RESTRICTED SELECTION INDICES[1]

OSCAR KEMPTHORNE AND ARNE W. NORDSKOG

Iowa State College
Ames, Iowa, U.S.A.

Introduction

Smith [1936] developed a discriminant function for the selection of varieties according to their genotypic values in the presence of errors of observations. Smith's argument is applicable to the selection of varieties which are self-perpetuating and, essentially, involves no genetic theory. Hazel [1943] extended this technique to the case when one wishes to select individuals whose progeny will be of superior merit by assuming that each individual has a true unknown "breeding value" (half of which is expressed in its offspring) and the correlations of "breeding values" with observed phenotypic expressions are known. It is not our purpose to expound the assumptions necessary for utility of such a genetic index, but it appears that predictions of increase in population mean from selection and mating of selected individuals will be correct apart from sampling and estimation error if the following conditions hold:

(1) the phenotypic value (P_i) defined as the observed value of attribute i for an individual shall be made up additively of two parts, a genotypic value (G_i) defined as the average of the phenotypic values possible over a population of environments and an environmental contribution (E_i), i.e., $P_i = G_i + E_i$. It is possible to permit interactions of genotype and environment provided genotypes and environments are associated entirely at random and any such interaction then is incorporated into E_i.

(2) the genotypic value G_i is composed entirely of additive effects of genes and is then, appropriately, also called the breeding value.

(3) with attributes denoted by $i = 1, 2, \cdots, m$, the genotypic economic value of an individual is $H = \sum a_i G_i$, where the a_i

[1]Journal Paper No. J-3462 of the Iowa Agricultural and Home Economics Experiment Station Ames, Iowa. Projects 890 and 1039.

are known economic constants, and G_i, $i = 1, 2, \cdots, m$, are the genotypic values of the individual for the m attributes.

(4) the quantities P_i and H are such that the regression of H on any linear function of the P_i is linear.

Condition (2) above will rarely hold and for development of an index it is assumed that G_i represents the breeding value which is the least squares prediction of G_i on the basis of a model containing only additive effects of genes, and that deviations of actual G_i from predicted values behave completely as random variables. With this assumption the offspring of a mating will have a genotypic value equal to the average of the breeding values of the parents. The symbol G_i is used in this paper for both genotypic value and breeding value, but it should be borne in mind that this usage is based on an approximation.

Under these circumstances, if selection based on a linear function $I = \sum_i b_i P_i$ is made, the gain in H is equal to $\beta_{HI} \Delta_I$ where Δ_I is the difference in mean I resulting from the selection, i.e., the difference of the mean of I after selection and the mean before selection, which is the selection differential. This can also be written as

$$\frac{\text{Cov}\,(H, I)}{\sigma_I^2}\,\Delta_I = \rho_{HI}\,\frac{\sigma_H}{\sigma_I}\,\Delta_I = \rho_{HI}\sigma_H\left(\frac{\Delta_I}{\sigma_I}\right).$$

The above expression shows that the genetic change due to selection is proportional to ρ_{HI} (not ρ_{HI}^2, as might at first be thought) and to Δ_I/σ_I which is the selection differential in the index in standard deviation units. If I is in addition normally distributed, the b_i's are known without error and truncation selection with a saving of the top p per cent is used, then the relationship of p to Δ_I/σ_I is given by the two equations

$$p = \int_k^\infty \frac{1}{\sqrt{2\pi}}\, e^{-\frac{1}{2}x^2}\, dx$$

which defines k as the normit (or probit minus 5) corresponding to p and

$$\frac{\Delta_I}{\sigma_I} = \frac{1}{p}\int_k^\infty \frac{1}{\sqrt{2\pi}}\, x e^{-\frac{1}{2}x^2}\, dx = \frac{e^{-\frac{1}{2}k^2}}{\sqrt{2\pi}\, p} = \frac{z}{p}$$

where z is the ordinate of the normal distribution above which there is the fraction p of the distribution. The choice of selection index then reduces to finding that linear function $I = \sum_i b_i P_i$ which correlates best with the index H. As shown by Smith and Hazel, the equations determining the b's are $Pb = Ga$ where P is the matrix of phenotypic variances and covariances, the ijth element P_{ij} being the covariance of P_i and P_j, and G is the matrix of genotypic variances and covariances so that G_{ij} is the covariance of G_i and G_j and a is the column array of economic constants or weights.

