

Estimation of variance components based on triallel mating design

M. R. Srinivasan* and K. N. Ponnuswamy

Department of Statistics, University of Madras, Chepauk, Madras-600005, India

Received August 10, 1992; Accepted September 3, 1992

Communicated by A. R. Hallauer

Summary. The simple mating designs provide unbiased estimates for genetic components of variance (additive genetic variance and dominance variance) under the assumption of no epistatic effect. There is empirical evidence, however, that suggests the existence of epistatic gene effects. The triallel and double cross mating designs permit the estimation of epistatic gene effects. A systematic and mathematical approach is suggested for the estimation of variance components based on the alternate model for triallel mating design.

Key words: Triallel crosses – Design components – Genetic components

Introduction

There are two types of designs involved in any breeding experiment, mating design and experimental or environmental design. The systematic manner in which progenies are developed and used in the breeding experiment is known as the mating design; for example, diallel and triallel mating designs. The manner in which progenies are evaluated in environments is known as an experimental design; for example, completely randomized design and randomized complete block design. Cockerham (1961) emphasized that one should be judicious in the choice of mating design and the experimental design to be used. The type and nature of information obtained from the breeding experiment depend on the mating and experimental design adopted, the precision of the estimates, and the validity of the tests of hypotheses. The design

components of variance are estimated based on the linear model adopted. The genetic components of variance are obtained on the basis of the relationship between the design and genetic components using the concept of covariance between the relatives under certain broad genetic assumptions. Also, a statistical test for the design components serves as a test for genetic hypotheses of relevance.

The simple diallel mating designs provide unbiased estimators for additive variance σ_A^2 and dominance variance σ_D^2 only under the assumption that the epistatic effects are absent. If the assumption fails, then the estimates would be biased and inferences would not be valid. There is sufficient empirical evidence available in the literature to show that the above assumption is an illusory one. Hence, the development of mating designs that could estimate the epistatic gene effects would provide not only unbiased estimators for σ_A^2 and σ_D^2 but also for the epistatic components of variance. The higher order triallel and double cross mating designs introduced by Rawlings and Cockerham (1962 a, b) provide unbiased estimators for the genetic components of variance (σ_A^2 , σ_D^2 , σ_{AA}^2 , σ_{AD}^2 , and σ_{DD}^2) and only one of the second order epistatic components (σ_{AAA}^2 , σ_{AAD}^2 , σ_{ADD}^2 , and σ_{DDD}^2). Hinkelmann (1965) suggested an alternative model, and Ponnuswamy (1971) developed the analysis based on the alternative model. In this paper, the concept of triallel crosses is first introduced before describing the model and analysis of triallel mating design.

A cross between an F_1 hybrid (AB) involving the distinct lines A and B and unrelated line C is the three-way hybrid (or three-way cross) (AB)C, where A and B are the grand-parents (or half-parents) and the unrelated line C is the parent. In the absence of reciprocal effects and maternal-paternal interactions, the crosses (AB)C and (BA)C would represent the same three-way hybrid.

* Present address: Center for Demographic and Population Genetics, University of Texas, Graduate School of Biomedical Sciences, P.O. Box 20334, Houston, TX, 77225, USA
Correspondence to: M. R. Srinivasan

The triallel mating design can be defined as a set of all possible three-way hybrids based on a given set of lines constituting the triallel crosses. If p is the number of lines in the triallel crosses, there would be $P = p(p-1)(p-2)/2$ distinct three-way crosses. Each line will occur in $r_H = (p-1)(p-2)$ three-way crosses as a grand-parent and in $r_P = (p-1)(p-2)/2$ three-way crosses as a parent. Similarly, each pair of lines will occur in $r_d = (p-2)$ three-way crosses with both as grand-parents and in $r_s = (p-2)$ three-way crosses with one as a grand-parent and the other as a parent.

There are two basic models available for the triallel mating design. The Rawlings and Cockerham (1962a) model is an orthogonal model, whereas the Hinkelmann (1965) model is a nonorthogonal model. Simulation studies conducted by Radhakrishnan (1981) and Ponnuswamy and Srinivasan (1988) have shown that the Hinkelmann (1965) model of ten design components of variance is a more reasonable model to study the specific nature of gene effects. There exist, however, many possible candidates as the estimates of design components of variance for the Hinkelmann (1965) model because it is a nonorthogonal model. Ponnuswamy (1971) has presented one such choice.

