

Selection Accuracy

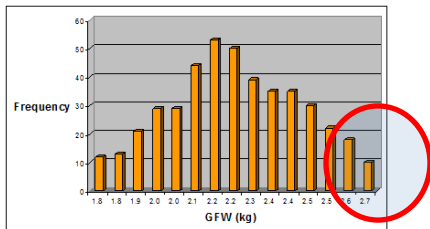
refers to the accuracy of selecting on breeding value

- Estimated Breeding Values (*EBVs*) are estimates of **the True Breeding Values** (TBVs), which cannot be observed directly
- Quality of EBV is measured by **Accuracy**
- Accuracy **is correlation between EBV and TBV** (Acc 0-1)
- The closer the EBV is to the true breeding value the higher the accuracy, the greater the selection response

How much genetic change?

Breeder's Equation

$$\text{Selection Response} = i r_{IA} \sigma_A$$



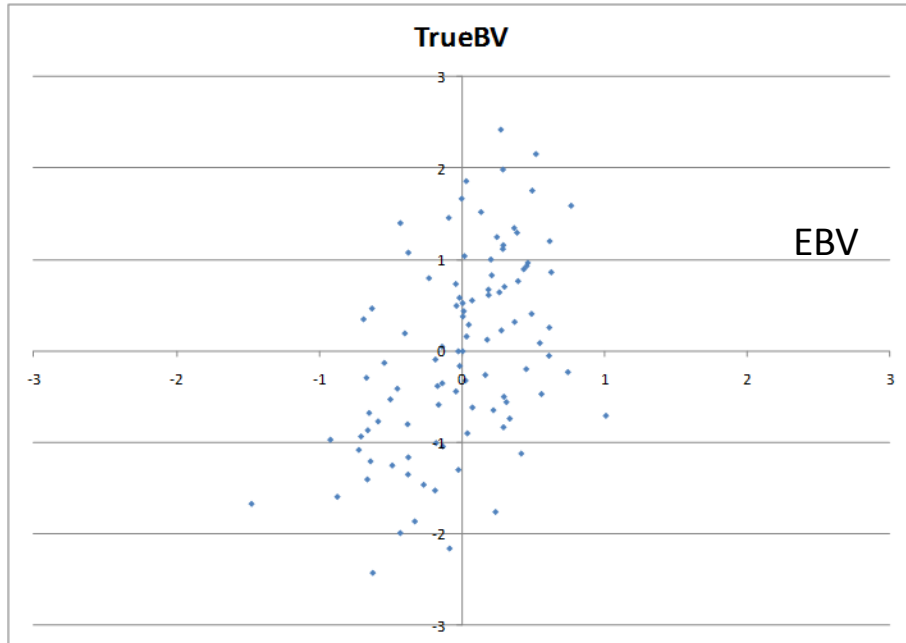
Selection intensity

Selection accuracy

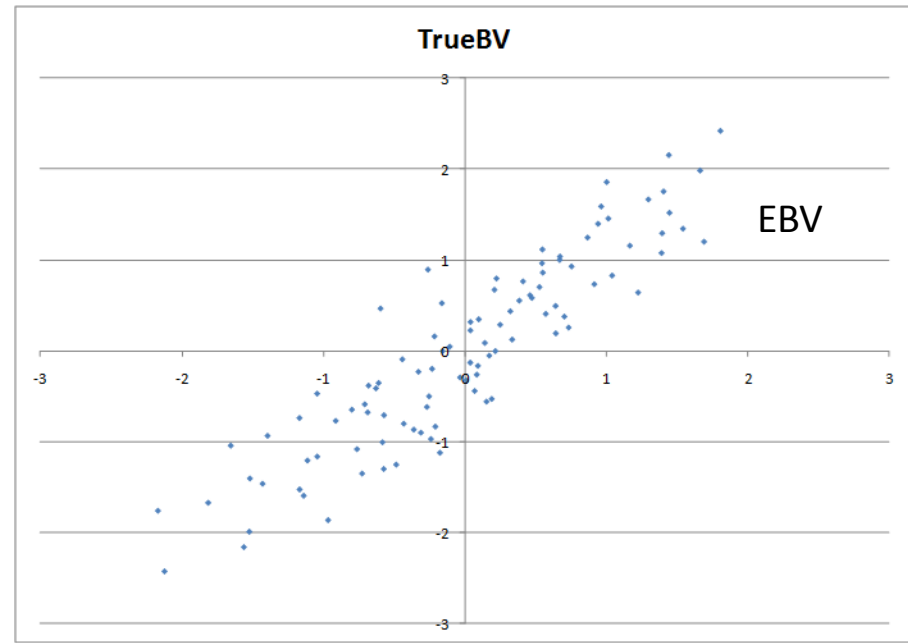
Genetic variation



Accuracy of EBV = correlation with True BV



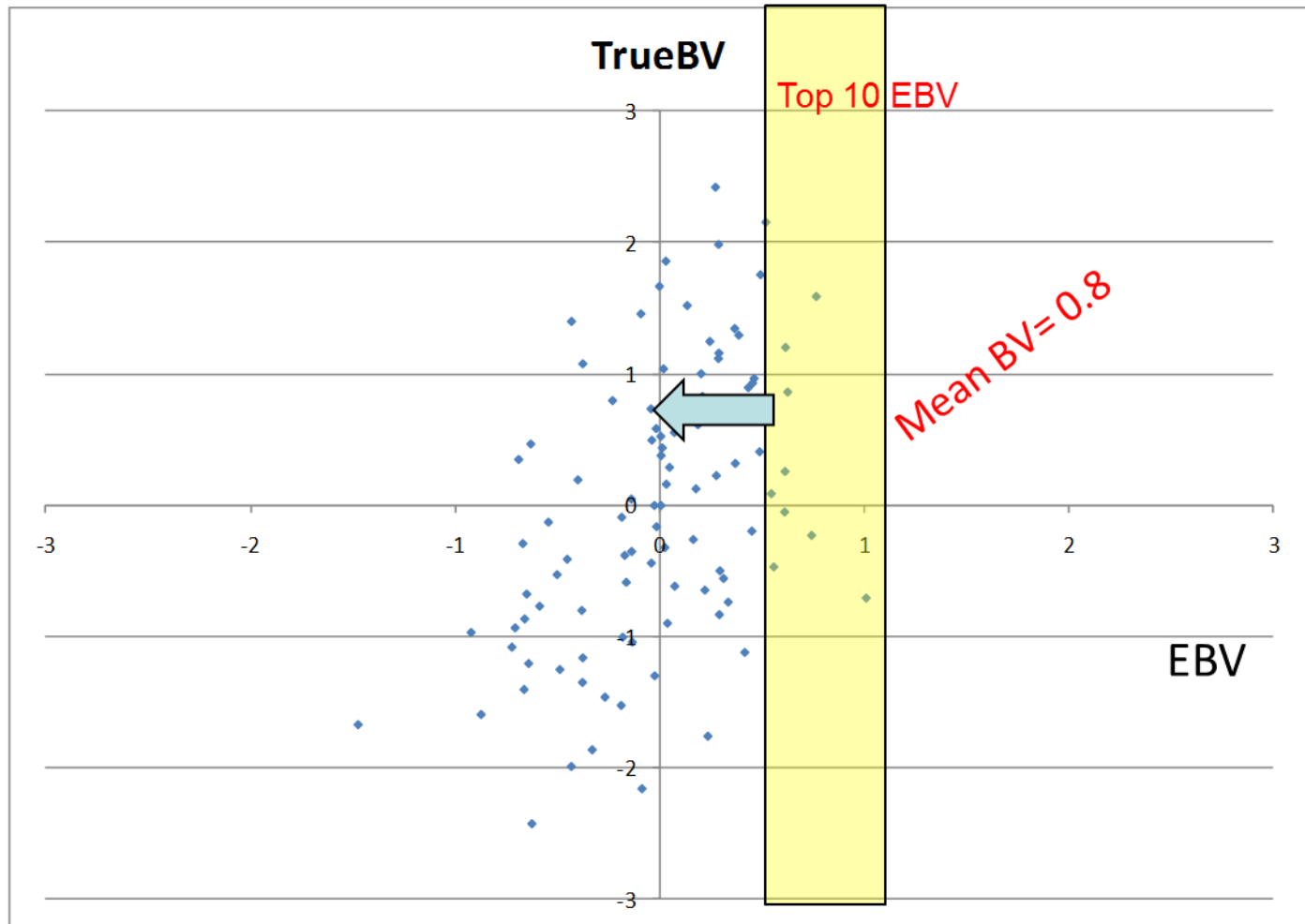
Accuracy = 45%



Accuracy = 90%

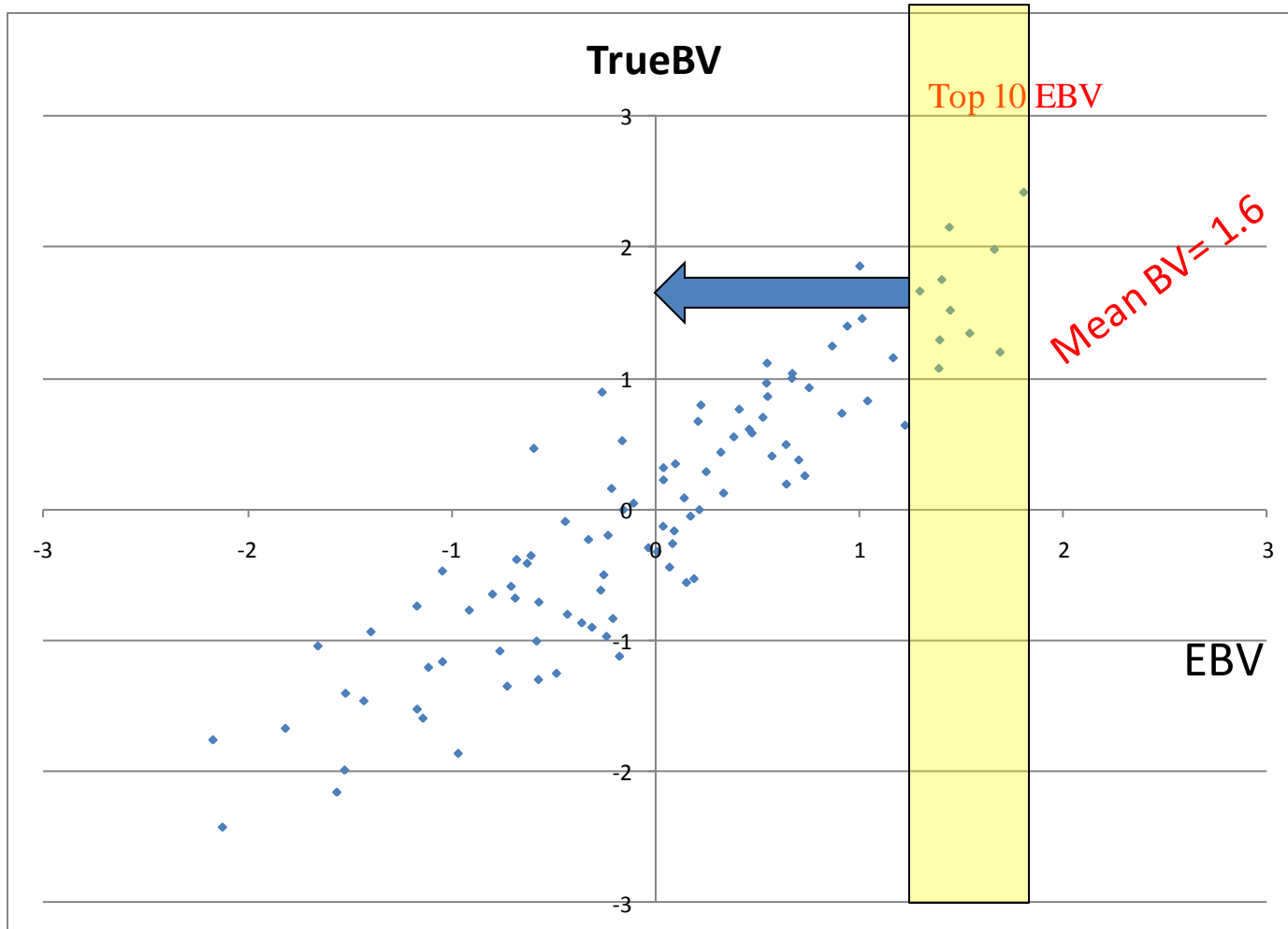
Select on EBV: accuracy related to response

Accuracy = 45%



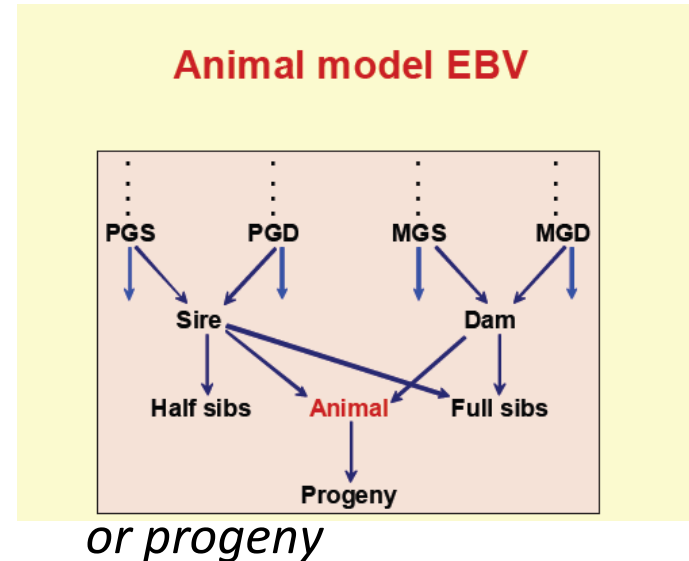
Double accuracy gives double selection response!

Accuracy = 90%



Sources of information to get EBV

- own phenotype, as $EBV_i = h^2P$
- But we can have a number of sources of information, such as information from relatives
