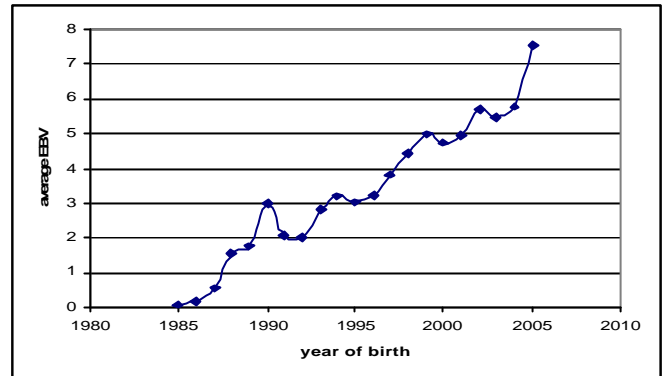
- It corrects for selection.
Since the better parents have usually (more) offspring, we expect that the average breeding value goes up in later generations (in case of selection).

13: The Animal Model and Selection

- The genetic trend is estimated from the average EBVs over time, i.e. the EBVs are plotted against the birth year of the animals.

- There is a correction a correction for assortative mating

- BLUP allow comparison of individuals across age classes (as it accounts for trend) This allows optimal selection of the best animals, irrespective of their age class, and practically this means that selection based on BLUP EBV optimizes generation interval in breeding programs. If genetic trends are higher, relatively more young animals will be selected.



13: The Animal Model and Selection

As an illustration, compare the EBV with and without accounting for relationships across years. Offspring 1 and 2 are from sire 1 and offspring 3 is from sire 2. A heritability of 0,25 is assumed here. The simple example not only illustrates the estimation of genetic trend, it also shows how BLUP separates between genetic and non-genetic trends.

1) No relationships across years: EBV's within a year sum to zero.
Year effect is overestimated

year	sire 1	sire 2	sire 3
1 300	350	300	250
	13	0	-13
	offspr. 1	offspr. 2	offspr. 3
2 333	365	325	310
	8	-2	-6

2)BLUP with relationships across years: EBV's in subsequent year are above zero.
Year effect estimate now only reflects environmental trend.

year	sire 1	sire 2	sire 3
1 300	350	300	250
	14	-2	-13
	offspr. 1	offspr. 2	offspr. 3
2 329	365	325	310
	13	5	-4

BLUP accounting for changes of variance due to selection

It has been shown that variance in a population change drastically after one round of selection. Since this change is large and immediate, it is important that the models that we use for genetic evaluation and for prediction of response account for the Bulmer effect. This section describes how BLUP and REML account for the Bulmer effect using the mixed model.

The genetic model assumed for showing these properties is (again) the infinitesimal model, i.e. the only changes in genetic variance due to selection are due the gametic disequilibrium and inbreeding, but not to change of gene frequencies. This assumption is reasonable when short-term responses are considered (Bulmer 1971).

Kennedy and Sorensen (1988) have given an excellent explanation of the genetic properties of the mixed model, and showed that the mixed model accounts for changes of genetic variance after selection. They refer to Henderson's (1975) paper on properties of BLUP under a selection model, where he used an argument of Pearson (1903) to show that the mixed model equations yield BLUP in case of selection, given that all the data used in making selection decisions are included in the analysis. Pearson's result gives the variance of a variable x after selection on a correlated variable y as

$$\mathbf{s}_{sx}^2 = \mathbf{s}_x^2 - \text{cov}(x, y)^2 \mathbf{s}_y^{-2} (1 - (\mathbf{s}_{ys}^2 / \mathbf{s}_y^2)) = \mathbf{s}_x^2 - \text{cov}(x, y)^2 \mathbf{s}_y^{-2} \mathbf{s}_x^{-2} (1 - (1 - k)) \mathbf{s}_x^2 = (1 - r^2 k) \mathbf{s}_x^2$$

where r is the correlation between x and y and k is the proportional reduction in phenotypic variance in the selected group. Using Pearson's result, we can write the genetic variance in a selected group (after mass selection) as $(1 - h^2 k) \mathbf{s}_A^2$, since the correlation between additive genetic value and phenotype is equal to h (the square root of the heritability). Using this formula and assuming equal selection intensities gives the genetic variance in the progeny of selected parents as $(1 - \frac{1}{2} h^2 k) \mathbf{s}_A^2$, a result also found by Bulmer (1971).

