IMPROVEMENT THROUGH SELECTION AT SUCCESSIVE STAGES

By

J. P. JAIN

Dissertation submitted in partial fulfilment of the requirements for the award of Diploma in Agricultural & Animal Husbandry Statistics of the Institute of Agricultural Research Statistics, New Delhi, 1961.

ACKNOWLEDGEMENT

I am very much indebted to Shri V.N. Amble, Assistant Statistical Adviser (Animal Husbandry), Institute of Agricultural Research Statistics, New Delhi, for his valuable guidance, constant encouragement and constructive criticism in the preparation of this dissertation.

CONTENTS

- : o : -

1. INTRODUCTION

Plant and animal breeders are often faced with the problem of finding the optimal values of intensities of selection at various stages, the optimal values considered being those which result in maximizing the genetic advance. With dairy cows under selection for milk yield, for example, each successive lactation provides new data on the milk yielding capacity of the animal. These lactation yields would form the basis for the successive stages of selection - selection at the rth stage being made on the evidence of yield in the latest lactation, combined with the available information on previous lactation yields. A usual feature in selection problems is that we cannot assess directly the genetic value of the character which we wish to improve. In the present case selection is aimed at securing cows with superior genotypic value for milk yield but it has to be based on the observed or phenotypic values in successive lactations. The problem of improving some character y, which is not directly measurable, by means of indirect selection that is made from a group of tests or measurements x_1, x_2 ....x_r at successive stages and of measuring the rate of improvement in terms of the genetic gain for different intensities of selection at various stages is essentially a statistical one.

contd..... page 2

The object of the present dissertation is two fold :

(i) To deal with the extension of the selection programme to r-stages utilizing information collected in the previous r - 1 stages.

(ii) To highlight the difficulties which the experimenter would encounter in its application for higher r.

For overcoming these two simpler approximations which are much easier in their application have been suggested.

Selection aspect of a breeding programme of dairy cattle has been discussed at length to illustrate the working of the above method.

- : o : -

2. REVIEW OF LITERATURE

Smith (1936) has discussed the genetic gain for one stage selection assuming linear relationship between the genotypic and phenotypic values. Another form of the same expression has been presented by Hazel (1943) while Panse (1946) has discussed its application taking several characters simultaneously with particular reference to selection in poultry.

Sieben (1954) and Keuls and Sieben (1955) who have discussed a similar problem with reference to plant selection follow a different scheme of selection. After arbitrarily partitioning the whole population of varieties into a few 'good' ones (high yielders), a few 'bad' ones (low yielders) and a large number of intermediate varieties whose yields are such that it is immaterial whether they are retained or rejected, the rule of selection is based on the consideration which aims at minimising the probability of rejection of good varieties and of selection of bad ones.

All these authors, however, are concerned primarily with one stage selection. Dickerson and Hazel (1944) have gone one stage further. The application of the formula which they have given is however restricted in that, the values of proportions retained after second culling among those retained after first, must not be either much larger or smaller than 0.5. For these restricted values, the exact value of the selection differential expected after second culling does not differ appreciably from that expected from a normal distribution.

Cochran (1951) has discussed the optimum rule of selection which maximizes the gain in y. He has also derived the general form for gain in y expected after two-stage selection for the case when the variates y, x_1 and x_2 follow a multivariate normal distribution.

Finney (1957) has advanced a theory for two-stage selection programme and discussed the implications of its extension to r stage. This assumes that selection at stage r would be based solely on the evidence of yields in that stage. Consequently this would mean sacrificing information on the previous (r - 1) records which might have been usefully utilized in accelerating the pace of genetic gain. Nevertheless the results are of interest in that they provide a lower limit to the gains that may accrue from different rules of selection. It may be mentioned that although in theory the methods used for computing the consequences of two-stage selection could be extended to any number of stages, the complexity of the formulae and the limits of accuracy of various mathematical tables that are employed make this impracticable even for $r = 3$".

- : o : -

3. APPROACH TO THE PROBLEM

The line of approach followed in the present work is the same as suggested by Cochran (1951), which may be summarised as follows :

Assume the variates $y, x_1, x_2 \ldots\ldots x_r$ to follow a distribution whose frequency function is known. If the regression $\eta(x)$ of y on the x's exists, Cochran has shown firstly that $\eta(x)$ is the best selection index i.e. the regressions $\eta_1(x)$ of y on x_1; $\eta_2(x)$ of y on x_1 and x_2 etc. will constitute the optimum selection indices at different stages of selection. If the proportions selected $\alpha_1, \alpha_2, \ldots\ldots \alpha_r$ at different stages have been decided in advance, the units at the first stage will be selected whenever $\eta_1 \geqslant k_1$, where k_1 is the truncation point corresponding to the frequency of selection α_1; the units selected at the second stage will be those for which $\eta_2 \geqslant k_2$, where, given k_1, k_2 is the truncation point corresponding to the frequency of selection $\alpha_1\alpha_2$. The same argument will be true for further stages of selection. Secondly the gain in y is a linear function of gains in η's.

In the light of the above two fundamental results, the gain in y due to selection over r stages has been worked out in the next section.

4. SELECTION IN r STAGES

For convenience without loss of generality we may take deviations of all variates $y, \eta_1, \eta_2 \ldots\ldots \eta_r$ from their respective means and effect a scaling transformation so that in the population all the variates have zero means and unit variances. ~~being linear functions of x's will also be normally distributed.~~ Let the parameters $\rho_1, \rho_2, \ldots\ldots \rho_r$ denote the simple correlations between y and η_1, y and η_2 and so on. For fixed $\alpha_1, \alpha_2, \ldots\ldots \alpha_r$ the points of truncation $k_1, k_2 \ldots\ldots k_r$ will be given by the following r equations :

$$\alpha_1 = \frac{1}{\sqrt{2\pi}} \int_{k_1}^{\infty} e^{-\eta_1^2/2} \, d\eta_1 \; ; \tag{1}$$

$$\alpha_1 \alpha_2 = \frac{1}{2\pi\sqrt{1-\rho_{12}^2}} \int_{k_1}^{\infty} d\eta_1 \int_{k_2}^{\infty} e^{-\frac{1}{2}\left[\frac{1}{1-\rho_{12}^2}\right]\left[\eta_1^2 - 2\rho_{12}\eta_1\eta_2 + \eta_2^2\right]} \, d\eta_2 \; ; \tag{2}$$

- - - - - - - - - - - - -

- - - - - - - - - - - - - -

- - - - - - - - - - - -

Finally,

$$\alpha_1 \alpha_2 \cdots \alpha_r = \frac{1}{(2\pi)^{r/2} \Delta_r^{1/2}} \int_{k_1}^{\infty} d\eta_1 \int_{k_2}^{\infty} d\eta_2 \cdots\cdots \int_{k_r}^{\infty} e^{-\frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} A_{ij}\eta_i\eta_j} \, d\eta_r \tag{3}$$

where Δ_r is the determinant of the variance-covariance matrix of order r

i.e.

$$\Delta_r = |a_{ij}|_r = \begin{vmatrix} 1 & \rho_{12} & \rho_{13} & \cdot & \cdot & \rho_{1r} \\ \rho_{21} & 1 & \rho_{23} & \cdot & \cdot & \rho_{2r} \\ \rho_{31} & \rho_{32} & 1 & \cdot & \cdot & \rho_{3r} \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot & \cdot & \cdot \\ \rho_{r1} & \rho_{r2} & \rho_{r3} & \cdot & \cdot & 1 \end{vmatrix}$$

$$A_{ij} = \frac{\text{Cofactor of } a_{ij} \text{ in } |a_{ij}|_r}{\Delta_r} = \frac{B_{ij}}{\Delta_r} \text{ say}$$

and α_1 ,($\alpha_1 \alpha_2$),...... and ($\alpha_1 \alpha_2 \cdots\cdots \alpha_r$) are the proportions retained at first, second ...... and rth stage of selection respectively from the unselected units.

