Theor. Appl. Genet. 54, 273-275 (1979)

THEORETICAL AND APPLIED GENETICS

An Interval Estimation of Expected Response to Selection

G.C.C. Tai
Agriculture Canada Research Station, Fredericton, New Brunswick (Canada)

Summary. A method is presented here for obtaining an interval estimate of expected response to selection based on results of a progeny test experiment. The structure of the constructed confidence limits is then examined for the influence of the numbers of lines and replicates on the precision of predicting the expected response to selection.

Key words: Interval estimation – Confidence interval – Response to selection

Introduction

Predicting the response to selection is one of the most useful concepts in quantitative genetics and can be applied to plant and animal breeding. The size of the expected response of a quantitative trait can be used to assess the efficiency of a given selection scheme and to determine the optimum intensity of selection pressure. The response to selection is the product of the heritability and the selection differential (Lerner 1958; Falconer 1960). Depending on the selection schemes, various formulae have been proposed for measuring the expected response to selection (e.g., Falconer 1960; Sprague 1966; Empig et al. 1971). However, to date their application has been restricted to point estimates of the expected response. The aim of this paper is to derive an expression that will provide for an interval estimate. Although the derivation is based on the simplest experimental design for progeny testing, it can also be extended to more complicated cases.

Derivation of the Interval

Let a group of n progeny lines be tested in a completely randomized experiment with r replicates. A one-way analysis of variance can be performed on data of a quantitative trait of the lines. The genotypic (σ_g^2) and error (σ_e^2) variance components are estimated as follows:

$$\hat{\sigma}_g^2 = (\mathrm{MSL} - \mathrm{MSE})/r$$

$$\hat{\sigma}_e^2 = \mathrm{MSE}$$

where MSL is the mean square among lines with degrees of freedom (n-1), and MSE is the mean square of error with degrees of freedom n(r-1). The expected response to selection is expressed as
$R = ih^2\sigma_p = ih\sigma_g$ (Falconer 1960), where $h^2 = r\sigma_g^2/(r\sigma_g^2 + \sigma_e^2)$, $\sigma_p^2 = \sigma_g^2 + \sigma_e^2/r$ and i is the intensity of selection. We shall attempt to construct a confidence interval for $h^2\sigma_g^2$. An interval for R will follow naturally when we know the interval of $h^2\sigma_g^2$.

Let (1-a) be the confidence coefficient. The (1-a) confidence limits on $r\sigma_g^2 + \sigma_e^2$ are given by Graybill (1976), i.e.,

$$P[L_1 \leq r\sigma_g^2 + \sigma_e^2 \leq U_1] \geq (1\text{-}a) \text{ where}$$
$$L_1 = S_1^2/\chi^2_{a_1:n_1} \tag{1}$$
$$U_1 = S_1^2/\chi^2_{1-a_2:n_1}$$

$a_1 + a_2 = a$, $n_1 = (n\text{-}1)$, $S_1^2 = (n\text{-}1)$ MSL, and $\chi^2_{a_1:n_1}$ and $\chi^2_{1-a_2:n_1}$ are tabulated χ^2 values with n_1 degrees of freedom.

Williams (1962) and Graybill (1976) give a general method of constructing a confidence interval from experimental data by combining two or more known intervals which contain functions of the parameter of interest and other parameters. The method is adapted here in order to obtain a (1-a) interval for $h^2\sigma_g^2$.

Let the confidence limits of $r\sigma_g^4/(r\sigma_g^2 + \sigma_e^2)$ (= $h^2\sigma_g^2$) be L and U such that

$$P[L \leq r\sigma_g^4/(r\sigma_g^2 + \sigma_e^2) \leq U] \geq (1\text{-}a) \tag{2}$$

0040-5752/79/0054/0273/$1.00

It is seen that the limits in (1) and (2) justify the following inequality based on the Bonferroni inequality (Graybill 1976):

$$P[L_1 \leq r\sigma_g^2 + \sigma_e^2 \leq U_1 \text{ and } L \leq r\sigma_g^4/(r\sigma_g^2 + \sigma_e^2) \leq U] \geq (1\text{-}2a) \tag{3}$$

This relationship can be used to obtain a (1-2a) confidence interval for σ_e^2 using William's (1962) method. Let L_2 and U_2 be the confidence limits,

$$P[L_2 \leq \sigma_e^2 \leq U_2] \geq (1\text{-}2a) \tag{4}$$

The (1-2a) confidence limits on σ_e^2, in fact, are known (Graybill 1976):

$$L_2 = S_2^2/\chi^2_{2a_1:n_2}$$

$$U_2 = S_2^2/\chi^2_{1\text{-}2a_2:n_2}$$

where $n_2 = r(n\text{-}1)$, $a_1 + a_2 = a$, $S_2^2 = n(r\text{-}1)$ MSE, and $\chi^2_{2a_1:n_2}$ and $\chi^2_{1\text{-}2a_2:n_2}$ are tabulated χ^2 values with n_2 degrees of freedom. This suggests that (3) and (4) can be used instead to determine the unknown L and U.