 - Performance on dam, sire, siblings,
 - Genomic test
- More information results in a **more accurate EBV**
 - And therefore more selection response



Use Selection Index / BLUP

What part of observed phenotypic differences in parents is passed on to progeny?

Parent has a phenotypic deviation (e.g. +10)

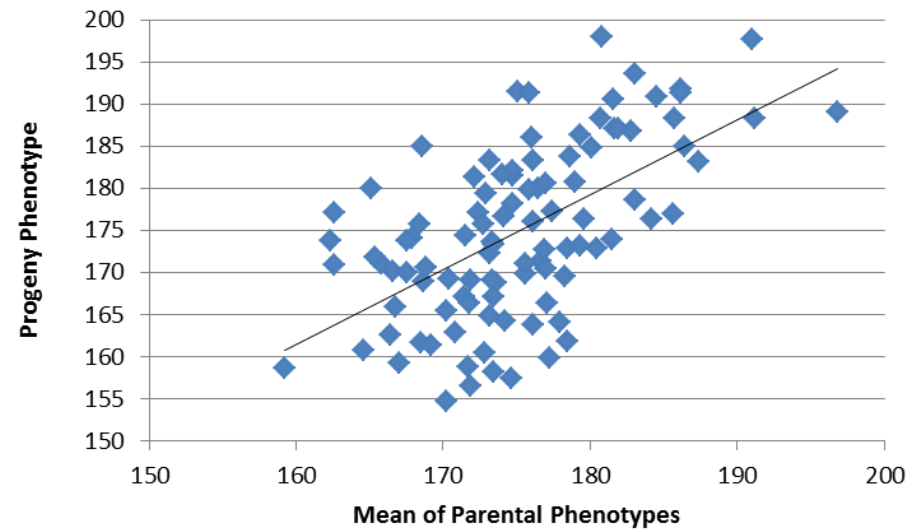
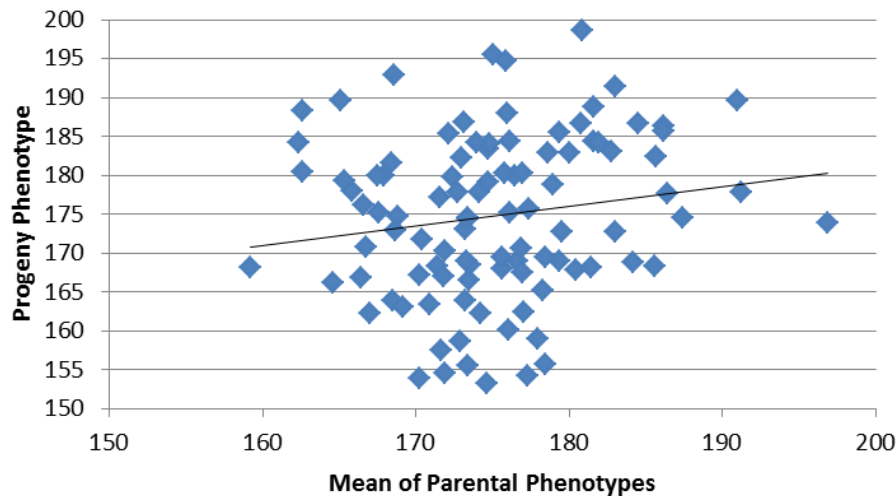
Progeny gets $0.5 * \text{heritability} * \text{phenotypic deviation}$

heritability = 5%

progeny gets + 0.25

heritability = 75%

progeny gets + 3.75



Breeding Value

What part of differences in parents is passed on to progeny?

Parent has *a phenotypic deviation e.g. +10*

Progeny gets $0.5 * \text{heritability} \times \text{phenotypic deviation}$

heritability = 5%

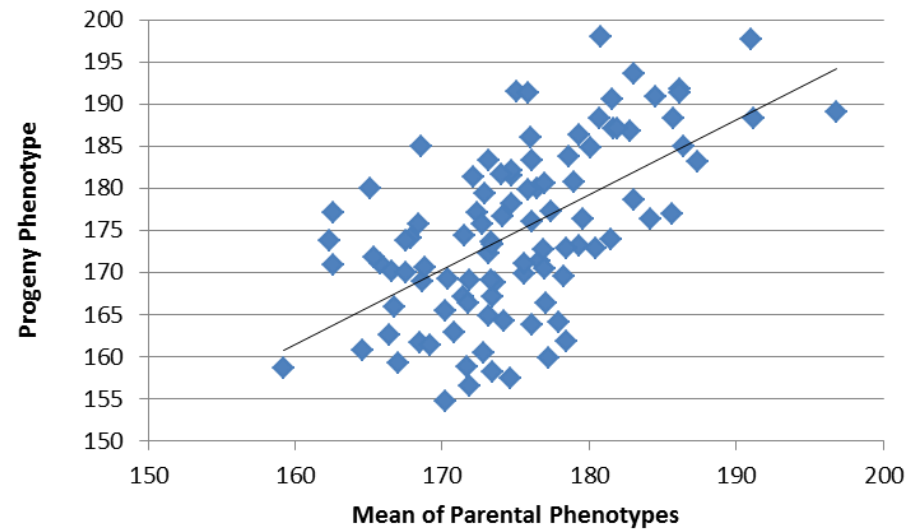
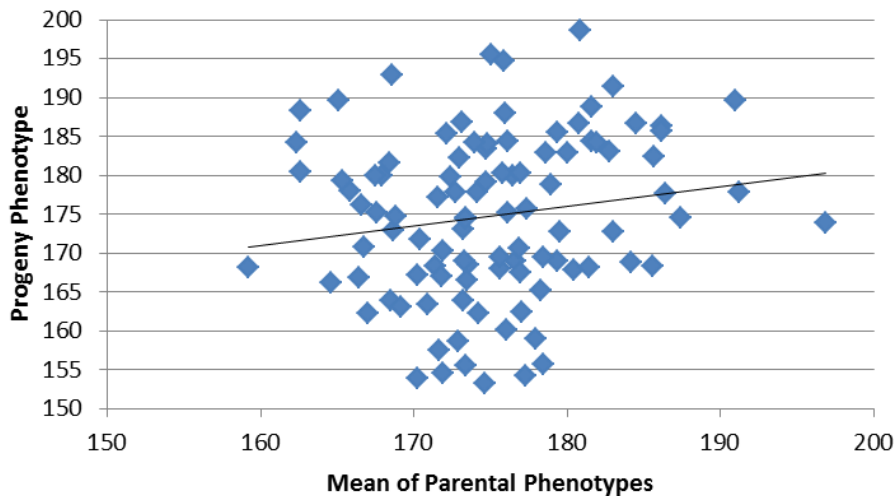
parent EBV = + 0.50

parent EPD = + 0.25

heritability = 75%

parent EBV = + 7.50

parent EPD = + 3.75



Estimating Breeding Values

- Principle
 - how is it estimated?
- Using information from relatives
- Properties
 - Accuracy
 - Variance
 - Prediction Error
- Selection Response
 - select on EBV

How to use information on an animals' own performance?

