

Selection Theory - Predicting Genetic Change

Expected genetic mean of progeny = mean of (selected) parents

$$\begin{aligned} E(g_{\text{progeny}}) &= \frac{1}{2} g_{\text{sires}} + \frac{1}{2} g_{\text{dams}} \\ &= \frac{1}{2} \bar{g}_{\text{sires}} + \frac{1}{2} \bar{g}_{\text{dams}} + \frac{1}{2} S_{\text{sires}} + \frac{1}{2} S_{\text{dams}} \end{aligned}$$

Genetic mean of
parental generation

Mean Selection Differential
of selected parents (S)

S = Genetic Superiority or Selection Differential

Selection Response = Mean of progeny generation – Mean of parent generation

$$= \frac{1}{2} S_{\text{sires}} + \frac{1}{2} S_{\text{dams}}$$

Predicting Selection response

$$\text{Selection Response} = \frac{1}{2} S_{\text{sires}} + \frac{1}{2} S_{\text{dams}}$$

Predict S from breeding value = additive genetic effect
(what a parent passes on to offspring)

Predicting S from average EBV of selected parents, as regression $b_{a,EBV} = 1$

$$S = \bar{\hat{a}}_{\text{selected sires}} - \bar{\hat{a}}_{\text{contemporaries}}$$

$$\text{Predict as } S = i \sigma_{EBV} = i r_{IA} \sigma_a$$

i = selection intensity

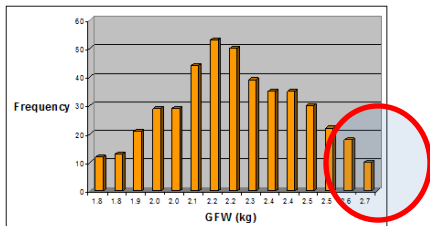
r_{IA} = selection accuracy

σ_a = additive genetic SD

How much genetic change?

Breeder's Equation

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



Selection intensity

Selection accuracy

Genetic variation



Selection intensity

selint.xls

Superiority of selected group (in SD units)

