

Lecture 3

Index Selection

Bruce Walsh lecture notes
Tucson Winter Institute
9 - 11 Jan 2013

Selection on an Index of Traits

A common way to select on a number of traits at once is to base selection decisions on a simple **index** of trait values,

$$I = \sum b_j z_j = \mathbf{b}^T \mathbf{z}$$

The resulting phenotypic and additive-genetic variance for this synthetic trait are;

$$\sigma_I^2 = \sigma(\mathbf{b}^T \mathbf{z}, \mathbf{b}^T \mathbf{z}) = \mathbf{b}^T \sigma(\mathbf{z}, \mathbf{z}) \mathbf{b} = \mathbf{b}^T \mathbf{P} \mathbf{b}$$

$$\sigma_{A_I}^2 = \sigma_A(\mathbf{b}^T \mathbf{z}, \mathbf{b}^T \mathbf{z}) = \mathbf{b}^T \sigma_A(\mathbf{z}, \mathbf{z}) \mathbf{b} = \mathbf{b}^T \mathbf{G} \mathbf{b}$$

Heritability of an Index

$$\sigma_I^2 = \sigma(\mathbf{b}^T \mathbf{z}, \mathbf{b}^T \mathbf{z}) = \mathbf{b}^T \sigma(\mathbf{z}, \mathbf{z}) \mathbf{b} = \mathbf{b}^T \mathbf{P} \mathbf{b}$$

$$\sigma_{A_I}^2 = \sigma_A(\mathbf{b}^T \mathbf{z}, \mathbf{b}^T \mathbf{z}) = \mathbf{b}^T \sigma_A(\mathbf{z}, \mathbf{z}) \mathbf{b} = \mathbf{b}^T \mathbf{G} \mathbf{b}$$

This gives the resulting heritability of the index as

$$h_I^2 = \frac{\sigma_{A_I}^2}{\sigma_I^2} = \frac{\mathbf{b}^T \mathbf{G} \mathbf{b}}{\mathbf{b}^T \mathbf{P} \mathbf{b}}$$

Predicted response in the index is

$$R_I = h_I^2 S_I = h_I \sigma_{A_I} i_I$$

Thus, we can think of an index as just a new univariate trait, with its own heritability

Class problem

Example 33.1. A convenient dataset we use through this chapter is that of Brim et al. (1959), who estimated the genetic and phenotypic covariances for several characters in soybeans. Consider three of these traits, z_1 = oil content, z_2 = protein content, and z_3 = yield. For these characters, Brim et al. estimated the covariance matrices as

$$\mathbf{P} = \begin{pmatrix} 287.5 & 477.4 & 1266 \\ 477.4 & 935 & 2303 \\ 1266 & 2303 & 5951 \end{pmatrix}, \quad \mathbf{G} = \begin{pmatrix} 128.7 & 160.6 & 492.5 \\ 160.6 & 254.6 & 707.7 \\ 492.5 & 707.7 & 2103 \end{pmatrix}$$

For $\mathbf{b} = \begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix}$

Compute additive and phenotypic variance, and h^2 for the index

Enter P, G, and b

```
> P<-matrix(c(287.5,477.4,1266,477.4,935,2303,1266,2303,5951),nrow=3)
> P
      [,1] [,2] [,3]
[1,] 287.5 477.4 1266
[2,] 477.4 935.0 2303
[3,] 1266.0 2303.0 5951
> G<-matrix(c(128.7,160.6,492.5,160.6,254.6,707.7,492.5,707.7,2103),nrow=3)
> G
      [,1] [,2] [,3]
[1,] 128.7 160.6 492.5
[2,] 160.6 254.6 707.7
[3,] 492.5 707.7 2103.0
> b<-matrix(c(2,-3,4),nrow=3)
> b
      [,1]
[1,] 2
[2,] -3
[3,] 4
```

Compute phenotypic ($b^T P b$) and additive ($b^T G b$) for index

```
> t(b) %*% P %*% b
      [,1]
[1,] 64036.2
> t(b) %*% G %*% b
      [,1]
[1,] 25422.2
```

$$\begin{aligned}\text{Heritability} &= 25422.2 / 64026.2 \\ &= 0.397\end{aligned}$$

Thus, the response in the index due to selection is

$$R_I = \bar{i} h_I^2 \sigma_I = \bar{i} \cdot \frac{\mathbf{b}^T \mathbf{G} \mathbf{b}}{\mathbf{b}^T \mathbf{P} \mathbf{b}} \sqrt{\mathbf{b}^T \mathbf{P} \mathbf{b}} = \bar{i} \cdot \frac{\mathbf{b}^T \mathbf{G} \mathbf{b}}{\sqrt{\mathbf{b}^T \mathbf{P} \mathbf{b}}}$$

How does selection on I translate into selection on the underlying traits?

$$S_j = \frac{\bar{i}}{\sigma_I} \sum_k b_k P_{jk} \quad \mathbf{S} = \frac{\bar{i}}{\sigma_I} \mathbf{P} \mathbf{b}$$

Since $\mathbf{R} = \mathbf{G}\mathbf{P}^{-1}\mathbf{S}$, the response vector for the underlying means becomes

$$\mathbf{R} = \mathbf{G}\mathbf{P}^{-1}\mathbf{S} = \frac{\bar{l}}{\sigma_I} \mathbf{G}\mathbf{b} = \bar{l} \cdot \frac{\mathbf{G}\mathbf{b}}{\sqrt{\mathbf{b}^T \mathbf{P} \mathbf{b}}}$$

Hence, we can not only compute the response in the weighted index, but also the individual response in each of the component traits (the vector \mathbf{R}).

Class problem

For a selection intensity of $i = 2.06$ (upper 5% selected) compute the response in the index and the vectors of selection differentials and responses (S and R) for the vector of traits.

