

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 V_A for additive genetic variance and V_P for phenotypic variance.

Note that for a single trait prediction you can also substitute these by $V_P = 1$ and $V_A = h^2$

Answers:

1) One own performance record

P-matrix: $\text{var}(X_1) = \sigma_P^2$ is the phenotypic variance

G-vector: $\text{Cov}(X_1, A) = \text{Cov}(A+E, A) = \text{Cov}(A, A) + \text{cov}(E, A) = \sigma_A^2 + 0 = \sigma_A^2$

$b = P^{-1}G = (\sigma_P^2)^{-1} \sigma_A^2 = h^2$

accuracy² = $b'G/\sigma_A^2 = h^2 \cdot \sigma_A^2/\sigma_A^2 = h^2$. Hence, accuracy = equal to h.

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

Information sources: X_1 = own performance
 X_2 = performance of sire

- variance and covariance of information sources:

$$\text{var} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = P = \begin{pmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) \\ \text{cov}(X_2, X_1) & \text{var}(X_2) \end{pmatrix}$$

$\text{var}(X_1) = \sigma_P^2$ is the phenotypic variance

$\text{var}(X_2) = \sigma_P^2$ is the phenotypic variance

$\text{Cov}(X_1, X_2) = \text{Cov}(A+E, A_s + E_s)$
 $= \text{Cov}(A, A_s) + \text{cov}(A, E_s) + \text{cov}(E, A_s) + \text{cov}(E, E_s)$
 $= \frac{1}{2}\sigma_A^2 + 0 + 0 + 0.$

- covariance between information sources and the animal's breeding value

$$\text{cov} \left(\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, A \right) = G = \begin{pmatrix} \text{cov}(X_1, A) \\ \text{cov}(X_2, A) \end{pmatrix}$$

$\text{Cov}(X_1, A) = \sigma_A^2$

$\text{Cov}(X_2, A) = \text{Cov}(A_s + E_s, A) = \text{Cov}(A_s, A) + \text{cov}(E_s, A) = \frac{1}{2}\sigma_A^2 + 0.$

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

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = P^{-1}G = \begin{pmatrix} \sigma_P^2 & \frac{1}{2}\sigma_A^2 \\ \frac{1}{2}\sigma_A^2 & \sigma_P^2 \end{pmatrix}^{-1} \begin{pmatrix} \sigma_A^2 \\ \frac{1}{2}\sigma_A^2 \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{2}h^2 \\ \frac{1}{2}h^2 & 1 \end{pmatrix}^{-1} \begin{pmatrix} h^2 \\ \frac{1}{2}h^2 \end{pmatrix}$$

3 Information known on own performance and an EBV of the sire (accuracy = 0.9)

- variance and covariance of information sources:

$$\text{var} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix} = P = \begin{pmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) \\ \text{cov}(X_2, X_1) & \text{var}(X_2) \end{pmatrix}$$

$P(1,1) = \text{var}(X_1) = \sigma_P^2$ is the phenotypic variance

$P(2,2) = \text{var}(X_2) = r^2 \sigma_A^2$ where r is accuracy and r^2 is reliability of the sire's EBV

$P(2,1) = P(1,2) = \text{Cov}(X_1, X_2) = \frac{1}{2} r^2 \sigma_A^2$

- covariance between information sources and the animal's breeding value

$$\text{cov} \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}, A = G = \begin{pmatrix} \text{cov}(X_1, A) \\ \text{cov}(X_2, A) \end{pmatrix}$$

$G(1,1) = \text{Cov}(X_1, A) = \sigma_A^2$

$G(2,1) = \text{Cov}(X_2, A) = \frac{1}{2} r^2 \sigma_A^2$

4) 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

$P(3,3) = \text{var}(X_3) = r^2 \sigma_A^2$ where r is accuracy and r^2 is reliability of the dam's EBV