If $f(y, x_1, x_2 \ldots\ldots x_r)$ is the joint frequency function of the variates $y, x_1, x_2 \ldots\ldots\ldots x_r$, the gain in y due to selection over r stages will be

$$G(y) = \frac{1}{\alpha_1\alpha_2\cdots\alpha_r}\int_{-\infty}^{\infty} dy \int_{\eta_1 \geq k_1} \int_{\eta_2 \geq k_2} \cdots\cdots \int_{\eta_r \geq k_r} y\, f(y, x_1, x_2, \cdots x_r)\, dx_1 dx_2 \cdots dx_r$$

$$= \frac{1}{\alpha_1\alpha_2\cdots\alpha_r}\int_{-\infty}^{\infty} dy \int_{\eta_1 \geq k_1} \int_{\eta_2 \geq k_2} \cdots\cdots \int_{\eta_r \geq k_r} y\, \varphi(y|x_1, x_2, \cdots x_r)\, f_1(x_1, x_2, \cdots x_r)\, dx_1 dx_2 \cdots dx_r$$

$$= \frac{1}{\alpha_1\alpha_2\cdots\alpha_r}\int_{\eta_1 \geq k_1} \int_{\eta_2 \geq k_2} \cdots\cdots \int_{\eta_r \geq k_r} f_1(x_1, x_2, \cdots x_r)\, dx_1 dx_2 \cdots dx_r \int_{-\infty}^{\infty} y\, \varphi(y|x_1, x_2, \cdots x_r)\, dy$$

$$= \frac{1}{\alpha_1 \alpha_2 \cdots \alpha_r} \int\limits_{\eta_1 \geq k_1} \int\limits_{\eta_2 \geq k_2} \cdots\cdots \int\limits_{\eta_r \geq k_r} \eta(x_1, x_2, \cdots x_r) f_1(x_1, x_2, \cdots x_r)\, dx_1 dx_2 \cdots dx_r$$

where $\varphi(y \mid x_1, x_2 \ldots\ldots x_r)$ is the conditional frequency function of y, given the x's and $f_1(x_1, x_2 \ldots\ldots x_r)$ is the joint frequency function of the x's.

η's being linear functions of x's, the gain in y due to selection on η_1, followed by selection on η_2 and so on, will be a linear function of the gains in $\eta_1, \eta_2 \ldots\ldots\ldots$ and η_r.

If $y = \beta_1 \eta_1 + \beta_2 \eta_2 + \beta_3 \eta_3 + \ldots\ldots\ldots\ldots + \beta_r \eta_r + e$ where $(\beta_1 \eta_1 + \beta_2 \eta_2 + \beta_3 \eta_3 + \ldots\ldots\ldots + \beta_r \eta_r)$ is the multiple regression of y on the η's in the unselected population, then the expected value of y, the expectation being taken over the selected part of the universe, will be the gain in y since the variates are measured from their respective means and can be written as

$$G(y) = \beta_1 G(\eta_1) + \beta_2 G(\eta_2) + \beta_3 G(\eta_3) + \cdots\cdots\cdots + \beta_r G(\eta_r) \qquad (4)$$

We, therefore, need only find $G(\eta_1)$, $G(\eta_2)$, $G(\eta_3), \cdots\cdots\cdots$ and $G(\eta_r)$

For that consider

$$\alpha_1\alpha_2\cdots\alpha_r G\left[\frac{\partial g}{\partial \eta_r}\right] = \frac{1}{(2\pi)^{r/2}\Delta_r^{1/2}} \int_{k_1}^{\infty} d\eta_1 \int_{k_2}^{\infty} d\eta_2 \cdots \int_{k_r}^{\infty} \left[\frac{\partial g}{\partial \eta_r}\right] e^{-\frac{1}{2\Delta_r}\sum_{i=1}^{r}\sum_{j=1}^{r} B_{ij}\eta_i\eta_j} d\eta_r$$

where

$$g = \frac{1}{2}\sum_{i=1}^{r}\sum_{j=1}^{r} B_{ij}\eta_i\eta_j$$

$$= \frac{\Delta_r^{1/2}}{(2\pi)^{r/2}} \int_{k_1}^{\infty} d\eta_1 \int_{k_2}^{\infty} d\eta_2 \cdots \int_{k_{r-1}}^{\infty} e^{-\frac{1}{2\Delta_r}\left[\sum_{i=1}^{r-1}\sum_{j=1}^{r-1} B_{ij}\eta_i\eta_j + 2k_r\sum_{i=1}^{r-1} B_{ir}\eta_i + B_{rr}k_r^2\right]} d\eta_{r-1}$$

Let $u_i = \eta_i - a_{ir}k_r \qquad (i = 1, 2, \ldots\ldots r-1)$

$$|J| = \left|\frac{\partial(\eta_1, \eta_2, \ldots\ldots \eta_{r-1})}{\partial(u_1, u_2, \ldots\ldots u_{r-1})}\right| = 1$$

then $d\eta_1\, d\eta_2 \cdots\cdots d\eta_{r-1} = du_1\, du_2 \cdots\cdots du_{r-1}$

and noting

$$\begin{cases} \sum_{i=1}^{r} a_{ij} B_{ij} = \sum_{j=1}^{r} a_{ij} B_{ij} = \Delta_r \\ a_{11} = a_{22} = \cdots\cdots = a_{rr} = 1 \\ \sum_{i=1}^{r} a_{ir} B_{ji} = 0 \end{cases}$$

the above integral reduces to

$$\alpha_1 \alpha_2 \cdots\cdots \alpha_r \; G\left[\sum_{1}^{r} B_{ir}\eta_i\right] -$$

$$- - = \frac{\Delta_r^{1/2}}{(2\pi)^{r/2}} \int_{D_1} du_1 \int_{D_2} du_2 \cdots\cdots \int_{D_{r-1}} e^{-\frac{1}{2\Delta_r}\left[\sum_{i=1}^{r-1}\sum_{j=1}^{r-1} B_{ij} u_i u_j + \Delta_r k_r^2\right]} du_{r-1}$$

$$= \left(\frac{e^{-\frac{1}{2}k_r^2}}{\sqrt{2\pi}}\right) \frac{\Delta_r^{1/2}}{(2\pi)^{\frac{r-1}{2}}} \int_{D_1} du_1 \int_{D_2} du_2 \cdots\cdots \int_{D_{r-1}} e^{-\frac{1}{2\Delta_r}\sum_{i=1}^{r-1}\sum_{j=1}^{r-1} B_{ij} u_i u_j} du_{r-1} \qquad (5)$$

