



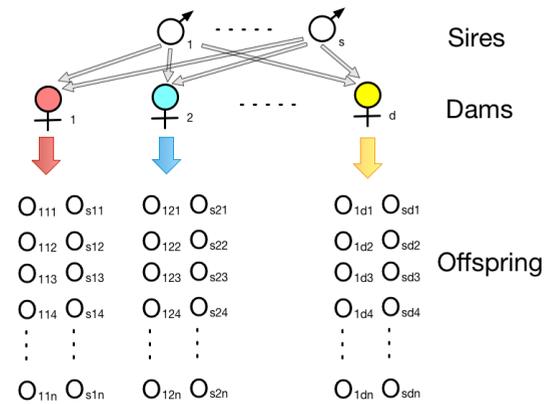
Precision of heritability and selection

Jinliang Yang
Nov. 2nd, 2018

Precision and design

If we want to design a balanced experiment:

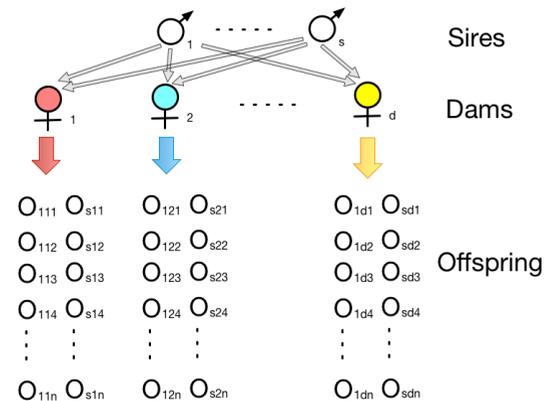
- s **sires** each mated to d **dams**
- Each dam has n **progenies**



Precision and design

If we want to design a balanced experiment:

- s **sires** each mated to d **dams**
- Each dam has n **progenies**



Questions before experimental design?

1. Parent-offspring, half-sib, full-sib, or others?
2. How many families?
3. Numbers of progeny?
4. What if it is unbalanced?

Sampling variance of b

F & M page 178, the sampling variance of the parent-offspring regression is approximately:

$$SV_b = \frac{k[1 + (n - 1)t]}{nN}$$

- N families (offspring and parents)
- k parents (1 or 2) for each family
- n offspring per family
- t the intra-class correlation between offspring in a family

Sampling variance of b

F & M page 178, the sampling variance of the parent-offspring regression is approximately:

$$SV_b = \frac{k[1 + (n - 1)t]}{nN}$$

- N families (offspring and parents)
- k parents (1 or 2) for each family
- n offspring per family
- t the intra-class correlation between offspring in a family

Parent-offspring regression: One parent

$$k = 1$$

$$SV_b = \frac{[1 + (n - 1)t]}{nN}$$

Sampling variance is minimal when $n = 1$, i.e. $(n - 1)t = 0$.

Parent-offspring regression: One parent

The **most efficient** design:

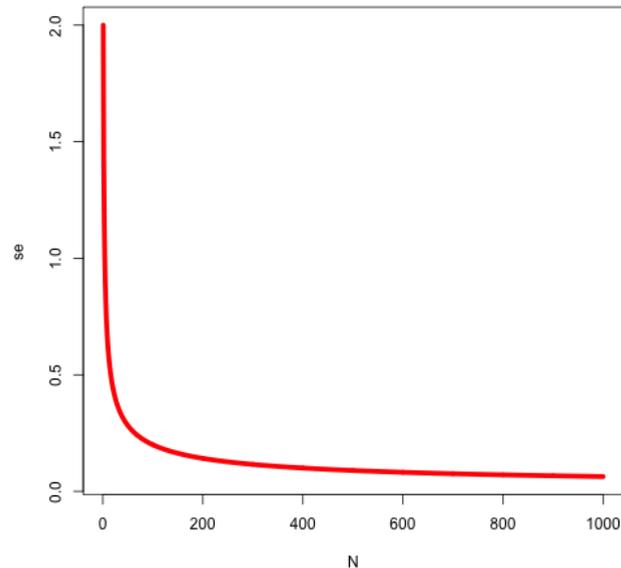
- as many as families as possible
- measure only one offspring per family

```
N=1:1000  
se <- 2/sqrt(N)  
  
plot(N, se, type="l",  
      lwd=5, col="red")
```

