Theor Appl Genet (1989) 78: 798–800

© Springer-Verlag 1989

A proposal to improve the efficiency of index selection by "rounding"

G. C. C. Tai

Agriculture Canada Research Station, Fredericton, N. B., E3B 4Z7 Canada

Received February 20, 1989; Accepted June 27, 1989
Communicated by A. R. Hallauer

Summary. This paper describes a rounding procedure to improve the efficiency of index selection. The procedure involves performing canonical variate analysis on the phenotypic and genotypic variances of a group of traits estimated from a progeny test experiment. The eigenvectors corresponding to the significant eigenvalues are used to transform the original traits into a set of independent variables. The selection index is then constructed based on the new set of variables. The efficiency of the new index is expected to be improved by rounding off the variables associated with the insignificant eigenvalues.

Key words: Index selection – Rounding

Introduction

One of the problems asssociated with index selection (Smith 1936; Hazel 1943) is that population parameters involved in calculating the selection indices have to be estimated from a sample of genetic material. Many authors (Williams 1962a, b; Harris 1964; Sales and Hill 1976a, b; Hayes and Hill 1980; Tai 1986) have studied the effect of sample size on the efficiency of index selection. Hayes and Hill (1981) proposed a "bending" method to modify the estimates of genotypic and phenotypic variance-covariance (VAR-COV) matrices used in the construction of selection indices. This method is especially effective in dealing with defective estimates of population parameters. The present paper proposes a method, termed "rounding", to improve the efficiency of index selection. The basic idea is to round off the noises associated with the estimates of parameters. This then leads to improved selection indices.

Theory

We assume that data of m traits are available from g genotypes grown in a progeny test experiment with r replicates. Let $\mathbf{T}$ and $\boldsymbol{\varepsilon}$ be the matrices of "between" (genotypes) and "within" mean squares and products obtained from the analyses of variance and covariance of the m traits. The degrees of freedom associated with $\mathbf{T}$ and $\boldsymbol{\varepsilon}$ are $n_1 = g-1$ and $n_2 = r(g-1)]$ or $(r-1)(g-1)$ if the experiment has a block effect], respectively. The phenotypic (**P**), genotypic (**G**) and error (**E**) VAR-COV matrices are estimated by $\hat{\mathbf{P}} = (\mathbf{T} - \boldsymbol{\varepsilon})/r + \boldsymbol{\varepsilon}$, $\hat{\mathbf{G}} = (\mathbf{T} - \boldsymbol{\varepsilon})/r$ and $\hat{\mathbf{E}} = \boldsymbol{\varepsilon}$, respectively when selection is based on individual observations.

Let **a** be the vector of economic weights of the m traits. The coefficients used in index selection are estimated by

$$\hat{\mathbf{b}} = \hat{\mathbf{P}}^{-1}\hat{\mathbf{G}}\mathbf{a}.$$

Selection is based on the index $I = \hat{\mathbf{b}}'\mathbf{x}$ where $\mathbf{x}$ is the vector of observed values of m traits of a genotype.

It is possible to transform $\mathbf{x}$ into a new array of variables. Let $\mathbf{C}$ be an abitrary full rank $(m * m)$ square matrix of constants. We can transform $\mathbf{x}$ into $\mathbf{y} = \mathbf{C}'\mathbf{x}$. Let $\mathbf{d}$ be the coefficients associated with $\mathbf{y}$ in the index equations. We have

$$I = \hat{\mathbf{b}}'\mathbf{x} = \mathbf{d}'\mathbf{y} = \mathbf{d}'\mathbf{C}'\mathbf{x}.$$

This relationship indicates that $\hat{\mathbf{b}} = \mathbf{C}\mathbf{d}$ or $\mathbf{d} = \mathbf{C}^{-1}\hat{\mathbf{b}}$. The expected response to selection based on I is estimated by

$$\hat{R}_I = i\,\mathbf{a}'\hat{\mathbf{G}}\hat{\mathbf{b}}/[\hat{\mathbf{b}}'\hat{\mathbf{P}}\hat{\mathbf{b}}]^{1/2} = i\,\mathbf{e}'\mathbf{H}\mathbf{d}/[\mathbf{d}'\mathbf{Q}\mathbf{d}]^{1/2}$$

