Exercises Day 1

Part 1: EBV accuracy

**Exercise 1.1 Effect of using relatives’ information on selection accuracy**

For single trait prediction of breeding value, write out the P-Matrix and the G-vector for the following cases:

1. One own performance record
2. Information known on own performance and performance of sire (1 record each)
3. Information known on own performance and an EBV of the sire (accuracy = 0.9)
4. Information known on own performance and an EBV of the sire (acc = 0.9) and dam (acc =0.5)
5. Information on own performance, EBV of the sire (acc = 0.9) and the mean of 25 half sibs
6. Information on own performance, mean of 25 half sibs and mean of 50 progeny

*Use the symbols VA for additive genetic variance and VP for phenotypic variance.*

*Note that for a single trait prediction you can also substitute these by VP = 1 and VA = h2*

Answers:

1) **One own performance record**

P-matrix: var(X1) = σ2P is the phenotypic variance

G-vector: Cov(X1, A) = Cov(A+E, A) = Cov(A, A) + cov(E,A) = σ2A + 0 = σ2A

b = P-1G = (σ2P )-1 σ2A = h2

accurcay2 = Var(EBV)/Var(BV) = b’Pb/σ2A = b’G/σ2A = h2. σ2A/ σ2A = h2. Hence, accuracy = equal to h.



2) **Information known on own performance and performance of sire (1 record each)**

Information sources: X1 = own performance

 X2 = performance of sire

* variance and covariance of information sources:

var(X1) = σ2P is the phenotypic variance

var(X2) = σ2P is the phenotypic variance

Cov(X1,X2) = Cov(A+E, As +Es)

 = Cov(A, As) + cov(A,Es) + cov(E,As) +cov(E,Es)

 = ½σ2A+ 0 + 0 + 0.

* covariance between information sources and the animal’s breeding value

Cov(X1, A) = σ2A

Cov(X2, A) = Cov(As +Es , A) = Cov(As, A) + cov(Es,A) = ½σ2A + 0 .

such that index weights obtained by regression = covariance/variance:



1. **Information known on own performance and an EBV of the sire (accuracy = 0.9)**
* variance and covariance of information sources:

P(1,1) = var(X1) = σ2P is the phenotypic variance

P(2,2) = var(X2) = r2σ2A where r is accuracy and r2 is reliability of the sire’s EBV

P(2,1) = P(1,2) = Cov(X1,X2) = ½ r2. σ2A

* covariance between information sources and the animal’s breeding value

G(1,1) = Cov(X1, A) = σ2A

G(2,1) = Cov(X2, A) = ½ r2 σ2A



1. **Information known on own performance and an EBV of the sire (acc=0.9) and dam (acc=0.5)**

P is a 3 by 3 matrix, G a 3 by 1 vector with elements the same as in the previous (3), and additional elements

P(3,3) = var(X3) = r2σ2A where r is accuracy and r2 is reliability of the dam’s EBV

P(1,3) = P(3,1) = Cov(X1,X3) = ½ r2dam. σ2A

P(2,3) = P(3,2) = Cov(X2,X3) = 0

G(3,1) = Cov(X3, A) = ½ r2dam σ2A



Note that the EBV of the sire and the dam have the same weights, eventhough the sire is more accurate. The weights are the same because for each EBV, equally much is transferred to the progeny. The sire EBV will still have a lot more influence on the ranking of the progeny, because var(EBV) is higher for the sires than for the dam, Think of an extreme case where the dam accurcay is close to zero. We would still have the same weighting, but since all dams have more or less the same EBV (var(EBV)=0 because acc=0) they would not affect the EBVs of their progny very much.

If there is no own perfrance record (we can mimic this by turning the h2 to zero) then the weights for both parent EBVs will be 0.5. Wih an own performance record on the progeny, some of the parents EBV is also in the own performance record of the animals.so the weight is less than 0.5 to avoid double counting.

1. **Information on own performance, EBV of the sire (acc = 0.9) and the mean of 25 half sibs**

P is a 3 by 3 matrix, G a 3 by 1 vector with elements the same as in the (3), and additional elements

P(3,3) = var(X3) = tHS σ2P + ((1-tHS)/n) σ2P where tHS = 1/4 h2 is the intra class correlation

P(1,3) = Cov(X1,X3) = tHS

P(2,3) =P(3,2)= Cov(X2,X3) = ½\*r2\*σ2A

G(3,1)= Cov(X3, A) = tHS



1. **Information on own performance, mean of 25 half sibs and mean of 50 progeny**