Recall (from previous results)

$$\mathbf{b}^T \mathbf{P} \mathbf{b} = 64036.2; \quad \mathbf{b}^T \mathbf{G} \mathbf{b} = 25422.2$$

$$R_I = \bar{i} \cdot \frac{\mathbf{b}^T \mathbf{G} \mathbf{b}}{\sqrt{\mathbf{b}^T \mathbf{P} \mathbf{b}}}$$

```
> Vp <- 64036.2
> Va <- 25422.2
> i <- 2.06
> i*Va/sqrt(Vp)
[1] 206.9510
```

Index increases by ~ 207 8

Responses in each trait

$$\mathbf{R} = \bar{i} \cdot \frac{\mathbf{Gb}}{\sqrt{\mathbf{b}^T \mathbf{P} \mathbf{b}}}$$

```
> (i/sqrt(Vp))* G %*% b
      [,1]
[1,] 14.21017 Oil
[2,] 19.44129 Protein
[3,] 59.21364 Yield
```

Selection differentials on each trait

$$\mathbf{S} = \frac{\bar{i}}{\sigma_I} \mathbf{P} \mathbf{b}$$

```
> (i/sqrt(Vp))* P %*% b
      [,1]
[1,] 34.24572 Oil
[2,] 59.92920 Protein
[3,] 158.14672 Yield
```

Selection gradients, $\beta = \mathbf{P}^{-1}\mathbf{S}$

```
> S<-(i/sqrt(Vp))* P %*% b
> solve(P) %*% S
      [,1]
[1,] 0.01628113 Oil
[2,] -0.02442169 Protein
[3,] 0.03256225 Yield
```

The Smith-Hazel Index

Suppose we wish to maximize the response on the linear combination $\mathbf{a}^T \mathbf{z}$ of traits

This is done by selecting those individuals with the largest breeding value for this index, namely

$$H = \mathbf{a}^T \mathbf{g} = \sum a_i g_i$$

Hence, we wish to find a phenotypic index $I = \mathbf{b}^T \mathbf{z}$ so that the correlation between H and I in an individual is maximized.

Key: This vector of weights \mathbf{b} is typically rather different from \mathbf{a} !

The Smith-Hazel Index

Smith and Hazel show that the maximal gain is obtained by selection on the index $\mathbf{b}^T \mathbf{z}$, where $\mathbf{b}_s = \mathbf{P}^{-1} \mathbf{G}^T \mathbf{a}$

Thus, to maximize the response in $\mathbf{a}^T \mathbf{z}$ in the next generation, we choose those parents with the largest breeding value for this trait, $H = \mathbf{a}^T \mathbf{g}$. This is done by choosing individuals with the largest values of $I = \mathbf{b}_s^T \mathbf{z}$.

$$\begin{aligned} I_s &= \mathbf{b}_s^T \mathbf{z} \\ &= \left(\mathbf{P}^{-1} \mathbf{G}^T \mathbf{a} \right)^T \mathbf{z} \\ &= \mathbf{a}^T \mathbf{G} \mathbf{P}^{-1} \mathbf{z} \end{aligned}$$

In-class problem

We wish to improve our previous index

$$I = 2*\text{oil} - 3*\text{protein} + 4*\text{yield}$$

$$\mathbf{a} = \begin{pmatrix} 2 \\ -3 \\ 4 \end{pmatrix}$$

Compute the Smith-Hazel weights (using previous \mathbf{P}, \mathbf{G})

$$\mathbf{b}_s = \mathbf{P}^{-1} \mathbf{G}^T \mathbf{a}$$

```
> a<-matrix(c(2,-3,4),nrow=3)
> a
      [,1]
[1,]    2
[2,]   -3
[3,]    4
> solve(P) %*% t(G) %*% a
      [,1]
[1,]  5.205520
[2,] -8.262400
[3,]  3.312388
```

Hence, SH index is $I = 5.2*\text{oil} - 8.26*\text{protein} + 3.31 * \text{yield}$

Key: Because G rotates the vector of responses away from the desired gains b , selection in the Smith-Hazel direction optimizes gain in desired direction

Response on the component traits under a Smith-Hazel index are

$$R = \bar{z} \cdot \frac{G\mathbf{b}_s}{\sqrt{\mathbf{b}_s^T \mathbf{P} \mathbf{b}_s}}$$

Hence, the response in the index $I_a = \mathbf{a}^T \mathbf{z}$ is just $\mathbf{a}^T \mathbf{R}$, or

$$\bar{z} \cdot \frac{\mathbf{a}^T G\mathbf{b}_s}{\sqrt{\mathbf{b}_s^T \mathbf{P} \mathbf{b}_s}}$$

Now let's compare response under Smith-Hazel weights (\mathbf{b}_s) versus the response selecting on an index using direct weights (\mathbf{a})

Given selection using $I_b = \mathbf{b}^T \mathbf{z}$, the response in another index $I_a = \mathbf{a}^T \mathbf{z}$ is just $R_b = \mathbf{a}^T \mathbf{R}$, where

$$\mathbf{R} = \bar{l} \cdot \frac{\mathbf{G}\mathbf{b}}{\sqrt{\mathbf{b}^T \mathbf{P} \mathbf{b}}}$$

Response in I_a using SH weights,

```
> bs <- solve(P) %*% t(G) %*% a    Compute weights  $\mathbf{b}_s$ 
> t(bs) %*% P %*% bs              Compute variance of index,  $\mathbf{b}_s^T \mathbf{P} \mathbf{b}_s$ 
      [,1]
[1,] 13448.47
> t(a) %*% G %*% bs              Compute  $\mathbf{a}^T \mathbf{G} \mathbf{b}_s$ 
      [,1]
[1,] 13448.47
> 2.06*13448.47/sqrt(12448.47)    Compute Reponse
[1] 248.3031
```

From before, response in I_a given selection using I_a is ~ 207 . Hence $238.9/207 = 1.15$, a 15% gain in I_a by selecting on a DIFFERENT index, I_b , whose weights are given by the Smith Hazel result

Direct vs. indirect response

If we select using the index $I = 2 \cdot \text{oil} - 3 \cdot \text{protein} + 4 \cdot \text{yield}$, (i.e., choose individuals with the highest value of I), then the direct response in I is 207

Conversely, if we select using the SH index
 $I_{SH} = 5.2 \cdot \text{oil} - 8.26 \cdot \text{protein} + 3.31 \cdot \text{yield}$
then the correlated response in the index I is 238

Hence, the correlated response in I when selection is based on I_{SH} gives a 15% larger response than that direct response from selection using I

Problem with Smith-Hazel

Require accurate estimate of P, G !

