Genet. Res., Camb. (1962), **3**, *pp.*417–423
Printed in Great Britain

Genetic gains under three methods of selection

By D. J. FINNEY

Department of Statistics and Agricultural Research Council Unit of Statistics, University of Aberdeen

(*Received* 24 *January* 1962)

1. INTRODUCTION

Young (1961) has developed formulae relating to genetic gains to be expected when a population of animals or plants on which several traits have been measured is subjected to selection. He considered tandem selection, independent culling levels, and index selection. The purpose of the present paper is to generalize his results so that the only restrictions on their validity are that: (i) the genetic values of the traits follow a multivariate normal distribution, (ii) the phenotypic values are the genetic with the addition of independent environmental components that also follow a multivariate normal distribution.

Unfortunately, Young's notation is not very convenient for the general theory, and has had to be substantially modified.

2. NOTATION

$\mathbf{a}$	Column vector of economic weights of traits (elements a_i).
$\mathbf{b}$	Column vector of coefficients b_i of arbitrary linear function of the y_i
$\mathbf{B}$	$= \mathbf{b}'\mathbf{y}$.
$\mathbf{e}$	Column vector of environmental effects (elements e_i).
G	Genetic value of individual, $= \sum a_i x_i$.
h_i^2	Heritability of trait i, $= \sigma_{ii}/\omega_{ii}$.
H	Expected genetic gain in G as result of selection, with suffix to indicate type of selection.
i, j	Suffixes having range 1, 2, ..., n.
$\mathbf{m}$	Column vector of covariances of x_i with G (elements m_i).
n	Number of traits considered.
p_i	Fraction of population selected by taking $y_i \geqslant Y_i$ for one particular i:

$$p_i = \int_{Y}^{\infty} (2\pi\omega_{ii})^{-1/2} \exp\left(-\frac{y_i^2}{2\omega_{ii}}\right) dy_i.$$

P	Fraction of whole population to be selected.
$\mathbf{u}$	Column vector of elements $u_i = \omega_{ii}^{-1/2} z_i \Phi_{n-1.i}$.
$\mathbf{x}$	Column vector of genetic values (elements x_i).
$\mathbf{y}$	Column vector of phenotypic values (elements y_i); $\mathbf{y} = \mathbf{x} + \mathbf{e}$.
$\mathbf{Y}$	Column vector of cut-off points (elements Y_i).

$Y_{i.r} \quad = \left(Y_i - \frac{\omega_{ir}}{\omega_{rr}} Y_r\right) \omega_{ii.r}^{-1/2}.$

z_i Ordinate of standardized normal curve for truncation of area p_i, which

$$= (2\pi)^{-1/2} \exp\left(-\frac{Y_i^2}{2\omega_{ii}}\right).$$

Z Ordinate of standardized normal curve for truncation of area P.

$\boldsymbol{\theta}$ Column vector of means of the y_i conditional on $y_i \geqslant Y_i$ for all i, having elements θ_i.

$\phi_n(y_i; \omega_{ij})$ n-variate normal frequency function for the y_i.

$$\Phi_n \quad = \int_{Y_1}^{\infty} \int_{Y_2}^{\infty} \dots \int_{Y_n}^{\infty} \phi_n(y_{ij}; \omega_{ij})\, dy_1 dy_2 \dots dy_n.$$

$\Phi_{n-1.r}$ Integral corresponding to Φ_n, but taken over $(n-1)$ variates *excluding* y_r, having the $Y_{i.r}$ as lower limits and the $\omega_{ij.r}$ as the variances and covariances of the $(n-1)$-variate normal integrand.

$\boldsymbol{\sigma}$ Variance–covariance matrix for $\mathbf{x}$ (elements σ_{ij}).