Kennedy and Sorensen (1988) point out that Henderson (1975) has shown that BLUP correctly accounts for selection because Pearson's rules for conditional variances apply. That is, if the base population is unselected, and the next generation descends from the best parents of the previous generation, than both generations can be evaluated in an unbiased fashion. Although the 2nd generation animals are not random animals (but from selected parents), they are unselected within their families. In statistical terms, the distribution (i.e. mean as well as variance) of the second generation conditional on the first generation is not affected by selection. In quantitative genetic terminology, this conditional variance is equal to the within family variance, or the Mendelian sampling variance. The parental contributions are evaluated unbiasedly if their contemporaries that they were selected from are in the model

In an earlier chapter we have written A as TDT' where T describes the flow of genes over the generations and D described the variance of the part of the breeding value of each animal that is not explained by its ancestors. The breeding values can be written as $a = T\phi$, i.e. they are a linear combination of values in ϕ , which are not affected by selection. The matrix T describes how a given breeding value is a linear function of effects of ancestors that each was not affected by selection. This assumes that unknown ancestors not in the model (not in a) were unselected. For our example, T was

$$T = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0.5 & 0.5 & 0 & 1 & 0 & 0 & 0 \\ 0.25 & 0.5 & 0.5 & 0 & 1 & 0 & 0 \\ .375 & .25 & .75 & 0 & .5 & 1 & 0 \\ .4375 & 0.125 & 0.875 & 0 & 0.25 & 0.5 & 1 \end{pmatrix}$$

Therefore, as long as the base population is unselected, and the Mendelian sampling terms are not affected by selection, as is the case with an infinitesimal model, then $\text{var}(a) = A\mathbf{S}_A^{-2}$, even though selection may have operated in subsequent generations, and the effects of linkage disequilibrium and inbreeding are accommodated (Kennedy and Sorensen, 1988).

An example might illustrate that Pearson's rules for conditional variance lead to the same results as Bulmer (1971). The example is (slightly) adapted from Henderson (1982).

Consider a model $y = a + e$, being phenotype explained by additive genetic and environmental effects. Heritability is 0.30. In the genetic part of the model we can replace the additive genetic value by the value of the 2 gametes: $y_i = g_i^p + g_i^m + e$.

Now consider 2 sires and we are interested in selecting one of them (based on its phenotype)

The variance of the gametic effects before selection is:

$$G = \text{var}(g) = \text{var} \begin{pmatrix} g_1^p \\ g_1^m \\ g_2^p \\ g_2^m \end{pmatrix} = \begin{pmatrix} 0.15 & 0 & 0 & 0 \\ 0 & 0.15 & 0 & 0 \\ 0 & 0 & 0.15 & 0 \\ 0 & 0 & 0 & 0.15 \end{pmatrix}$$

Now suppose we select on the phenotypes. After selection among the 2, the variance of the phenotypes

after selection is (based on order statistics) equal to $\text{var}(y_s) = V_s = \begin{pmatrix} 0.6817 & 0.3183 \\ 0.3183 & 0.6817 \end{pmatrix}$.

Using Pearson's rule, the variance among the gametes after selection is equal to

$$G_s = G - \text{cov}(g,y) \cdot \text{var}(y)^{-1} \cdot (\text{var}(y) - \text{var}(y_s)) \cdot \text{var}(y)^{-1} \cdot \text{cov}(y,g) =$$

$$= G - B V^{-1} (V - V_s V^{-1} B) \quad \text{where } B = \text{cov}(y,g)$$

$$= \begin{pmatrix} .1428 & -.0072 & .0072 & .0072 \\ -.0072 & .1428 & .0072 & .0072 \\ .0072 & .0072 & .1428 & 0.0072 \\ .0072 & .0072 & -.0072 & .1428 \end{pmatrix}$$

Note that in this result, the covariance between gametes on the same animals is negative after selection.

Furthermore, the variance of breeding values is $\text{var}(a_i) = \text{var}(g_i^p) + \text{var}(g_i^m) + 2\text{cov}(g_i^p, g_i^m) = .1428 + .1428 + 2*(-.0072) = 0.2712$. This agrees with the variance reduction expected from Bulmer's formula. For selecting 1 out of 2, $i = .564$ and $x = 0$, therefore $k = 0.318$. The variance of a_i after

selection on phenotype is $(1-h^2k)S_A^2 = (1-.3*.318)*.3 = 0.271$. The additive genetic variance in the progeny (a_p) of this selected animal (giving only the paternal gamete, to mate with an unselected maternal gamete) is $\text{var}(a_p) = \text{var}(g^p) + \text{var}(g^m) = 0.2928$. Hence, 50% of the loss in genetic variance due to selection is recovered in the next generation.

Which variance should be used in BLUP?