Sampling variance of b

$$SV(h^2 = 2b) = 4/N$$

$$s.e. (h^2) = 2/\sqrt{N}$$



Sampling variance of b

F & M page 178, the sampling variance of the regression is approximately:

$$SV_b = \frac{k[1 + (n - 1)t]}{nN}$$

- N families (offspring and parents)
- k parents (1 or 2) for each family
- n offspring
- t the intra-class correlation between offspring in a family

Sampling variance of b

F & M page 178, the sampling variance of the regression is approximately:

$$SV_b = \frac{k[1 + (n - 1)t]}{nN}$$

- N families (offspring and parents)
- k parents (1 or 2) for each family
- n offspring
- t the intra-class correlation between offspring in a family

Parent-offspring regression: both parents

When $k = 2$ (use mid-parent values), the $SV(h^2 = b) = 2/N$. So the standard error, $se(h^2) = \sqrt{2/N}$

Parent-offspring regression: Both-parents

Sampling variance of b

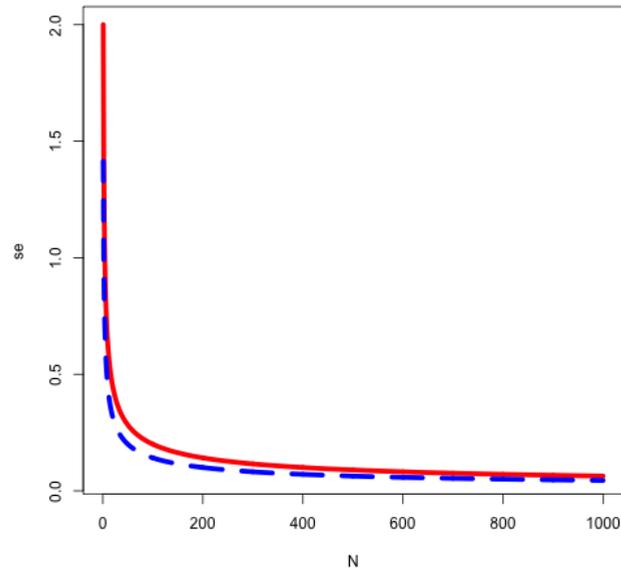
One parent:

$$s.e.(h^2) = 2/\sqrt{N}$$

Both parents:

$$s.e.(h^2) = \sqrt{2/N}$$

```
N=1:1000  
se <- 2/sqrt(N)  
se2 <- sqrt(2/N)  
plot(N, se, type="l",  
      lwd=5, col="red")  
lines(N, se2, type="l",  
       lwd=5, col="blue", lty=2)
```



Sib analyses

The **correlation** between A and P, r_{AP} :

$$\begin{aligned}r_{AP} &= \frac{Cov(A, P)}{\sqrt{\sigma_A^2 \sigma_P^2}} \\ &= \sqrt{\frac{\sigma_A^2}{\sigma_P^2}} \\ &= h\end{aligned}$$

Sib analyses

The **correlation** between A and P, r_{AP} :

$$\begin{aligned} r_{AP} &= \frac{Cov(A, P)}{\sqrt{\sigma_A^2 \sigma_P^2}} \\ &= \sqrt{\frac{\sigma_A^2}{\sigma_P^2}} \\ &= h \end{aligned}$$

Therefore, the **intraclass correlation (t)**, is also a function of h .