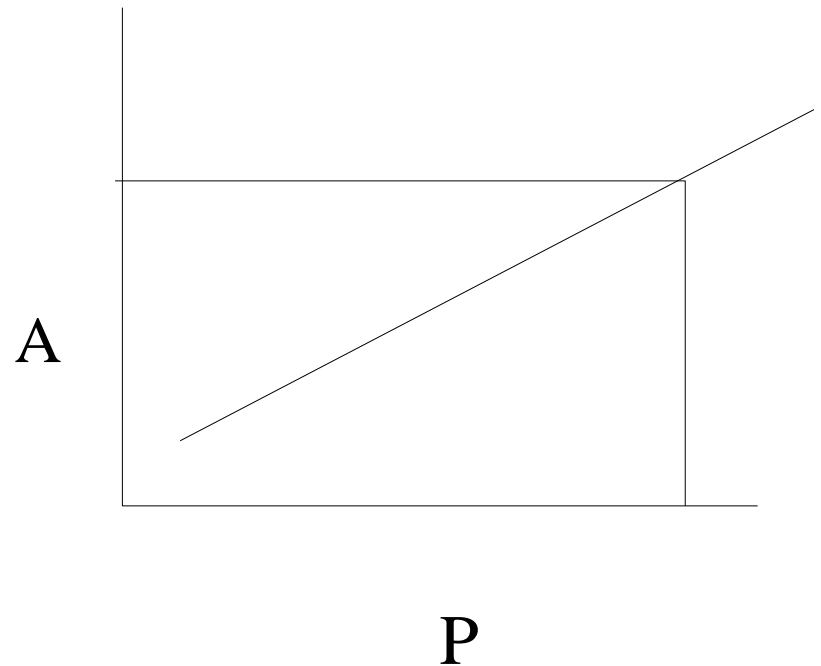
Animal	<u>X1</u> own perform.	<u>EBV</u>
1	3.6	?
2	-8.3	?
3	3.5	?
4	16.7	?
5	-14.4	?
6	6.6	?
7	-1.4	?
8	-6.2	?
9	6.2	?
10	15.0	?

Principle of EBV estimation

Regression of breeding value on phenotype

*How much do we expect the BV
to be better if we know that P
is one unit better*

$$\text{Slope} = \text{cov}(A,P)/\text{var}(P)$$



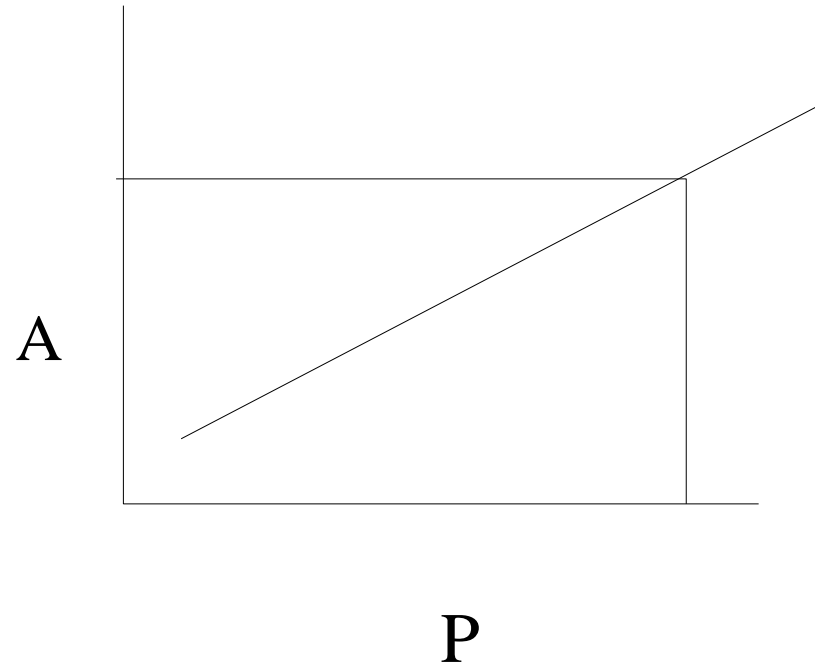
EBV estimation: regression

- If $P = \text{Own Performance}$

$$\text{Slope} = \text{cov}(A, P) / \text{var}(P)$$

$$= V_A / V_P$$

$$= h^2$$



How to use information from relatives?

<u>Animal</u>	<u>sire's perform.</u>	EBV
1	9.2	?
2	3.2	?
3	-15.4	?
4	7.1	?
5	-4.9	?
6	16.1	?
7	-12.2	?
8	8.7	?
9	7.4	?
10	-4.0	?

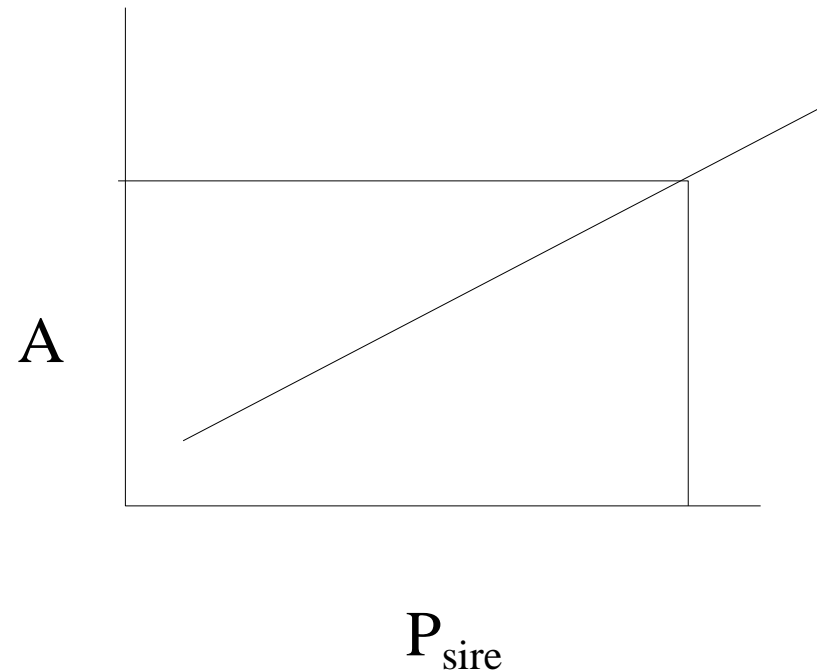
EBV estimation: regression

- If P = Performance of Sire

$$\text{Slope} = \text{cov}(A, P_{\text{sire}}) / \text{var}(P_{\text{sire}})$$

$$= 1/2 V_A / V_P$$

$$= 1/2 h^2$$



Here we need some basic Quantitative Genetic theory

$$P = A + E$$

→ General Model

$$\text{Var}(P) = \text{var}(A) + \text{var}(E) = V_A + V_E$$

no cov. between A and E

$$\text{cov}(A_i, P_i) = \text{cov}(A_i, A_i) + \text{cov}(A_i, E_i) = V_A$$

if A same animal as P

$$\text{cov}(A_i, P_j) = \text{cov}(A_i, A_j) + \text{cov}(A_i, E_j) = a_{ij} V_A$$

a_{ij} = additive genetic
relationship between i and j

$$\text{cov}(P_i, P_j) = a_{ij} V_A$$

as E's are uncorrelated if not same animal

Finding the optimal index weights

Regression of breeding value on a phenotype **EBV = b . P**

$$\text{Regression coefficient} = \frac{\text{cov}(P_i, A)}{\text{var}(P_i)}$$

Examples:

if P = Own Performance

$$b = \frac{V_A}{V_A + V_E} = h^2$$

if P = Sire Performance

$$b = \frac{\frac{1}{2}V_A}{V_A + V_E} = \frac{1}{2}h^2$$

What does this all mean?

Regression of y on x :

- If we know x, what does that tell us about y?

if we know an animal performed well, what does that tell us about his breeding value?

A: Depends on heritability

$$b = h^2$$

if we know an animal's sire performed well, what does that tell us about his breeding value?

A: It carries only half of Sire's genes + it depends on heritability

$$b = \frac{1}{2} h^2$$

if we know an animal's great granddam performed well, what does that tell us about his breeding value?