Denote l_1, u_1, l_2, u_2, l and u as the calculated values of the variables L_1, U_1, L_2, U_2, L and U respectively. Then the following two sets of simultaneous equations can be established based on the inequalities (3) and (4):

$$\left\{\begin{array}{l} r\sigma_g^2 + \sigma_e^2 = l_1 \\ \sigma_e^2 = l_2 \\ l(r\sigma_g^2 + \sigma_e^2) = r\sigma_g^4 \end{array}\right. \tag{5}$$

$$\left\{\begin{array}{l} r\sigma_g^2 + \sigma_e^2 = u_1 \\ \sigma_e^2 = u_2 \\ u(r\sigma_g^2 + \sigma_e^2) = r\sigma_g^4 \end{array}\right. \tag{5$'$}$$

Solving (5) and (5)′ for l and u,

$$\begin{array}{l} l = (l_1 - l_2)^2/rl_1 \\ u = (u_1 - u_2)^2/ru_1 \end{array} \tag{6}$$

The general formulae for the confidence limits L and U are derived by bringing the expressions of the confidence limits in (1) and (4) into (6), i.e.,

$$\begin{array}{l} L = \left[\dfrac{S_1^2}{\chi^2_{a_1:n_1}} - \dfrac{S_2^2}{\chi^2_{2a_1:n_2}}\right]^2 \Bigg/ \dfrac{rS_1^2}{\chi^2_{a_1:n_1}} \\ U = \left[\dfrac{S_1^2}{\chi^2_{1\text{-}a_2:n_1}} - \dfrac{S_2^2}{\chi^2_{1\text{-}2a_2:n_2}}\right]^2 \Bigg/ \dfrac{rS_1^2}{\chi^2_{1\text{-}a_2:n_1}} \end{array} \tag{7}$$

We are, finally, able to construct a (1-a) confidence interval for the expected response to selection. Let L_R and U_R be the lower and upper confidence limits for R,

$$P[L_R < R < U_R] \geq (1\text{-}a)$$

$$\begin{array}{l} L_R = iL = i\left[\dfrac{S_1^2}{\chi^2_{a_1:n_1}} - \dfrac{S_2^2}{\chi^2_{2a_1:n_2}}\right] \Bigg/ \sqrt{\dfrac{rS_1^2}{\chi^2_{a_1:n_1}}} \\ U_R = iU = i\left[\dfrac{S_1^2}{\chi^2_{1\text{-}a_2:n_1}} - \dfrac{S_2^2}{\chi^2_{1\text{-}2a_2:n_2}}\right] \Bigg/ \sqrt{\dfrac{rS_1^2}{\chi^2_{1\text{-}a_2:n_1}}} \end{array} \tag{8}$$

We see that L_R is replaced by zero when $S_1^2/\chi^2_{a_1:n_1} < S_2^2/\chi^2_{2a_1:n_2}$ and R is taken as zero when both L_R and $U_R < 0$.

The above procedure can be used to establish a generalized confidence interval with any number of replicates r′ based on data from an experiment of r replicates. By replacing r with r′ in the third equation in (5) and (5)′, the generalized limits are given as

$$\begin{array}{l} L'_R = L_R\left[1 - \dfrac{r'\text{-}r}{r'}\,\dfrac{S_2^2\chi^2_{a_1:n_1}}{S_1^2\chi^2_{2a_1:n_2}}\right]^{-\frac{1}{2}} \\ U'_R = U_R\left[1 - \dfrac{r'\text{-}r}{r'}\,\dfrac{S_2^2\chi^2_{1\text{-}a_2:n_1}}{S_1^2\chi^2_{1\text{-}2a_2:n_2}}\right]^{-\frac{1}{2}} \end{array} \tag{9}$$

An Example

An hypothetical example of a one-way ANOVA for a progeny test experiment with 41 lines and 3 replicates (n = 41, r = 3) is used for illustration. Assuming 10% of the best lines are selected, MSL = 40, and MSE = 10, the expected response to selection is estimated as

$$R = ih\sigma_g = 4.66$$

where i = 1.70, h^2 = (MSL-MSE)/MSL = .75, and σ_g^2 = (MSL-MSE)/r = 10. The tabulated χ^2 values for $n_1 = n\text{-}1 = 40$, $n_2 = n(r\text{-}1) = 82$, and $a_1 = a_2 = .05$ are $\chi^2_{.05:40} = 55.8$, $\chi^2_{.10:82} = 98.8$, $\chi^2_{.95:40} = 26.5$ and $\chi^2_{.90:82} = 66.1$. Also $S_1^2 = 1600$ and $S_2^2 = 820$. Bring all the values in (8),

$$L_R = 1.70\left[\frac{1600}{55.8} - \frac{820}{98.8}\right] \Bigg/ \sqrt{\frac{3 \cdot 1600}{55.8}} = 3.74$$

$$U_R = 1.70\left[\frac{1600}{26.5} - \frac{820}{66.1}\right] \Bigg/ \sqrt{\frac{3 \cdot 1600}{26.5}} = 6.06$$