$\boldsymbol{\omega}$ Variance–covariance matrix for $\mathbf{y}$ (elements ω_{ij}).

$$\omega_{ij.r} \quad = \omega_{ij} - \frac{\omega_{ir}\omega_{jr}}{\omega_{rr}}.$$

3. INDEPENDENT CULLING

The distributions of $\mathbf{x}$, $\mathbf{e}$, $\mathbf{y}$ will be supposed n-variate normal, and without loss of generality all means can be taken to be zero. The economic value, G, of an individual is assumed to be represented by a linear function of the genetic values for the separate traits:

$$G = \mathbf{x}'\mathbf{a}. \tag{1}$$

If selection is to be by independent culling levels, all individuals are selected for which

$$y_i \geqslant Y_i \quad \text{for all } i,$$

the Y_i being constrained in such a way as to make

$$\Phi_n = P. \tag{2}$$

The mean values of the y_i subject to this condition have been obtained by Tallis (1961), whose result can be written

$$\boldsymbol{\theta} = \boldsymbol{\omega}\mathbf{u}/P. \tag{3}$$

Now because $\mathbf{x}$ and $\mathbf{y}$ jointly have a $2n$-variate normal distribution, each x_i can be regarded as having a linear regression on the set of y_i. Evidently

$$\mathbf{m} = E(G\mathbf{y}) = \boldsymbol{\sigma}\mathbf{a}, \tag{4}$$

https://doi.org/10.1017/S0016672300003256 Published online by Cambridge University Press

and therefore $\boldsymbol{\beta}$, the column vector of partial regression coefficients of G on the y_i is the solution of

$$\boldsymbol{\omega\beta} = \mathbf{m}. \tag{5}$$

Hence the mean value of G in the selected portion of the population is

$$\begin{aligned} \boldsymbol{\theta}'\boldsymbol{\beta} &= \mathbf{u}'\boldsymbol{\omega\beta}/P \\ &= \mathbf{u}'\mathbf{m}/P. \end{aligned} \tag{6}$$

This quantity is H_c, the expected genetic gain in G as a result of the selection. It can be written

$$H_c = \sum_{i=1}^{n} m_i \omega_{ii}^{-1/2} z_i \Phi_{n-1.i}/P. \tag{7}$$

Both the genetic and the phenotypic correlations enter into H_c; the σ_{ij} appear in the m_i and the ω_{ij} affect the Φ functions. The formula for H_c is in a different form from Young's formula, and is perhaps a somewhat simpler expression, but the two are of course algebraically identical for $n = 2$. If the traits are genetically uncorrelated,

$$m_i = a_i \sigma_{ii}, \tag{8}$$

and H_c then depends on the genetic parameters only through the heritabilities

$$h_i^2 = \frac{\sigma_{ii}}{\omega_{ii}}. \tag{9}$$

On the other hand, if the traits are phenotypically uncorrelated, P is simply the product of fractions selected for each trait separately; that is to say

$$P = \prod_{i=1}^{n} p_i \tag{10}$$

and also

$$P = p_i \Phi_{n-1.i}. \tag{11}$$

Hence

$$H_c = \sum_{i=1}^{n} m_i \omega_{ii}^{-1/2} \left(\frac{z_i}{p_i}\right). \tag{12}$$

Young did not distinguish between the consequences of genetic and phenotypic independence. If genetic independence is a reasonable approximation, m_i from equation (8) may be substituted in equation (7). On the other hand, phenotypic independence alone is unlikely, for it would require exact compensation between the genetic and the purely environmental components of covariance. Hence, although equation (12) is appropriate to phenotypic independence even with the general form of m_i, equation (4), it is unlikely to be needed except when genetic independence permits m_i to be given by equation (8).