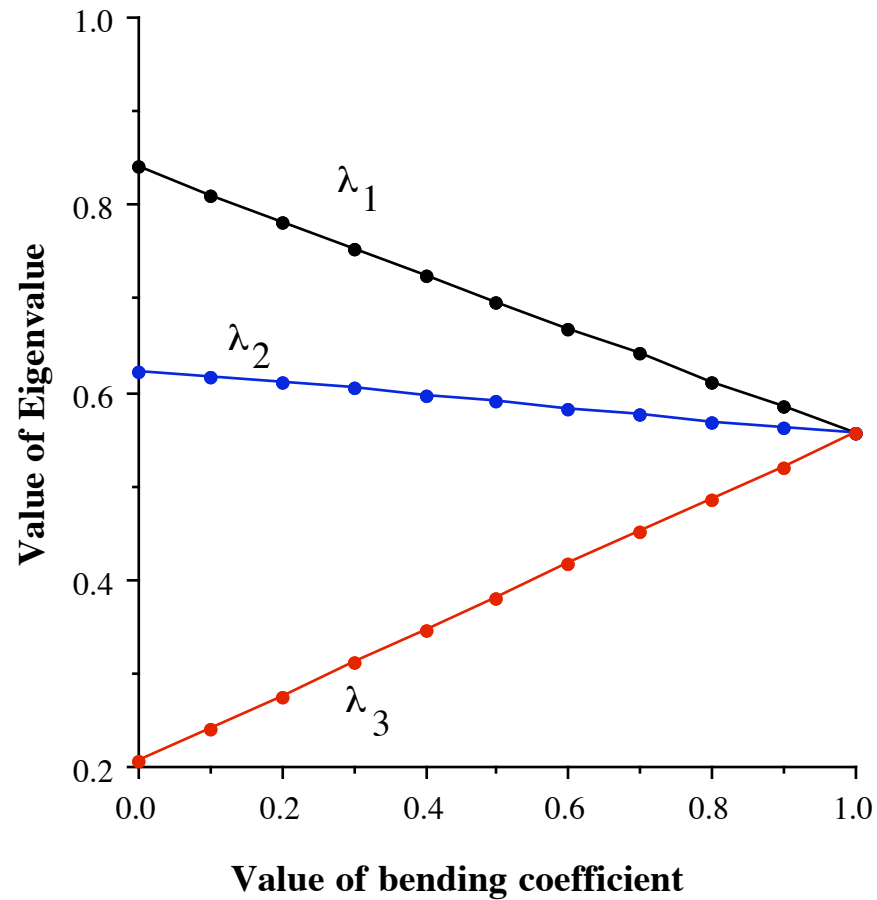
Values of P, G change each generation from LD!

Dealing with negative eigenvalues of $H = P^{-1}G$

Bending: $\hat{H}^* = (1 - \gamma) \cdot \hat{H} + \gamma \cdot \bar{\lambda} \mathbf{I}$ for $0 \leq \gamma \leq 1$

Net result is to regress eigenvalues back to their mean

The effect of bending is to regress all eigenvalues back to their mean



Other indices

The Smith-Hazel weights require accurate estimates of \mathbf{P} and \mathbf{G} . The **estimated index** is an approximation of the SH index based on estimated \mathbf{P} and \mathbf{G} values.

Estimated index:
$$\hat{I}_s = \hat{\mathbf{P}}^{-1} \hat{\mathbf{G}} \mathbf{a}^T \mathbf{z}$$

The **base index** is simply using the economic, as opposed to Smith-Hazel, weights $I_b = \sum a_i z_i$

The **heritability index** is based on the idea that traits with higher heritabilities provide more information as to their breeding values than do low heritability traits, weights are $b_i = a_i h^2_i$

Other indices

The **Elston** (or **weight-free** or **multiplicative**) **index** has no economic weights, but instead uses a minimal acceptable value m_i for each trait,

$$I_e = (z_1 - m_1)(z_2 - m_2) \cdots (z_n - m_n) = \prod_{j=1}^n (z_j - m_j)$$

The **retrospective index** finds the weights that were used to obtain the observed response. Since $R = Gb$,

$$b = G^{-1}R$$

Restricted and Desired-gains Indices

Instead of maximizing the response on a index, we may (additionally) wish to restrict change in certain traits or have a desired gain in specific traits in mind

Suppose for our soybean bean data, we wish no response in trait 1

Class Problem: Suppose weights are $a^T = (0,1,1)$, i.e., no weight on trait 1. Compute the response in trait one under this index.

Morely (1955): Change trait z_1 while z_2 constant

$$\begin{pmatrix} R_1 \\ R_2 \end{pmatrix} = \begin{pmatrix} G_{11} & G_{12} \\ G_{21} & G_{22} \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

Want b_1, b_2 such that
no response in trait 2

$$R_2 = G_{21}b_1 + G_{22}b_2 = 0$$

Setting $b_1 = 1$ and
solving gives weights

$$\begin{pmatrix} b_1 \\ b_2 \end{pmatrix} = \begin{pmatrix} 1 \\ -\frac{G_{21}}{G_{22}} \end{pmatrix}$$

$$I_r = z_1 - \left(\frac{\sigma_{g_1, g_2}}{\sigma_{g_2}^2} \right) \cdot z_2$$

Kempthorne-Nordskog restricted index

Suppose we wish the first k traits to be unchanged.

Since response of an index is proportional to $G\mathbf{b}$, the constraint is $\mathbf{C}G\mathbf{b}_r = \mathbf{0}$, where \mathbf{C} is an $k \times m$ matrix with ones on the diagonals and zeros elsewhere.