P is a 3 by 3 matrix, G a 3 by 1 vector with elements the same as in the (3), and additional elements

var(X3) = tHS σ2P + ((1-tHS)/n) σ2P where tHS = 1/4 h2 is the intra class correlation

Cov(X1,X3) = tHS

Cov(X2,X3) = ½h2

And Cov(X3, A) = tHS

P is a 3 by 3 matrix, G a 3 by 1 vector with elements the same as in the (3), and additional elements

P(3,3) = var(X3) = tHS σ2P + ((1-tHS)/n) σ2P where tHS = 1/4 h2 is the intra class correlation

P(1,3) = Cov(X1,X3) = tHS

P(2,3) = Cov(X2,X3) = ½h2

G(3,1)= Cov(X3, A) = tHS



**Exercise 1.2 Correlations between relatives’ EBV**

Consider the following cases, and for each case, calculate the correlation between EBVs on full sibs and half sibs.

You can use STEBVaccuracy.XLS (using the STSELIND tab) and use the P-matrix and the index weights to work out this problem.

1. Information known on EBV of the sire (acc=0.9) and dam (acc=0.5)
2. One own performance record
3. Information known on own performance and an EBV of the sire (acc=0.9) and dam (acc=0.5)
4. Information on own performance, EBV of the sire (acc=0.9), dam (acc=0.5) and 50 progeny

Answer:

Cov(EBV1,EBV2) = cov(bX1, bX2) = b’cov(X1,X2)b = b’P\*b’

and the correlation between these is b’P\*b’/b’Pb. The denominator is the variance of the EBV, we assume both sibs have an EBV based on the same information, so σEBV1= σEBV1 and σEBV1 . σEBV1 = var(EBV) = b’Pb.

P\* is like the usual P matrix, but rather than variances and covariances of all information sources, it contains all covariances between the information sources of the 2 sibs. Many of these could be in common. For example, for 2 full sibs, the information on the sire, the dam, and the half sib mean will be the same.

1. an index is based on just sire and dam, P\* = P and the correlation between the EBV of 2 FS is 1.
2. If an index is based on an own record, the P=1 whereas P\* = tFS , the eight b = h2 and the accuracy is h (square root of heritability) so the correlation is h4.tFS/h4 = tFS. You can check this with the STEBVaccuracy.xls (STSELIND tab)
3. For 3 and 4 please check the P\* matrices in STEBVaccuracy.xls (STSELIND tab)

**Exercise 1.3 Pseudo BLUP**

In real life, parents have not just their own records, but they have an estimated breeding value with certain accuracy, using BLUP. This accuracy is based on ancestor information, their own siblings and perhaps their offspring. Also, BLUP corrects for the records of the mates of sires, when their progeny are evaluated. The amount of ancestral information can be derived from a given population structure.

The STEBVaccuracy.XLS (using the PseudoBLUP tab) program does a full Pseudo-BLUP prediction of EBV accuracy, given genetic parameters, and a certain population structure (Half-sib and full-sib family size). You can follow the steps in more detail in BLUP\_EBV.XLS.

For 2 cases:

h2=0.25, c2=0.15 FS family size =3, HS family size = 12 and

h2=0.10, c2=0.0  FS family size =4, HS family size = 80

1. explain the negative weight on EBV of mates;
2. vary h2 and look at weights on parental EBV
3. compare Pseudo BLUP accuracy with that of a that simple selection index approach assuming just a single record for parents (STEBVaccuracy.XLS; using the STSELIND tab)

Answers

1. BLUP corrects for assortative mating. So if the dams of the half sib group are above average, it would be unfair to us a high HS mean for the prediction of breeding value before correcting that for the contributions of te dams to these half sibs.
2. 1st case: h2=0.25, c2=0.15 FS family size =3, HS family size = 12

  

2nd case: h2=0.10, c2=0.0  FS family size =4, HS family size = 80

  

SO with low heritability: Accuracy is lower, in spite of larger families

The own performance is a lot less valuable, and weight is lower

Correlations between EBV of sibs is a lot higher (also due to larger families)

With Simple STSE LIND:

1. 1st case: h2=0.25, c2=0.15 FS family size =3, HS family size = 12 



2nd case: h2=0.10, c2=0.0  FS family size =4, HS family size = 80



With simple SELIND, the accuracies are a bit lower than pseudo BLUP, as in BLUP more ancestral information is accounted for

Exercises Day 1

Part 2: Selection response

**Exercise 2.1**

Consider an ongoing nucleus breeding program for trait with heritability equal to 0.25, a phenotypic standard deviation equal to 20, and a mean of 100. The trait is expressed in females only (sex-limited trait) when they are 1 year of age. Each round 10 sires are mated to 5 dams each, and each dam has 4 male and 4 female progeny.