proportion selected

p

10.00%

Selection threshold

x

1.282

Height at threshold

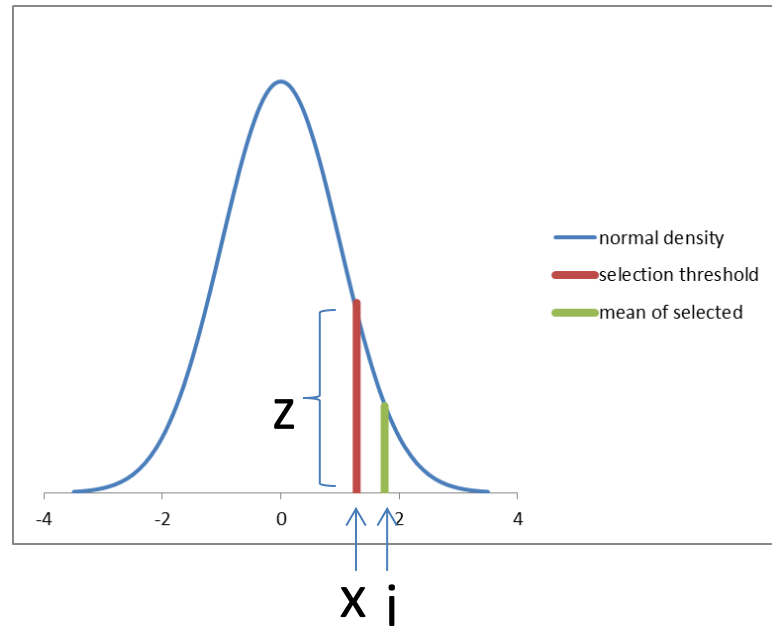
z

0.1755

Selection intensity

i

1.755



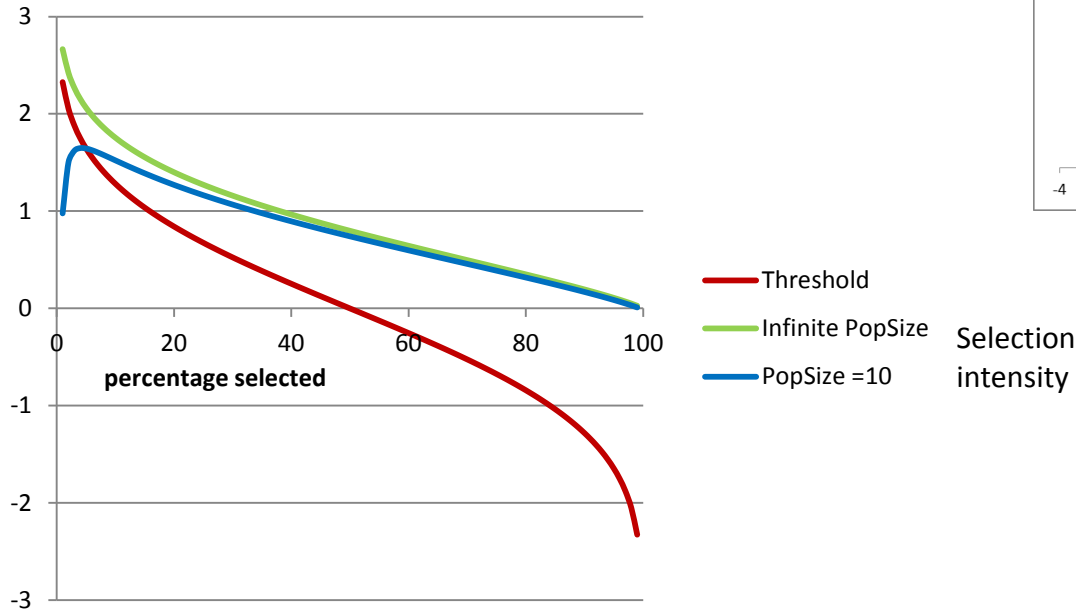
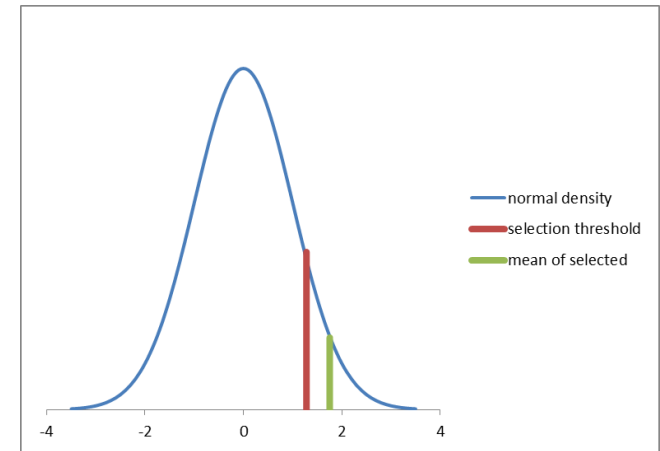
X: cumulative density function($x > \infty$) = p

x = -NORMSINV(p) in excel

$$z = \frac{e^{-0.5x^2}}{\sqrt{2\pi}}$$

$$i = z/p$$

Selection intensity correct for population size



➔ i is reduced in smaller populations, more so for small p and N

Burrows $i_{\text{corrected}} = i - \frac{N - Np}{2Np(N+1).i}$ or $p_{\text{corrected}} = \frac{Np + \frac{1}{2}}{N + (\frac{Np}{2N})}$ (approx)