With the stated assumptions, the index $\sum b_i P_i$ obtained maximizes gain in the genotypic economic value H when errors of estimation are ignored. While use of I will result in best progress in H, the means of the G_i will change in either a positive or negative direction, so that a breeder may well be interested in increasing H as much as possible with a restriction that some G_i or some linear functions of the G_i will *not* change. For example, a poultry breeder may feel that he should keep mean egg size at a constant intermediate level while using an index to maximize progress in genetic economic value based on egg weight, body weight, and production. It was in fact such a situation which led to the present note.

The Mathematical Solution

We shall suppose that the breeder wishes to maximize progress in H given by

$$H = a_1 G_1 + a_2 G_2 + \cdots + a_m G_m ,$$

but he wishes no change in $r(< m)$ quantities V_i given by

$$V_1 = \sum_{j=1}^{m} c_{j1} G_j , \qquad V_2 = \sum_{j=1}^{m} c_{j2} G_j , \qquad \cdots \qquad V_r = \sum_{j=1}^{m} c_{jr} G_j ,$$

in which the coefficients a_j and c_{ij} , $i = 1, 2, \cdots , r; j = 1, 2, \cdots , n$ are given. A common circumstance would be that in which there is one linear function V_1 and only one coefficient c_{ij} is unity, the others being zero.

We are given $P_1 , P_2 , \cdots , P_m$, and are to determine the linear function $I = \sum b_i P_i$ so that the correlation of I with H is a maximum subject to the conditions

$$\text{Cov}(I, V_k) = 0, \qquad k = 1, 2, \cdots , r.$$

Now

$$\rho_{IH}^2 = \frac{\text{Cov}^2(I, H)}{\sigma_H^2 \sigma_I^2} = \frac{(\sum_{ij} a_i b_j G_{ij})^2}{(\sum_{ij} a_i a_j G_{ij})(\sum_{ij} b_i b_j P_{ij})}.$$

For brevity of manipulation it is convenient to utilize matrix notation as follows:

$$a = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_m \end{bmatrix}, \quad b = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}$$

$$c_1 = \begin{bmatrix} c_{11} \\ c_{21} \\ \vdots \\ c_{m1} \end{bmatrix}, \quad c_2 = \begin{bmatrix} c_{12} \\ c_{22} \\ \vdots \\ c_{m2} \end{bmatrix}, \cdots, c_r = \begin{bmatrix} c_{1r} \\ c_{2r} \\ \vdots \\ c_{mr} \end{bmatrix}.$$

Also, since the correlation ρ_{IH} is unaffected by multiplication of the b's by a constant, we take

$$\sum_{ij} b_i b_j P_{ij} = b'Pb = 1.$$

Then

$$\rho_{IH}^2 = \frac{(a'Gb)^2}{(a'Ga)},$$

$$\mathrm{Cov}(I, V_k) = b'Gc_k = c_k'Gb, \qquad k = 1, 2, \cdots, r,$$

and we have to maximize ρ_{IH}^2 subject to the r conditions

$$\mathrm{Cov}(I, V_k) = 0$$

and the condition $b'\,Pb = 1$.

Introducing $(r + 1)$ undetermined Lagrange multipliers, $2\lambda_k$ and ν, we have to maximize without restriction the quantity

$$\frac{(a'Gb)^2}{(a'Ga)} - 2 \sum_k \lambda_k(c_k'Gb) - \nu(b'Pb - 1)$$

with regard to the b's.