A: It carries even less of great grandma's genes

$$b = \frac{1}{8} h^2$$

Variance of a mean

$$\text{Var}(\text{mean}) = \text{var}(\text{sum}/n) = (1/n)^2 \text{var}(\text{sum})$$

$$\text{Var}(\text{sum}) = \text{var}(\text{all variances and covariances})$$

$$\text{e.g. } \text{var}(x+y) = \text{var}(x) + \text{var}(y) + 2\text{cov}(x,y)$$

An easy way to workout variances of means

$$\text{Mean} = \text{Sum}/n$$

$$\text{Var}(\text{Mean}) = \text{var}(\text{Sum})/n^2$$

1	t	..	t
t	1		t
.
t	t	..	1

σ_p^2

$$\text{Var}(\text{Sum}) = \{ n \cdot 1 + n(n-1) \cdot t \} \sigma_p^2$$

Var-cov matrix of all elements that make up the mean

$$\text{Var}(\text{Mean}) = \{ [n \cdot 1 + n(n-1)t] / n^2 \} \sigma_p^2$$

$$= \{ [1 + (n-1)t] / n \} \sigma_p^2$$

Variance of a mean

$$\frac{1 + (n-1)t}{n} = \frac{1-t + t + (n-1)t}{n}$$

$$= \frac{(1-t)}{n} + t$$

Unique to each
averaged out

Shared among all
not averaged out

$$V_{\text{mean}} = \left\{ \frac{(1-t)}{n} + t \right\} \sigma_p^2$$

An easy way to workout co-variances with means

$$\text{Mean} = \text{Sum}/n$$

$$\text{CoVar}(X, \text{Mean}) = \text{CoVar}(X, \text{Sum})/n$$

$$\text{CoVar}(X, \text{Mean}) = \{ n.t \} \sigma_p^2 / n$$

$$\text{CoVar}(X, \text{Mean}) = t\sigma_p^2$$

t

.

t

σ_p^2

Vector with covariances of a variable X with all elements that make up the mean, X could be either phenotype, or a breeding value
 $\text{Covar}(X, P_i) = t\sigma_p^2$

Covariance of a variable with the mean is same as covariance of a variable with each element that make up the mean

Some more basic Quantitative Genetics theory

Using information based on a mean (e.g. mean of sibs)

$$\text{Var}(\text{mean } P) = \frac{1+(n-1)t}{n} \sigma_p^2$$

FS

HS

$$t = \frac{1}{2} h^2 + c^2$$

$$t = \frac{1}{4} h^2$$

$$\text{Cov}(\text{mean } P, A) = a_{ij} \sigma_a^2$$

$$a_{ij} = \frac{1}{2} h^2$$

$$a_{ij} = \frac{1}{4} h^2$$

a_{ij} = genetic relationship between members of group and subject (A)

Weight for Mean of

n Full Sibs

n Half Sibs

$$\frac{\text{Cov}}{\text{Var}}$$

=

$$\frac{\frac{1}{2}.n.h^2}{1+(n-1)t_{FS}}$$

$$\frac{n.t}{1+(n-1)t_{HS}}$$

$$= \frac{n}{n+(1-t_{HS})/t_{HS}}$$

Some more Quantitative Genetics theory

Using a progeny mean

var(mean) = common variance + specific/n
equal to all Specific bit is averaged out

Mean performance of n progeny (= HS mean):

$$P_{HS} = \frac{1}{2} A_{sire} + \left\{ \frac{1}{2} A_{dam} + \text{Mendelian Sampling term} + \text{Residual} \right\} / n$$

$$\text{Var}(P_{HS}) = \frac{1}{4} V_A + \left(\frac{3}{4} V_A + V_E \right) / n = t \cdot \sigma_p^2 + (1-t) \sigma_p^2 / n \quad t = \frac{1}{4} h^2$$

Weight for Mean of n Progeny

$$\frac{\text{Cov}}{\text{Var}} = \frac{\frac{1}{2} \cdot n \cdot h^2}{1 + (n-1) \frac{1}{4} h^2} = \frac{2nt}{(1-t) + nt} = \frac{2n}{n + (1-t)/t} = \frac{2n}{n + \lambda}$$

alternatively

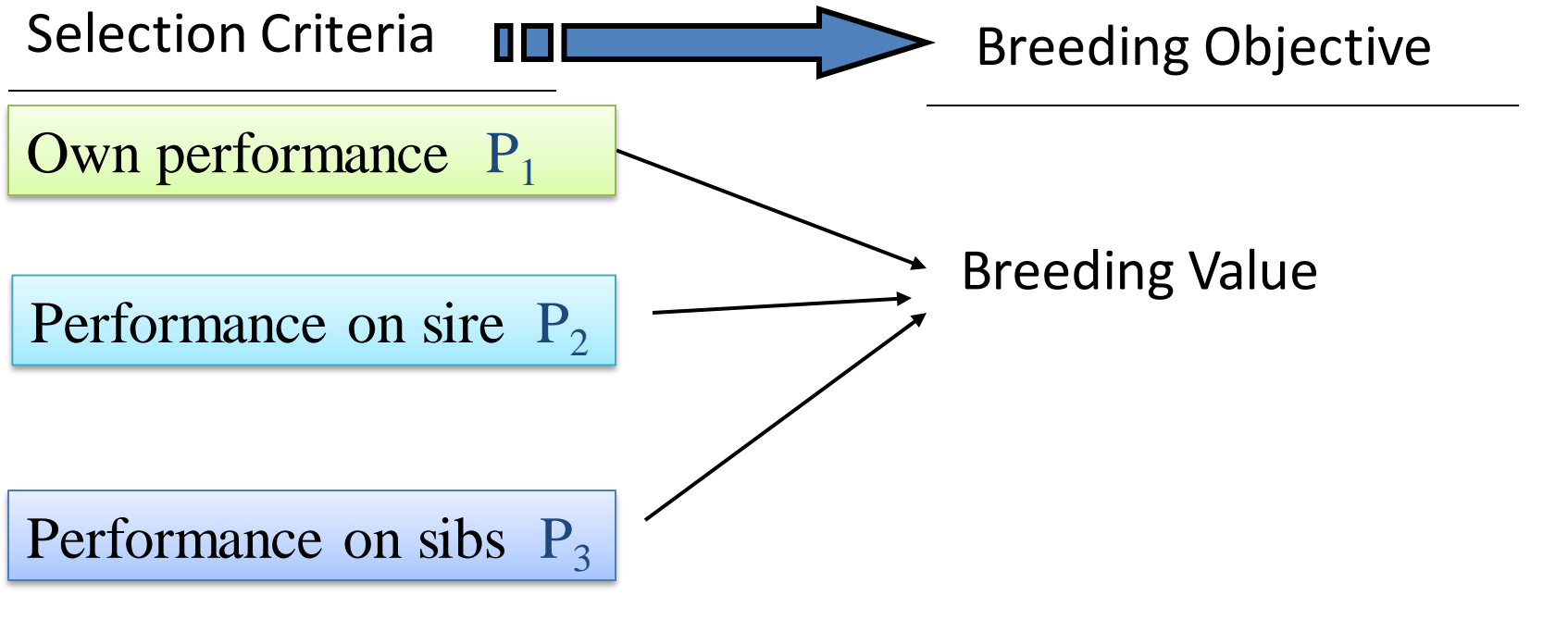
$$\frac{\text{Cov}}{\text{Var}} = \frac{\frac{1}{2} V_A}{\frac{1}{4} V_A + \left(\frac{3}{4} V_A + V_E \right) / n} = \frac{2n}{n + \left(\frac{3}{4} V_A + V_E \right) / \frac{1}{4} V_A} = \frac{2n}{n + \lambda}$$

How to combine information from relatives?

Animal	X1	X2	EBV
	own perform.	sire's perform.	Index
1	3.6	9.2	?
2	-8.3	3.2	?
3	3.5	-15.4	?
4	16.7	7.1	?
5	-14.4	-4.9	?
6	6.6	16.1	?
7	-1.4	-12.2	?
8	-6.2	8.7	?
9	6.2	7.4	?
10	15.0	-4.0	?

Who should we select, animal 10 or animal 6?

Combining different sources of information



Selection Index (multiple regression)

$$EBV = \text{Index} = b_1P_1 + b_2P_2 + b_3P_3 + \dots + b_nP_n$$

Selection index with more information sources

$$\text{multiple regression: Index} = b_1 X_1 + b_2 X_2$$

Need to account for covariance between various sources of information

X = vector with information sources

(phenotypic deviations, or P_i = Sel. Criteria, we call these now $X = [X_1 \ X_2]$)

A = True breeding Value

$$\text{var}(X) = P - \text{matrix} = P = \begin{bmatrix} \text{var}(X_1) & \text{cov}(X_1, X_2) \\ \text{cov}(X_2, X_1) & \text{var}(X_2) \end{bmatrix}$$

$$\text{cov}(X, A) = G - \text{vector} = G = \begin{bmatrix} \text{cov}(X_1, A) \\ \text{cov}(X_2, A) \end{bmatrix}$$

$$\text{Index weights} \begin{bmatrix} b_1 \\ b_2 \end{bmatrix} = P^{-1}G \quad \text{is like "cov}(X,A)/\text{var}_X"$$

This is prediction of a single trait BV, tomorrow MT prediction

Example

X1 = Own Performance
X2 = Sire's Performance

h ²	0.3
varP	100
varA	30
varE	70

$$\begin{array}{cc|c} P & & -1 \\ \hline 100 & 15 & \\ 15 & 100 & \\ \hline G & & \\ \hline 30 & & \\ 15 & & \\ \hline = & & \\ \hline b & & \\ \hline 0.284 & & \\ 0.107 & & \end{array}$$

How do we use such weights?

weights 0.284 0.107

Animal	<u>X1</u>	<u>X2</u>	EBV Index
	own perform.	sire's perform.	
1	3.6	9.2	2.00
2	-8.3	3.2	-2.01
3	3.5	-15.4	-0.67
4	16.7	7.1	5.50
5	-14.4	-4.9	-4.63
6	6.6	16.1	3.60
7	-1.4	-12.2	-1.72
8	-6.2	8.7	-0.82
9	6.2	7.4	2.54
10	15.0	-4.0	3.83

Var-covar of actual x1 and x2	98	17	104
----------------------------------	----	----	-----

This is like the P-matrix but based on realized values, usually we use population parameters (phenotypic (co-)variances)

e.g. use STEBVaccuracy.xls

Single Trait selection index calculation

Parameters	
Heritability	0.25
Repeatability of subsequent records	0.5
c-squared (among full sibs)	0
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)	20
nr. of progeny	20

Run	Index weight
	0.111
	0.048
	0.025
	0.063
	0.179
	0.887

Why use selection index?