Selection intensity - correct for correlated EBV

Criteria for selection often correlated amongst selection candidates

e.g. if EBV= Parent Average, all FS will have same EBV

with 10 FS families: selecting 10 out of 100 is really selecting 1 out of 10

$$i_{\text{corrected}} = \sqrt{(1-t_{\text{av}})} \cdot I$$

t_{av} is average correlation between EBV
across all candidates

Rawlings (1976)

$$t_{\text{av}} = \frac{t_{fs}(n_w - 1) + t_{hs}n_w(n_{fs} - 1)}{n_w n_{fs} n_{hs} - 1}$$

t_{fs} = correlation among FS

t_{hs} = correlation among HS

n_{hs} sires

n_{fs} dams per sire

n_w progeny per dam

Selection intensity - correct for correlated EBV

Example

Finite Population Size =

100

Nr Sires nhs 10

Nr dams/sire nfs 1

nw offspring per dam 10

FS EBV correlation 1

HS EBV Correlation 0.4

Corrected for population size

1.730

t_average 0.090909

Corrected for related EBV

1.649

Adjustment has limited effect,
and irrelevant with optimal contribution selection

Response per year

- Response per generation

$$\frac{1}{2} S_{\text{sires}} + \frac{1}{2} S_{\text{dams}}$$

Superiority of parents
averaged over males and females

- Response per year

$$\frac{\frac{1}{2} S_{\text{sires}} + \frac{1}{2} S_{\text{dams}}}{\frac{1}{2} L_{\text{sires}} + \frac{1}{2} L_{\text{dams}}} = \frac{S_{\text{sires}} + S_{\text{dams}}}{L_{\text{sires}} + L_{\text{dams}}}$$

Generation interval (in years)
averaged over males and females

Generation interval is
average age of sires (dams)
when their progeny are born

Derivation of $\frac{S_{\text{sires}} + S_{\text{dams}}}{L_{\text{sires}} + L_{\text{dams}}}$

Mean of offspring generation	\bar{g}_0
Mean of sires' generation	$\bar{g}_0 - L_S \cdot R_{\text{yr}}$
Mean of dams' generation	$\bar{g}_0 - L_D \cdot R_{\text{yr}}$
Mean of sires	$\bar{g}_0 - L_S \cdot R_{\text{yr}} + S_s$
Mean of dams	$\bar{g}_0 - L_D \cdot R_{\text{yr}} + S_d$

Mean of offspring

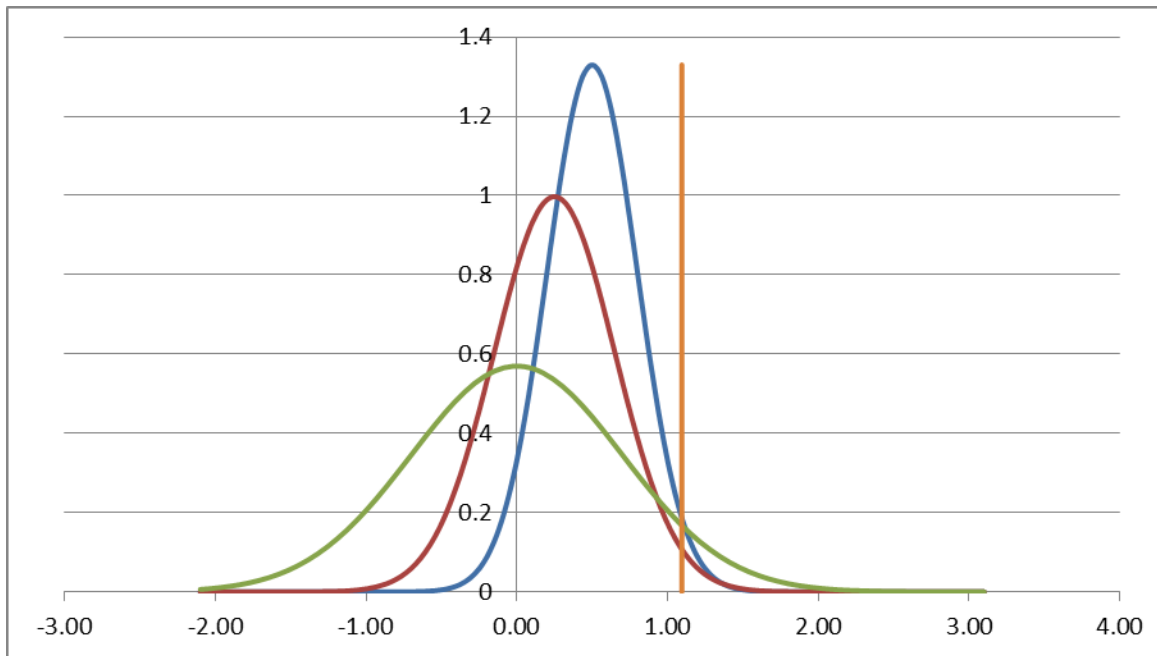
$$\frac{1}{2}(\bar{g}_0 - L_S \cdot R_{\text{yr}} + S_s) + \frac{1}{2}(\bar{g}_0 - L_D \cdot R_{\text{yr}} + S_d) = \bar{g}_0$$