Sib analyses

Now we look at the sampling variance of the **intra-class correlation (t)**. According to F & M page 180, the sampling variance of **t** is:

$$SV_t = \frac{2[1 + (n - 1)t]^2 \times (1 - t)^2}{n(n - 1)(N - 1)}$$

- N families (offspring and parents)
- n offspring per family
- t the intra-class correlation between offspring in a family

Sib analyses

Now we look at the sampling variance of the **intra-class correlation (t)**. According to F & M page 180, the sampling variance of **t** is:

$$SV_t = \frac{2[1 + (n - 1)t]^2 \times (1 - t)^2}{n(n - 1)(N - 1)}$$

- N families (offspring and parents)
- n offspring per family
- t the intra-class correlation between offspring in a family

Optimal family size

- If we let $T = nN$, the total number measured in a generation
- The sampling variance is minimized when **n (number of offspring per family) = 1/t**, approximately

Optimal family size

With $n = 1/t$, approximately,

In the simplest cases with no common environmental effect in families, then

- For half-sibs: $t_{HS} = h^2/4$, then $n = 4/h^2$
- For full-sibs: $t_{FS} = h^2/2$, then $n = 2/h^2$

Optimal family size

With $n = 1/t$, approximately,

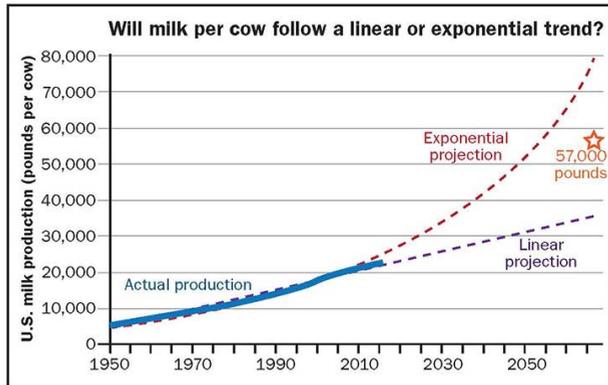
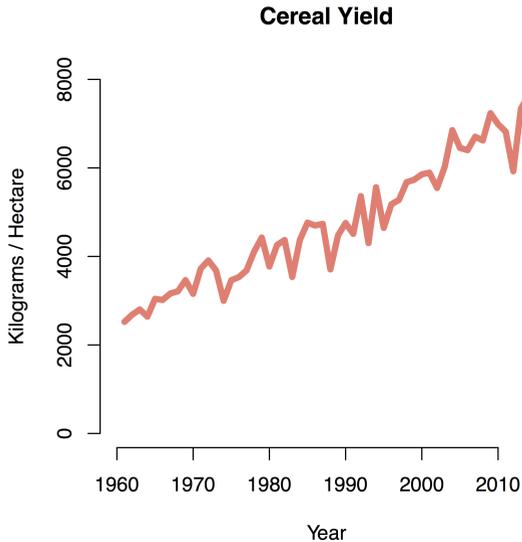
In the simplest cases with no common environmental effect in families, then

- For half-sibs: $t_{HS} = h^2/4$, then $n = 4/h^2$
- For full-sibs: $t_{FS} = h^2/2$, then $n = 2/h^2$

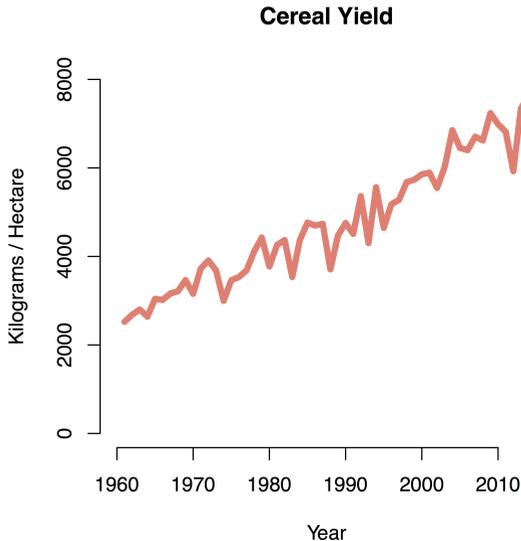
Caveats:

- Some **assumption of h^2 ahead of time** is essential in planning the data collection!
- Note that in the case of half-sibs, we are assuming that only one offspring per dam is measured. so n is the **optimal number of dams per sire**.
 - in the absence of prior knowledge, half-sib family size 20-30. (Robertson 1959)

The improvement through breeding

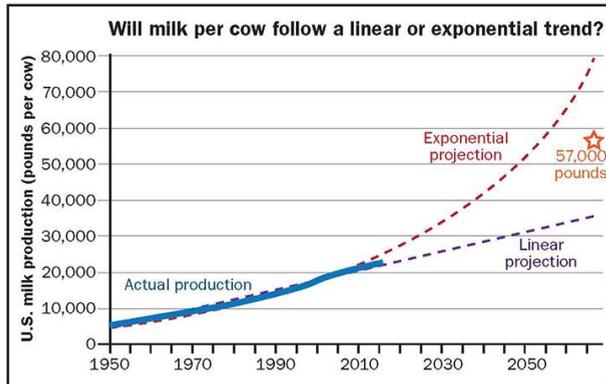


The improvement through breeding



Response to selection

1. Why has a character changed over generations?
2. Can we predict a change?



Response to selection

Response to selection

1. Why has a character changed over generations?
2. Can we predict a change?

Improve the body weight of chicken



Increase the cob length

Response to selection

Change, or response (R), is given by the basic equation:

$$R = h^2 S$$

Here S is the **selection differential**, which is equal to the mean value of the selected parents (μ_S) minus the population mean (μ).

Improve the body weight



Response to selection

Change, or response (R), is given by the basic equation:

$$R = h^2 S$$

Here S is the **selection differential**, which is equal to the mean value of the selected parents (μ_S) minus the population mean (μ).

Improve the body weight



1. Base population (N=100)
mean = 1,000g
2. Selected 5 chicken as the
parents, with mean =1,050g.

What is S ?

Standardized selection differential

With the assumption that the phenotypic distributions are normal.

Selection differential in standard deviation units:

$$\frac{S}{\sigma_P}$$

Standardized selection differential

With the assumption that the phenotypic distributions are normal.

Selection differential in standard deviation units:

$$\frac{S}{\sigma_P}$$

Intensity of selection (i)

$$i = \frac{S}{\sigma_P}$$

Intensity of selection (i)

- With the assumption that the phenotypic distributions are normal.
- And standardize the selection differential:

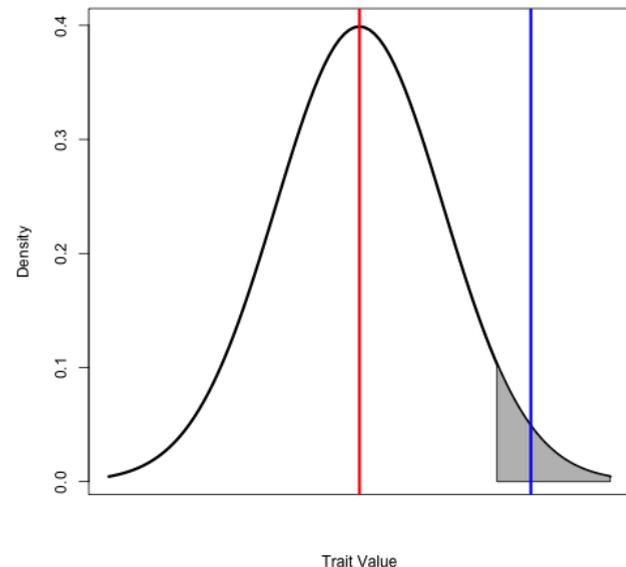
$$i = S/\sigma_P$$

Intensity of selection (i)