$P(1,3) = P(3,1) = \text{Cov}(X_1, X_3) = \frac{1}{2} r_{\text{dam}}^2 \sigma_A^2$

$P(2,3) = P(3,2) = \text{Cov}(X_2, X_3) = 0$

$G(3,1) = \text{Cov}(X_3, A) = \frac{1}{2} r_{\text{dam}}^2 \sigma_A^2$

5) 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

$P(3,3) = \text{var}(X_3) = t_{\text{HS}} \sigma_P^2 + ((1-t_{\text{HS}})/n) \sigma_P^2$ where $t_{\text{HS}} = 1/4 h^2$ is the intra class correlation

$P(1,3) = \text{Cov}(X_1, X_3) = t_{\text{HS}}$

$P(2,3) = P(3,2) = \text{Cov}(X_2, X_3) = \frac{1}{2} h^2$

$G(3,1) = \text{Cov}(X_3, A) = t_{\text{HS}}$

6) 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

$\text{var}(X_3) = t_{\text{HS}} \sigma_P^2 + ((1-t_{\text{HS}})/n) \sigma_P^2$ where $t_{\text{HS}} = 1/4 h^2$ is the intra class correlation

$\text{Cov}(X_1, X_3) = t_{\text{HS}}$

$\text{Cov}(X_2, X_3) = \frac{1}{2} h^2$

And $\text{Cov}(X_3, A) = t_{\text{HS}}$

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

$P(3,3) = \text{var}(X_3) = t_{\text{HS}} \sigma_P^2 + ((1-t_{\text{HS}})/n) \sigma_P^2$ where $t_{\text{HS}} = 1/4 h^2$ is the intra class correlation

$P(1,3) = \text{Cov}(X_1, X_3) = t_{\text{HS}}$

$P(2,3) = \text{Cov}(X_2, X_3) = \frac{1}{2} h^2$

$G(3,1) = \text{Cov}(X_3, A) = t_{\text{HS}}$

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:

$$\text{Cov}(\text{EBV}_1, \text{EBV}_2) = \text{cov}(bX_1, bX_2) = b' \text{cov}(X_1, X_2) b = b' P^* b'$$

and the correlation between these is $b' P^* b' / b' P b$. The denominator is the variance of the EBV, we assume both sibs have an EBV based on the same information, so $\sigma_{\text{EBV}_1} = \sigma_{\text{EBV}_2}$ and $\sigma_{\text{EBV}_1} \cdot \sigma_{\text{EBV}_2} = \text{var}(\text{EBV}) = b' P b$.

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 $P^* = t_{FS}$, the weight $b = h^2$ and the accuracy is h (square root of heritability) so the correlation is $h^4 \cdot t_{FS} / h^4 = t_{FS}$. 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:

$h^2=0.25, c^2=0.15$ FS family size =3, HS family size = 12 and

$h^2=0.10, c^2=0.0$ FS family size =4, HS family size = 80

- 1) explain the negative weight on EBV of mates;
- 2) vary h^2 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: $h^2=0.25, c^2=0.15$ FS family size =3, HS family size = 12

Information used	Nr.Records	Index weight		value of variate:
nr of own records	1	0.195	1 own	18%
nr. of dams per sire	4	0.380	EBV dam	7%
nr of progeny per dam	3	0.300	EBV sire	4%
nr. of progeny recorded on	0	0.044	2 FS	1%
	both sexes	0.160	9 HS	2%
		-0.080	4 Mates	0%
		-	-	-

Accuracy of EBV **0.6241**

correlation EBV FS **0.567**

correlation EBV HS **0.340**

2nd case: $h^2=0.10, c^2=0.0$ FS family size =4, HS family size = 80

Information used	Nr.Records	Index weight		value of variate:
nr of own records	1	0.075	1 own	9%
nr. of dams per sire	20	0.430	EBV dam	9%
nr of progeny per dam	4	0.191	EBV sire	1%
nr. of progeny recorded on	0	0.066	3 FS	2%
	both sexes	0.478	76 HS	11%
		-0.239	20 Mates	0%
		-	-	-