There are several methods available for the estimation of variance components from the analysis of variance (ANOVA), including maximum likelihood (ML), restricted maximum likelihood (REML), minimum norm quadratic unbiased estimation (MINQUE), and modified symmetric sum (MSS). The methods involve large corpuses of matrix algebra, numerical and computing problems like the inversion of large-order matrices, and the solving of the relevant non-linear equations subject to certain constraints. The complexity of using any of the methods would be unimaginable when considering a ten-component triallel model with nonorthogonal data, as there exists no unified theory of the estimation of variance components based on linear models.

Seely (1969) considered some general results on the estimation in finite dimensional linear spaces wherein he treated the quadratic unbiased estimation in mixed linear model as a special case. The general mixed model with linear structure is

$$Y = X\beta + \sum_{i=1}^m U_i \varepsilon_i, \quad (1)$$

where Y is a $n \times 1$ vector of observations, X is a $n \times p$ matrix of known constants, and ε_i are $q_i \times 1$ vectors of random variables with

$$E(\varepsilon_i) = 0; \quad \text{Var}(\varepsilon_i) = C_i \sigma_i^2, \quad (2)$$

where C_i 's are $q_i \times q_i$ matrices of known constants and $E(\varepsilon_i \varepsilon_j) = 0$ for $i \neq j$. From model (1) it follows,

$$E(Y) = X\beta = \sum_{i=1}^p X_i \beta_i,$$

and

$$\text{Var}(Y) = \sum_{i=1}^m U_i C_i U_i' \sigma_i^2 = \sum_{i=1}^m V_i \sigma_i^2. \quad (3)$$

By denoting $W = YY'$,

$$E(W) = \sum_{i=1}^p \sum_{j=1}^p B_{ij} \theta_{ij} + \sum_{i=1}^m V_i \sigma_i^2, \quad (4)$$

where $B_{ij} = X_i X_j' + X_j X_i'$; $i \neq j$; $B_{ii} = X_i X_i'$; $\theta = \beta_i \beta_j$; and $M = p(p+1)/2 + m$. Following the general idea of the usual least squares normal equations ($X'X\hat{\beta} = X'Y$), Seely (1969) arrived at the equation

$$H^* H \hat{\theta} = H^* W, \quad (5)$$

where, H^* is the adjoint of H . These equations can be expressed as an $M \times M$ system of equations with

$$H^* H = \begin{pmatrix} (B_{11}, B_{11}) \dots (B_{11}, B_{pp}) & (B_{11}, V_1) \dots (B_{11}, V_m) \\ \vdots & \vdots \\ (B_{pp}, B_{11}) \dots (B_{pp}, B_{pp}) & (B_{pp}, V_1) \dots (B_{pp}, V_m) \\ (V_1, B_{11}) \dots (V_1, B_{pp}) & (V_1, V_1) \dots (V_1, V_m) \\ \vdots & \vdots \\ (V_m, B_{11}) \dots (V_m, B_{pp}) & (V_m, V_1) \dots (V_m, V_m) \end{pmatrix}$$

$$(H^* W)' = ((B_{11}, YY') \dots (B_{pp}, YY') (V_1, YY') \dots (V_m, YY'))$$

and $\theta = [\theta_{11} \dots \theta_{pp} \sigma_1^2 \dots \sigma_m^2]'$. It should be noted that in the above matrices the elements of the form (A, B) represent the trace of (AB) . Yuan (1977), motivated by the works of Seely (1969) and Koch (1967, 1968), has partitioned

$$E(W) = H\theta = H_1 \theta_1 + H_2 \theta_2, \quad (6)$$

where $H_1 \theta_1 = \sum_{i=1}^p \sum_{j=1}^p B_{ij} \theta_{ij}$, and $H_2 \theta_2 = \sum_{i=1}^m V_i \sigma_i^2$, to facilitate the estimation of variance components. Yuan (1977) obtained the least square estimate $H\hat{\theta}$ or $H\theta$ as an orthogonal projection of YY' on the range space of H and has given the explicit form of the projection. The present paper deals with the estimation of variance components based on Hinkelmann's triallel model by adopting the general approach outlined by Seely (1969) and Yuan (1977).