where $D_i \equiv k_i - a_{ir} k_r$ to ∞ $\quad (i = 1, 2, \cdots\cdots r-1)$

At this stage we define a new variance-covariance matrix Δ'_{r-1} of order r - 1 formed by all the possible combinations of the first order partial correlation coefficients of variates ranging from 1 to r - 1, keeping rth variate constant,

$$\text{viz.} \quad \Delta'_{r-1} = \left|a'_{ij}\right|_{r-1} = \begin{vmatrix} 1 & \rho_{12\cdot r} & \rho_{13\cdot r} & \cdot\;\cdot & \rho_{1,r-1\cdot r} \\ \rho_{21\cdot r} & 1 & \rho_{23\cdot r} & \cdot\;\cdot & \rho_{2,r-1\cdot r} \\ \rho_{31\cdot r} & \rho_{32\cdot r} & 1 & \cdot\;\cdot & \rho_{3,r-1\cdot r} \\ - & - & - & \cdot\;\cdot & - \\ - & - & - & \cdot\;\cdot & - \\ \rho_{r-1,1\cdot r} & \rho_{r-1,2\cdot r} & \rho_{r-1,3\cdot r} & \cdot\;\cdot & 1 \end{vmatrix}$$

and B'_{ij} = cofactor of a'_{ij} in $\left|a'_{ij}\right|_{r-1} = \Delta'_{r-1} A'_{ij}$

Now effecting the transformation

$$u_i = \sqrt{1-\rho_{ir}^2}\; t_i \qquad (i = 1, 2, \cdots\cdots r-1)$$

and noting

(i) $\Delta'_{r-1} = \dfrac{\Delta_r}{\prod_{i=1}^{r-1}(1-\rho_{ir}^2)}$

(ii) $B'_{ii} = \dfrac{B_{ii}}{\prod_{j}^{r-1}(1-\rho_{jr}^2)}$

(iii) $B'_{ij} = B_{ij}\sqrt{\dfrac{B'_{ii}\,B'_{jj}}{B_{ii}\,B_{jj}}}$

$\left.\right\}\ i = j = 1, 2, \ldots r-1, \quad i \neq j$

(iv) $\dfrac{\Delta_r}{\Delta'_{r-1}}\,\dfrac{B'_{ii}}{B_{ii}} = (1-\rho_{ir}^2) \qquad (i = 1, 2, \ldots\ldots r-1)$

integral (5) reduces to

$$\alpha_1\alpha_2\cdots\alpha_r\, G\left[\sum_{i=1}^{r} B_{ir}\eta_i\right] = \Delta_r\left(\frac{e^{-\frac{1}{2}k_r^2}}{\sqrt{2\pi}}\right)\left(\frac{1}{(2\pi)^{\frac{r-1}{2}}\Delta'^{\frac{1}{2}}_{r-1}}\int_{D'_1} dt_1\int_{D'_2} dt_2\cdots\int_{D'_{r-1}} e^{-\frac{1}{2}\sum_{i=1}^{r-1}\sum_{j=1}^{r-1}A'_{ij}t_it_j}\,dt_{r-1}\right)$$

where $D'_i \equiv \dfrac{k_i - a_{ir}k_r}{\sqrt{1-\rho_{ir}^2}}$ to $\infty \quad (i = 1, 2, \ldots\ldots r-1)$

Thus $\alpha_1\alpha_2\cdots\cdots\alpha_r\, G\left[\sum_{i=1}^{r} B_{ir}\eta_i\right] = \Delta_r\, Z_{(k_r)}\, I_{12\cdots\cdots r-1}$

where Z denotes the ordinate of the univariate normal curve and I the incomplete volume of the r - 1 variate normal surface respectively.

similarly

$$\alpha_1 \alpha_2 \cdots \alpha_r G\left[\sum_{i=1}^{r} B_{i\,r-1} \eta_i\right] = \Delta_r Z_{(k_{r-1})} I_{12\ldots\ldots(r-2)\,r}$$

\- - - - - - - - - - - - - - - - - - - -

\- - - - - - - - - - - - - - - - - - -

\- - - - - - - - - - - - - - - - - - -

$$\alpha_1 \alpha_2 \cdots \alpha_r G\left[\sum_{i=1}^{r} B_{i1} \eta_i\right] = \Delta_r Z_{(k_1)} I_{23\ldots\ldots r}$$

Solving these eqations for $G(\eta_1)$, $G(\eta_2)$ etc., we get

$$\alpha_1 \alpha_2 \cdots \alpha_r \begin{bmatrix} G(\eta_1) \\ G(\eta_2) \\ \vdots \\ G(\eta_r) \end{bmatrix} = \left[a_{ij}\right]_r \begin{bmatrix} (Z_{(k_1)} I_{23\ldots r}) \\ (Z_{(k_2)} I_{13\ldots r}) \\ \vdots \\ (Z_{(k_r)} I_{12\ldots r-1}) \end{bmatrix}$$

Substituting the values of $G(\eta)'s$ in (4) we have

$$\begin{aligned} \alpha_1 \alpha_2 \cdots \alpha_r G(y) = {} & (\beta_1 + \rho_{12}\beta_2 + \rho_{13}\beta_3 + \cdots + \rho_{1r}\beta_r) Z_{(k_1)} I_{23\ldots r} \\ & + (\rho_{21}\beta_1 + \beta_2 + \rho_{23}\beta_3 + \cdots + \rho_{2r}\beta_r) Z_{(k_2)} I_{13\ldots r} \\ & + - - - - - - - - - - - - - - - - \\ & + - - - - - - - - - - - - - - - - \\ & + (\rho_{r1}\beta_1 + \rho_{r2}\beta_2 + \rho_{23}\beta_3 + \cdots + \beta_r) Z_{(k_r)} I_{12\ldots r-1} \end{aligned}$$

But by definition

$$\rho_1 = \mathrm{Cov}\,[y\eta_1] = \mathrm{Cov}\left[\left(\beta_1\eta_1 + \beta_2\eta_2 + \cdots\cdots + \beta_r\eta_r\right)\eta_1\right]$$

$$= \beta_1 + \rho_{12}\,\beta_2 + \cdots\cdots + \rho_{1r}\,\beta_r$$

Similarly

$$\rho_2 = \rho_{21}\,\beta_1 + \beta_2 + \cdots\cdots + \rho_{2r}\,\beta_r$$

— — — — — — — — — —

— — — — — — — — — —

$$\rho_r = \rho_{r1}\,\beta_1 + \rho_{r2}\,\beta_2 + \cdots\cdots + \beta_r$$

Therefore

$$G(y) = \frac{\rho_1\, Z_{(k_1)}\, I_{23\ldots r} + \rho_2\, Z_{(k_2)}\, I_{13\ldots r} + \cdots\cdots + \rho_r\, Z_{(k_r)}\, I_{12\ldots r-1}}{\alpha_1\,\alpha_2\cdots\cdots\alpha_r} \qquad (6)$$

If y does not have unit standard deviation the only change needed is to multiply the right hand side of (6) by genetic standard deviation σ_y.