Kempthorne-Nordskog showed that \mathbf{b}_r is obtained by

$$\mathbf{b}_r = \left[\mathbf{I} - \mathbf{P}^{-1}\mathbf{G}_r^T \left(\mathbf{G}_r\mathbf{P}^{-1}\mathbf{G}_r^T \right)^{-1} \mathbf{G}_r \right] \mathbf{P}^{-1}\mathbf{G}^T \mathbf{a}$$

where $\mathbf{G}_r = \mathbf{C}G$

Example 33.8. Consider the soybean data from Example 33.1. Suppose that character z_1 (oil content) is at its optimal value, but we wish to optimize the sum of protein content and yield, $H = g_2 + g_3$. Here,

$$\mathbf{C} = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix} \quad \text{and} \quad \mathbf{a} = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}$$

Applying Equation 33.36a gives

$$\mathbf{b}_r = \begin{pmatrix} -3.3 \\ -0.8 \\ 1.1 \end{pmatrix}$$

Hence $\sigma(I_r) = \sqrt{\mathbf{b}_r^T \mathbf{P} \mathbf{b}_r} \simeq 23.2$ and from Equation 33.5 the expected change in the vector of character means is

$$\mathbf{R} = \frac{\bar{z}}{\sigma(I_r)} \cdot \mathbf{G} \mathbf{b}_r = \bar{z} \cdot \begin{pmatrix} 0 \\ 4.0 \\ 11.2 \end{pmatrix}$$

The resulting response in merit becomes

$$H = \bar{z} (1 \cdot 4.0 + 1 \cdot 11.2) = \bar{z} \cdot 15.2$$

If instead of restricting the change in character one, the Smith-Hazel index is applied with \mathbf{a} as above so that no weight is placed on changes in z_1 (i.e., we do not care what values they take), then \mathbf{a} is as above and

$$\mathbf{b}_s = \mathbf{G}\mathbf{P}^{-1}\mathbf{a} = \begin{pmatrix} 0.38 \\ 0.16 \\ 1.17 \end{pmatrix} \quad \text{giving} \quad \mathbf{R} = \frac{\bar{z}}{\sigma(I_s)} \cdot \mathbf{G}\mathbf{b}_s = \bar{z} \cdot \begin{pmatrix} 6.4 \\ 9.2 \\ 27.3 \end{pmatrix}$$

The response in H is now $36.5 \cdot \bar{z}$ under the Smith-Hazel index, as compared with only $15.2 \cdot \bar{z}$ under the restricted index. The price paid for no change in character one is that the expected response in H is only 42 percent of that under no restrictions. This result is fairly typical, in that using a restricted index often results in a rather significant decrease in the expected merit.

Tallis restriction index

More generally, suppose we specify the desired gains for k combinations of traits,

$$L_1 = \sum_{j=1}^k c_{j1} R_j = d_1, \quad L_2 = \sum_{j=1}^k c_{j2} R_j = d_2, \quad \dots, \quad L_k = \sum_{j=1}^k c_{jk} R_j = d_k$$

Here the constraint is $CGb_r = d$

Solution is

$$b_r = \left[I - P^{-1} G_r^T \left(G_r P^{-1} G_r^T \right)^{-1} G_r \right] P^{-1} G^T a + P^{-1} G_r^T \left(G_r P^{-1} G_r^T \right)^{-1} d$$

Desired-gain indices

Suppose we desire a gain of R , then since $R = Gb$, the **desired-gains index** is $b = G^{-1}R$

Example 33.9. Once again using the soybean data from Example 33.1, suppose we wish to increase $(z_1, z_2, z_3) = (\text{oil content, protein content, yield})$ by relative amounts $(1 : 1 : 1)$ giving the vector of desired gains as

$$\Delta_d = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

Using Equation 33.39, and rescaling b_d so that its first element is one gives

$$b_d = \begin{pmatrix} 1.0 \\ 0.8 \\ -0.5 \end{pmatrix}$$

Selection on the index $I_d = z_1 + 0.8 \cdot z_2 - 0.5 \cdot z_3$ thus gives the same response in each character. To verify this, first note that $\sigma(I_d) = \sqrt{b_d^T P b_d} \simeq 5.45$, giving

$$R = \left(\frac{\bar{z}}{\sigma(I_d)} \right) \cdot G b_d = 2.22 \bar{z} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

To distinguish between a desired-gains and Smith-Hazel index, compare this response with that expected from the Smith-Hazel index on the merit function $z_1 + z_2 + z_3$. Under the desired-gains index, the response in merit is $\Delta\mu_1 + \Delta\mu_2 + \Delta\mu_3 = 6.66 \cdot \bar{z}$, only 13 percent of the expected response of $49.67 \cdot \bar{z}$ under the Smith-Hazel index (Example 33.7), illustrating the cost of specifying response in each character as opposed to just being concerned with the maximal response in merit. Now suppose that while all three characters are measured, we are only interested in having equal response in oil and protein content and are unconcerned with yield (z_3). Here

$$\Delta_d = \begin{pmatrix} 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \mathbf{G} = \begin{pmatrix} 128.7 & 160.6 & 492.5 \\ 160.6 & 254.6 & 707.7 \end{pmatrix}$$

Applying Equation 33.40,

$$\mathbf{b}_d = \mathbf{P}^{-1} \mathbf{G}^T (\mathbf{G} \mathbf{P}^{-1} \mathbf{G}^T)^{-1} \Delta_d = \begin{pmatrix} 0.0131 \\ -0.0049 \\ 0.0002 \end{pmatrix}$$

This gives $\sigma^2(I_d) = \mathbf{b}_d^T \mathbf{P} \mathbf{b}_d = 0.01286$, so that the vector of responses is

$$\mathbf{R} = \left(\frac{\bar{z}}{\sigma(I_d)} \right) \cdot \mathbf{G} \mathbf{b}_d = 8.82 \bar{z} \begin{pmatrix} 1.0 \\ 1.0 \\ 3.4 \end{pmatrix}$$

giving the change in the merit as $\bar{z} \cdot 8.82 \cdot (1 + 1 + 3.4) = 47.63 \cdot \bar{z}$, 96 percent of the response under the Smith-Hazel index.