1. Calculate the accuracy of BLUP EBV of young male and female selection candidates (use STEBVaccuracy.XLS without Bulmer)
2. Predict the genetic superiority of selected males and females, assuming selection of the top 10% of males and the top 50% of females. Use the SELINT.XLS (correlated EBV tab) to account for correlated EBVs.
3. Predict the expected phenotypic performance of the progeny generation (generation 1).

Solutions

First use *STEBVaccuracy.xls* to work out accuracy

Males and female will have different accuracy as females have an own performance info

We need to select ‘females only’ and in result we see it uses only half the number in each sex,

we get for females:



And for males



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |
|  |  | prop | selInt | acc | sigma\_A |  | Sup |
| superiority males |  | 10% | 1.755 | 0.559 | 10 |  | 9.81 |
| superiority females |  | 50% | 0.798 | 0.664 | 10 |  | 5.30 |
|  |  |  |  |  |  |  |  |
| response per generation |  |  |  |  |  |  | 7.55 |
|  |  |  |  |  |  |  |  |
| mean of next generation |  |  |  |  |  |  | 107.55 |

Note that there are 200 males and females born every generation, and we need only 10 and 50.

So selected proportions could be 5% and 25%. However, they are 10% and 50%, because we assume only half of the males and females born are available for selection. This is only an assumption. It is quite realistic that breeders discard many of their potential selection candidates, e.g. because i) they cull them for ‘other reasons’, ii) thy only measure half of all possible candidates.

**Exercise 2.2 Selection across age groups**

Consider selection of females in dairy cattle for milk yield (heritability = 0.3, genetic SD = 550 kg). Three age groups of females are available, with numbers, ages, trait means, and accuracies of selection as given below. Our aim is to select a total of 50 females for breeding.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Age when | Number | Age group | Accuracy |
| Age group | progeny born (yr) | available candidates | traitmean (kg) | Ofselection |
| 1 | 2 | 500 | 12,000 | 0.55 |
| 2 | 3 | 300 | 11,700 | 0.68 |
| 3 | 4 | 200 | 11,400 | 0.72 |

Predict the genetic superiority and generation interval for the following two situations:

1. The 50 females are selected by selecting the best 10, 20, and 20 from age groups 1, 2, and 3
2. The 50 females are selected by truncation selection across age groups (use truncsel.xls)
3. We the predefined age structure
4. 
5. We can use trncsel.xls

input is



and results are



Note that in both cases we assume the male superiority is zero and male generation interval is 2. This might not be realistic but the difference =between the scenarios (comapring female age structure) is the same.**Exercise 2.3**

The response per year can be given by the formula of Rendel and Robertson: Ryr = S/L.

For simplicity assume equal selection intensities in males and females).

We can maximize selection response by truncation section across age classes, (assuming the selection criteria are comparable across age classes). Truncation selection across age classes maximizes the mean of the selected parents.

Show algebraically that maximizing the mean of selected parents results in maximizing the response per year (hence, optimizes selection across age classes)

Answer:

We have

Mean of offspring generation ḡo

Mean of parents from age class i (with age Li) ḡo – LiRyr

Mean of parents selected from age class i ḡo – LiRyr + Si

 Where Si = superiority of animals selected within age class i

Proportion of parents originating from age class i pi with (Σpi = 1)

The mean of all selected parents from n different age classes is $\sum\_{i=1}^{n}p\_{i}\overbar{(g\_{o}}-L\_{i}R\_{yr}+S\_{i})$.

This is the same as ḡo because the mean of all selected parents is the same as the mean of the current generation

So maximizing the mean of all selected parents is the same as maximizing the mean of the progeny.

We use the same proportions selected (set of optimal pi values) in the following

We can rearrange ḡo = $\sum\_{i=1}^{n}p\_{i}(\overbar{g\_{o}}-L\_{i}R\_{yr}+S\_{i})$

So ḡo = $\sum\_{i=1}^{n}p\_{i}\overbar{g\_{o}}$ -$ \sum\_{i=1}^{n}p\_{i}L\_{i}R\_{yr}$ +$ \sum\_{i=1}^{n}p\_{i}S\_{i}$

 = $\overbar{g\_{o}}$ -$ \sum\_{i=1}^{n}p\_{i}L\_{i}R\_{yr}$ +$ \sum\_{i=1}^{n}p\_{i}S\_{i}$ because (Σpi = 1)