Differentiating with regard to b_i we have

$$2\,\frac{(a'Gb)}{(a'Ga)}\,(a'G)_i - 2 \sum_k \lambda_k(c_k'G)_i - 2(b'P)_i = 0.$$

Arranging these equations in matrix form we have

$$\frac{(a'Gb)}{(a'Ga)}\,(a'G) = \sum_k \lambda_k(c_k'G) + \nu b'P.$$

So

$$\nu Pb = \frac{(a'Gb)}{(a'Ga)} Ga - \sum_k \lambda_k Gc_k$$

or

$$b = \frac{1}{\nu} \frac{(a'Gb)}{(a'Ga)} P^{-1} Ga - \frac{1}{\nu} \sum_k \lambda_k P^{-1} Gc_k .$$

But

$$c_l' Gb = 0 \quad \text{for} \quad l = 1, 2, \cdots, r,$$

so

$$0 = \frac{(a'Gb)}{(a'Ga)} c_l' GP^{-1} Ga - \sum_k \lambda_k c_l' GP^{-1} Gc_k .$$

If we use the matrices $C = (c_1, c_2, \cdots, c_r)$ so that C' is $r \times m$ and C is $m \times r$ and $\lambda' = (\lambda_1, \lambda_2, \cdots, \lambda_r)$, then

$$C'GP^{-1}GC\lambda = \frac{(a'Gb)}{(a'Ga)} C'GP^{-1}Ga$$

so that

$$\lambda = (C'GP^{-1}GC)^{-1} \frac{(a'Gb)}{(a'Ga)} C'GP^{-1}Ga.$$

Also the equation defining b can be written

$$b = \frac{1}{\nu} \frac{(a'Gb)}{(a'Ga)} p^{-1}Ga - \frac{1}{\nu} P^{-1}GC\lambda$$

$$= \frac{1}{\nu} \frac{(a'Gb)}{(a'Ga)} [1 - P^{-1}GC(C'GP^{-1}GC)^{-1}C'G]P^{-1}Ga.$$

We may now relax the condition that $b'Pb$ equals unity and drop the term $(a'Gb)/\nu(a'Ga)$ which is merely a constant multiplier to all the b's to give the index in final form

$$b = [I - P^{-1}GC(C'GP^{-1}GC)^{-1}C'G]\, P^{-1}Ga.$$

We note that if there were no restrictions the index would be $b = P^{-1}Ga$, so that the effect of the restrictions is to multiply this unrestricted index by a matrix. Also we find

$$\rho_{IH}^2 = \frac{a'G[I - P^{-1}GC(C'GP^{-1}GC)^{-1}C'G]P^{-1}Ga}{a'Ga}.$$

We verify that the conditions $c_k'Gb = 0$ or matrix-wise $C'GB = 0$ are satisfied since

$$C'Gb = C'G[I - P^{-1}GC(C'GP^{-1}GC)^{-1}C'G]P^{-1}Ga$$

$$= [C'G - (C'GP^{-1}GC)(C'GP^{-1}GC)^{-1}C'G]P^{-1}Ga$$
$$= (C'G - C'G)P^{-1}Ga$$
$$= 0.$$

The results of using the index are as follows. The change in G_i resulting from a change Δ_I in the index is equal to

$$\beta_{G_i I}\Delta_I = \frac{\mathrm{Cov}(G_i, I)}{\sigma_I^2}\Delta_I = \frac{\mathrm{Cov}(G_i, I)}{\sigma_i}\frac{\Delta_I}{\sigma_I}.$$

This is equal to

$$\frac{\sum_j b_j G_{ij}}{\sigma_I}\frac{\Delta_I}{\sigma_I}$$

which can be worked out in any particular case.

Examples

The methods are illustrated below for a situation with poultry. The estimates of genetic parameters used are given in Tables 1 and 2.