where $\mathbf{H} = \mathbf{C}'\hat{\mathbf{G}}\mathbf{C}$, $\mathbf{Q} = \mathbf{C}'\hat{\mathbf{P}}\mathbf{C}$, $\mathbf{e} = \mathbf{C}^{-1}\mathbf{a}$, $\mathbf{d} = \mathbf{C}^{-1}\hat{\mathbf{b}}$ and i is the selection intensity. Also, $\mathbf{Q}\,\mathbf{d} = \mathbf{H}\,\mathbf{e}$ because $\hat{\mathbf{P}}\,\hat{\mathbf{b}} = \hat{\mathbf{G}}\,\mathbf{a}$. Thus,

$$\hat{R}_I = i\,[\hat{\mathbf{b}}'\hat{\mathbf{P}}\hat{\mathbf{b}}]^{1/2} = i\,[\mathbf{d}'\mathbf{Q}\mathbf{d}]^{1/2}.$$

A suitable way of transformation is to establish the following likelihood equation (Hayes and Hill 1980):

$$\mathrm{L}(\mathbf{c}) = \mathbf{c}'\,\hat{\mathbf{G}}\,\mathbf{c} - \phi[\mathbf{c}'\,\hat{\mathbf{P}}\,\mathbf{c} - 1].$$

A set of m eigenvalues, $\phi_1 \geq \phi_2 \geq \ldots \geq \phi_m$, are obtained by maximizing $\mathrm{L}(\mathbf{c})$ with respect to $\mathbf{c}$, i.e. $d[\mathrm{L}(\mathbf{c})]/d\mathbf{c} = 0$. The eigenvectors $\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_m$ corresponding to the eigenvalues are used to obtain a set of m independent variables, i.e. $\mathbf{y} = \mathbf{C}'\,\mathbf{x}$ where $\mathbf{C} = [\mathbf{c}_1\ \mathbf{c}_2\ \ldots\ \mathbf{c}_m]$. The phenotypic variances of all $\mathbf{y}$ are set to unity, i.e. $\mathbf{Q} = 1$. $\mathbf{H}$ is a diagonal matrix. Its elements represent the heritability estimates of $\mathbf{y}$, i.e. $\mathbf{H} = \mathrm{diag}[h_1^2\ h_2^2 \ldots h_m^2]$.

The above transformation is actually equivalent to the one used in canonical variate analysis, i.e.

$$L^*(\mathbf{c}) = \mathbf{c}'\,\mathbf{T}\,\mathbf{c} - v[\mathbf{c}'\,\boldsymbol{\varepsilon}\,\mathbf{c} - 1]$$

where v is the eigenvalue. This is because the roots of the determinantal equations $|\mathbf{G} - \phi\,\mathbf{P}| = 0$ bear the relation with those of $|\mathbf{T} - v\,\boldsymbol{\varepsilon}| = 0$ as $\phi_i = (v_i - 1)/(v_i + r - 1)$, $i = 1, 2, \ldots, m$. Therefore, the first variable is derived by maximizing the differences between genotypes. The subsequent variables are derived by maximizing the remaining differences between genotypes not accounted for by the preceding ones. All variables are independent of one another.

Selection based on $I = \mathbf{d}'\,\mathbf{y}$ gives exactly the same results as that based on the original traits, i.e. $I = \mathbf{b}'\,\mathbf{x}$. In canonical variate analysis, a majority of the between genotype variability is accounted for by the first few canonical variables. A Chi-square test, in fact, is available to identify the first k significant eigenvalues (see, e.g. Seal 1965).

$$\chi^2_{df} = [n_2 - (m - n_1 + 1)/2] \cdot ln \left\{ \prod_{i=k+1}^{m} [n_1 + n_1(r-1)\,\phi_i]/[n_2(1 - \phi_i)] \right\}$$

in which $df = (m - k)\,(n_1 - k)$.

The canonical variable corresponding to a non-significant eigenvalue most likely represents the variability in the original data caused by non-genetical factors. Thus, the precision of index selection is improved if only the first k canonical variables are involved in index selection. An improved index is proposed as

$$J = d_1\,y_1 + d_2\,y_2 + \ldots + d_k\,y_k = \mathbf{d}'\,\mathbf{y}_r$$

in which $\mathbf{y}_r = \mathbf{C}_r'\,\mathbf{x}$ and $\mathbf{C}_r'$ is an $(m * m)$ matrix, i.e. $\mathbf{C}_r = [\mathbf{c}_1\ \mathbf{c}_2 \ldots \mathbf{c}_r\ \mathbf{0} \ldots \mathbf{0}]$. Also, let $\mathbf{d}_r$ be a $(m * 1)$ vector, $\mathbf{d}_r' = [d_1\ d_2 \ldots d_r\ 0 \ldots 0]$, we have $J = \mathbf{d}_r'\,\mathbf{y}$ because $\mathbf{d}_r'\,\mathbf{y} = \mathbf{d}'\,\mathbf{y}_r$.