Class problem

Compute b for desired gains (in our soybean traits)
of $(1, -2, 0)$

Lush's index:

- Suppose you have information on both individual values and family means. What do you selection on?
 - Within-family deviations ($z - \bar{z}$), where \bar{z} is family mean (n sibs)
 - Between-family deviations (families with largest \bar{z} s)
 - Individual selection (within + between, $(z - \bar{z}) + \bar{z} = z$)
 - All examples of the index $I = b_1^*(z - \bar{z}) + b_2 \bar{z}$, which can also be written as $I = b_1 z + b_2 \bar{z}$
- The Lush index uses the Smith-Hazel weights to obtain the optimal values of b_1 and b_2 (derivation in WL Chapter 33)

Define the family intra-class correlation t by

$$t = r_A h^2 + c^2 \quad \text{with} \quad \frac{c^2}{\sigma_z^2} = \begin{cases} \sigma_{Ec(HS)}^2 & \text{for half-sibs} \\ \sigma_D^2/4 + \sigma_{Ec(FS)}^2 & \text{for full-sibs} \end{cases}$$

$$t_n = t + \frac{1-t}{n} \quad \text{and} \quad r_{A,n} = r_A + \frac{1-r_A}{n}$$

The Lush index is given by

$$I = (z - \bar{z}_f) + \left(\frac{r_{A,n}}{t_n} \right) \left(\frac{1-t}{1-r} \right) \bar{z}_f$$

The increase in response over mass selection (on z only) is

$$\sqrt{1 + \frac{(n-1)(t-r_A)^2}{(1-t)[1+t(n-1)]}}$$

Information for relatives

- One can also improve the prediction of the breeding value of an individual by also using information from all their relatives (details in WL Chapter 33).
 - $I_1 = b_1 z_1 + b_2 z_2 + \dots + b_n z_n$, where z_i is the value of i th relative of individual one
 - Predicted breeding value (of individual i) of the form $h^2 \mathbf{g}^T \mathbf{P}^{-1}(\mathbf{z} - \mathbf{u})$
- This is a prelude to BLUP, which uses all the information from measured relatives and allows for general fixed effects

Using Information from markers

- Marker-assisted selection (MAS) is using a selection index incorporating marker information to improve estimation of breeding value
 - Neimann-Sorensen & Robertson (1961), Smith (1967) developed single-marker indices
 - Lande & Thompson (1990) developed a general index over k markers. They regressed breeding value (g) over the marker values (\mathbf{m}) from an individual

$$g = a + \sum_{i=1}^n \alpha_i \cdot m_i = a + \boldsymbol{\alpha}^T \mathbf{m}$$

Lande and Thompson defined an individual's marker score m as

$$m = \sum_{i=1}^n \alpha_i \cdot m_i = \boldsymbol{\alpha}^T \mathbf{m}$$

Let ρ be the total fraction of genetic variance accounted for by the marker information

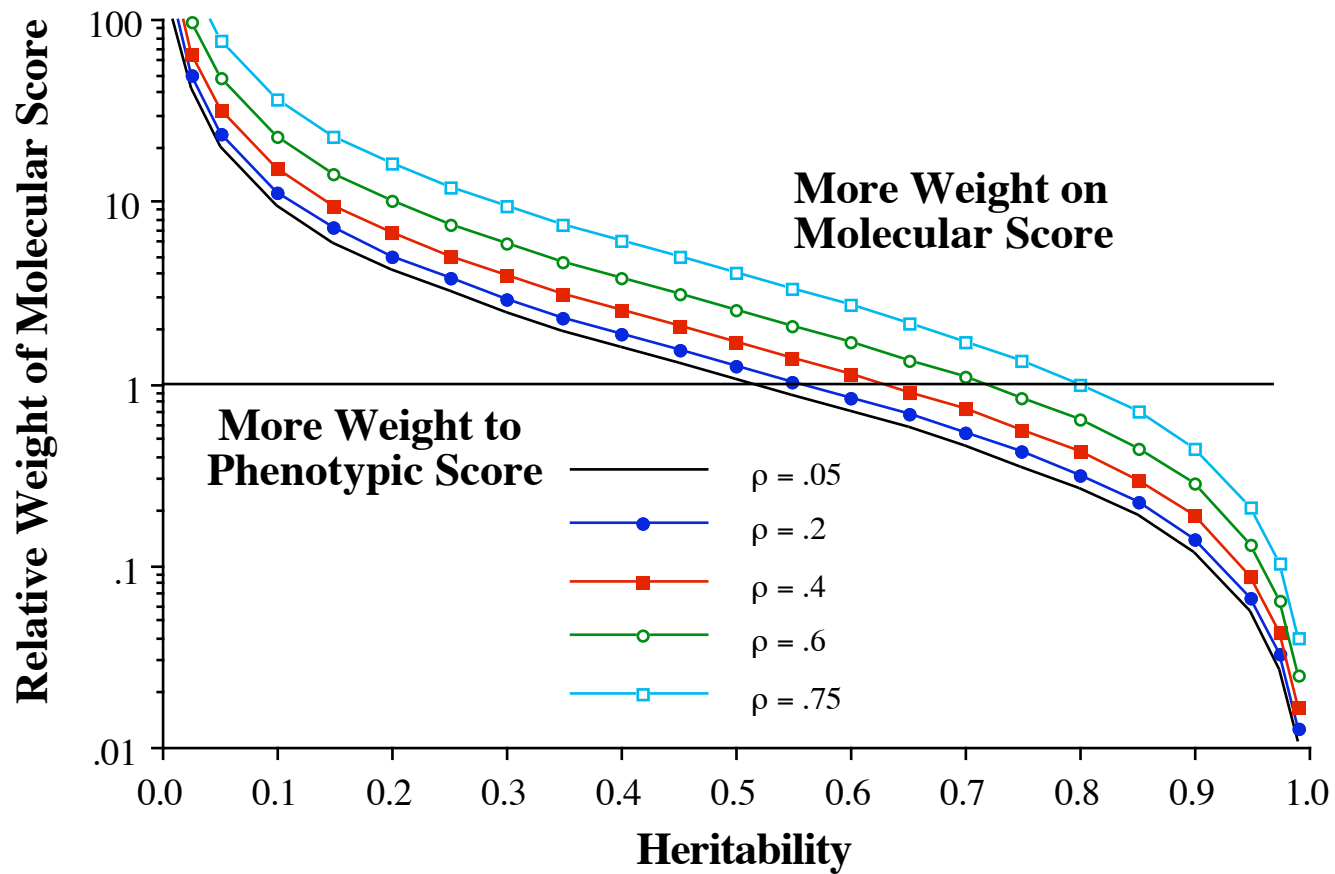
Lande and Thompson found that the Smith-Hazel weights on phenotype (z) and marker score (m) gives the optimal index as