Triallel model

Hinkelmann (1965) proposed the following linear model for the triallel crosses to estimate the variance components:

$$Y_{ijkl} = \mu + h_i + h_j + g_k + d_{ij} + s_{ik} + s_{jk} + t_{ijk} + r_l + e_{ijkl}, \quad (7)$$

where Y_{ijkl} is the yield of the three-way cross $(ij)k$ of the l -th replicate, μ is the general mean, r_l is the l -th replicate effect, h_i and h_j are the general line effect of the grand-parents i and j , g_k is the general line effect of the line k as parent, d_{ij} is the two-line specific effect of the grand-parental lines i and j , s_{ik} and s_{jk} are the two-line specific

effect of the grand-parental lines i and j and parental line k , t_{ijk} is the three-line specific effect of the grand-parental lines i and j and the parental line k , and e_{ijkl} is the error associated with $(ij)k$ cross in the l -th replicate. The general mean and the replicate effects are fixed effects, and the family of random variables h_i , g_i , d_{ij} , s_{ik} , and t_{ijk} are expected to have zero means and the following variance-covariance structure:

$$\begin{aligned} E(h_i^2) &= \sigma_h^2, E(g_i^2) = \sigma_g^2 \\ E(g_i h_i) &= \sigma_{gh}, E(d_{ij}^2) = \sigma_d^2 \\ E(d_{ij} s_{ij}) &= E(s_{ij} d_{ij}) = \sigma_{ds}, E(s_{ij}^2) = \sigma_s^2 \\ E(s_{ij} s_{ji}) &= \sigma_{ss}, E(t_{ijk}^2) = \sigma_t^2 \\ E(t_{ijk} t_{ikj}) &= E(t_{ijk} t_{jki}) = \sigma_{tt}, E(e_{ijkl}^2) = \sigma_e^2 \end{aligned} \quad (8)$$

and all other covariances are zero. The ten components of variances and covariances are designated as design components of variance. The trialallel crosses are superimposed on a completely randomized design with each cross replicated r times constituting the breeding experiment. However, the crosses can be evaluated in randomized complete block (RBD), latin square design (LSD), and incomplete block designs.

Estimation of design components of variance

The trialallel model (7) can be rewritten in the matrix notation as:

$$Y = 1_n \mu + Z_1 R + U_1 H + U_2 G + U_3 D + U_4 S_1 + U_5 S_2 + U_6 L_1 + U_7 L_2 + U_8 L_3 + U_9 \varepsilon, \quad (9)$$

where Y is a $n \times 1$ vector of observations, $n = P \times r$, 1_n is a $n \times 1$ vector of unit elements, Z_1 is a $n \times l$ matrix of binary elements 0 or 1, depending on the replicates, R is the corresponding $l \times 1$ vector of replicates effects, and U_1 is the model matrix corresponding to the vector H , of random variables h_i . The matrix U_1 is of the order $n \times p$, with elements either 0 or 1 depending on the configuration of the grand-parental lines; U_2 is the model matrix corresponding to the vector G , of random variables g_i , and is of the same order as U_1 , but the elements depend on the configuration of the parental lines; U_3 is the model matrix of order $n \times q$ with elements again 0 or 1, $q = pP_1/2$, depending on the configuration of the grand-parental lines i and j ; and D is the corresponding vector of random variables d_{ij} . U_4 and U_5 are also model matrices corresponding to the vectors S_1 , of random variables s_{ij} , and S_2 , of random variables s_{ji} , respectively, having the same order as U_3 , but the configuration depends on the grand-parental line i (or j) and parental line j (or i), where j is less than i . U_6 , U_7 , and U_8 are the matrices of the same order $n \times P/3$ with elements depending on the configuration of the grand-parental lines (i, j) , (i, k) , and (j, k) and with parental lines k, j , and i , respectively. L_1 , L_2 , and L_3 are the vectors of order $P/3 \times 1$ with random variables t_{ijk} , t_{ikj} , and t_{jki} , respectively;

and U_9 is the identity matrix of order n , with ε the corresponding $n \times 1$ vector of error variables.