TABLE 1

VARIANCES AND HERITABILITIES

		Variance	Heritability
BW	Mature body wt. (lbs.)	.34	.45
EW	Mature egg wt. (oz/doz.)	2.4	.50
SM	Age 1st egg (days)	13.	.40
BS	Blood spots (fraction)	50 $\times 10^{-6}$	.15
EP	Egg production (to 72 weeks)	512	.20

TABLE 2

PHENOTYPIC AND GENOTYPIC CORRELATIONS

	BW	*EW*	*SM*	*BS*	*EP*
BW		.25	0	0	−.05
EW	.30		0	0	−.05
SM	0	0		−.02	−.40
BS	0	0	−.07		+.03
EP	.15	.15	−.50	.08	

Phenotypic correlations are above the diagonal and genetic correlations below.

To ease the purely computational problem, the variables were coded to have variances of approximately the same order of magnitude as follows:

Attribute	Equal to
1	10 *BW*
2	3 *EW*
3	*SM*
4	*EP*/3
5	10^3 *BS*

This leads to matrices P and G which are of course symmetric:

$$P = \begin{bmatrix} 34 & 6.77504 & 0 & 0 & -2.19881 \\ 6.77504 & 21.6 & 0 & 0 & -1.75256 \\ 0 & 0 & 13 & -.50991 & -10.87709 \\ 0 & 0 & -.50991 & 50 & 1.59986 \\ -2.19881 & -1.75256 & -10.87706 & 1.59986 & 56.88 \end{bmatrix}$$

$$G = \begin{bmatrix} 15.30 & 3.855027 & 0 & 0 & 1.97825 \\ 3.85503 & 10.80 & 0 & 0 & 1.66260 \\ 0 & 0 & 5.20 & -.43716 & -3.84567 \\ 0 & 0 & -.43716 & 7.50 & .73894 \\ 1.97825 & 1.66260 & -3.84567 & .73894 & 11.376 \end{bmatrix}$$

The matrices P and G are the primary ingredients and the only requirement for obtaining the unrestricted index is to determine the economic weights for attributes 1 to 5.

The economic weights were chosen as indicated in Table 3.

The entry in the last column is obtained by noting, for example, that a change of 1 unit in 3 *EW* is equal to a change of $\frac{1}{3}$ unit in *EW* which is worth $21.6/3 = 7.2$. Alternatively, the index of value is

$$\begin{aligned} &-25.0\,BW + 21.6\,EW + 0\,SM - 7.9\,BS + 3.6\,EW \\ = &-2.50\,(10BW) + 7.20\,(3EW) + 0\,SM \\ &- 0.0079\,(10^3BS) + 10.80\,(EW/3). \end{aligned}$$

TABLE 3

ECONOMIC WEIGHTINGS

Trait	Unit	Real value (cents)	Coded trait	Value of unit change in coded trait
BW	1 lb.	−25.0	$x_1 = 10\ BW$	−2.50
EW	1 oz/doz.	21.6	$x_2 = 3\ EW$	7.20
SM	1 day	0	$x_3 = SM$	0
BS	1%	− 7.9	$x_4 = 10^3\ BS$	− .0079
EP	1 egg	3.6	$x_5 = EP/3$	10.80

The following calculations are presented in terms of coded traits and the index and changes may be decoded by substituting $10BW$ for x_1, $3EW$ for x_2, and so on.

The unrestricted selection index has coefficients which are the elements of $P^{-1}Ga$ and is

$$-0.3959x_1 + 4.2819x_2 - 1.4111x_3 + 0.0759x_4 + 2.1281x_5 .$$

Next we compute an index designed to lead to no change in attribute 1. For this case

$$C'G = (15.30 \quad 3.85503 \quad 0 \quad 0 \quad 1.97825),$$

i.e., the matrix consisting of the first row of G. The matrix $C'GP^{-1}GC$ has dimensions 1×1 and is a scalar equal to 71.09014 and insertion of the appropriate matrices in the expression for the coefficients of the index leads to the index

$$-1.3143x_1 + 4.1913x_2 - 1.5206x_3 + 0.1102x_4 + 1.9970x_5 .$$

As a second example, an index maximizing the gain in economic value but keeping x_1 and x_2 constant was constructed. For this

$$C'G = \begin{bmatrix} 15.30 & 3.85503 & 0 & 0 & 1.97825 \\ 3.85503 & 10.80 & 0 & 0 & 1.66261 \end{bmatrix}$$

consists of the first two rows of G and

$$C'G'P^{-1}GC = \begin{bmatrix} 7.10901 & 2.29706 \\ 2.29706 & 5.54508 \end{bmatrix}.$$