- With the assumption that the phenotypic distributions are normal.
- And standardize the selection differential:

$$i = S/\sigma_P$$

```
curve(dnorm(x,0,1), xlim=c(-3,3), xaxp=c(1,0,1), yaxp=c(1,0,1))
fromd <- qnorm(.95); tod <- 3
S.x <- c(fromd, seq(fromd, tod, 0.01))
S.y <- c(0, dnorm(seq(fromd, tod, 0.01)))
polygon(S.x,S.y, col="grey")
abline(v=mean(S.x**S.y)*2, col="blue")
```



Intensity of selection (i)

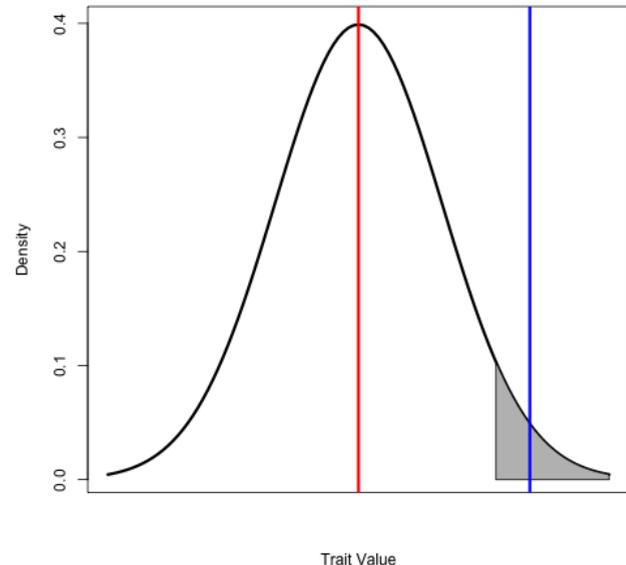
- With the assumption that the phenotypic distributions are normal.
- And standardize the selection differential:

$$i = S/\sigma_P$$

- If p is **the proportion selected**, i.e. the proportion of the population falling beyond the point of truncation.
- And z is the **height of the ordinate** at the point of truncation.

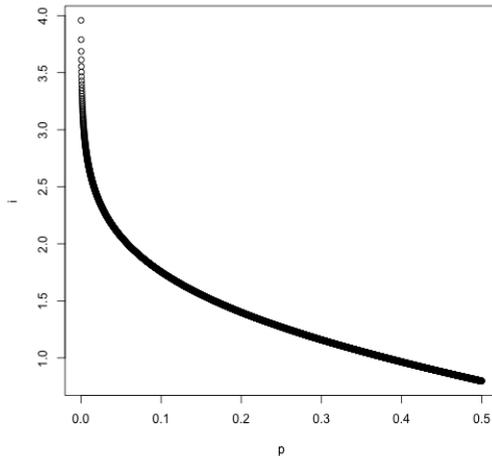
$$i = \frac{z}{p}$$

```
curve(dnorm(x,0,1), xlim=c(-3,3), xaxp=c(1,1,1), yaxp=c(1,1,1))
fromd <- qnorm(.95); tod <- 3
S.x <- c(fromd, seq(fromd, tod, 0.01))
S.y <- c(0, dnorm(seq(fromd, tod, 0.01)))
polygon(S.x,S.y, col="grey")
abline(v=mean(S.x**S.y)*2, col="blue")
```



