LEAST SQUARES METHOD FOR COMPARING PROGRESS AMONG RECURRENT SELECTION METHODS

S. A. Eberhart

Progress from selection for quantitative traits is usually assumed to be linear during early cycles of selection. As the genetic variance changes because of inbreeding and changes in gene frequencies due to selection, progress per cycle is expected to be nonlinear. When the populations from many cycles of selection are tested simultaneously and the gene frequencies for the desired alleles are above 0.5 at most loci controlling the character under selection, a model which expresses the relation expected among populations is as follows:

\[ Y_{ij} = \mu_0 + \beta_1X_{1i} + \beta_2X_{2i} + \delta_{ij} \]

where \( \delta_{ij} \) is the deviation from regression, \( X_{1i} \) and \( X_{2i} \) indicates the cycle of selection; \( i = 0, 1, \ldots, c \). The least squares regression analysis \((1, 3, 4)\) is used to estimate the mean of the base population \((\mu_0)\) and the linear and quadratic coefficients \((\beta_1, \beta_2)\), to partition the variation among populations into sums of squares due to linear and quadratic effects and to deviations from regression, and to make the necessary tests of significance. However, if the frequencies of most desired alleles are very low in the base population, a cubic response may be expected if enough cycles of selection have been made so that the genetic variance first increased and then later decreased as frequencies of the desired alleles increased. If this cubic response is expected, the model should be extended to include the cubic term.

A somewhat different model and analysis are appropriate when two or more distinct populations are developed from the same base population or variety by different methods of recurrent selection. For instance, several different testers may be used on the same base population to evaluate the efficiency of different types of testers. Different methods of testing such as selection based on single line performance versus selection based on testcross progeny performance may also be evaluated. In such situations an estimate of the progress per cycle with its standard error as well as a statistical test for differences in rates of progress among the various methods being evaluated is desired for each tester.

The Proposed Model

When only one base population exists, a single model with one parameter for the base population mean is desired in order to partition the variation among the base population and all other populations developed from this population by the various methods of recurrent selection. Assuming that a quadratic model is adequate, an appropriate model would be as follows:

\[ Y_{ij} = \mu_0 + \beta_1X_{1i} + \beta_2X_{2i} + \delta_{ij} \]

where \( X_{1i} \) and \( X_{2i} \) indicates the cycle of selection for the jth method; \( i = 0, 1, \ldots, c \) and \( j = 0, 1, \ldots, m \). The \( X_{1ij} \)'s and \( X_{2ij} \)'s are the coefficients for the design matrix.

\[ \begin{align*}
Y_{ij} & = \mu_0 + \beta_1X_{1ij} + \beta_2X_{2ij} + \delta_{ij} \\
\end{align*} \]

The parameters are \( \mu_0, \beta_1 \) and \( \beta_2 \), where \( \mu_0 \) is the mean of the base population, \( \beta_1 \)'s are the linear coefficients for the jth method, and \( \beta_2 \)'s are the quadratic coefficients for the jth method, and the \( \delta_{ij} \)'s is the deviation.

When all populations are compared simultaneously in randomized block design in several environments, the usual replication and environmental parameters can be added and the usual analysis computed. When entries are orthogonal with replications and environments, the following four models could then be fitted successively to partition the variation among entry means and to estimate the desired parameters:

\[ \begin{align*}
Y_{ij} & = \mu_0 + \beta_1X_{1ij} + \beta_2X_{2ij} + \delta_{ij} \\
Y_{ij} & = \mu_0 + \beta_1X_{1ij} + \beta_2X_{2ij} + \delta_{ij} \\
Y_{ij} & = \mu_0 + \beta_1X_{1ij} + \beta_2X_{2ij} + \delta_{ij} \\
Y_{ij} & = \mu_0 + \beta_1X_{1ij} + \beta_2X_{2ij} + \delta_{ij} \\
\end{align*} \]

The unweighted analysis of variance shown in Table 1 was computed from entry means for each of the 7 environments since the number of replications was not the same.

### An Example for High and Low Selection

Although two cycles of selection will provide only preliminary information on the effectiveness of recurrent selection in a population, the data from high and low recurrent selection for yield in the Alpha variety of maize \((2)\) will illustrate the usefulness of this method of analysis. The \( X_4 \) matrices for the kth model and the Y vector are as follows:

\[ X_1 = \begin{bmatrix} 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 1 & 0 \\ \end{bmatrix}, \quad Y = \begin{bmatrix} Y_0 \\ Y_1 \\ Y_2 \\ Y_3 \\ \end{bmatrix} \]

Then \( B_k = A^{-1}G_k \) where \( B_k \) is the vector of estimated parameters for the kth model, \( A_k = X_k'X_k \), and \( G_k \) = \( X_k'Y_k \).

The unweighted analysis of variance shown in Table 1 was computed from entry means for each of the 7 environments since the number of replications was not the same.