If the proportion selected, P, is fixed, the optimal system of independent culling will be that which maximizes the expectation of genetic gain. Unfortunately no

analytic procedure for maximizing H_c in equation (7) is available. Young & Weiler (1960) have provided a system of nomographs for $n = 2$ and for several values of the phenotypic correlation coefficient, but for any larger n purely numerical methods of trial and error would be needed. In the special case of no phenotypic correlation, if all the $m_i \omega_{ii}^{-1/2}$ (Young's λ_i) are equal, symmetry indicates that the maximization will have all $Y_i \omega_{ii}^{-1/2}$ equal, so making

$$p_i = P^{1/n} \tag{13}$$

and

$$H_c = n m_0 \omega_{00}^{-1/2} \left(\frac{z_0}{p_0} \right), \tag{14}$$

where the suffix 0 indicates a common value for all traits.

4. TANDEM SELECTION

Young has stated the well-known and completely general results applicable when all selection is concentrated on a single trait; in the present notation, the expectation of gain in G is

$$H_T = m_i \omega_{ii}^{-1/2} \left(\frac{Z}{P} \right). \tag{15}$$

Tandem selection involves selecting in any one generation for one trait only. If trait i is the one chosen, equation (15) represents the expectation of genetic gain. Naturally one would wish to take for i the trait that maximizes $m_i \omega_{ii}^{-1/2}$. Young appears to assume for his calculations that this is achieved, thereby setting an upper limit to the gain from tandem selection. If little advance information on the σ_{ij} and ω_{ij} is available, possibly a more reasonable value to take for the expectation of genetic gain is the average over all possible traits:

$$\bar{H}_T = \frac{Z}{nP} \sum_{i=1}^{n} m_i \omega_{ii}^{-1/2}. \tag{16}$$

5. INDEX SELECTION

Suppose now that

$$B = \mathbf{b}' \mathbf{y} \tag{17}$$

is an arbitrary linear function of the y_i. Then the expectation of genetic gain from selection by taking the fraction P of highest values of B is easily seen to be

$$H_B = \frac{\mathbf{b}' \mathbf{g}}{(\mathbf{b}' \boldsymbol{\omega} \mathbf{b})^{1/2}} \cdot \frac{Z}{P}. \tag{18}$$

If the index is to be chosen so as to maximize H_B, the vector $\mathbf{b}$ must satisfy

$$\boldsymbol{\omega} \mathbf{b} = \alpha \mathbf{m}$$

https://doi.org/10.1017/S0016672300003256 Published online by Cambridge University Press

where α is a constant which can be taken as unity without loss of generality. Hence

$$\mathbf{b} = \boldsymbol{\omega}^{-1}\mathbf{m}.$$

The optimal selection index is therefore

$$B_I = \mathbf{m}'\boldsymbol{\omega}^{-1}\mathbf{y}, \tag{19}$$

with expectation of genetic gain in G

$$H_I = (\mathbf{m}'\boldsymbol{\omega}^{-1}\mathbf{m})^{1/2}\frac{Z}{P}, \tag{20}$$

which is Young's equation (5).

6. COMPARISON OF SELECTION METHODS

Ratios between the expressions for H in equation (7) and its simplifications (15), (16), and (20) enable the relative efficiencies of the three methods of selection to be compared in terms of expectations of total genetic gain. The formulae are essentially the same as those given by Young, but with somewhat greater generality. The approach used here emphasizes the central role of the m_i, the covariances of the genetic values of the separate traits, with G, the genetic value of the whole individual. All the formulae are expressed in terms of the m_i, which appear as more fundamental than the

$$\lambda_i = a_i h_i^2 \omega_{ii}^{1/2} = a_i \sigma_{ii} \omega_{ii}^{-1/2}$$

used by Young. The introduction of the m_i would enable some of Young's interesting numerical comparisons to be given wider applicability.

Of course, the assumption underlying the comparisons made here is that the aim of selection is to maximize a single linear combination of the x_i, and this need not always be appropriate. The wish of the investigator may indeed be to maximize two or more of the x_i (or two or more functions of the x_i) simultaneously; recognizing the impossibility of this, he may compromise by seeking to maximize a particular x_i subject to conditions that there shall be only small probabilities that certain others of the x_i fall below specified limits. In such circumstances, independent culling and tandem selection may compare more favourably with index selection.