$$\begin{aligned} E(W) &= E(YY') = Z \theta \theta' Z' + U_1 U_1' \sigma_h^2 + U_2 U_2' \sigma_g^2 \\ &\quad + (U_1 U_2' + U_2 U_1') \sigma_{gh} + U_3 U_3' \sigma_d^2 \\ &\quad + (U_4 U_4' + U_5 U_5') \sigma_s^2 + (U_3 U_4' + U_4 U_3' + U_5 U_3' \\ &\quad + U_3 U_5') \sigma_{ds} + (U_4 U_5' + U_5 U_4') \sigma_{ss} \\ &\quad + (U_6 U_6' + U_7 U_7' + U_8 U_8') \sigma_t^2 + (U_6 U_7' + U_7 U_6' \\ &\quad + U_6 U_8' + U_8 U_6' + U_7 U_8' + U_8 U_7') \sigma_{tt} + U_9 U_9' \sigma_e^2 \end{aligned} \quad (10)$$

$$= Z \theta \theta' Z' + V_1 \sigma_h^2 + V_2 \sigma_g^2 + V_3 \sigma_{gh} + V_4 \sigma_d^2 + V_5 \sigma_s^2 + V_6 \sigma_{ds} + V_7 \sigma_{ss} + V_8 \sigma_t^2 + V_9 \sigma_{tt} + V_{10} \sigma_e^2. \quad (11)$$

Equation (11) adheres to the Yuan (1977) specification for the estimation of variance components based on the trialallel model (7). The structure of V_i 's; $i = 1, \dots, 10$, can be obtained in the case of the trialallel model. Also Seely's (1969) condition for the estimability is satisfied in the case of the trialallel model for the estimation of the ten design components of variance. The estimating equation for obtaining the variance components based on the trialallel model is thus given by

$$H_2^* T_2 H_2 \beta = H_2^* T_2 W, \quad (12)$$

wherein $\beta = [\sigma_h^2 \sigma_g^2 \sigma_{gh} \sigma_d^2 \sigma_s^2 \sigma_{ds} \sigma_{ss} \sigma_t^2 \sigma_{tt} \sigma_e^2]'$, and the matrix of $H_2^* T_2 H_2$ for the trialallel model is a 10×10 matrix with the (i, j) -th element as

$$(V_i, V_j - Q V_j Q); \quad i = 1, 2, \dots, 10 \text{ and } j = 1, 2, \dots, 10.$$

Similarly, the 10×1 vector $H_2^* T_2 W$ for the trialallel model with the i -th element is

$$(V_i, W - Q W Q); \quad i = 1, 2, \dots, 10.$$

The orthogonal projection matrix Q in the case of the trialallel model is given by

$$Q = Z(Z'Z)^{-1}Z' = \text{Diag}(J_P/P) = \frac{1}{P} \begin{pmatrix} J_P & 0 & \dots & 0 \\ 0 & J_P & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & J_P \end{pmatrix}$$

where J_P is a matrix of order $P \times P$ with unit elements. Since the structure of the matrices V_i ; $i = 1, 2, \dots, 10$ are known in terms of U_i ; $i = 1, 2, \dots, 9$, the elements of the 10×10 symmetric matrix $H_2^* T_2 H_2$ can be obtained explicitly in terms of the number of lines p and replications r using the trace properties of the matrix. For example, the $(1, 1)$ -th element of the matrix can be obtained as follows:

$$\begin{aligned} (V_1, V_1 - Q V_1 Q) &= \text{Trace}(V_1 V_1) - \text{Trace}(V_1 Q V_1 Q) \\ &= \text{Trace}[U_1 U_1' U_1 U_1'] \\ &\quad - \text{Trace}[U_1 U_1' D(J_P/P) U_1 U_1' D(J_P/P)] \\ &= \text{Trace}[U_1' U_1 U_1' U_1] \\ &\quad - \text{Trace}[U_1' D(J_P/P) U_1 U_1' D(J_P/P) U_1] \\ &= p^2 p_1^2 p_2^2 r^2 - 4 p_1^2 p_2^2 r^2 \\ &= p_1^2 p_2^4 r^2. \end{aligned}$$