Another way to justify the rounding procedure is to examine the relation $\hat{\mathbf{P}}\,\hat{\mathbf{b}} = \hat{\mathbf{G}}\,\mathbf{a}$, i.e. $\mathbf{Q}\,\mathbf{d} = \mathbf{H}\,\mathbf{e}$. We know that

$$[d_1 \ldots d_{k+1} \ldots d_m]' = [e_1\,h_1^2 \ldots e_k\,h_k^2\,e_{k+1}\,h_{k+1}^2 \ldots e_m\,h_m^2]'.$$

The non-significance of the last $(m - k)$ eigenvalues indicate a lack of genotypic variability of the last $(m - k)$ canonical variables (i.e. $y_{k+1} \ldots y_m$). Thus, it is reasonable to set $h_{k+1}^2 = \ldots = h_m^2 = 0$. This leads to the construction of a new selection index, J.

Numerical example

Data of a random sample of 30 genotypes were extracted from a breeding population of *Solanum tuberosum* L. group Andigena. The genotypes were tested in eight-hill plots in 1983, 1984 and 1985 at the Benton Ridge Potato Breeding Substation, Benton Ridge, N. B., Canada. Three traits were used in constructing the selection index. They were specific gravity (SG) of the tubers recorded by the weight-in-air and weight-in-water method and converted to (SG-1)*1000, total number of tubers per plot, and mean tuber weight in g/tuber obtained by total tuber yield/total tuber number. The goal of selection was to increase specific gravity and tuber size and reduce tuber number. The economic weights of the three traits were set as $\mathbf{a}' = [1 - 0.5\ 1]$. Analyses of variance and covariance were carried out for the three traits by treating the 3 years as replicates. Thus, we have $g = 30$, $m = 3$ and $r = 3$. Selection was assumed on a per plot basis. The estimates of the phenotypic ($\mathbf{P}$) and genotypic ($\mathbf{G}$) variances and covariances of the three traits are as follows:

$$\hat{\mathbf{P}} = \begin{bmatrix} 136.9050 & 90.2126 & -72.4825 \\ 90.2126 & 1365.4800 & -331.0910 \\ -72.4825 & -331.0910 & 708.3390 \end{bmatrix}$$

$$\hat{\mathbf{G}} = \begin{bmatrix} 90.7387 & 167.5630 & -38.0007 \\ 167.3630 & 564.5250 & -436.1010 \\ -38.0007 & -436.1010 & 295.1720 \end{bmatrix}$$

Heritability estimates for specific gravity, tuber number and mean tuber weight are 55.4%, 26.7% and 41.7%, respectively, based on the above results. The three eigenvalues were obtained by solving $|\hat{\mathbf{G}} - \phi\,\hat{\mathbf{P}}| = 0$. They are $\phi_1 = 0.7342$, $\phi_2 = 0.4782$ and $\phi_3 = -0.1227$. The corresponding eigenvectors are $\mathbf{c}_1' = [-0.046011\ -0.014900\ 0.009670]$, $\mathbf{c}_2' = [0.060084\ -0.005202\ 0.028570]$ and $\mathbf{c}_3' = [-0.027785\ -0.005202\ 0.026920] \cdot$ Chi-square test indicated highly significant results for the first ($\chi^2 = 218.8$, $df = 87$) and second ($\chi^2 = 95.4$, $df = 56$), but not the third ($\chi^2 = 20.4$, $df = 27$) eigenvalues. The three eigenvalues are also the heritability estimates of the three transformed variables based on $\mathbf{c}_1$, $\mathbf{c}_2$ and $\mathbf{c}_3$. The third variable has a negative estimate of heritability but not significant. The coefficients for the index I were estimated by $\hat{\mathbf{b}} = \hat{\mathbf{P}}^{-1}\,\hat{\mathbf{G}}\,\mathbf{a}$, i.e. $\hat{\mathbf{b}}' = [0.2137\ -0.2814\ 0.5612]$. Thus,

$$I = 0.2173\,x_1 - 0.2814\,x_2 + 0.5612\,x_3$$

where x_1, x_2 and x_3 are observed values of specific gravity, tuber number and mean tuber weight of a progeny.