$$I_s = z + \left(\frac{1 - h^2}{h^2(1 - \rho)} \right) \cdot m$$

Increase in response using index as opposed to mass selection

$$\sqrt{1 + \frac{(1 - h^2)^2 \rho}{h^2(1 - h^2 \rho)}}$$

Relative weights on phenotype and marker score



Non-linear indices

Care is required in considering the improvement goals when using a nonlinear merit function, as apparently subtle differences in the desired outcomes can become critical

With a linear index, the mean of the index equals the index evaluated at the mean, e.g.,

$$E[H(\mathbf{g})] = \mathbf{a}^T \boldsymbol{\mu} = H(\boldsymbol{\mu}) = H(E[\mathbf{g}])$$

With a nonlinear index, $E[H(\mathbf{g})] \neq H(E[\mathbf{g}])$, and thus we must decide if our the goal to improve the *average merit of all individuals* in the population $E[H]$ or to improve the *merit of the population average* $H(E[\mathbf{z}]) = H(\boldsymbol{\mu})$. These goals are equivalent under a linear merit function (as $E[H] = H[\boldsymbol{\mu}]$), but are generally different when nonlinear merit functions are used. For example, suppose $H = g^2$ so that $E[g^2] = \sigma_g^2 + \mu^2 \geq \mu^2 = (E[g])^2$.

A related concern is whether we wish to **maximize the additive genetic value in merit in the parents or in their offspring**. Again, with a linear index these are equivalent, as the mean breeding value of the parents μ equals the mean value in their offspring.

This is not the case with nonlinear merit.

Quadratic selection index

$$\begin{aligned} H &= c + \sum_{i=1}^k a_i (\mu_i + g_i) + \sum_{i=1}^k b_{ii} (\mu_i + g_i)^2 + \frac{1}{2} \sum_{i=1}^k \sum_{j=1}^k b_{ij} (\mu_i + g_i) (\mu_j + g_j) \\ &= c + \mathbf{a}^T (\boldsymbol{\mu} + \mathbf{g}) + (\boldsymbol{\mu} + \mathbf{g})^T \mathbf{A} (\boldsymbol{\mu} + \mathbf{g}) \end{aligned}$$

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12}/2 & \cdots & a_{1k}/2 \\ a_{12}/2 & a_{22} & \cdots & a_{2k}/2 \\ \vdots & \vdots & \ddots & \vdots \\ a_{1k}/2 & a_{2k}/2 & \cdots & a_{kk} \end{pmatrix}$$

$$I = \alpha + \mathbf{b}^T \mathbf{z} + \mathbf{z}^T \mathbf{B} \mathbf{z}$$

$$\mathbf{b} = \mathbf{P}^{-1} \mathbf{G}^T (\mathbf{a} + 2\mathbf{A}\boldsymbol{\mu})$$

$$\mathbf{B} = \mathbf{P}^{-1} \mathbf{G}^T \mathbf{A} \mathbf{G} \mathbf{P}^{-1}$$

$$H = c + \mathbf{a}^T (\boldsymbol{\mu} + \mathbf{g}) + (\boldsymbol{\mu} + \mathbf{g})^T \mathbf{A} (\boldsymbol{\mu} + \mathbf{g})$$

Let's expand H in a first-order Taylor series (Equation A5.6). Taking the vector of first derivatives of the linear term with respect to \mathbf{g} gives

$$\nabla_{\mathbf{g}} [\mathbf{a}^T (\boldsymbol{\mu} + \mathbf{g})] = \mathbf{a}$$

which follows from Equation A5.1a. Equation A5.1e gives the derivative for the quadratic term as

$$\nabla_{\mathbf{g}} [(\boldsymbol{\mu} + \mathbf{g})^T \mathbf{A} (\boldsymbol{\mu} + \mathbf{g})] = 2\mathbf{A}(\boldsymbol{\mu} + \mathbf{g})$$

When evaluated $\mathbf{g} = 0$, the linearized contribution from the quadratic becomes $2\mathbf{A}\boldsymbol{\mu}$. Hence, the linear approximation of the merit function has the form

$$H(\mathbf{g}) \simeq H(\mathbf{0}) + \nabla_{\mathbf{g}}^T(H) \Big|_{\mathbf{g}=\mathbf{0}} \mathbf{g} = H(\mathbf{0}) + (\mathbf{a} + 2\mathbf{A}\boldsymbol{\mu})^T \mathbf{g} \quad (33.43)$$

and this are the best linear approximation of the weights. Thus the linear term of the quadratic index (Equation 33.42b) is just a Smith-Hazel index now using the best linear approximation for the function, Equation 33.43, as the weights. As the population mean changes under selection, so to do the weights. Note that the best linear approximation is a *local* one, and no longer holds if one considers large differences in \mathbf{g} from around the mean. In such cases, the first-order Taylor series may no longer be a good approximation and other approaches are required to find the best linear weights (as detailed below).

Fitting the best linear index to a nonlinear merit function

How might the best linear index be obtained? The simplest situation is when the nonlinear index can be transformed into a linear index, which can then be maximized using the standard linear theory. For example, if the merit function is a simple polynomial, say $H(\mathbf{g}) = g_1 + g_2^2 + g_3$, by defining $\tilde{g}_2 = g_2^2$ the index becomes linear, viz. $H(\mathbf{g}) = g_1 + \tilde{g}_2 + g_3$. This approach was first hinted at by Smith (1936) and formally suggested by Kempthorne and Nordskog (1959), but requires the phenotypic and additive genetic covariances of the transformed variables.