Accuracy of EBV **0.5737**

correlation EBV FS **0.799**

correlation EBV HS **0.561**

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 L ND:

1) 1st case: $h^2=0.25$, $c^2=0.15$ FS family size =3, HS family size = 12

Information used	Nr.Records
nr of own records	1
nr. of records on dam	1
nr of records on sire	1
nr of full sib records	2
nr. of half sib records (excl. full sibs)	9
nr. of progeny	0

Index weight		value of variate:
0.202	own	22%
0.092	dam	5%
0.070	sire	3%
0.060	FS	1%
0.174	HS	3%
-	-	-

Accuracy of EBV	0.5976
correlation EBV FS	0.766
correlation EBV HS	0.338

2nd case: $h^2=0.10$, $c^2=0.0$ FS family size =4, HS family size = 80

Information used	Nr.Records
nr of own records	1
nr. of records on dam	1
nr of records on sire	1
nr of full sib records	3
nr. of half sib records (excl. full sibs)	76
nr. of progeny	0

Index weight		value of variate:
0.080	own	12%
0.042	dam	3%
0.016	sire	0%
0.082	FS	4%
0.519	HS	18%
-	-	-

Accuracy of EBV	0.5290
correlation EBV FS	0.931
correlation EBV HS	0.606

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

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 group	Age when progeny born (yr)	Number available candidates	Age group trait mean (kg)	Accuracy Of selection
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:

- The 50 females are selected by selecting the best 10, 20, and 20 from age groups 1, 2, and 3
- The 50 females are selected by truncation selection across age groups (use trncsel.xls)

1) See spreadsheet Response_across_age_groups_2.2_Answers.xls

Selection across 3 age groups for trait with genetic SD =

550

Items in green cells are given

Age group	Age when progeny born (yr)	Number available	Proportion in age group	Age group trait mean	Proportion selected	Intensity	Accuracy of selection	Genetic Superiority	Mean of selected	$p_i w_i$	$p_i w_i \bar{g}_i$	St.dev. Selection Criter (genetic SD)	
i			w_i	μ_i	p_i	i_i	r_i	S_i	\bar{g}_i				
1	2	500	0.5	12,000	0.020	2.421	0.55	732.3	12,732	0.010		302.5	
2	3	300	0.3	11,700	0.067	1.940	0.68	725.4	12,425	0.020		374	
3	4	200	0.2	11,400	0.100	1.755	0.72	695.0	12,095	0.020		396	
Overall prop. selected = P =										0.05			
Pooled genetic superiority =										714.62		Mean generation int=	3.2 years
Mean selected group =										12354.6		11640	
												714.6	

2) We can use trncsel.xls

input is

Candidates Age Class	Nr Candidates in age group	mean	SD
1	500	12000	302.50
2	300	11700	374.00
3	200	11400	396.00

and results are

Proportion Selected	Selection Intensity	Nr Selected	Mean of selected	age of selected
8.19%	1.841	40.94	12556.84	1.00
2.69%	2.273	8.07	12549.96	2.00
0.50%	2.708	0.99	12472.38	3.00
			0.00	4.00
			0.00	5.00
		50.00	12554.06	1.20
			mean of selected	Age of selected (GenInt)

Exercise 2.3

The response per year can be given by the formula of Rendel and Robertson: $R_{yr} = S/L$.

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

We can maximize selection response by truncation selection 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	\bar{g}_o
Mean of parents from age class i (with age L_i)	$\bar{g}_o - L_i R_{yr}$
Mean of parents selected from age class i	$\bar{g}_o - L_i R_{yr} + S_i$

Where S_i = superiority of animals selected within age class i

Proportion of parents originating from age class i	p_i	with ($\sum p_i = 1$)
--	-------	-------------------------

The mean of all selected parents from n different age classes is $\sum_{i=1}^n p_i (\bar{g}_o - L_i R_{yr} + S_i)$.