The coefficients for the index J were estimated by $\mathbf{d} = \mathbf{C}^{-1}\,\hat{\mathbf{b}}$, where $\mathbf{C} = [\mathbf{c}_1\,\mathbf{c}_2\,\mathbf{c}_3]$, i.e. $\mathbf{d}' = [14.2309\ 14.5769\ 0.2658]$. The third transformed variable was derived from non-significant eigenvalue and had negative heritability estimate. It is, thus, rounded off from the new index J. Let $\mathbf{d}'_r = [14.2309\ 14.5769\ 0]$. We have

$$J = \mathbf{d}'_r\,\mathbf{y} = 14.2309\,y_1 + 14.5769\,y_2$$

where $y_1 = \mathbf{c}'_1\,\mathbf{x}$ and $y_2 = \mathbf{c}'_2\,\mathbf{x}$ and $\mathrm{x}' = [x_1\,x_2\,x_3]$.

The expected response to selection based on I is estimated by $\hat{R}_I = i(\mathbf{d}'\,\mathbf{d})^{1/2} = 20.3733$ and that based on J by $\hat{R}_J = i(\mathbf{d}'_r\,\mathbf{d}_r)^{1/2} = 20.3717$.

Thus, $\hat{R}_J$ is almost equal to $\hat{R}_I$. The estimate of R_J is based on lesser number of variables with higher heritabilities. Hence, $\hat{R}_J$ is espected to be more precise and, consequently, selection based on J would be more reliable than based on I.

Discussion

The new selection index, J, is established by rounding off the non-significant canonical variables. Both the economic weights (**e**) and index coefficients (**d**) associated with J are derived by transforming the original weights (**a**) and coefficients (**b**). J still reflects the economic merit of a genotype. Hayes and Hill (1981) did suggest eliminating the number of transformed variables by setting the negative roots of $\hat{\mathbf{G}}$ or $\hat{\mathbf{P}}^{-1}\,\hat{\mathbf{G}}$ to zero. However, they preferred the bending procedure. A problem of the bending procedure is that the bending factor has to be chosen on a trial and error basis. The present paper gives a more complete treatment of the rounding procedure. It is noted that the elimination of variable(s) is carried out on the basis of a formal statistical test. The concept of rounding off canonical variables associated with small eigenvalues has also been considered by Yang and Dai (1983) in their approach to construct canonical selection characters.

The rounding off of non-significant canonical variables from participating in the construction of J would undoubtedly increase the precision of the expected response to selection based on J. This is because a large proportion of the observed variability of the original traits due to non-genetical factors is eliminated from the index equation. The new indices after rounding are derived from fewer transformed "traits" with higher heritabilities than the original traits. It is inevitable that a variable but small proportion of the genetic variability of the original traits would be lost due to the rounding procedure. This would cause a loss of expected response to selection. It is expected, however, that the improvement in precision would outweigh the loss in genetic gain.

The genetic gains of individual traits would also be expected to have various degrees of increase in precision and reduction in size of the expected response to selection. The magnitude of changes for a specific trait would probably depend on its heritability and genetic correlation with other traits.

References

Harris DL (1964) Expected and predicted progress from index selection involving estimates of population parameters. Biometrics 20:46–72

Hayes JF, Hill WG (1980) A reparameterization of a genetic selection index to locate its sampling properties. Biometrics 36:237–248

Hayes JF, Hill WG (1981) Modification of estimates of parameters in the construction of genetic selection indices ("bending"). Biometrics 37:483–493

Hazel LN (1943) The genetic basis for constructing selection indexes. Genetics 28:476–490

Sales J, Hill WG (1976a) Effect of sampling errors on efficiency of selection indices. 1. Use of information from relatives for single trait improvement. Anim Prod 22:1–17

Sales J, Hill WG (1976b) Effect of sampling errors on efficiency of selection indices. 2. Use of information on associated traits for improvement of a single important trait. Anim Prod 23:1–14

Seal HL (1965) Multivariate statistical analysis for biologists. Methuen, London

Smith HF (1936) A discrimination function for plant selection. Ann Eugen 7:240–250

Tai GCC (1986) A method to construct confidence interval for expected response to multi-trait selection. Theor Appl Genet 71:595–599

Williams JS (1962a) Some statistical properties of a genetic selection index. Biometrika 49:325–337

Williams JS (1962b) The evaluation of a selection index. Biometrics 18:375–393

Yang D, Dai J (1983) Approaches to canonical correlations of multiple quantitative traits. IV. Canonical selection character and genetic selection index. J Hunan Agric Coll 2:28–52