- : o : -

5. SELECTION IN THREE STAGES

It is not possible to utilise the general formula derived in the last section for the case of more than three stages of selection since the tables of multi-variate normal integrals is limited to trivariates only. We may consider the case of three variables in more specific detail. This case is of particular importance with dairy cows under selection for improvement in their level of milk production for the reasons detailed in section 7.

In this case the general formula reduces to

$$G(y) = \frac{\rho_1 Z_{(k_1)} I_{23} + \rho_2 Z_{(k_2)} I_{13} + \rho_3 Z_{(k_3)} I_{12}}{\alpha_1 \alpha_2 \alpha_3} \qquad (7)$$

where k_1, k_2 and k_3 will be found from the normal tables satisfying the following three equations :

$$\alpha_1 = \frac{1}{\sqrt{2\pi}} \int_{k_1}^{\infty} e^{-\eta_1^2/2} \, d\eta_1 \; ;$$

$$\alpha_1 \alpha_2 = \frac{1}{2\pi\sqrt{1-\rho_{12}^2}} \int_{k_1}^{\infty} d\eta_1 \int_{k_2}^{\infty} e^{-\left[\frac{1}{2(1-\rho_{12}^2)}\right]\left[\eta_1^2 - 2\rho_{12}\eta_1\eta_2 + \eta_2^2\right]} \, d\eta_2 \; ;$$

$$\alpha_1 \alpha_2 \alpha_3 = \frac{1}{(2\pi)^{3/2} \Delta^{1/2}} \int_{k_1}^{\infty} d\eta_1 \int_{k_2}^{\infty} d\eta_2 \int_{k_3}^{\infty} e^{-\frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3} A_{ij}\eta_i\eta_j} \, d\eta_3 \, .$$

where

$$A_{11} = \frac{1-\rho_{23}^2}{\Delta}, \quad A_{22} = \frac{1-\rho_{13}^2}{\Delta}, \quad A_{33} = \frac{1-\rho_{12}^2}{\Delta}$$

$$A_{12} = \frac{\rho_{13}\rho_{23}-\rho_{12}}{\Delta}, \quad A_{13} = \frac{\rho_{12}\rho_{23}-\rho_{13}}{\Delta}, \quad A_{23} = \frac{\rho_{12}\rho_{13}-\rho_{23}}{\Delta}.$$

and $\Delta = 1-\rho_{12}^2-\rho_{13}^2-\rho_{23}^2+2\rho_{12}\rho_{13}\rho_{23}$.

$Z(k_1)$, $Z(k_2)$ and $Z(k_3)$ are the ordinates of the univariate normal curve corresponding to k_1, k_2 and k_3; and I's the incomplete volumes of the bivariate normal surface,

where

$$I_{12} = I\left(\frac{k_1-k_3\rho_{13}}{\sqrt{1-\rho_{13}^2}}, \frac{k_2-k_3\rho_{23}}{\sqrt{1-\rho_{23}^2}}; \rho_{12.3}\right)$$

$$I_{23} = I\left(\frac{k_2-k_1\rho_{21}}{\sqrt{1-\rho_{21}^2}}, \frac{k_3-k_1\rho_{31}}{\sqrt{1-\rho_{31}^2}}; \rho_{23.1}\right)$$

and

$$I_{31} = I\left(\frac{k_3-k_2\rho_{32}}{\sqrt{1-\rho_{32}^2}}, \frac{k_1-k_2\rho_{12}}{\sqrt{1-\rho_{12}^2}}; \rho_{31.2}\right)$$

k_1 will be obtained from the univariate normal tables (Pearson, 1931) corresponding to the frequency of selection α_1.

The value of k_2 can be got from the bivariate normal tables given by Pearson (1931). But the use of these tables involves considerable amount of interpolation work. Tables computed by Owen (1956) overcome this difficulty to some extent. However, the method of S.C. Das (1956) which consists in reducing bivariate integral to a single integral which is then to be evaluated numerically seems more suitable for fixing the truncation point k_2. This method can be summarized briefly as follows :

Evaluating

$$I = \frac{1}{2\pi\sqrt{1-\rho^2}} \int_{k_1}^{\infty}\int_{k_2}^{\infty} e^{-\left[\frac{1}{2(1-\rho^2)}\right]\left[y_1^2 - 2\rho y_1 y_2 + y_2^2\right]} dy_1 dy_2$$

is equivalent to evaluating numerically J defined as

$$J = \sqrt{\pi}\, I = h \sum_{n=-\infty}^{\infty} \exp(-n^2 h^2)\, P(anh + b_1)\, P(\pm anh + b_2)$$

the plus sign being taken when ρ is positive and the minus sign when ρ is negative.

Here, n denotes the number of segments into which the range has been divided, h is the width of the interval

$$P(x) = \frac{1}{\sqrt{2\pi}} \int_{x}^{\infty} \exp(-\tfrac{1}{2}t^2)\, dt$$

and a, b_1, b_2 are determined from the relations

$$|\rho| = \frac{a^2}{2+a^2}, \quad k_1 = \frac{b_1\sqrt{2}}{\sqrt{2+a^2}}, \quad k_2 = \frac{b_2\sqrt{2}}{\sqrt{2+a^2}}$$

The value of k_3 can be fixed with the help of T-function tabulated by Owen (1956) and S-function tabulated by Steck (1958) coupled with univariate normal tables. However, these tables which are better suited for evaluating the volume of the trivariate distribution given the range of integration are not very helpful for the reverse procedure of reading k_3.

Another general approach to the problem is by means of the tetrachoric series which has been generalised by M.G. Kendall (1941). From a theoretical point of view this solves the problem; but in practice, since the tetrachoric series converges very slowly for large ρ_{ij} it is of little use.

Method of Plackett (1954) which expresses the trivariate integral as a sum of lower dimensional normal integrals and an integral which is to be evaluated by numerical integration, too, is not suited to our problem.

The procedure given by S.C. Das (1956) which consists in reducing the trivariate normal integral to a single integral which is then to be evaluated numerically meets this situation. But this method is also limited in its scope since for it implies that the correlations ρ_{12}, ρ_{13} and ρ_{23} are such that their joint product is positive and each one is numerically greater than the product of the other two. For problems of selection in dairy cattle breeding ρ_{13} is always equal to the product of the other two as has been shown in a later section. The method is thus ruled out.

Peter (1959) has discussed a numerical solution of multivariate normal integrals, This is also restricted to cases where covariance matrix is equal to the sum of a diagonal matrix, say D, and the product of a raw vector with its transpose.

In the words of Peter Ihm (1959), who was concerned with the evaluation of multivariate normal integral, "the most satisfying general method seems to be the Monte Carlo method by use of an electronic computor".

In special cases where the units under selection are all retained at one of the stages, the problem is much simplified. It reduces to two stage selection scheme. A problem of this nature has been examplified in section 7.