This leads to the index

$$-0.1685x_1 - 0.2007x_2 - 1.6007x_3 + 0.0861x_4 + 1.6944x_5 .$$

As a final example an index was computed, solely for illustrative purposes, keeping $x_1 + 0.1x_2$ constant. For this

$$C'G = (15.68550 \quad 4.93503 \quad 0 \quad 0 \quad 2.14451)$$

and $C'GP^{-1}GC = 7.62388$. This leads to the index

$$-1.5386x_1 + 4.0421x_2 - 1.5583x_3 + 0.0801x_4 + 1.9519x_5 .$$

The different cases are summarized in Table 4. These examples indicate the general effects of restriction of index. For instance, with increasing restriction the correlation of index with economic value decreases. The weights attached to some attributes are essentially unchanged with the imposition of restrictions while others change because of the particular correlational structure.

The important question of connection between the economic weights and the restrictions has been raised by a referee. One can imagine taking account of a desire to maintain egg weight, for example, at a nearly constant value of 24 by including in the total economic value

TABLE 4

SUMMARY OF EXAMPLES

	Unrestricted		x_1 constant		x_1 , x_2 constant		$x_1+0.1x_2$ constant	
	b*	Δ**	b	Δ	b	Δ	b	Δ
x_1	−0.39	0.55	−1.31	0	−0.17	0	−1.54	−0.16
x_2	4.28	1.82	4.19	1.68	−0.20	0	4.04	+1.60
x_3	−1.41	−0.59	−1.52	−0.31	−1.60	−0.92	−1.56	−0.61
x_4	0.08	0.10	0.11	0.11	0.09	0.16	0.08	0.11
x_5	2.13	1.36	2.00	1.28	1.69	1.54	1.95	1.25
ρ	.1873		.1832		.1174		.1805	

*The coefficient in the index.

**The expected genetic change in the attribute with selection differential equal 1 standard deviation unit of the index.

function a term such as $-k(\text{egg weight} - 24)^2$ where k is a prechosen number, instead of a term such as 21.6 egg weight, or $6EW$ or $18EW/3$. This would permit some change in egg weight but changes in either direction from the optimum (24 in the above illustration) would carry a negative value and the decrease in value would increase as the square of the deviation from the optimum value. This could be handled by the ordinary selection index process since the term

$$-k\ (\text{egg weight} - 24)^2$$

is equal to

$$-k\ (\text{egg weight})^2 + 48k\ \text{egg weight} - 576.$$

To carry out the estimation of index we would need to know the correlations, phenotypic and genotypic, of egg weight and (egg weight)^2 with each other and with the other attributes, as well as the phenotypic and genotypic variances of egg weight and (egg weight)^2. It would of course be a routine procedure to obtain the necessary estimates.

The procedure given here permits no genetic change in the chosen attribute, attributes, or linear functions of attributes, and may be unduly restrictive from some points of view. It is not entirely obvious but can be shown that if the restricted index is to keep say attribute 1 constant, then the weight associated with attribute 1 in the genotypic economic value is irrelevant.

Conclusions

The purpose of the present note is to give a derivation and examples of restricted selection indices. A further development of indices requiring some genetic changes to be of particular sign will be presented in a later paper.

REFERENCES

1. Smith, H. Fairfield [1936]. A discriminant function for plant selection. *Ann. Eugen. Lond.* *7*, 240–250.
2. Hazel, L. N. [1943]. The genetic basis for constructing selection indexes. *Genetics* *28*, 476–490.