Intensity of selection (i)

```
ifun <- function(p=0.5){  
  x=qnorm(p=(1-p)) # get the truncated  
  z=dnorm(x) # get z  
  return(z/p) # get i  
}  
  
p <- seq(0.0001, 0.5, by=0.0001)  
i <- ifun(p)  
plot(p, i)
```



```
head(data.frame(p, i), 20)
```

```
##           p           i  
## 1  0.0001  3.958480  
## 2  0.0002  3.789212  
## 3  0.0003  3.686955  
## 4  0.0004  3.612829  
## 5  0.0005  3.554381  
## 6  0.0006  3.505980  
## 7  0.0007  3.464589  
## 8  0.0008  3.428376  
## 9  0.0009  3.396149  
## 10 0.0010  3.367090  
## 11 0.0011  3.340611  
## 12 0.0012  3.316274  
## 13 0.0013  3.293746  
## 14 0.0014  3.272767  
## 15 0.0015  3.253129  
## 16 0.0016  3.234665  
## 17 0.0017  3.217235  
## 18 0.0018  3.200726  
## 19 0.0019  3.185040  
## 20 0.0020  3.170097
```

Breeder's equation

$$R = h^2 S$$

$$S = i\sigma_P$$

Breeder's equation

$$R = h^2 S$$

$$S = i\sigma_P$$

So now, we have, as a prediction:

$$R = h^2 i\sigma_P$$

$$i = \frac{z}{p}$$

Breeder's equation

$$R = h^2 i \sigma_P$$
$$i = \frac{z}{p}$$

Genders

With differential selection opportunities between genders in sexual reproduction, we need to account for this in our predictions:

$$i = \frac{1}{2}(i_m + i_f)$$

Breeder's equation

$$R = h^2 i \sigma_P$$
$$i = \frac{z}{p}$$

Genders

With differential selection opportunities between genders in sexual reproduction, we need to account for this in our predictions:

$$i = \frac{1}{2}(i_m + i_f)$$

Generation interval

Generation interval (L) is the **"average age of the parents at the birth of their selected offspring"**.

- it is when they effectively leave the next generation that is then sampled to repeat the selection process.

$$L = \frac{1}{2}(L_m + L_f)$$

Breeder's equation

Now, on a time-constant basis to allow comparisons of alternatives:

$$\begin{aligned} R &= \frac{i}{L} h^2 \sigma_A \\ &= \frac{i_m + i_f}{L_m + L_f} h^2 \sigma_A \end{aligned}$$

With the h and σ_A terms possibly constant, we can simply compare the i/L portions under different scenarios of selection and reproduction.

Example: selection for body weight in chicken

A chicken breeder is selecting for 56-day body weight in chicken.

- **Base population:** a random mating population of 154 males and 155 females. Mean value = 1,000g; sd = 50g.
- **Selecion scheme:** 8 males and 48 females with the highest body weight to found the next generation.
- **h²:** from previous parent-offspring regression and half-sib analysis, h²=0.45.

Example: selection for body weight in chicken

A chicken breeder is selecting for 56-day body weight in chicken.

- **Base population:** a random mating population of 154 males and 155 females. Mean value = 1,000g; sd = 50g.
- **Selecion scheme:** 8 males and 48 females with the highest body weight to found the next generation.
- **h2:** from previous parent-offspring regression and half-sib analysis, h2=0.45.

What is the predicted response to selection?

$$R = \frac{i_m + i_f}{L_m + L_f} h^2 \sigma_P$$

Example: selection for body weight in chicken

A chicken breeder is selecting for 56-day body weight in chicken.

- **Base population:** a random mating population of 154 males and 155 females. Mean value = 1,000g; sd = 50g.
- **Selection scheme:** 8 males and 48 females with the highest body weight to found the next generation.
- **h²:** from previous parent-offspring regression and half-sib analysis, h²=0.45.

What is the predicted response to selection?

$$R = \frac{i_m + i_f}{L_m + L_f} h^2 \sigma_P$$

```
i_m <- ifun(p= 8/154) #2.05  
i_f <- ifun(p= 48/155) # 1.14  
(i_m + i_f)/2*50*0.45
```

```
## [1] 35.8358
```

Example: selection for body weight in chicken

A chicken breeder is selecting for 56-day body weight in chicken.