$$\rightarrow \frac{1}{2}(-L_S \cdot R_{\text{yr}} + S_s) + \frac{1}{2}(-L_D \cdot R_{\text{yr}} + S_d) = 0$$

$$\rightarrow \frac{1}{2}(S_s) + \frac{1}{2}(S_d) = \frac{1}{2}(L_S) + \frac{1}{2}(L_D) R_{\text{yr}}$$

$$\rightarrow (S_s + S_d) / (L_S + L_D) = R_{\text{yr}}$$

Selection across age classes

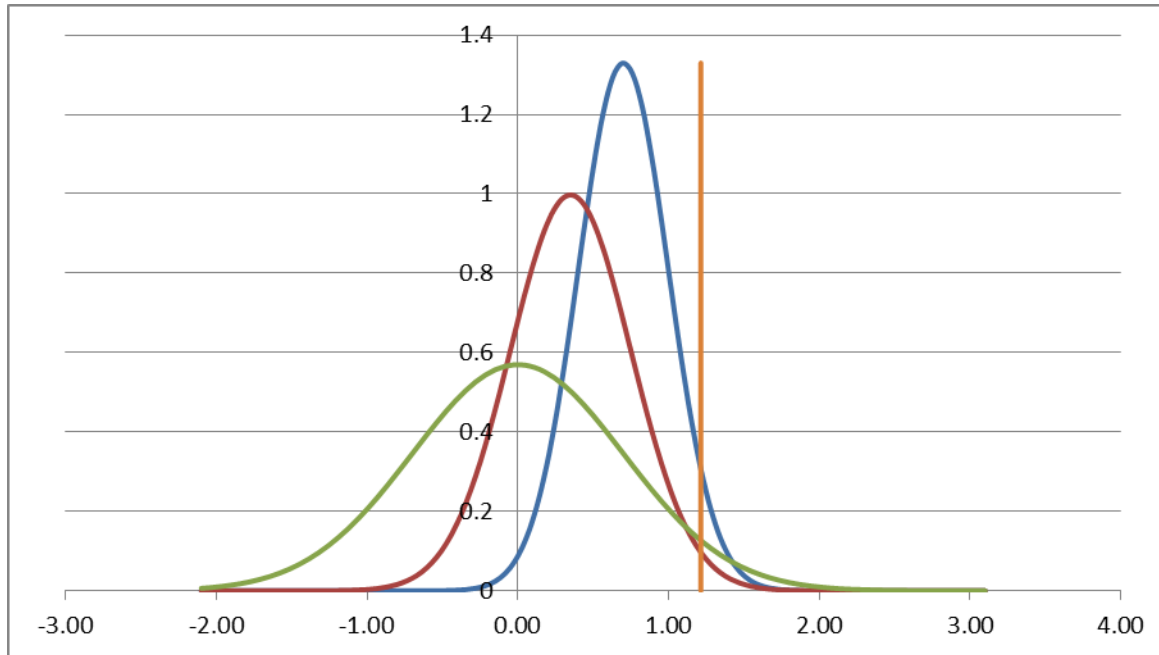


Select 20 animals
from 3 age classes

Candidates Age Class	Nr Candidates in age group	mean	SD
1	200	0.5	0.30
2	200	0.25	0.40
3	200	0	0.70