A more general approach is to first obtain the best linear approximation of a nonlinear merit using a first-order Taylor series about the population mean,

$$H(g_1, g_2, \dots, g_n) \simeq H(\boldsymbol{\mu}) + \sum_{i=1}^n a_i (g_i - \mu_i) \quad \text{where} \quad a_i = \left. \frac{\partial H}{\partial g_i} \right|_{\mathbf{g}=\boldsymbol{\mu}} \quad (33.44a)$$

This provides the (current) economic weights for the best linear approximation of the merit. The resulting optimal index is then given by a Smith-Hazel index using these weight (Moav and Hill 1966, Harris 1970), namely

$$\mathbf{a} = \nabla_{\mathbf{g}}[H] \big|_{\mathbf{g}=\boldsymbol{\mu}} \quad (33.44b)$$

Example 33.11. Moav and Hill (1966) introduced the following nonlinear merit function:

$$H = a - bg_1 - \frac{c}{g_2}$$

Here,

$$\nabla_{\mathbf{g}}(H) = \begin{pmatrix} \partial H / \partial g_1 \\ \partial H / \partial g_2 \end{pmatrix} = \begin{pmatrix} -b \\ c/g_2^2 \end{pmatrix}$$

Hence $\nabla_{\mathbf{g}}^T(H)|_{\boldsymbol{\mu}} = (-b \quad c/\mu_2^2)^T$ giving the best linear approximation of H around the current population mean as

$$H \simeq a - bg_1 + \frac{c}{\mu_2^2} g_2$$

The resulting vector of weights for the Smith-Hazel index become

$$\mathbf{a} = \begin{pmatrix} -b \\ c/\mu_2^2 \end{pmatrix}$$

In the next generation, we would update with the new mean and proceed. In a real experiment, would estimate the new mean each generation and use this for the new value of μ_2 . If our goal instead is to predict the expected response after some number of generations, then Equation 33.19 can be used to compute the change in the (linearized) merit, while Equation 33.20 predicts the change in the components (g_1, g_2) of the merit. Thus, in the next generation the weight on g_2 would be $c/(\mu_2 + R_2)^2$ where R_2 is obtained from Equation 33.20. Proceeding in this fashion, one could iterate the expected response out to any desired generation.

Optimal multi-generation response on nonlinear index

- Goddard (1983) noted that a simple solution allows one to optimize the total response in any nonlinear index by constant selection on an approximate linear index
- Key: Given the total selection intensity (k generations * i per generation) defines a surface of potential responses. This allows for a graphical solution of the optimal linear index to achieve this response

Sets of response vectors and selection differentials

Given a fixed selection intensity i , we would like to know the set of possible \mathbf{R} (response vectors) or \mathbf{S} (selection vectors)

$$\mathbf{R} = \frac{\bar{i}}{\sigma_I} \mathbf{G} \mathbf{b}, \quad \text{implying} \quad \mathbf{b} = \frac{\sigma_I}{\bar{i}} \mathbf{G}^{-1} \mathbf{R}$$

Substituting the above value of \mathbf{b} into $\sigma_I^2 = \mathbf{b}^T \mathbf{P} \mathbf{b}$

$$\text{recovers} \quad \sigma_I^2 = \left(\frac{\sigma_I}{\bar{i}} \right)^2 \mathbf{R}^T \mathbf{G}^{-1} \mathbf{P} \mathbf{G}^{-1} \mathbf{R}$$

Sets of response vectors and selection differentials

Hence, $\bar{i}^2 = \mathbf{R}^T \mathbf{G}^{-1} \mathbf{P} \mathbf{G}^{-1} \mathbf{R}$

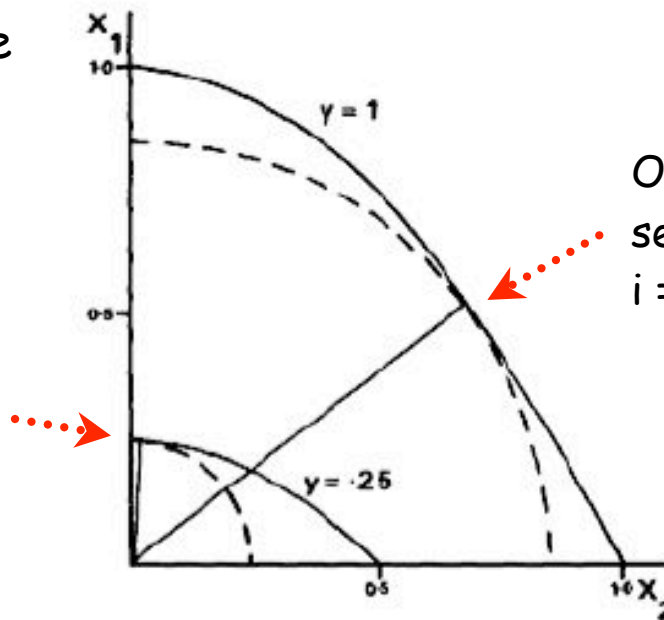
This equation describes a quadratic surface of possible, R values,

$$\sum_i \sum_j R_i R_j A_{ij} = \bar{i}^2, \quad \text{where} \quad \mathbf{A} = \mathbf{G}^{-1} \mathbf{P} \mathbf{G}^{-1}$$

Similarly, $\bar{i}^2 = \mathbf{S}^T \mathbf{P}^{-1} \mathbf{G} \mathbf{G}^{-1} \mathbf{P} \mathbf{G}^{-1} \mathbf{G} \mathbf{P}^{-1} \mathbf{S} = \mathbf{S}^T \mathbf{P}^{-1} \mathbf{S}$

Solid curves indicate surfaces (sets of equal value) of the nonlinear index

Optimal weights when selection intensity $i = 0.25$



Dashed lines denote the response surface for a given selection intensity

Figure 33.3. Two contours (solid lines) of the merit function $H = g_1 + g_2^2$ are plotted (corresponding to values of 0.25 and 1). The dashed lines are the response surfaces corresponding to two different intensities of selection. The intersection between the merit and response surfaces gives the optimal linear weights on both traits. Note that under weak selection, most of the weight is on g_1 , while with stronger selection, there is roughly equal weight on g_1 and g_2 . Thus, the optimal index weights are a function of the intensity of selection. After Goddard (1983).