7. SELECTION OVER SEVERAL GENERATIONS

Many writers on selection, including Young, and also Hazel & Lush (1942), write of the genetic gain from one stage of selection as though it is a gain to be maintained over a number of generations. In theory, however, even if the initial genetic and phenotypic distributions are exactly normal, any of the methods of selection discussed here (and indeed almost any method likely to be used) will destroy the normality at its first application. The distortion of normality will be greatest if heritabilities are high and the fraction to be selected is small, but the recombinations that occur in the production of a new generation and the accumulation of new

environmental variation will perhaps tend towards the partial restoration of normality. Possibly in many practical situations, the development of non-normality (particularly skewness) as a consequence of selection is sufficiently slow for the use of formulae based on normality not to be seriously wrong over several generations. This will be more true with domestic animals, because of the limitations on intensity of selection imposed by low multiplication rates, than with insects or plants. It must not be forgotten however that the conditions that conduce to the validity of normal formulae over several successive stages of selection (high p_i and low h_i^2) are precisely those that are unfavourable to success in achieving substantial genetic gains. I have published elsewhere (Finney, 1956, 1961) methods of adjusting mean values after selection in order to take account of the non-normality induced by previous selections, but even for two stages the series required are laborious in use. Qualitatively, the general consequence of the negative skewness produced by selection is that the assumption of normality in calculating expected gains under further selection usually overestimates the true gains (cf. Hazel & Lush).

Whether or not normality may safely be assumed for more than one generation of selection, it must not be forgotten that any selection applied in one generation will reduce the genetic variances of all traits correlated with the function used as the basis of selection and will also modify the covariances between these traits; the phenotypic variances and covariances in subsequent generations will therefore also be changed. Relevant formulae are well known for the situation in which the individual genetic values are reproduced exactly in each generation of those selected, as with a population of selfed homozygotes (Pearson, 1902, 1912; Tallis, 1961; Finney, 1956, 1961, 1962), but may need to be substantially more complicated if there is much genetic segregation and recombination between generations. When selective intensities are low, several generations might elapse without the reductions in genetic variances and correlations becoming very important, but if P were less than 0·1 in each generation both this effect and non-normality could rapidly complicate the formulae relevant to selection.

SUMMARY

Results obtained by Young for the expectation of genetic gain in an arbitrary linear function of several traits under selection by independent culling levels, under tandem selection, and under index selection have been obtained in slightly more general form and their dependence on basic genetic and phenotypic parameters exhibited. A warning is given about the effects of selection in modifying the distribution of traits; when the distribution has become appreciably non-normal, any calculation of genetic gains from formulae based on normality will tend to overestimation.

REFERENCES

FINNEY, D. J. (1956). The consequences of selection for a variate subject to errors of measurement. *Bull. Inst. int. Statist.* **24**, 1–10.

FINNEY, D. J. (1961). The transformation of a variate under selection. *Sankhyā*, **A23**, 309–324.

FINNEY, D. J. (1962). The statistical evaluation of educational allocation and selection. *J. R. statist. Soc.* (in press)

HAZEL, L. N. & LUSH, J. L. (1942). The efficiency of three methods of selection. *J. Hered.* **33**, 393–399.

PEARSON, K. (1902). On the influence of natural selection on the variability and correlation of organs. *Phil. Trans.* **A200**, 1–66.

PEARSON, K. (1912). On the general theory of the influence of selection on correlation and variation. *Biometrika*, **8**, 437–443.

TALLIS, G. M. (1961). The moment generating function of the truncated multi-normal distribution. *J. R. statist. Soc.* **B23**, 223–229.

YOUNG, S. S. Y. (1961). A further example of the relative efficiency of three methods of selection for genetic gains under less-restricted conditions. *Genet. Res.* **2**, 106–121.

YOUNG, S. S. Y. & WEILER, H. (1960). Selection for two correlated traits by independent culling levels. *J. Genet.* **57**, 329–358.