- BLUP does it 'automatically'
but
- Can use selection index theory to work out:
 - Accuracy for a given set of information
 - The weighting of various information sources
 - The 'value' of various information sources
 - The weighting on own vs family information
 - Correlation between EBV of sibs

Single Trait selection index calculation

STEBVaccurcay.xls

Parameters	
Heritability	0.25
Repeatability of subsequent records c-squared (among full sibs)	0.5 0

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)	20
nr. of progeny	20

Run

Index weight		value of variate:
0.111	own	3%
0.048	dam	1%
0.025	sire	0%
0.063	FS	1%
0.179	HS	1%
0.887	Prog	22%

Accuracy of EBV **0.817**

correlation EBV FS **0.525**
 correlation EBV HS **0.418**

Selection Index (single trait objective)

First summarize some definitions

$$I = b_1x_1 + b_2x_2 + \dots + b_nx_n \quad \text{var}(X) = P$$

Single trait breeding objective:

$$H = A \quad (\text{breeding value})$$

$\text{Cov}(X, A) = G$ (a vector with ST objective)

Optimal weights are $b = P^{-1}G$

$$\text{Var}(I) = \text{var}(b'X) = b' \text{var}(X)b = b'Pb = \sigma_I^2$$

$$\text{Var}(H) = \sigma_a^2$$

$$\text{Cov}(I, H) = \text{cov}(b'X, A) = b' \text{cov}(X, A) = b'G = b'Pb = \sigma_I^2$$

$$b = P^{-1}G \rightarrow Pb = G$$

Accuracy of selection index (single trait)

r_{IA} = correlation between Index (=EBV) and A

$$= \frac{\text{cov}(I,A)}{\sigma_I \sigma_A} = \frac{\sigma_I^2}{\sigma_I \sigma_A} = \frac{\sigma_I}{\sigma_A} = \sqrt{(b'Pb / \sigma_a^2)}$$

Because $\text{cov}(I,A) = \text{var}(I)$

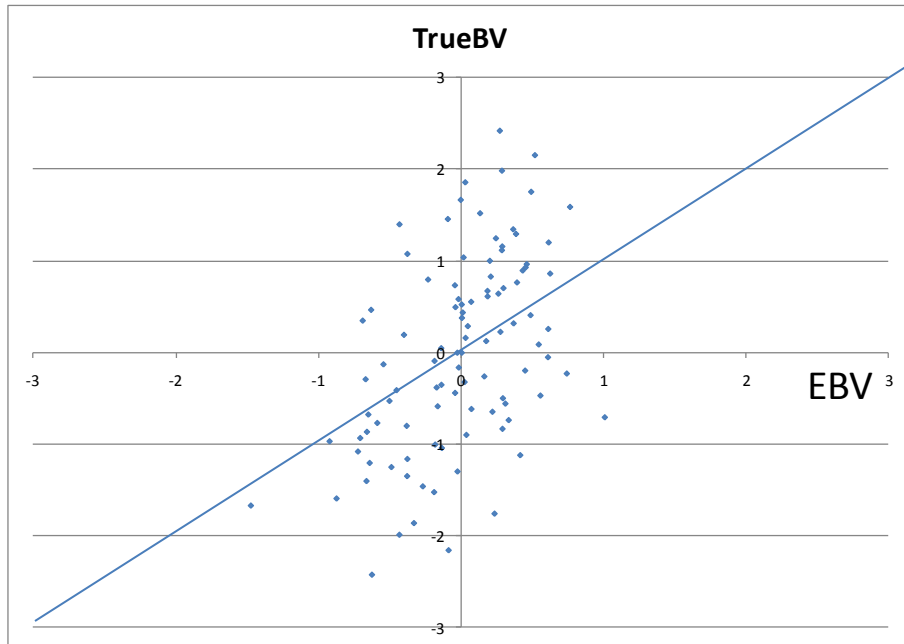
Selection Index = Best Linear Prediction BLP

Index (I) is best estimate of breeding value: $I = E(A|X) = \text{cov}(X,A)/\text{var}(X)$

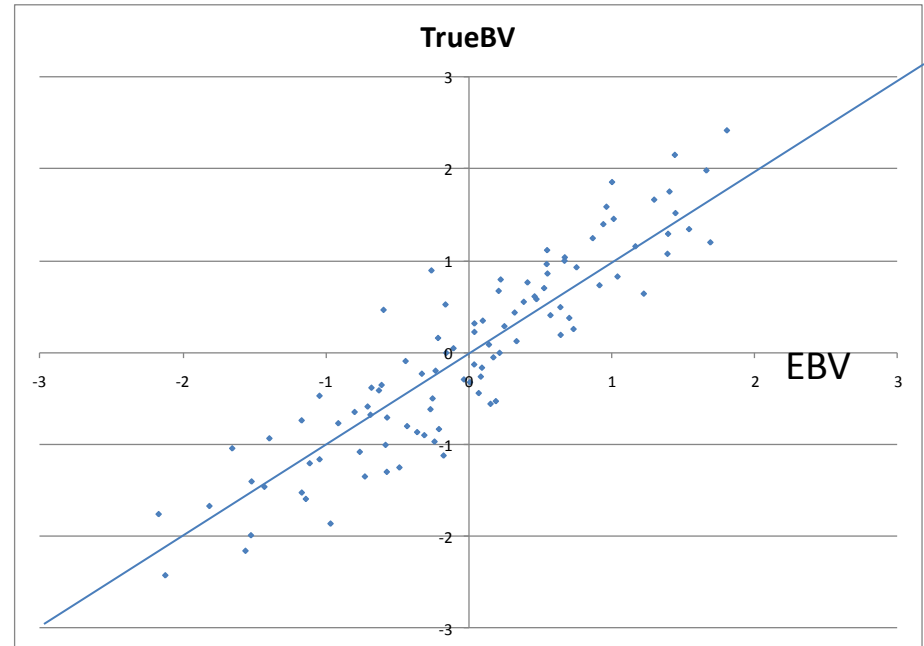
Same as BLUP, but without fixed effects.

$$\text{Var}(I) = \text{var}(EBV) = r_{IA}^2 \cdot \text{var}(BV) = r_{IA}^2 \sigma_a^2 \quad r_{IA}^2 \text{ also known as reliability}$$

accuracy of EBV = correlation with True BV



Accuracy = 45%



Accuracy = 90%

$$\text{Regression A on I} = \text{cov}(I, A) / \sigma^2_I = 1$$

Because $\text{cov}(I, A) = \text{var}(I)$

Suppose

EBV_A +56 $r = 0.50$

EBV_B +56 $r = 0.95$

select A or B?

Examples of accuracies

	$h^2=0.1$	$h^2=0.3$	
own information	0.32	0.55	<i>equal to sqrt h^2</i>
mean of 10 half sib	0.23	0.33	
mean of 1000 half-sibs	0.49	0.50	<i>max is sqrt 0.25 = 0.5</i>
mean of 1000 full-sibs	0.70	0.71	<i>max is sqrt 0.5 = 0.71</i>
mean of 100 progeny	0.85	0.94	<i>max approaches 1.0</i>

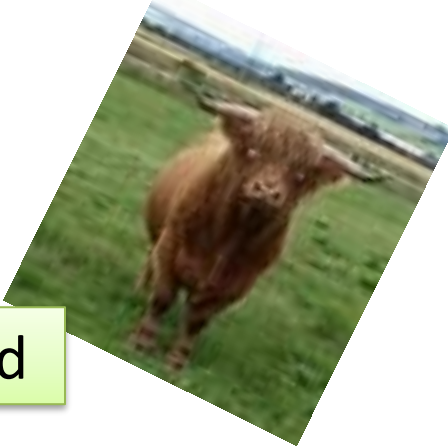
Accuracies of animal increase as they get older → (more info)