-: o :-

6. REDUCTION TO LOWER STAGES

In the formula for three stage selection if we omit the suffix 2 wherever it is occuring and replace 3 by 2 we get the corresponding form of (6) for two stage selection given by Cochran.

$$G(y) = \frac{\rho_1 Z_{(k_1)} I_2 + \rho_2 Z_{(k_2)} I_1}{\alpha_1 \alpha_2}$$

where

$$I_1 = I\left(\frac{k_1 - k_2 \rho_{12}}{\sqrt{1-\rho_{12}^2}}\right)$$

$$I_2 = I\left(\frac{k_2 - k_1 \rho_{12}}{\sqrt{1-\rho_{12}^2}}\right)$$

The same rule can be applied for stepping one stage down say from rth to (r - 1)th stage by omitting the suffix r - 1 wherever it occurs and replacing r by r - 1.

- : o : -

7. AN APPLICATION OF THE METHOD

The application of the foregoing method may be illustrated with reference to a plan for evolving a new breed of dairy cattle by crossing an Indian breed with a suitable foreign breed and inter-breeding and selecting among the F_2 and subsequent generations. A breeding programme of this nature is necessarily of a long duration and very expensive. It is therefore of paramount importance that such a programme is drawn up with great care taking into consideration both the resources available and relative efficiencies in terms of the rate of genetic gain achieved through alternative programmes. Of the different facets of the programme, the discussion has been confined to the selection of breeding cows only.

Once the choice among breeds, both exotic and indigenous to be used for cross-breeding, is decided, the breeding plan will consist of inter-breeding the cross-bred progeny coupled with intensive selection among the F_2 and subsequent generations, the selection being made among the breeding cows on the consideration of their own lactation performance.

To permit sufficient scope for intensive selection among the females, it is necessary to raise a reasonable number of adult females in each generation. This number should be maintained at a constant level so that the herd strength fluctuates only between narrow limits from year to year. Any steady increase in the number from generation to generation would be a strain on the resources. Any steady decrease would ultimately leave only a very small herd of the new breed evolved under the programme. After specification of the number of F_2 females and envisaging the stable condition of raising the same number of females for breeding in each generation, the pattern of selection after the first and successive lactations among F_2's and later generations has to be considered.

Let N be the number of F_2 females completing their first lactation. Out of these N cows, a fraction α_1 having the highest yields is selected, the rest being discarded. From among those $\alpha_1 N$ cows which complete the second lactation, a fraction α_2 having the highest yields is selected, the remainder being culled. A fraction α_3 is selected from $\alpha_1\alpha_2 N$ cows on the basis of their first three lactation records and the remainder is discarded and similarly selection being made for further stages of selection. The above process is repeated for the successive generations of cows till we are left with improved quality stock of a new breed.

It is not advisable to retain even the better animals excepting the few outstanding ones, if any, for more than four lactations for the case when the herd strength from generation to generation is envisaged to be more or less constant. This is so because the culling of all the cows after only one or two lactations would mean a continuous reduction in the herd strength, and on the other hand retaining selected cows for a larger number of lactations would mean an increase in the generation interval and a corresponding decrease in the rate of genetic improvement per year.

The problem, in case of one stage selection programme while envisaging a constant female strength from generation to generation, amounts to one of considering the optimal value of intensity of selection given by equation

$$pN + pL_1N = N$$

$$\text{or} \quad L_1 = \frac{1-p}{p} \qquad (8a)$$

where for every N cows bred, pN is the number of their daughters expected to complete their first lactation. This equation admits solution of L_1 for $p \geq 0.5$. But in dairy cows such a high rate of reproduction is not possible. It has been shown at a later stage that the value of p would be in the neighbourhood of 0.4 in most cases. For this value of p, equation (8a) gives an impossible value of 1.5 for i_1. This means that a permissible value of i_1 cannot be determined unless restriction of raising the same number of females in each generation is waived. In that case there will be a continuous decrease in the herd strength from generation to generation of the order of 1/5 to 3/5 times the previous generation number.

For two stage selection, equation corresponding to (8a) takes the form

$$i_1 + i_1 i_2 = \frac{1 - p}{p} \qquad (8b)$$

Although this equation is solvable for i_1 and i_2, the contribution to the expected percentage genetic advance will be quite low as compared to that under three stage selection. This can be seen from Table I vide sets 2, 3 and 7.

Keeping the above considerations in view the problem then reduces to one of considering the optimal values of intensities of selection at three different stages subject to the restriction that the same number of females are raised in each generation i.e. $i_1 + i_1 i_2 + i_1 i_2 i_3 = \frac{1 - p}{p}$ (8c)

One approach to the solution of the problem would be to consider a range of set of values for i_1, i_2 and i_3 satisfying equation (8c) and examining the values which maximize the average ~~percentage~~ genetic advance which is given by :

$$\text{Average genetic advance} = \frac{N\alpha_1 G'_1(y) + N\alpha_1\alpha_2 G'_2(y) + N\alpha_1\alpha_2\alpha_3 G'_3(y)}{N + N\alpha_1 + N\alpha_1\alpha_2 + N\alpha_1\alpha_2\alpha_3} \qquad (9)$$

where $G'_1(y)$, $G'_2(y)$ and $G'_3(y)$ are the amounts of average genetic advance expected after first, second and third selection respectively. These are given by

$$G'_1(y) = \frac{\rho_1 Z(k_1)}{\alpha_1}\,\sigma_1(y)$$

$$G'_2(y) = \frac{\rho_1 Z(k_1) I_2 + \rho_2 Z(k_2) I_1}{\alpha_1\alpha_2}\,\sigma_2(y)$$

$$G'_3(y) = \frac{\rho_1 Z(k_1) I_{23} + \rho_2 Z(k_2) I_{13} + \rho_3 Z(k_3) I_{12}}{\alpha_1\alpha_2\alpha_3}\,\sigma_3(y)$$

and

$\sigma_1^2(y)$ is the variance of y in the unselected population

$\sigma_2^2(y)$ is the variance of y among the units retained after first selection

$\sigma_3^2(y)$ is the variance of y among the units retained after second selection

Before we can proceed to find the average ~~percentage~~ genetic advance given by formula (9) for different sets of values of α_1, α_2 and α_3, we need the estimates of different parameters appearing therein.

Let y denote the genotypic value of lactation yield of a cow and x_1, x_2 and x_3 the phenotypic values for the first, second and third lactation yields respectively. These may be expressed as

$$x_1 = y + e_p + e_1$$
$$x_2 = y + e_p + e_2$$
$$x_3 = y + e_p + e_3$$

where e_p is the environmental error considered constant over different lactations and e's are the errors due to environmental factors varying from lactation to lactation.

The variates y, e_p and e's are assumed to be normally and independently distributed with zero means. We may also assume $\sigma^2_{e_1} = \sigma^2_{e_2} = \sigma^2_{e_3}$
In that case $\sigma^2_{x_1} = \sigma^2_{x_2} = \sigma^2_{x_3} = \sigma^2_P$ (say)

By the theory of least squares, η_1, η_2 and η_3 can then be shown to be equal to

$$\eta_1 = h_1^2 \, x_1$$
$$\eta_2 = \frac{h_1^2}{1+R} \, (x_1 + x_2)$$
$$\eta_3 = \frac{h_1^2}{1+2R} \, (x_1 + x_2 + x_3)$$

where h_1^2, the coefficient of heritability and R, the coefficient of repeatability are defined respectively as $\frac{\sigma^2_y}{\sigma^2_P}$ and $\frac{\sigma^2_y + \sigma^2_{e_p}}{(\sigma^2_y + \sigma^2_{e_p}) + \sigma^2_e}$.