Therefore $ $ -$ \sum\_{i=1}^{n}p\_{i}L\_{i}R\_{yr}$ +$ \sum\_{i=1}^{n}p\_{i}S\_{i}$=0

SO that $\sum\_{i=1}^{n}p\_{i}S\_{i}$ = $\sum\_{i=1}^{n}p\_{i}L\_{i}R\_{yr}$

And $R\_{yr} =\sum\_{i=1}^{n}p\_{i}S\_{i}/$ $\sum\_{i=1}^{n}p\_{i}L\_{i}$ so the same optimal proportions are used in optimizing Ryr

More intuitive:

**Would we pick a young bull that has an EBV of +11 or an old bull that has a +12 EBV?**

On one hand we want to select young bulls to keep a low generation interval

But the predicted progeny mean is the same as the mean of the selected bulls, so we should always select the bulls with the highest EBV, the old bull in this instance.

But remember that BLUP EBVs are comparable across age classes, they correct for genetic trend. On average, we would expect the young bulls to be better. But their EBVs will be less spread out than those of the old bulls, as SDEBV = accuracy x SDBV and older bulls will have more accurate EBVs. So the best old bulls will be competitive with the best young bulls, as illustrated in the pictre (red = young bulls distribution of EBV, blue is old bulls). We would more likely have more young bulls in the top 100 bull list if 10 the genetic trend was higher and 2) if young bulls have more accurate EBVs, e.g. because we use genomic predictions.

**Exercise 2.4**

Using the Excel Spreadsheet ‘Genetic\_gain.xls’, evaluate the impact of the percentage of cows inseminated by young bulls and progeny group size on genetic gain. Find the optimal combination of these two variables in order to maximize genetic gain

NO ANSWER GIVEN

**Exercises Day 1**

Part 3: Change of Variance

**Exercise 3.1 Response to selection with the Bulmer effect**

Consider the problem of Exercise 2.1.

1. Calculate the genetic variance and heritability among the individuals produced in generation 1
2. Calculate accuracy of BLUP EBV of male and female selection candidates from generation 1. For the accuracy of sires and dams, use the accuracies of EBV you obtained in Exercise 2.1 (i.e. from an unselected population).
3. Predict the mean, genetic variance, and heritability of individuals produced in generation 2
4. Derive the asymptotic genetic variance, accuracy, heritability and response to selection for this breeding program.
5. Compare results from 4. to those you get from using the program SelAction.

USE THE BULMER.XLS use the BLUP tab

The initial response of 7.55 is now reduced to5.60



**Exercise 3.2 Pseudo BLUP EBV with the Bulmer effect**

1. Use STEBVaccuracy.xls and compare accuracy and index weights of females of Exercise 2.1 with and without Bulmer. Do the same for males
2. Change the % selected and see how this changes index weights and accuracy of EBV.
3. Calculate the accuracy of a parental average EBV with and without selection (Bulmer correction) for different proportions selected. Look also at the variance of parental EBV.

PA = ½ EBVsire + ½ EBVdam

Var(PA) = ¼Var(EBVsire) + ¼Var(EBVdam)

Accuracy = sqrt(Var(PA)/VA)

1. Evaluate reduction in accuracy due to Bulmer with different heritabilities

**Matrix calculations using Excel**

You can do some basic matrix calculations with MS Excel.

First put in the values of your matrices

To multiply two matrices:

 - select an area of the size of the resulting matrix

 - type: =**MMULT(**

 - select the area of the first matrix

 - type a comma (,)

 - select area of the second matrix

 - type a close bracket )

 - on Windows press: Ctrl\_Shift\_Enter on Mac press: Cmnd\_Shift\_Enter

To add or subtract a matrix (vector):

 - select an area of the size of the resulting matrix

 - type: =(

 - select the area of the first matrix

 - type a **+** or **-**

 - select area of the second matrix

 - type a close bracket )

 - press: Ctrl\_Shift\_Enter

To invert a matrix:

 - select an area of the size of the resulting matrix

 - type: =**MINVERSE(**

 - select the area of the first matrix

 - type a close bracket )

 - press: Ctrl\_Shift\_Enter

To transpose a matrix (vector):

 - select an area of the size of the resulting matrix

 - type: =**TRANSPOSE(**

 - select the area of the first matrix

 - type a close bracket )

 - press: Ctrl\_Shift\_Enter

A more specialized matrix calculation program is MATLAB. It contains many more matrix functions and mathematical function than excel. MATLAB allows you to make and run programs, draw graphs, and run simulation). A MATLAB student version is very well suitable for animal breeding problems and quite easy to use.