The composition of the breeding value

$$A = \frac{1}{2} A_{\text{sire}} + \frac{1}{2} A_{\text{dam}} + MS$$

with no selection

$$V_A = \frac{1}{4} V_A + \frac{1}{4} V_A + \frac{1}{2} V_A$$



Dad



Mum



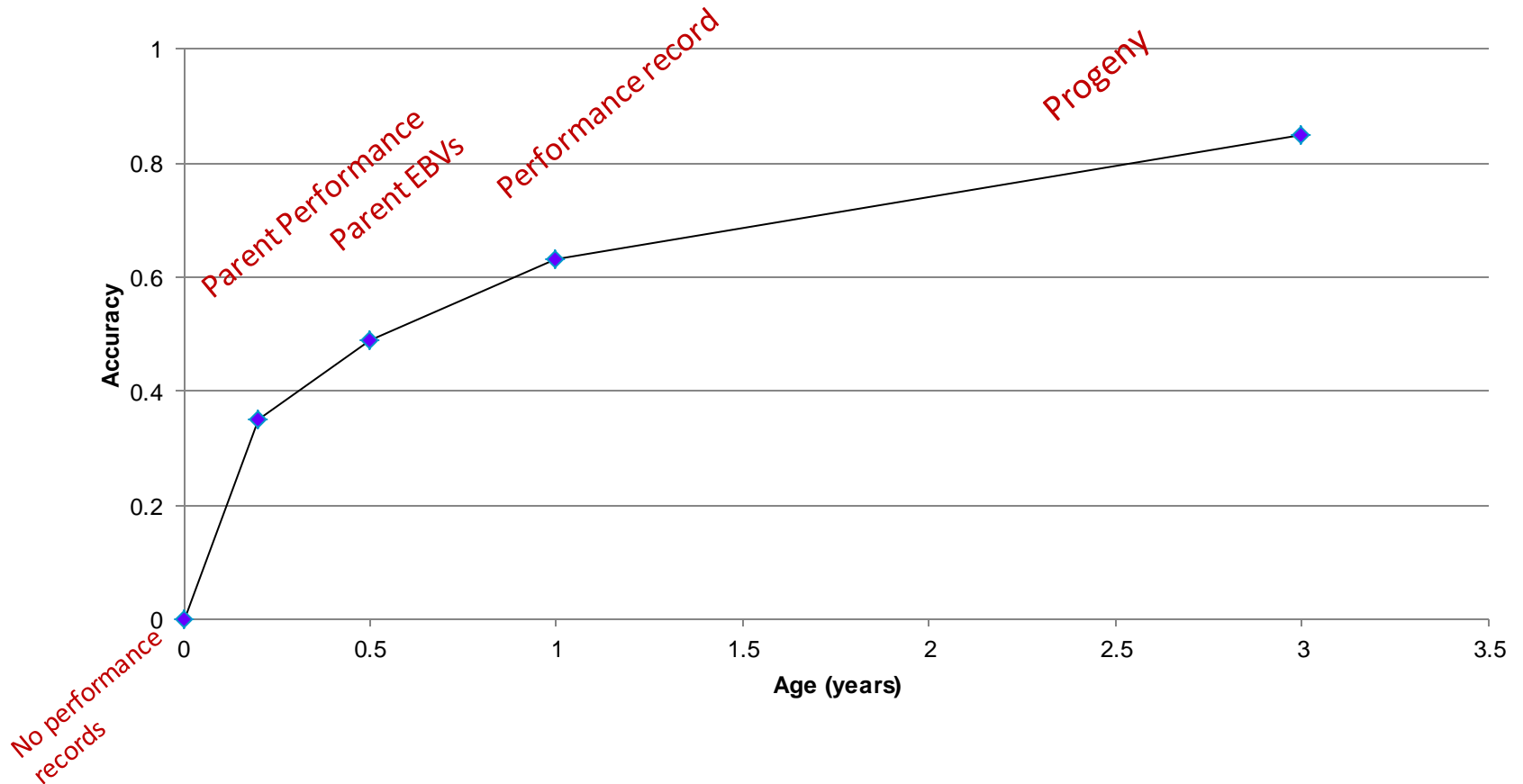
Examples of accuracies

	$h^2=0.1$	$h^2=0.3$	
own information	0.32	0.55	<i>equal to sqrt h^2</i>
mean of 10 half sib	0.23	0.33	
mean of 1000 half-sibs	0.49	0.50	<i>max is sqrt 0.25 = 0.5</i>
mean of 1000 full-sibs	0.70	0.71	<i>max is sqrt 0.5 = 0.71</i>
mean of 100 progeny	0.85	0.94	<i>max approaches 1.0</i>

Accuracies of animal increase as they get older → (more info)

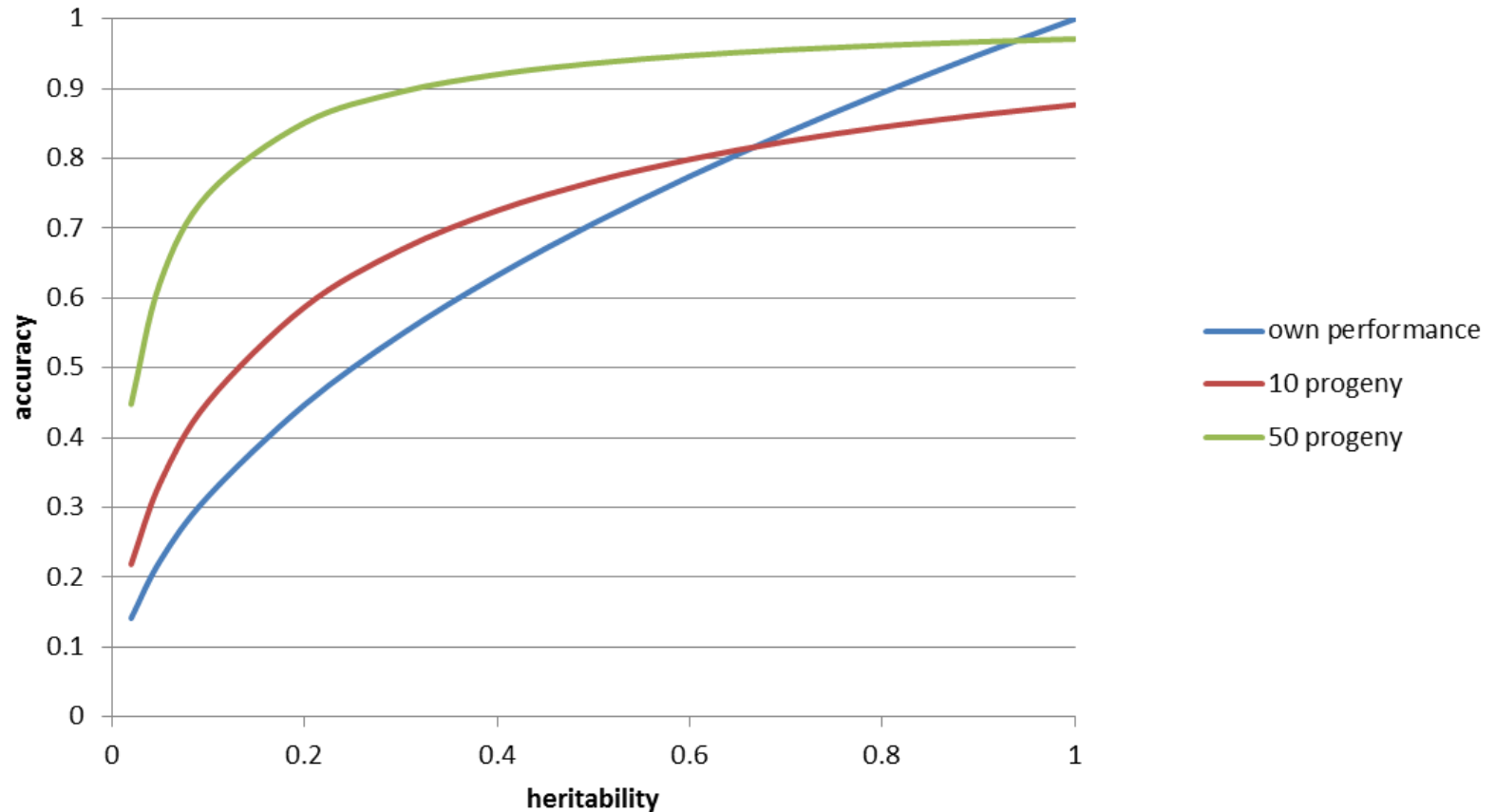
Accuracy of predicting a breeding value

- increases as an animal gets older



Assumed heritability = 25%

Accuracy of own performance vs progeny test



Progeny test gives usually more accurate EBV then own performance

EBV properties:

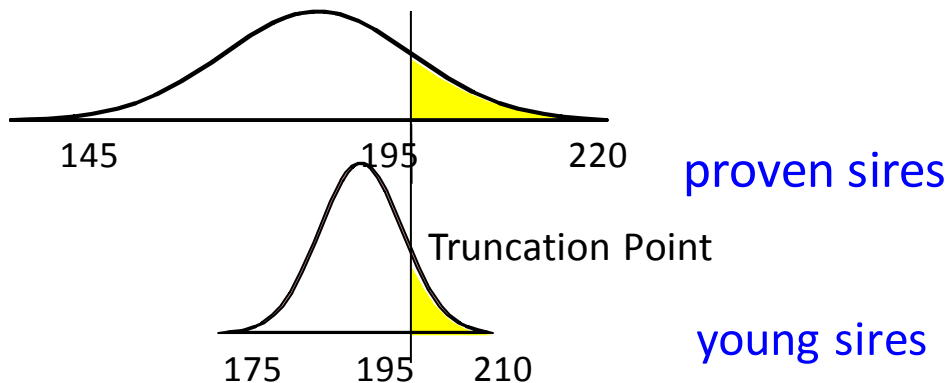
Variance of EBV's - *how much they differ*

- $\text{Var}(\text{EBV}) = r_{IA}^2 \sigma_a^2$
- $\text{SD}(\text{EBV}) = r_{IA} \sigma_A =$ Standard deviation of EBVs