Nr Selected
5
3
12

Selection across age classes



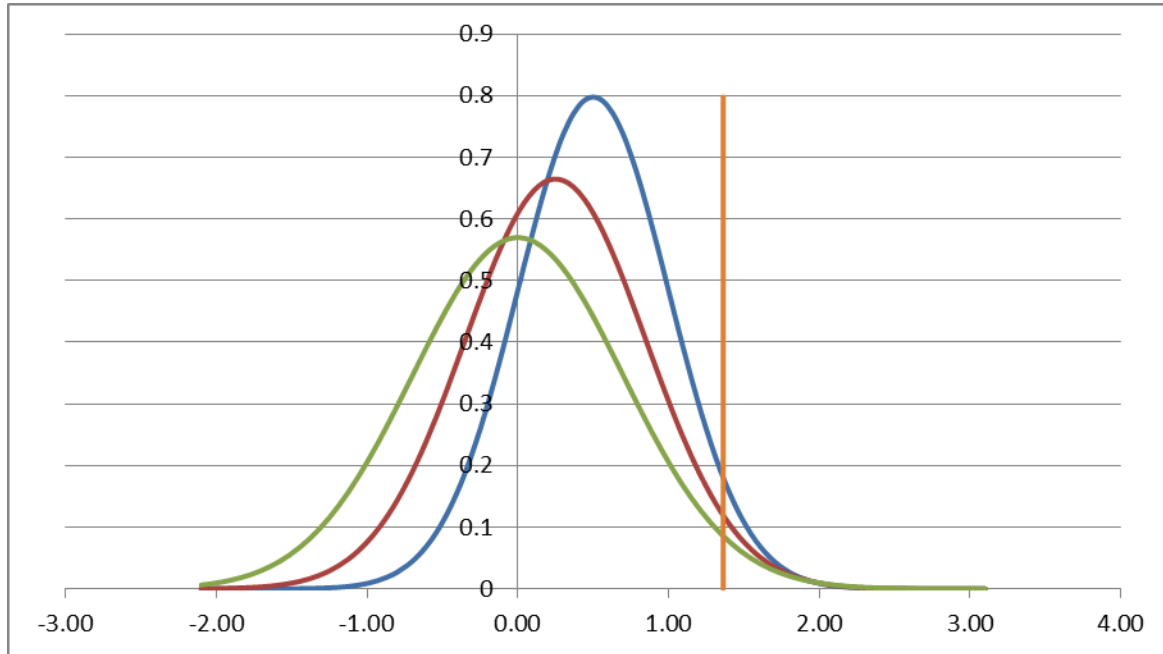
Select 20 animals
from 3 age classes

Candidates Age Class	Nr Candidates in age group	mean	SD
1	200	0.7	0.30
2	200	0.35	0.40
3	200	0	0.70