Illustration for η_2:

Let $y = ax_1 + bx_2$

as η_2 is the regression of y on x_1 and x_2.

Normal equations corresponding to a and b are

$$a \sum x_1^2 + b \sum x_1 x_2 = \sum yx_1$$

$$a \sum x_1 x_2 + b \sum x_2^2 = \sum yx_2$$

or

$$a \sigma_p^2 + bR \sigma_p^2 = \sigma_y^2 = h_1^2 \sigma_p^2$$

$$aR \sigma_p^2 + b \sigma_p^2 = \sigma_y^2 = h_1^2 \sigma_p^2$$

Soving for a and b we get

$$a = b = h_1^2/(1 + R)$$

$$\therefore \eta_2 = \frac{h_1^2}{1 + R} (x_1 + x_2)$$

ρ_1, ρ_2 and ρ_3 - the simple correlations between y and η_1; y and η_2 and y and η_3 respectively can be shown easily to be equal to

$$\rho_1 = h_1$$

$$\rho_2 = h_1 \sqrt{\frac{2}{1+R}}$$

$$\rho_3 = h_1 \sqrt{\frac{3}{1+2R}}$$

Then ρ_{12}, the correlation between η_1 and η_2

$$= \frac{\text{cov}(\eta_1 \eta_2)}{\sqrt{\text{Var}\,\eta_1 \,\text{Var}\,\eta_2}}$$

$$= \sqrt{\frac{1+R}{2}}$$

Likewise ρ_{13} and ρ_{23} can be shown to be

$$\rho_{13} = \sqrt{\frac{1+2R}{3}}$$

$$\rho_{23} = \sqrt{\frac{2(1+2R)}{3(1+R)}}$$

(It can be seen that $\rho_{13} = \rho_{23}\,\rho_{12}$).

To solve equation (8) for $\mathcal{L}_1$, $\mathcal{L}_2$ and $\mathcal{L}_3$ we need to know p. For that it is essential to assume values for vital statistics such as mortality rate, infertility rate, etc. for the cross-bred animals. The available literature could not provide much information on the subject excepting in papers by Kartha (1934), Littlewood (1933), Macguckin (1937) and Stonaker and others (1953). The figures reported in these papers are surprisingly high as compared to those for indigenous breeds at Government Livestock Farms, the breeding data of which have been examined at the I.A.R.S. Kartha studied the figures from data 1912 to 1930 at a number of military dairy farms. He gives figures for half-breds of 21 per cent for infertility and 14 per cent for abortions and still births. Macgukin studied the mortality of cattle in the military dairy farms in the Northern circle from 1935-37. He reports 46 deaths out of an average daily number of 479 adults during the period. The corresponding figure for calves is 28 out of 283. He also reports culling of 18 animals for sterility out of 479. Littlewood indicates the death of 13 calves and adults among the half-breds and gives a figure of 27 calves born. He mentions that during the period a total of 106 died out of which 66 were calves. Littlewood ascribed the high rate of mortality among calves due to their improper feeding and the findings of Stonaker and others are based on statistically insufficient data.

In the absence of consistent values, following figures were assumed for different vital characteristics for illustration :

1.	Sex ratio	50 : 50
2.	Infertility rate	5 per cent
3.	Abortions & still births	2 per cent
4.	Mortality upto one year	6 per cent
5.	Mortality one year to completion of Ist lactation	3 per cent
6.	Adult mortality	2 per cent
7.	Age at first calving	3 years.

In assuming the figures given above, the concept of infertility has been used in a special sense. A female not conceiving within about ten months after calving (in case of heifers after maturity i.e. 2¼ years) has been considered as infertile.

Under these assumptions, for every 100 cows bred 40 daughters would be expected to complete their first lactation. For this value of p i.e. 0.4, equation (8) reduces to

$$\ell_1 + \ell_1\ell_2 + \ell_1\ell_2\ell_3 = 1.5 \qquad (10)$$

For computation we may further assume the values of h_1^2, R and C, the coefficient of variation, as 0.3, 0.7 and 40 respectively. These are close to the values obtained in the course of extensive studies on breeding data of herds of Indian cattle at livestock farms.

---

* The genetic variance $\sigma^2(y)$ decreases with successive stages of selection, the magnitude of which depends upon the intensity of selection. However for computational convenience, $\sigma^2(y)$ has been assumed to remain unaltered under various stages of selection.

---

maximum limiting value viz. unity. This fixes the corresponding point of truncation as $-\infty$. After fixing one of the k's in the manner described the other two points could easily be found by following S.C. Das's method for two variates referred earlier. The other sets containing odd values of ℓ_1, ℓ_2 and ℓ_3 have been omitted as in those cases the fixation of k_3 would have involved unmanageably heavy computation (vide section 6).

Further computations are self explanatory and can be

Table I: Expected average percentage genetic advance

Serial No.	Proportions retained			Points of truncation			Normal ordinates at the points of truncation			$\frac{k_1 - \rho_{12} k_2}{\sqrt{1-\rho_{12}^2}}$ = L (say)	$\frac{k_2 - \rho_{12} k_1}{\sqrt{1-\rho_{12}^2}}$ = M (say)
	α_1	α_2	α_3	k_1	k_2	k_3	$Z(k_1)$	$Z(k_2)$	$Z(k_3)$		
1	2	3	4	5	6	7	8	9	10	11	12
1	1	0.25	1	$-\infty$	0.674490	$-\infty$	0	0.317776	0	$-\infty$	$-\infty$
2	1	0.50	-	$-\infty$	0	-	0	0.398942	-	$-\infty$	$+\infty$
3	0.75	1	-	-0.674500	$-\infty$	-	0.317776	0	-	$+\infty$	$-\infty$
4	0.50	1	1	0	$-\infty$	$-\infty$	0.398942	0	0	$+\infty$	$-\infty$
5	1	0.36	0.36	$-\infty$	0.358459	1.175000	0	0.374118	0.200040	$-\infty$	$+\infty$
6	0.58	1	0.58	-0.201900	$-\infty$	0.400000	0.390894	0	0.368269	$+\infty$	$-\infty$
7	0.82	0.82	-	-0.915400	-0.481700	-	0.262400	0.355237	-	-1.216900	0.935300
8.	0.65	0.65	1	-0.385300	0.166200	$-\infty$	0.370399	0.393470	0	-1.390400	1.346300

$\rho_1 = 0.547723$, $\rho_2 = 0.594089$, $\rho_3 = 0.612373$ $\rho_{12} = 0.921954$

Table I: Expected average percentage genetic advance (contd.).