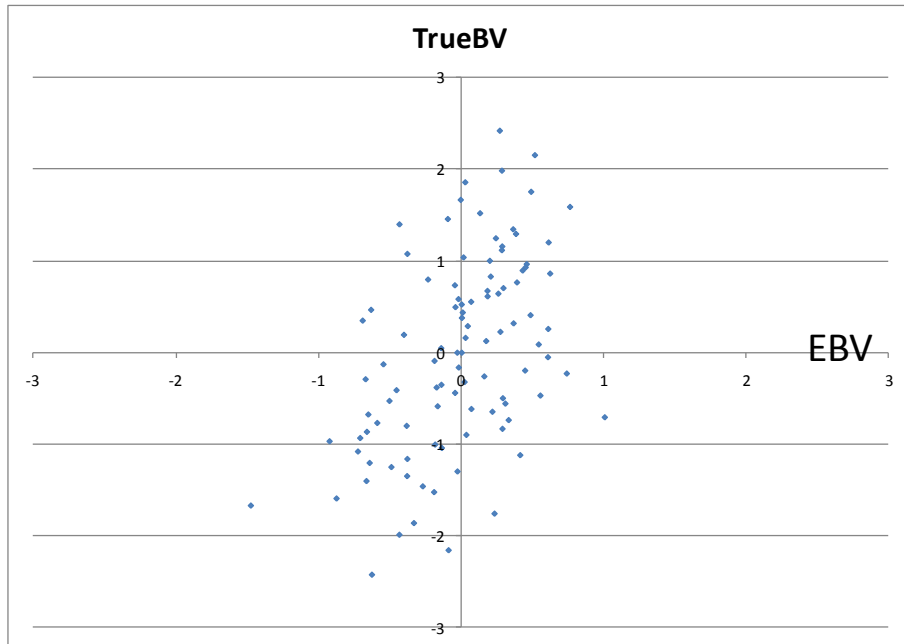
Simply a function of accuracy

Note the extremes of $\text{Var}(\text{EBV})$ if $r_{IA} = 0$ vs $r_{IA} = 1$

Important when selection
across age classes



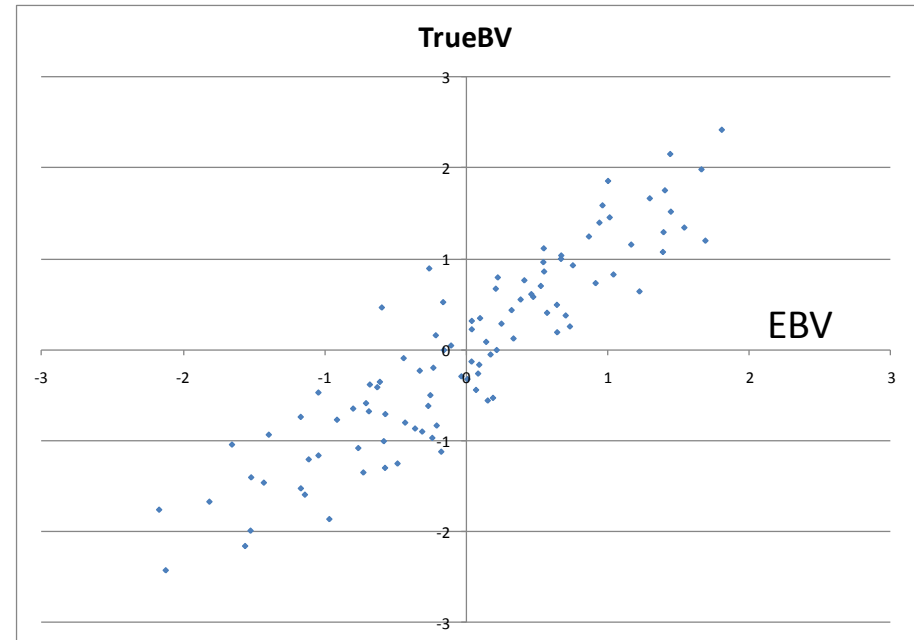
EBV properties



$$\begin{aligned} \text{Accuracy} = r &= 0.45 \\ \text{Reliability} = r^2 &= 0.2025 \end{aligned}$$

$$\text{Var}(\text{EBV}) = \text{low} = r^2 V_A$$

$$\text{Regression BV on EBV} = 1$$



$$\begin{aligned} \text{Accuracy} = r &= 0.90 \\ \text{Reliability} = r^2 &= 0.81 \end{aligned}$$

$$\text{Var}(\text{EBV}) = \text{high} = r^2 V_A$$

$$\text{Regression BV on EBV} = 1$$

EBV properties:

Prediction Error Variance- *how much they still may change*

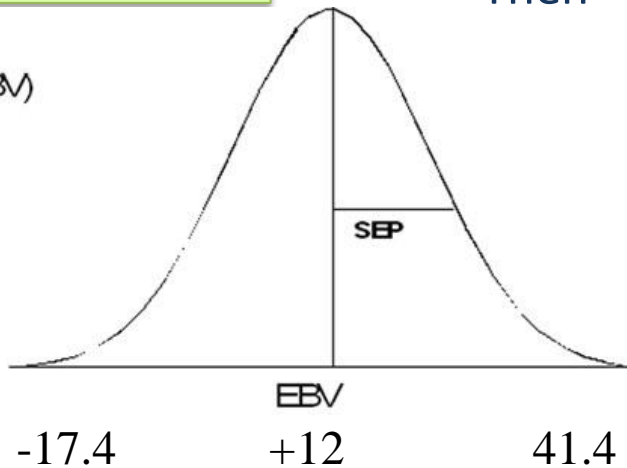
- $PEV = \text{var}(EBV-TBV) = (1-r^2_{IA})V_A$ Prediction Error Variance
- $SEP = \text{sqrt}(PEV) = \sqrt{(1-r^2_{IA})}\sigma_A$ Standard Error of Prediction
- $EBV \pm 1.96.SEP$ Confidence Interval

Probability density
of TBV | EBV

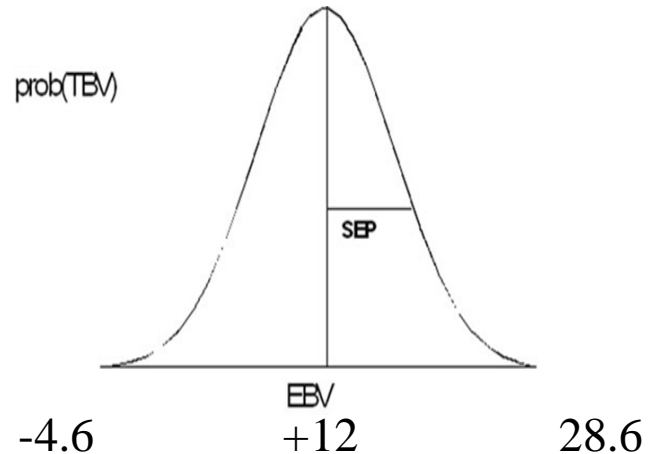
Let $\sigma_a = 19$ Kg, $EBV = +12$ Kg, $h^2 = 0.4$

Then

prob(TBV)



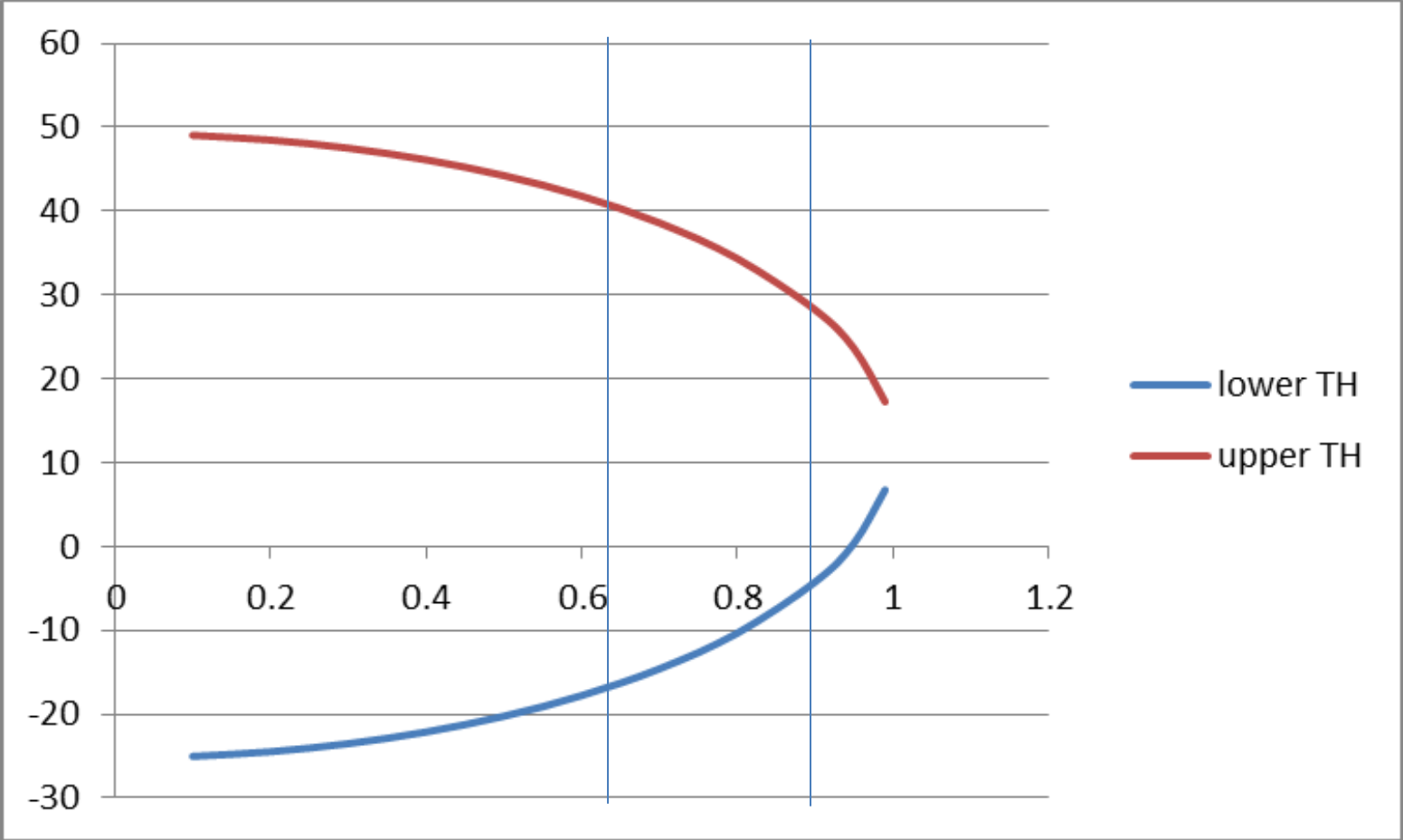
prob(TBV)



Own record: $r_{IA} = \sqrt{h^2} = 0.63$, $SEP = 14.7$

Progeny: $r_{IA} = \sqrt{0.81} = 0.9$, $SEP = 8.3$

Confidence interval of TBV | EBV



Selection on EBV vs inbreeding

X1 = Own Performance
X2 = Sire's Performance

h ²	0.3
varP	100
varA	30
varE	70

$$\begin{array}{cc|c|c|c}
 & & P & -1 & G & & b \\
 \hline
 & & 100 & 15 & 30 & = & 0.284 \\
 & & 15 & 100 & 15 & & 0.107 \\
 \hline
 \end{array}$$

Selection index weights for

heritability	X1 own perform.	X2 sire's perform.
0.1	0.098	0.045
0.3	0.284	0.107
0.5	0.467	0.133

Observe the weight on sire information, depending on h²

Lower h² means more emphasis on family information

Selection on EBV vs inbreeding

Simulation:

Belovsky and Kennedy, 1989

Selection on INDividual performance vs selection on BLUP

Heritability	Increase in genetic merit after 10 years of selection			Inbreeding Increase	
	IND	BLUP	BLUP/IND	IND	BLUP
0.1	0.78	1.41	1.81	0.174	0.383
0.6	5.16	5.31	1.03	0.205	0.293

1. BLUP more response than phenotypic selection (IND)

Difference is larger for smaller heritability.