- **Base population:** a random mating population of 154 males and 155 females. Mean value = 1,000g; sd = 50g.
- **Selecion scheme:** 8 males and 48 females with the highest body weight to found the next generation.
- **h²:** from previous parent-offspring regression and half-sib analysis, h²=0.45.

What is the selection differential for the male and female?

Example: selection for body weight in chicken

A chicken breeder is selecting for 56-day body weight in chicken.

- **Base population:** a random mating population of 154 males and 155 females. Mean value = 1,000g; sd = 50g.
- **Selecion scheme:** 8 males and 48 females with the highest body weight to found the next generation.
- **h2:** from previous parent-offspring regression and half-sib analysis, $h^2=0.45$.

What is the selection differential for the male and female?

```
Sm <- i_m*50 #102  
Sf <- i_f*50 # 57
```

Example: selection for body weight in chicken

A chicken breeder is selecting for 56-day body weight in chicken.

- **Base population:** a random mating population of 154 males and 155 females. Mean value = 1,000g; sd = 50g.
- **Selection scheme:** 8 males and 48 females with the highest body weight to found the next generation.
- **h^2 :** from previous parent-offspring regression and half-sib analysis, $h^2=0.45$.

What is the selection differential for the male and female?

```
Sm <- i_m*50 #102  
Sf <- i_f*50 # 57
```

Why response 36g not equal to average value of the $S_m + S_f$?

Improvement of response

$$R = ih^2\sigma_P/L$$

The form of the breeder's equation allows us to clearly see how to maximize response to selection per unit time.

Improvement of response

$$R = ih^2\sigma_P/L$$

The form of the breeder's equation allows us to clearly see how to maximize response to selection per unit time.

1. Reduce the generation interval

This factor is bound by the biology of the organisms, but ways to reduce the generation interval are possible.

- selection at juvenile stages
- selection during the off-season in plants
- genomic selection

Improvement of response

$$R = ih^2\sigma_P/L$$

The form of the breeder's equation allows us to clearly see how to maximize response to selection per unit time.

2. Increase the heritability of the trait

Depends on the genetic architecture, not always under the control of the breeder.

- maximizing the repeatability of trait evaluation.
- sound measurement methods and proper experimental design

Improvement of response

$$R = ih^2\sigma_P/L$$

The form of the breeder's equation allows us to clearly see how to maximize response to selection per unit time.

3. Increase the selection intensity

- reduce the number of individuals selected to serve as parents of the next generation
 - **Be cautious:** increase the inbreeding coefficient and hence loss of alleles through genetic drift.
- increase the population size from which selections are made

Improvement of response

$$R = ih^2\sigma_P/L$$

The form of the breeder's equation allows us to clearly see how to maximize response to selection per unit time.

3. Increase the selection intensity

- reduce the number of individuals selected to serve as parents of the next generation
 - **Be cautious:** increase the inbreeding coefficient and hence loss of alleles through genetic drift.
- increase the population size from which selections are made

Increase i is likely not the most efficient path! i.e. p from 10% to 5% $\Rightarrow i$ from 1.755 to 2.063.

Improvement of response

$$R = ih^2\sigma_P/L$$

The form of the breeder's equation allows us to clearly see how to maximize response to selection per unit time.

4. Increase additive genetic variance

$$h^2 = \frac{\sigma_A^2}{\sigma_P^2}; \quad h = \frac{\sigma_A}{\sigma_P}$$

Improvement of response

$$R = ih^2\sigma_P/L$$

The form of the breeder's equation allows us to clearly see how to maximize response to selection per unit time.

4. Increase additive genetic variance

$$h^2 = \frac{\sigma_A^2}{\sigma_P^2}; \quad h = \frac{\sigma_A}{\sigma_P}$$

$$R = ih\sigma_A/L$$

- The breeding population needs to contain adequate additive genetic variation for the trait of interest.
- If no difference in breeding values exist between individuals within the population, genetic gain through selection is not possible.