Nr Selected
9
3
8

More R_{yr}

Selection across age classes



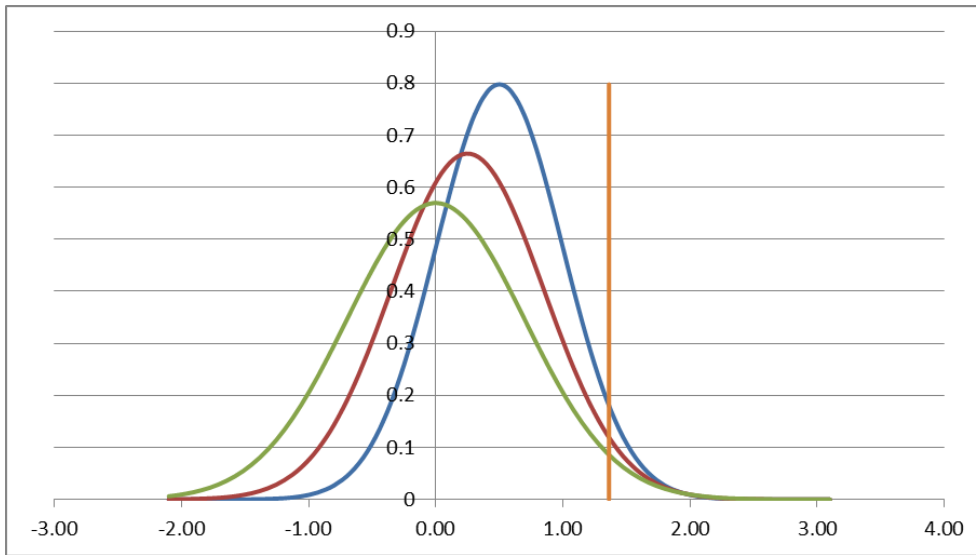
Select 20 animals
from 3 age classes

Candidates	Nr Candidates		
Age Class	in age group	mean	SD
1	200	0.5	0.50
2	200	0.25	0.60
3	200	0	0.70

Nr Selected
8
6
5

More accuracy
younger animals

Selection across age classes



If means and SD of
distributions are known

Algorithm

(Ducrocq and Quaas, 1989)