2. BLUP more inbreeding

Especially for low heritabilities

Co-selection of relatives

Selection on EBV vs inbreeding

Not only h^2 but also information available affects co-selection of relatives

$h^2 = 0.5$

Information used	Nr.Records
nr of own records	0
nr. of records on dam	1
nr of records on sire	1
nr of full sib records	0
nr. of half sib records (excl. full sibs)	40
nr. of progeny	0

Index weight		value of variate:
-	-	-
0.250	dam	20%
0.065	sire	1%
-	-	-
0.741	HS	15%
-	-	-

Accuracy of EBV **0.585**

correlation EBV FS **1.000**
 correlation EBV HS **0.633**

STEBVaccurcay.xls

$h^2 = 0.5$

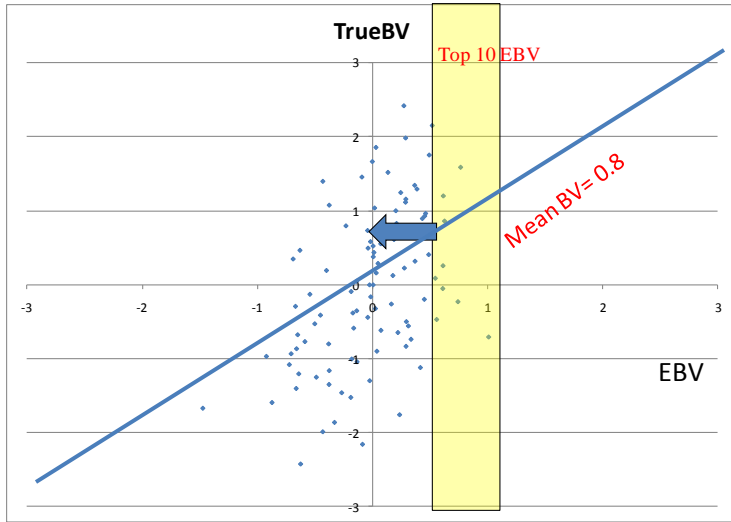
Information used	Nr.Records
nr of own records	1
nr. of records on dam	1
nr of records on sire	1
nr of full sib records	0
nr. of half sib records (excl. full sibs)	40
nr. of progeny	25

Index weight		value of variate:
0.164	own	2%
0.062	dam	0%
0.016	sire	0%
-	-	-
0.185	HS	0%
1.172	Prog	15%

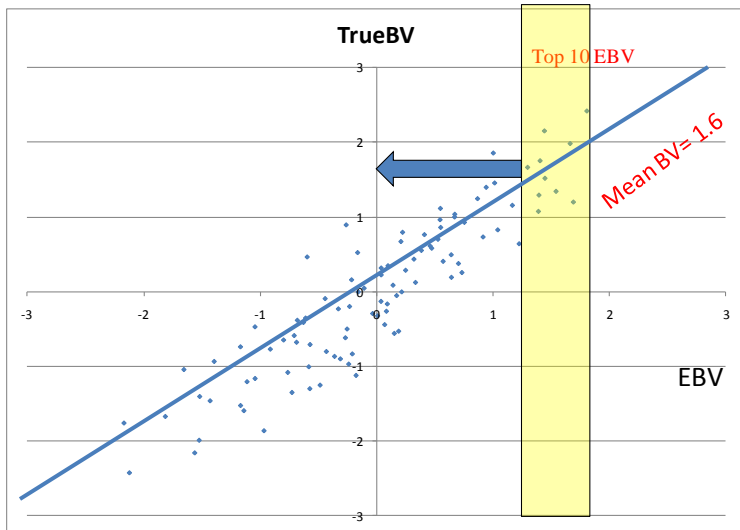
Accuracy of EBV **0.914**

correlation EBV FS **0.516**
 correlation EBV HS **0.455**

Note also that value of family info can be reduced due to Bulmer effect (reduced variance due to selection) – see later



Accuracy = 45%



Predicted Response

i = selection intensity (standard normal)

Regression of A on EBV = 1

i.e. slope is the same for accurate and inaccurate EBVs, see left

select on EBV's:

Response = $i * SD(EBV)$

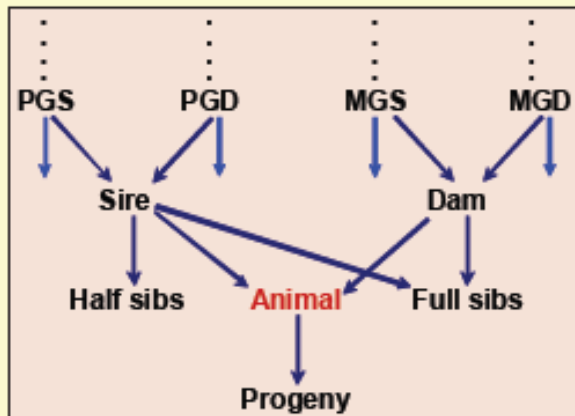
$$R = i * r * \sigma_a$$

Summary of this lecture

- Selection Index Theory can be used to work out weights and accuracy for a given set of information about an particular EBV
- Quantitative Genetic Theory and matrices (P, G) are used to work out such index weights (b) and accuracies
- In Genetic Evaluation we use BLUP where this all occurs 'automatically'
- Selection Index Theory still useful to predict what happens
 - Accuracy for a given amount of information
 - Accuracy if using genetic marker information
 - Importance of own vs family information for given situations

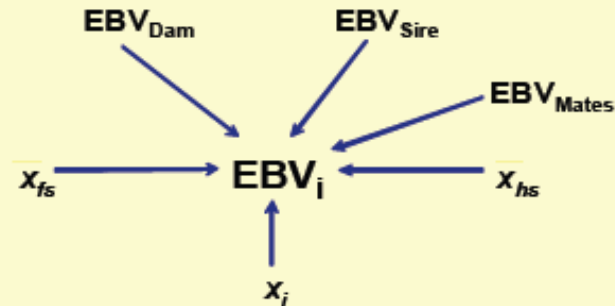
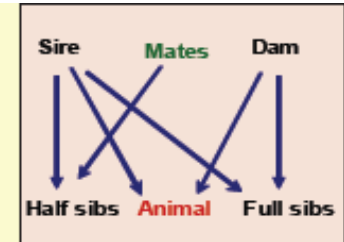
Modeling selection on Animal Model BLUP

Animal model EBV



Try to fit 'all relatives'
in selection index

Pseudo-BLUP Approximation of BLUP EBV by Selection Index



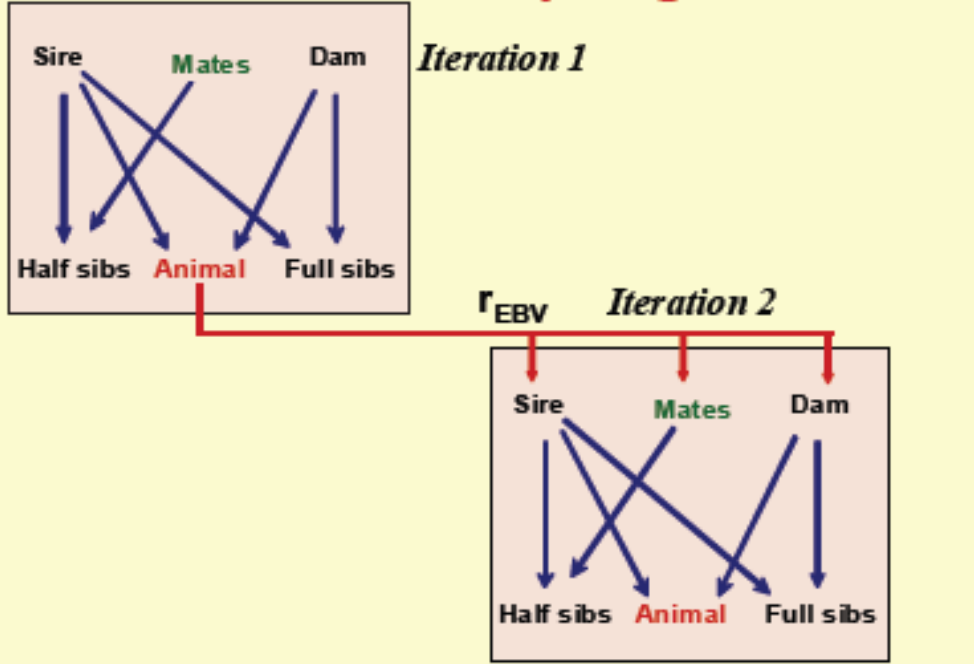
Fit sire and dam EBV
mates of sire (HS-dams)
Half sibs
Full sibs
Own info

Iterate

Pseudo BLUP

Iterate Selection Index

Iteration builds pedigree info



BLUP-EBV.xls
 STEBVaccuracy.xls

Iterative Procedure to Compute Accuracy of EBV

Building-up pedigree information

$$EBV_1 = b' [x_i, \bar{x}_{fs}, \bar{x}_{hs}, EBV_{Dam}, EBV_{Sire}, EBV_{Mates}]'$$

$$b = P^{-1}G \quad r_{EBV_1} = \sqrt{b'Pb/\sigma_g}$$

- 1) Set accuracy of $EBV_{Sire}=EBV_{Dam}=EBV_{Mates}=h$ (own record)
- 2) Set up index (P, G) and derive accuracy $r_{EBV_1} = \sqrt{b'Pb/\sigma_g}$
- 3) Set accuracy of $EBV_{Sire}=EBV_{Dam}=EBV_{Mates}$ equal to r_{EBV_1}
- 4) Repeat steps 2 and 3 until accuracy converges

Needs adaptation if selection after progeny

Example