Key: Optimal weights a function of intensity of selection

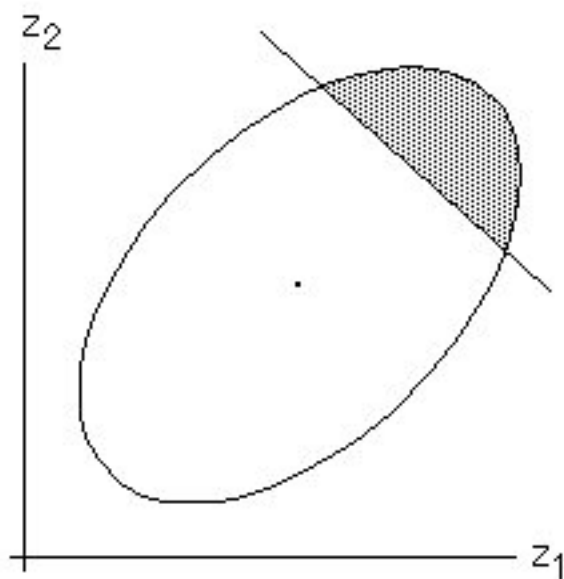
Sequential Approaches for Multitrait selection

Tandem Selection

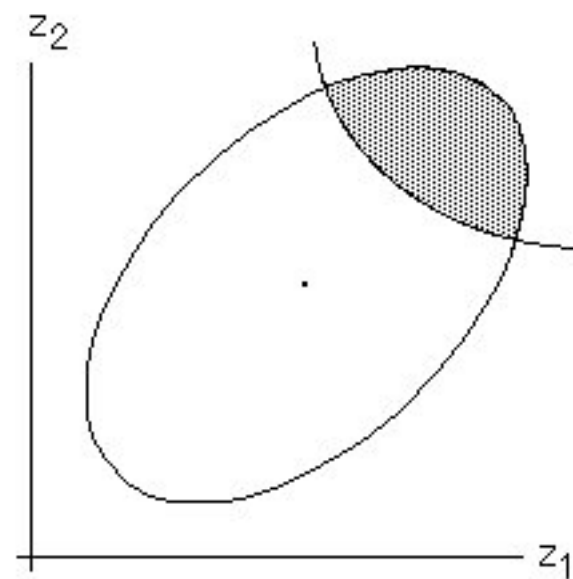
Independent Culling

Selection of extremes

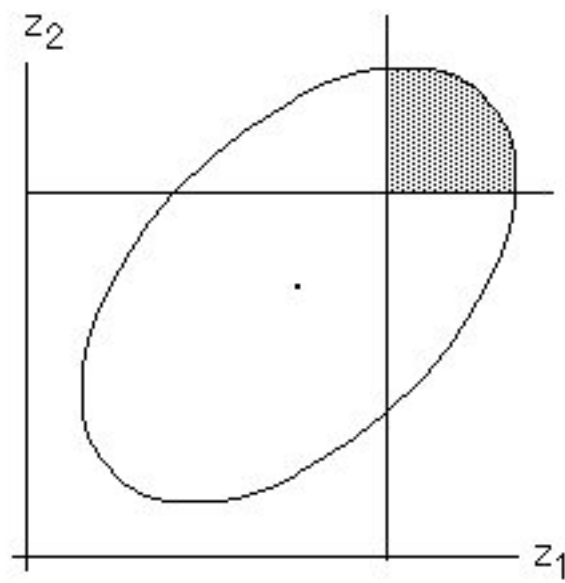
Index selection generally optimal, but may be economic reasons for other approaches



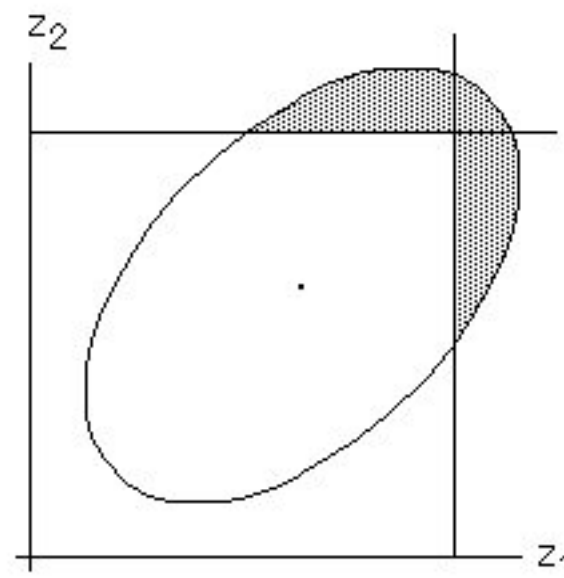
(Linear) Index Selection



Selection using the Elston Index



Independent Culling



Selection of extremes

Multistage selection

Selecting on traits that appear in different life stages

Optimal selection schemes

Cotterill and James optimal 2-stage. Select 1st on x , (save p_1), then select on y (save $p_2 = p/p_1$). Goal optimize the breeding value for merit g .

$$\frac{\Delta g}{\sigma(g)} = \bar{i}_1 \rho_{x,g} + \bar{i}_2 \left(\frac{\rho_{y,g} - \rho_{y,x} \rho_{x,g} \kappa_1}{\sqrt{1 - \rho_{x,y}^2 \kappa_1}} \right) \quad (33.50)$$

Equation 33.50 is then numerically maximized subject to the constraint that $p_1 p_2 = p$. Note that we can simply plot this as a function of p_1 , as p_1 determines both \bar{i}_1 and κ_1 , while $p_2 = p/p_1$ determines \bar{i}_2 .