target = 20; LowTH = - 5; UppTH = + 5

Loop until LowTH ~ UppTH

TH = (MaxTH + MinTH)/2

Calculate Nr. selected (Nsel) for current TH

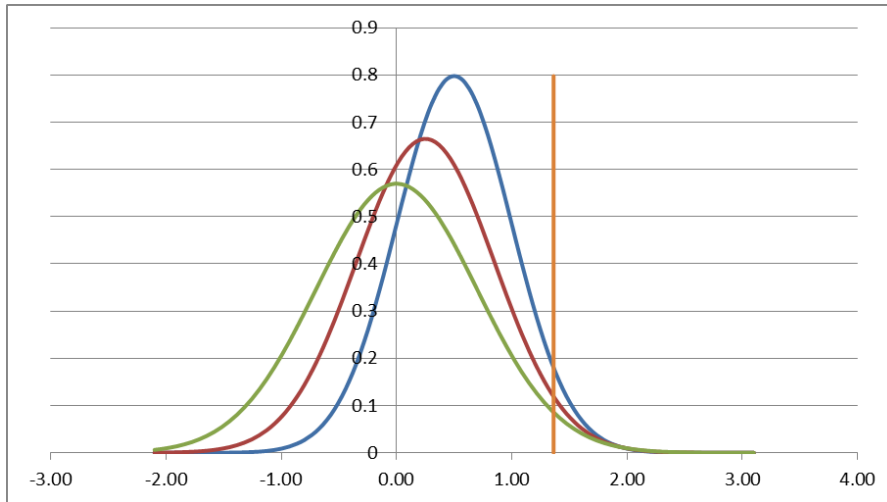
IF Nsel > target THEN LowTH=TH

ELSE UppTH=TH

Next

truncsel.xls

Optimize selection across age classes in a breeding program



Algorithm: Iterate

Start $R_{yr} = 0$

Loop until R_{yr} does not change

Calculate AC means from R_{yr}

Calculate TH for given means

Calculate S and L for given TH

Calculate R_{yr} from S and L

For males as well as females

If age class (AC) means and SD are NOT known

E.g. when optimizing a breeding program

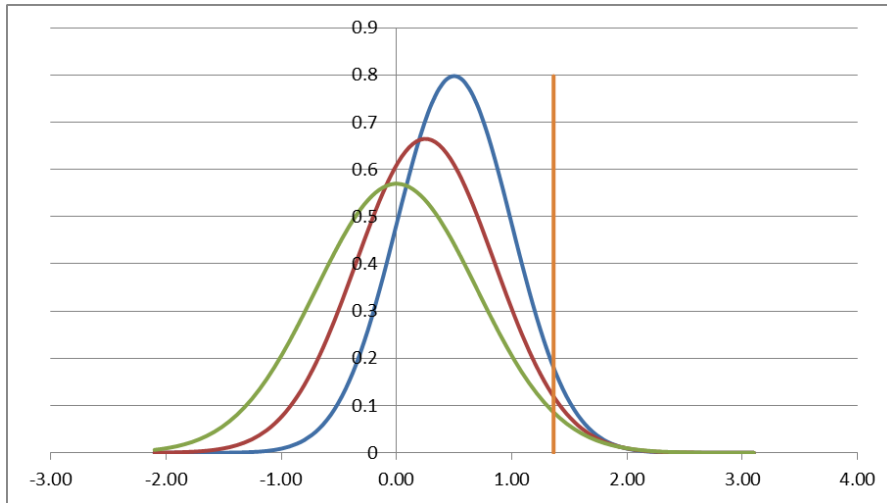
SD will depend on EBV accuracy
AC means will depend on R_{yr}

But R_{yr} will depend on S and L !

Next

truncsel.xls

Optimize selection across age classes in a breeding program



Just a reminder

In real breeding programs BLUP selection optimizes generation intervals automatically.

Truncation Selection modeling is needed for deterministic modeling and prediction of breeding program outcomes

Response with selection across age classes in a breeding program

Calculation of Response

Mean of offspring generation

$$\bar{g}_0$$

Mean of age class i (with age L_i)

$$\bar{g}_0 - L_i \cdot R_{yr}$$

Mean of animals selected from age class i

$$\bar{g}_0 - L_i \cdot R_{yr} + S_i$$

S_i = superiority of animals selected within age class i

Proportion of parents originating from age class i

$$p_i \quad (\sum p_i = 1)$$

Mean of offspring

(this is not split over sexes, assume same i , for sires and dams)

$$\sum p_i (\bar{g}_0 - L_i \cdot R_{yr} + S_i) = \bar{g}_0$$

$$\rightarrow \sum p_i (-L_i \cdot R_{yr} + S_i) = 0$$

$$\rightarrow \sum p_i S_i = \sum p_i L_i \cdot R_{yr}$$

$$R_{yr} = \sum p_i S_i / \sum p_i L_i$$

$\sum p_i S_i$ weighted mean of within age class superiority

$\sum p_i L_i$ weighted mean of ages = L

Response with selection across age classes in a breeding program

Calculation of Response

different i, L for males and females:

ps_i and pd_i are proportions and Sd_i and Ss_i are superiorities within age class for sires and dams resp.

Mean of offspring

$$\frac{1}{2} \sum ps_i (\bar{g}_0 - L_i \cdot R_{yr} + Ss_i) + \frac{1}{2} \sum pd_i (\bar{g}_0 - L_i \cdot R_{yr} + Sd_i) = \bar{g}_0$$

→

$$\rightarrow R_{yr} = \frac{\sum ps_i Ss_i + \sum pd_i Sd_i}{\sum ps_i L_i + \sum pd_i L_i}$$

= Weighted mean of within-AC-superiority (sum over sexes)

Weighted mean of AC-age (sum over sexes)

Response with selection across age classes in a breeding program

Example

Selection Across Age Groups



e.g. Selection of Bull Dams $\sigma_g = 550$ kg

Age Group	Age at Birth of Progeny	% of Bull Dams	% Selected	i	r	Genetic Superiority $ir\sigma_g$
Heifers	2 yr	50%	2.5%	2.34	.55	707.9
1 st Lact.	3 yr	30%	1.5%	2.53	.68	946.2
2 nd Lact.	4 yr	20%	1.5%	2.53	.72	1001.9