The .90 confidence interval for R is

$$P[3.74 \leq R \leq 6.06] \geq .90$$

Discussion

A comparison of the structure of the confidence limits of the interval estimate with that of the point estimate indicates the importance of both the numbers of lines (n) and replicates (r) used in the progeny test experiment. Let $A = \chi^2_{a_1:n_1}/n_1$ and $B = \chi^2_{2a_1:n_2}/n_2$. L_R can be expressed as

$$L_R = i\left[\frac{S_1^2}{n_1} - \frac{S_2^2}{n_2}\,\frac{A}{B}\right] \Big/ \sqrt{\frac{rS_1^2}{n_1}A}$$

Rewrite R in terms of the sums of squares,

$$R = i\left[\frac{S_1^2}{n_1} - \frac{S_2^2}{n_2}\right] \Big/ \sqrt{\frac{rS_1^2}{n_1}}$$

We see $L_R = R$ if and only if $A = B = 1$. In general, $A > 1$ and $B > 1$ since the probability levels a_1 and a_2 are always chosen to be less than .10 and thus $\chi^2_{a_1:n_1} > n_1$ and $\chi^2_{2a_1:n_2} > n_2$. The actual size of A is determined by n whereas that of B by both n and r. Obviously, the use of an increased number of replicates in an experiment would help to raise heritability of a trait and thus give higher estimates of R and L_R. But a larger r would also reduce the size of B and would increase the size of (rS_1^2/n_1) A to an extent greater than that of rS_1^2/n_1 when $A > 1$. This widens the gap between L_R and R especially when the error mean square is large. Therefore, the prediction of the expected response to selection is consequently less precise. In contrast, using a greater number of lines would narrow the deviations of both A and B from unity and therefore help to bring L_R closer to R. A similar conclusion can also be drawn on the influence of n and r on U_R. It appears the best way to predict the response to selection is to use large numbers of both lines and replicates in an experiment. A balance between n and r is needed for an experiment of limited size so that an acceptable compromise can be reached between the size and precision of the estimated response.

Expected response to selection is often calculated for a greater number of replicates than is used in an experiment. This is designed to determine whether or not further replication is warranted to increase the efficiency of a given selection scheme. We see from (8) $S_2^2\chi^2_{1-a_2:n_1}/S_1^2\chi^2_{1-2a_2:n_2} < S_2^2\chi^2_{a_1:n_1}/S_1^2\chi^2_{2a_1:n_2} < 1$ when $U_R > L_R > 0$. Bring the result to (9), we have $L'_R > L_R$, $U'_R > U_R$, and $U'_R - L'_R > U_R - L_R$ when $r' > r$.

In practice, the progeny test experiment for estimating heritability and expected response to selection requires a more complex experimental design than the one described above. Plant breeders, for example, often choose a random group of breeding lines for replicated experiments conducted over several years and locations (Allard 1960). Similarly, in animal breeding experiments, progenies of many families are tested together and expected responses to selection are compared between alternative selection schemes such as individual, family and within-family selections (Falconer 1960). The former represents a three factor crossed design and the latter a two factor nested design. By applying the general theory of linear models (Searle 1971; Graybill 1976), it is possible to obtain approximate confidence intervals for specific variance components and their linear combinations for such experiments. Once the formulae for these intervals are available, an interval estimate of the expected response to selection in relation to a specific selection scheme can be obtained quite simply by the method of Williams (1962).

Acknowledgement

I wish to thank Dr. K. MacRae and Dr. C.F. Wehrhahn for their critical review of an earlier draft of the paper, and to Dr. R.E. McQueen, Dr. H. De Jong and Dr. M.C. Clark for their review of the manuscript.

Literature

Allard, R.W.: Principles of Plant Breeding. New York: J. Wiley & Sons 1960

Empig, L.T.; Gardner, C.O.; Compton, W.A.: Theoretical gains for different population improvement procedures. Neb. Agr. Exp. Sta. Misc. Publ. 26 (1971)

Falconer, D.S.: Introduction to Quantitative Genetics. Edinburgh: Oliver and Boyd 1960

Graybill, F.A.: Theory and Application of the Linear Model. Belmont, California: Duxbury Press 1976

Lerner, I.M.: The Genetic Basis of Selection. New York: J. Wiley & Sons 1958

Searle, S.R.: Linear Models. New York: J. Wiley & Sons 1971

Sprague, G.F.: Quantitative genetics in plant improvement. In: Plant Breeding (ed. Frey, K.J.), chapter 8. Ames, Iowa: Iowa State Univ. Press 1966

Williams, J.S.: A confidence interval for variance components. Biometrika **49**, 278-281 (1962)

Received December 22, 1978
Communicated by R.W. Allard

Dr. G.C.C. Tai
P.O. Box 20280
Research Station
Agriculture Canada
Fredericton, N.B. E3B 4Z7
(Canada)