$I_1 = \frac{1}{\sqrt{2\pi}}\int_L^{\infty} e^{-t^2/2}\,dt$	$I_2 = \frac{1}{\sqrt{2\pi}}\int_M^{\infty} e^{-t^2/2}\,dt$	$\frac{k_1 - k_3\rho_{13}}{\sqrt{1-\rho_{13}^2}} = K_1$ (say)	$\frac{k_2 - k_3\rho_{23}}{\sqrt{1-\rho_{23}^2}} = K_2$ (say)	$\frac{k_2 - k_1\rho_{21}}{\sqrt{1-\rho_{12}^2}} = K_3$ (say)	$\frac{k_3 - k_1\rho_{31}}{\sqrt{1-\rho_{13}^2}} = K_4$ (say)	$\frac{k_3 - k_2\rho_{32}}{\sqrt{1-\rho_{23}^2}} = K_5$ (say)	$\frac{k_1 - k_2\rho_{12}}{\sqrt{1-\rho_{12}^2}} = K_6$ (say)	$I_{12} = I(K_1, K_2; \rho_{12.3})$	$I_{23} = I(K_3, K_4; \rho_{23.1})$	$I_{31} = I(K_5, K_6; \rho_{13.2})$
13	14	15	16	17	18	19	20	21	22	23
1	0	$-\infty$	∞	∞	$\infty - \infty$	$-\infty$	$-\infty$	0	0	1
1	0	-	-	-	-	-	-	-	-	-
0	1	-	-	-	-	-	-	-	-	-
0	1	∞	$\infty - \infty$	$-\infty$	$-\infty$	$-\infty$	∞	0	1	0
1	0	$-\infty$	-3.222500	∞	∞	3.411300	$-\infty$	1	0	0
0	1	-1.251200	$-\infty$	$-\infty$	1.298000	∞	∞	0.894569	0.097145	0
0.888178	0.174818	-	-	-	-	-	-	-	-	-
0.917796	0.089105	∞	∞	1.346300	$-\infty$	$-\infty$	-1.390469	0	0.089105	0.917807

$\rho_{13} = 0.894427$; $\rho_{23} = 0.970143$ $\rho_{12.3} = 0.50$; $\rho_{23.1} = 0.840$; $\rho_{31.2} = 0$

Table I: Expected average percentage genetic advance (contd.)

$\rho_1 Z_{(k_1)}(1 + I_2 + I_{23})$	$\rho_2 Z_{(k_2)}(I_1 + I_{13})$	$\rho_3 Z_{(k_3)} I_{12}$	$\rho_1 Z_{(k_1)}(1+I_2+I_{23}) + \rho_2 Z_{(k_2)}(I_1 + I_{13}) + \rho_3 Z_{(k_3)} I_{12}$ = (24) + (25) + (26)	$1+\alpha_1+\alpha_1\alpha_2 + \alpha_1\alpha_2\alpha_3$	Percentage Genetic Advance: $\frac{(27)}{(28)} \times 40\, h_1$	Rank
24	25	26	27	28	29	30
0	0.377574	0	0.377574	2.5000	3.31	4
0	0.237007	-	0.237007	2.5000	2.08	8
0.348106	0	-	0.348106	2.5000	3.05	6
0.655530	0	0	0.655530	2.5000	5.74	2
0	0.222259	0.122499	0.344768	2.4896	3.03	7
0.449003	0	0.201741	0.650744	2.4964	5.71	3
0.168847	0.187443	-	0.356290	2.4924	3.13	5
0.239030	0.429083	0	0.668113	2.4960	5.87	1

Table I : Expected average percentage genetic advance (contd.)

Percentage Genetic Advance by Method A	Rank	Percentage Genetic Advance by Method B	Rank
31	32	33	34
3.36	4	3.37	6
2.08	8	2.08	8
3.18	5	3.05	7
6.13	1	5.74	3
3.09	7	3.39	4
5.88	3	6.07	2
3.15	6	3.42	5
5.93	2	6.61	1

followed easily step by step. Finally the expected average percentage genetic gain in y has been calculated from the formula (9) which after simplification reduces to

$$\frac{\rho_1 Z_{(k_1)} (1 + I_2 + I_{23}) + \rho_2 Z_{(k_2)} (I_1 + I_{13}) + \rho_3 Z_{(k_3)} I_{12}}{1 + \alpha_1 + \alpha_1 \alpha_2 + \alpha_1 \alpha_2 \alpha_3} \times h_{i}c$$

From Table 1, column 29, it is seen that the scheme number 8 i.e. $\alpha_1 = 0.65 = \alpha_2$ and $\alpha_3 = 1$ is the best set to adopt for selection programme as this results in maximum average percentage genetic advance out of all the eight different sets considered here.

- : o : -

8. SIMPLER APPROXIMATIONS

It will be observed that the use of the foregoing formula even for three stage selection programme involves very cumbersome integrals which are not easy to evaluate. Beyond stage three, we require multivariate normal tables for fixing the values of truncation points. These are not available at present. Even if we resort to reduction method given by Plackett (1954) and other authors, the numerical integration becomes too complex to make it useful for higher dimensions. To overcome these difficulties, two empirical approximations were tried. They are detailed below.

Method A: According to this approximation method the genetic advance expected under different successive stages of selection can be taken to be equal to

$$G_1'(y) = h_1 \frac{Z_1'}{\alpha_1} \sigma_1(y)$$

$$G_2'(y) = h_2 \frac{Z_2'}{\alpha_1 \alpha_2} \sigma_2(y)$$

$$G_3'(y) = h_3 \frac{Z_3'}{\alpha_1 \alpha_2 \alpha_3} \sigma_3(y)$$

- - - - - - - - -

- - - - - - - - -

$$G_r'(y) = h_r \frac{Z_r'}{\alpha_1 \alpha_2 \cdots \alpha_r} \sigma_r(y)$$

where

Z_1' is the normal ordinate corresponding to α_1

z'_2 is the normal ordinate corresponding to $\alpha_1\alpha_2$

z'_3 is the normal ordinate corresponding to $\alpha_1\alpha_2\alpha_3$

- -

z'_r is the normal ordinate corresponding to $\alpha_1\alpha_2 \cdots\cdots \alpha_r$.

h_1^2, h_2^2 etc. and C are ~~as already~~ defined ~~in section~~ as per attached slip.

h_1^2 = the coefficient of heritability based on first lactation records.

h_2^2 = the coefficient of heritability based on first two lactation records $= 2h_1^2/1+R$

h_3^2 = the coefficient of heritability based on first three lactation records $= 3h_1^2/1+2R$

etc.

C = coefficient of variation of the lactation yield

R = the coefficient of repeatability

strictly holds when selection is practised on the basis of first r records and a proportion $\alpha_1 \alpha_2 \cdots\cdots \alpha_r$ of the best cows from the original population is retained while affecting selection at the rth stage. But in practice the selection will be based on a more limited information, in as much as the earlier cullings would have been made on the basis of fewer lactation records and as such the advance is likely to be smaller. However the approximation is of interest as it provides an upper limit to the gain that may accrue from selection on the compilation of successive lactations.