This is the same as \bar{g}_o 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 p_i values) in the following

We can rearrange $\bar{g}_o = \sum_{i=1}^n p_i (\bar{g}_o - L_i R_{yr} + S_i)$

$$\begin{aligned} \text{So } \bar{g}_o &= \sum_{i=1}^n p_i \bar{g}_o - \sum_{i=1}^n p_i L_i R_{yr} + \sum_{i=1}^n p_i S_i \\ &= \bar{g}_o - \sum_{i=1}^n p_i L_i R_{yr} + \sum_{i=1}^n p_i S_i \quad \text{because } (\sum p_i = 1) \end{aligned}$$

$$\text{Therefore} \quad - \sum_{i=1}^n p_i L_i R_{yr} + \sum_{i=1}^n p_i S_i = 0$$

$$\text{SO that} \quad \sum_{i=1}^n p_i S_i = \sum_{i=1}^n p_i L_i R_{yr}$$

$$\text{And} \quad R_{yr} = \sum_{i=1}^n p_i S_i / \sum_{i=1}^n p_i L_i \quad \text{so the same optimal proportions are used in optimizing } R_{yr}$$

More intuitive:

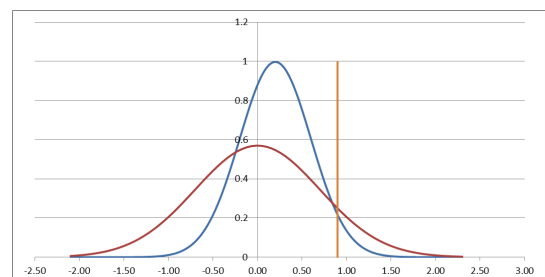
Would we pick a young bull that has an EBV of +11 or an old bull that has a +12 EBV?

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

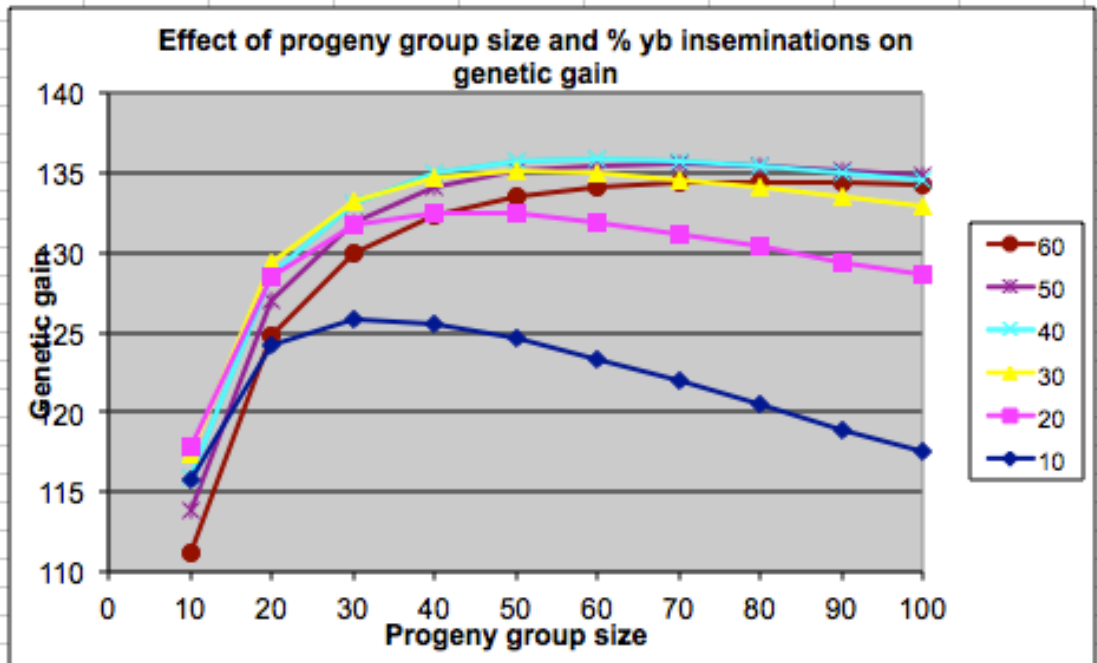
classes, they correct for genetic trend. On average, we would expect the young bulls to be better. But their EBVs will be less spread out than those of the old bulls, as $SD_{EBV} = \text{accuracy} \times SD_{BV}$ and older bulls will have more accurate EBVs. So the best old bulls will be competitive with the best young bulls, as illustrated in the picture (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 1) 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