In the previous we showed that the 'base population' variance is not the same as the variance in later generations. The genetic variance in a given generation can be divided in a part coming from variation between base parents and a part resulting from Mendelian sampling. This causes a somewhat complicated dilemma: what is the relevant genetic variance in a selection program. Is it the variance of the base population, or the variance at the current generation? The answer is that both are relevant. The base population variance remains important, because the Mendelian Sampling component of the variance (within FS family variance) remains to be 50% of the base population variance. The sire variance as well as the dam variance (i.e. the between family variance) are significantly reduced in a selection program. Estimates of genetic variance based on between family variance components are therefore not appropriate for populations under selection. Instead, an animal model is used, and variance components are estimated with REML. In principle, if all genetic relationships are included back to the base population, and if all data is used in the analysis that was included in the selection decisions, REML will provide estimates of the base population variance (Sorensen and Kennedy, 1984). It should be noted that including relationships does not only account for gametic disequilibrium due to selection, but also for reduction of variance due to inbreeding and the buildup of covariances of related animals. Of course a debate exists on what is the base population (how many generations do we need to go back), and we should realize that practically it is usually not possible to include relationships and data since the start of selection. The conclusion is that genetic variances that are estimated can be expected to be somehow biased by selection, generally more if more of the selection history is omitted from the analysis (Van der Werf and De Boer, 1990).

References

- Bulmer, M.G. 1971. The effect of selection on genetic variability. *American Nat.* 105:201.
 Bulmer, M.G. 1980. *The mathematical theory of quantitative genetics*. Clarendon Press. Oxford.
 Henderson, C.R. 1975. Best Linear Unbiased estimation and prediction under a selection model. *Biometrics* 31:423-447.
 Henderson, C.R. 1976. A simple method for computing the inverse of a numerator relationship matrix used in prediction of breeding values. *Biometrics* 32:69-83.
 Kennedy, B.W. and D. Sorensen. 1988. *Proc. of the Conference on Quantitative Genetics*.
 Meuwissen, T.H.E. and Luo.Z. 1992. Computing inbreeding coefficients in large populations. *Genet. Select. Evol.* 24:305-313
 Mrode, R.A. 1996. *Linear models for the prediction of animal breeding values*. CAB Int.
 Kennedy B.W. and D. Sorensen. 1987. Properties of mixed model methods for prediction of genetic merit. In: *Proc. Of the second. Int. Confer. On Quant. Genet.* Eds. Weir, B.S. et al. Sinauer Publ. Sunderland, Mass.
 Quaas, R.L., 1976 Computing the diagonal elements and inverse of a large numerator Relationship Matrix. *Biometrics* 32:949-953
 Quaas, R.L., 1988 Additive genetic model with groups and relationships. *J.Dairy Sci.* 71:1338-1345.
 Sorensen, D.A., and B.W. Kennedy. 1984. Estimation of genetic variance from selected and unselected populations. *J. Anim. Sci.* 59:1213.
 Van der Werf, J.H.J., and I.J.M. de Boer. 1990. Estimation of additive genetic variance when base populations are selected. *J. Animal Sci.* 68: 3124-3132.

Exercise 1

Given are pedigree and records on the following 8 animals

animal	sires	dam	weight
1	0	0	10
2	0	0	9
3	0	0	8
4	0	0	7
5	1	2	9
6	1	2	10
7	3	4	8
8	5	6	11

Set up the inverse of the direct inverse additive genetic relationship matrix
 Also construct the different matrices to build the NRM (A-matrix). Make P, T, D and L.

- Calculate BLUP EBV's for these animals (only account for the mean).
- Determine the average EBV for each generation
- Check that for animal 8, the EBV is estimated as:

$$\hat{a}_8 = \frac{1}{2}(\hat{a}_{sire} + \hat{a}_{dam}) + b(y - \hat{\mu} - \frac{1}{2}\hat{a}_{sire} - \frac{1}{2}\hat{a}_{dam})$$

- Work out the value for b from the MME
- Work out the value of b theoretically

Exercise 2: Relationship matrix accounting for selection

Consider the following data set, with 9 animals having yearling weight records

year	animal	sires	dam	weight
1	1	0	0	270
1	2	0	0	300
1	3	0	0	320
1	4	0	0	310
2	5	3	4	300
2	6	3	4	307.5
2	7	3	4	315
3	8	7	0	303
3	9	7	0	307;

Estimate breeding values (use h²=0.25) for the 9 animals.

For this, edit data in STBLUP.XLS, and run this program in excel.

Compare then the following models:

- no consideration of pedigree or generation effects (put all fixed effect codes to 1
- without consideration of pedigree, but account for selection differences by grouping animals by generation (as fixed effects)
- use only male pedigree (dams unknown), as well as groups
- use all known pedigree as well as groups