Selection Across Age Groups

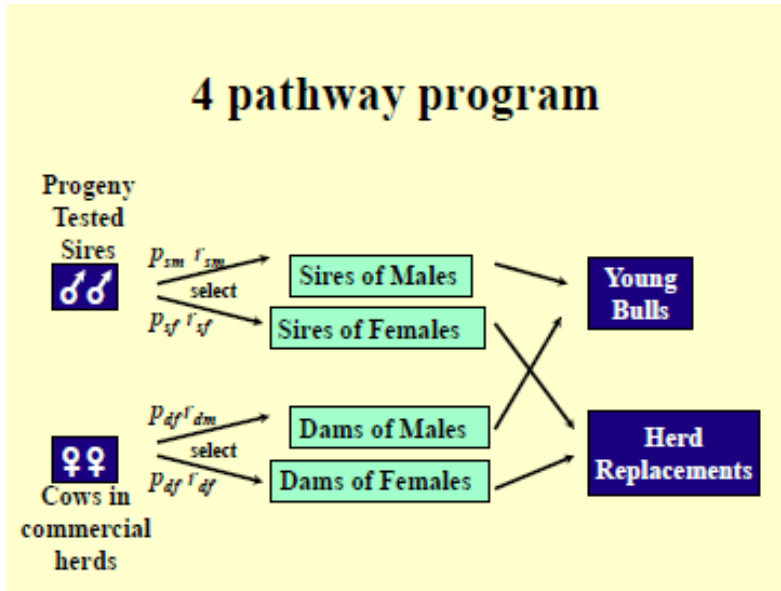
• Pooled Generation Interval

$$L_{dm} = 50\% * 2 \\ + 30\% * 3 \\ + 20\% * 4 \\ = 2.7 \text{ yr}$$

• Pooled Genetic Superiority

$$S_{dm} = 50\% * 707.9 \\ + 30\% * 946.2 \\ + 20\% * 1001.9 \\ = 838.2 \text{ kg}$$

Response with selection in a 4 pathway program



Rendel and Robertson 1950

$$\text{Asymptotic response/yr} = R_{yr} = \frac{S_{sm} + S_{dm} + S_{sf} + S_{df}}{L_{sm} + L_{dm} + L_{sf} + L_{df}}$$

Response with selection in a 4 pathway program

Example

Pathway	Proportion Selected (p_i)	Intensity (i_i)	Accuracy (r_i)	Genetic Superiority ($S_i = i r_i \sigma_g$)	Generation Interval (yr) (L_i)
Sires of males	2 %	2.42	0.90	$2.178 \sigma_g$	6
Sires of females	10 %	1.75	0.90	$1.575 \sigma_g$	7
Dams of males	0.5 %	2.89	0.60	$1.743 \sigma_g$	5
Dams of females	90 %	0.19	0.60	$0.114 \sigma_g$	6
TOTAL				$\Sigma S = 5.601 \sigma_g$	$\Sigma L = 24$

$$R_{yr} = \frac{(2.42 \times 0.9 + 1.75 \times 0.9 + 2.89 \times 0.6 + 0.19 \times 0.6)}{6 + 7 + 5 + 6} \sigma_g = 0.233 \sigma_g \text{ per yr}$$

genetic gain.xls

Some concluding comments

- In general this theory provides a good framework to predict outcomes of breeding strategies, including introducing new technologies (such as reproductive and genomic technologies)
- Much of the theory is applicable to a closed breeding nucleus, in reality we often have many of these nuclei interchanging breeding material (esp. in beef, sheep), and means, variances and genetic backgrounds between these might differ.
- The theory is often assuming one homogeneous population and it is about exploiting genetic variance within populations or breeds. In reality we can exploit genetic variation between lines, breeds, populations as well (esp. with genomics!)