Method B: Another approximate method which was tried for evaluating the average genetic advance can be put in the following form :

Average genetic advance

$$= \frac{N\alpha_1 \Delta D_1 + N\alpha_1\alpha_2(\Delta D_1 + \Delta D_2) + N\alpha_1\alpha_2\alpha_3(\Delta D_1 + \Delta D_2 + \Delta D_3) + \cdots\cdots}{N + N\alpha_1 + N\alpha_1\alpha_2 + N\alpha_1\alpha_2\alpha_3 + \cdots\cdots}$$

in which

$$\Delta D_1 = G_1'(y) = h_1' \frac{Z_1''}{\alpha_1} \sigma_1(y)$$

= average genetic superiority of units retained in the first selection.

$\Delta D_2 = h_2 \frac{Z_2''}{\alpha_2}\sigma_2(y)$ = additional genetic superiority of units obtained from the second culling of units retained in the first selection.

$\Delta D_3 = h_3 \frac{Z_3''}{\alpha_3}\sigma_3(y)$ = additional genetic superiority of units obtained from the third culling of units retained in the second selection.

etc.

where

Z_1'' is the normal ordinate corresponding to α_1

Z_2'' is the normal ordinate corresponding to α_2

Z_3'' is the normal ordinate corresponding to α_3

etc.

Other quantities are as already defined.

The average percentage genetic advance for three stage selection in this case becomes :

Av. percentage genetic advance

$$= \frac{h_1 Z_1''(1+\alpha_2+\alpha_2\alpha_3) + h_2 Z_2''(\alpha_1+\alpha_1\alpha_3) + h_3 Z_3''\alpha_1\alpha_2}{1+\alpha_1+\alpha_1\alpha_2+\alpha_1\alpha_2\alpha_3} \times c h_1$$

The expression for additional genetic superiority ΔD_2 will be realised when a proportion α_2 is retained from the original population. This, too, in practice is not feasible.

The expected contribution to the average percentage genetic advance by these two approximations have been worked out in Table 1, column 31 and 33 for the same sets of values of α_1, α_2 and α_3 as considered earlier. It is seen that the values obtained by method A are closer to those obtained by the more cumbersome approach than the values given by method B. As such method A is to be preferred.

It may be recalled that while discussing the application of the formula in section 7, the sets with odd values of α_1, α_2 and α_3 were omitted as they involve heavy calculations. The approximate methods are not only easily applicable for all possible values of intensities of selection at various stages but also can be extended to any stage.

It is seen from Table 1 that of all the schemes of selection, scheme number 8 if we follow the rigorous or the approximate approach B and scheme number 4 if we follow method A, gives the optimum values of α_1, α_2 and α_3. These schemes, however, are limited in their application as they do not allow scope for selection at all the three stages. Keeping this in view, other values of α's, which at the same time give fairly high value of genetic advance were examined by using methods A and B. With this cirterion, the near-optimum values of α's were found to be $\alpha_1 = 0.6$, $\alpha_2 = 0.8$ and $\alpha_3 = 0.875$. with the corresponding average percentage genetic advance of 6.03

and 6.81 for method A and method B respectively. Thus among the selection programmes spread over three stages the one which envisages selection of 60 per cent at the first stage, 80 per cent of the selected lot at the second stage and finally 87.5 per cent of selecting the units retained upto the second stage, may be considered the best from both the operational aspect as well as from the point of view of genetic improvement.

- : o : -

9. SUMMARY

An expression for the gain in genetic advance for multivariate normal populations under successive stages of selection has been derived. The problem of three stage selection has been dealt with in detail with particular reference to animal breeding and an example has been furnished to illustrate the working procedure. The difficulties in the wake of its application have been brought out.

Two simpler practical approximations for estimating genetic advance have also been discussed.

- : o : -

10. REFERENCES

1. Birnbaum, Z.W.(1950) "Effect of linear truncation on a multi-normal population", Annals of Math. Stat., 21, 272-279.

2. Birnbaum, Z.W. and Chapman, D.G. (1950) "On optimum selections from multi-normal populations", Annals of Math. Stat., 21, 443-447.

3. Cochran, W.G. (1951) "Improvement by means of selection", Proceedings of the Second Berkely Symposium on Mathematical Statistics and Probability, 449-470.

4. Das, S.C. (1956) "The numerical evaluation of a class of integrals II", Proc. Cambridge Phil. Soc., 52, 442-448.

5. David, F.N. (1953) "A note on the evaluation of the multivariate normal integral", Biometrika, 40, 458-459.

6. Dickerson, G.E. and Hazel, L.N. (1944) "Effectiveness of selection on progeny performance as a supplement to earlier culling in livestock", Jour. Agr. Res., 69, 459-476.

7. Finney, D.J. (1957) "Statistical problems of plant selection", Proceedings of the 30th session of the International Statistical Institute.

8. Hazel, L.N. (1943) "The genetic basis for constructing selection indices", Genetics, 28, 476-90.

9. Ihm, Peter (1959) "Numerical evaluation of certain multivariate normal integrals", Sankhya, 21, 363-366.

10. John, S. (1959) "On the evaluation of the probability integral of a multivariate normal distribution", ibid., 21, 367-370.

11. Kartha, K.P.R. (1934) "A note on the comparative economic efficiency of the Indian cow, the half-bred cow and the buffalo as producers of milk and butter fat", Agriculture & Livestock in India, 4, 605-623.

12. Kendall, M.G. (1941) "Proof of relations connected with the tetrachoric series and its generalization", Biometrika, 32, 196-198.

13. Keuls, M. and Sieben, J.W. (1955) "Two statistical problems in plant selection", Euphytica, 4, 34-44.

14. Littlewood, R.W. (1933) "Crossbreeding for milk", Ind. Jour. Vet. Sci. & Anim. Husb., 8, 325-33.

15. Macguckin, C.E. (1937) "Crossbred and grade dairy cattle in India", Ind. Jour. Vet. Sci. & Anim. Husb., 7, 263-272.

16. Moran, P.A.P. (1956) "The numerical evaluation of a class of integrals", Proc. Camb. Phil. Soc., 1956, 52, 230-233.

17. Owen, D.B. (1956) "Tables for computing bivariate normal probabilities", Ann. Math. Stat., 27, 1075-1090.

18. Panse, V.G. (1946) "An application of the discriminant function for selection in poultry", J. Genet., 47, 242-8.

19. Pearson, Karl (1931) "Tables for statisticians and Biometricians" Part II, Cambridge University Press.

20. Plackett, R.L. (1954) "A reduction formula for normal multivariate integrals", Biometrika, 41, 351-360.

21. Sieben, J.W. (1954) "Selectie bij de plantenveredeling op grond van proefveld resultaten", Statistica Neerlandica, 8, 179-189.

22. Smith, H.F. (1936) "A discriminant function for plant selection", Annals of Eugenics, 7, 240-250.

23. Steck, G.P. (1958) "A table for computing trivariate normal probabilities", Annals of Math. Stat., 29, 780-800.

24. Stonakar, H.H., Aggarwala, H.H. and Sundaresan, D. (1953) "Production characteristics of crossbred, backcross and purebred Red Sindhi cattle in the Gangetic Plains Region", 36, 678-687.

25. Tang, Y. (1938) "Certain statistical problems arising in plant breeding", Biometrika, 30, 29-56.

- : o : -