		% young bull inseminations					
		10	20	30	40	50	60
Progeny group size	10	115.77	117.81	117.41	115.96	113.85	111.23
	20	124.13	128.49	129.29	128.61	127.02	124.77
	30	125.82	131.66	133.25	133.08	131.88	129.92
	40	125.59	132.52	134.7	134.95	134.05	132.34
	50	124.64	132.45	135.1	135.69	135.05	133.56
	60	123.34	131.92	135	135.86	135.45	134.15
	70	121.9	131.2	134.6	135.74	135.53	134.4
	80	120.47	130.34	134.09	135.45	135.42	134.43
	90	118.92	129.42	133.48	135.04	135.19	134.34
	100	117.57	128.55	132.85	134.6	134.88	134.17



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.

- Calculate the genetic variance and heritability among the individuals produced in generation 1

	accuracy	% selected	intensity	superiority
males	0.5592	0.05	1.972	11.03
females	0.6639	0.25	1.238	8.22
		progeny mean =		109.62
		Response =		9.62

- 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).

	Truncation point	Height of Ordinate	Selection intensity	Var reduction factor	Genetic variance selected parents
	x	z	i = z/p	k=i(i-x)	$(1-r_{(t)}^2 k_s) \sigma_{g(t)}^2$
males	1.645	0.103	2.063	0.862	73.05
females	0.674	0.318	1.271	0.758	66.57
				Progeny genetic variance = 84.91	
				herit = 0.221	

- Predict the mean, genetic variance, and heritability of individuals produced in generation 2

Gener 2	Unselected accuracy	Genetic variance	True accuracy	Superiority	k	Genetic variance selected parents
			$\sqrt{\frac{(1-(1-r_{(t)}^2) \sigma_{g(t)}^2)}{\sigma_{g(t)}^2}}$			$(1-r_{(t)}^2 k_s) \sigma_{g(t)}^2$
males	0.5592	84.91	0.436	7.931	0.862	70.96
females	0.6639	84.91	0.584	6.665	0.758	62.93
		progeny mean =		116.92	Progeny genetic variance = 83.47	
		Response =		7.30	herit = 0.218	

- Derive the asymptotic genetic variance, accuracy, heritability and response to selection for this breeding program.

Gener 6	Unselected accuracy	Genetic variance	True accuracy	Superiority	k	Genetic variance selected parents
			$\sqrt{\frac{(1-(1-r_{(t)}^2) \sigma_{g(t)}^2)}{\sigma_{g(t)}^2}}$			$(1-r_{(t)}^2 k_s) \sigma_{g(t)}^2$
males	0.5592	83.32	0.418	7.533	0.862	70.74
females	0.6639	83.32	0.573	6.480	0.758	62.54
		progeny mean =		144.98	Progeny genetic variance = 83.32	
		Response =		7.01	herit = 0.217	

- Compare results from 4. to those you get from using the program SelAction.

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.

$$PA = \frac{1}{2} EBVsire + \frac{1}{2} EBVdam$$

$$Var(PA) = \frac{1}{4}Var(EBVsire) + \frac{1}{4}Var(EBVdam)$$

$$Accuracy = \sqrt{Var(PA)/V_A}$$

4. 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.