$$\begin{aligned}
& \sum_{i=1}^p Y_{i...} - 2p_1 p_2 r CF \\
& \sum_{i=1}^p Y_{..i}^2 - \frac{p_1 p_2}{2} r CF \\
& \sum_{i=1}^p Y_{i...} Y_{..i} - p_1 p_2 r CF \\
& \sum_{i,j=1}^p Y_{ij..}^2 - p_2 r CF \\
& \sum_{i,j=1}^p Y_{i.j}^2 + \sum_{i,j=1}^p Y_{j.i}^2 - 2p_2 r CF \\
& 2 \left[\sum_{i,j=1}^p Y_{i.j} Y_{ij..} + \sum_{i,j=1}^p Y_{j.i} Y_{ij..} - 2p_2 r CF \right] \\
& 2 \left[\sum_{i,j=1}^p Y_{i.j} Y_{j.i} - p_2 r CF \right] \\
& 2 \left[\sum_{i,j,k=1}^p Y_{ijk}^2 + \sum_{i,j,k=1}^p Y_{jki}^2 + \sum_{i,j,k=1}^p Y_{ikj}^2 - r CF \right] \\
& 2 \left[\sum_{i,j,k=1}^p Y_{ijk} Y_{jki} + \sum_{i,j,k=1}^p Y_{ijk} Y_{ikj} + \sum_{i,j,k=1}^p Y_{jki} Y_{ikj} - 2p_2 r CF \right] \\
& \sum_{i,j,k=1}^p \sum_{l=1}^r Y_{ijk l}^2 - \frac{\sum_{l=1}^r Y_{...l}^2}{P}.
\end{aligned}$$

Thus, for a given number of lines and replications, the ten design components of variance may be obtained directly using the explicitly obtained 10×10 matrix $H_2^* T_2 H_2$ and the 10×1 vector $H_2^* T_2 W$ for the trial data based on Hinkelmann's model. However, it is the genetic components of variance which are of interest to the breeder in his study on the nature of gene effects. Under certain broad genetic assumptions and when a proper genetic

model is considered the genetic components of variance are obtained using the established relationship between the design and genetic components of variance. The estimation of genetic components in the case of the triallel model is discussed in the following section. Ponnuswamy and Srinivasan (1988), from their limited simulation study, observed that Seely's method was quite sensitive in analyzing the triallel model.

Estimation of genetic components of variance

Cockerham (1961) determined the relationship between the design and genetic components of variance for triallel crosses. Ponnuswamy (1971) determined the relationship between the design and genetic components based on the Hinkelmann's model through covariance between relatives. The open genetic model for arbitrary number of loci is given by:

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2 + \sigma_{AA}^2 + \sigma_{AD}^2 + \sigma_{DD}^2 + \sigma_{AAA}^2 + \sigma_{AAD}^2 + \sigma_{ADD}^2 + \sigma_{DDD}^2 + \dots \quad (13)$$

To obtain unbiased estimators for genetic components of variance, Ponnuswamy (1971), Radhakrishnan (1981), and Srinivasan (1986) considered a restricted model:

$$\sigma_G^2 = \sigma_A^2 + \sigma_D^2 + \sigma_{AA}^2 + \sigma_{AD}^2 + \sigma_{DD}^2. \quad (14)$$

This model will be a closed model if the number of loci involved is only two or the higher order epistatic components are assumed to be negligible. If an arbitrary number of loci are involved and the higher order epistatic components are not negligible, the estimators will be biased. Hence, we restricted our attention to the closed model for the estimation of variance components based on triallel crosses. Ponnuswamy (1971) presented the explicit relationship between the design and genetic components of variance when the number of loci involved is exactly two. The five genetic components of variance, using the method of least squares, are:

$$\sigma_A^2 = \frac{1}{206 F^2} [440 \sigma_h^2 + 8 \sigma_g^2 + 588 \sigma_{gh}^2 - 292 \sigma_d^2 - 584 \sigma_{ss}]$$

$$\sigma_D^2 = \frac{1}{206 F^2} \left[416 \sigma_h^2 + 352 \sigma_g^2 + 496 \sigma_{gh}^2 - 336 \sigma_d^2 - 632 \sigma_{ds} \right. \\ \left. - \frac{1648}{3} \sigma_g^2 + \frac{4120}{3} \sigma_{ss} \right] - \frac{2}{F^2} \left[\sigma_t^2 - \frac{1}{3} \sigma_{tt} \right]$$

$$\sigma_{AA}^2 = \frac{1}{206 F^2} [-832 \sigma_h^2 + 704 \sigma_g^2 - 992 \sigma_{gh}^2 + 672 \sigma_d^2 - 1344 \sigma_{ds}]$$

$$\sigma_{AD}^2 = \frac{32}{3 F^3} [\sigma_s^2 - \sigma_{ss} + 2 \sigma_{tt}]$$

$$\sigma_{DD}^2 = \frac{1}{3 F^4} [-16 \sigma_s^2 + 16 \sigma_{ss} + 24 \sigma_t^2 - 32 \sigma_{tt}]. \quad (15)$$

Thus, the ten design components of variance obtained, using the Seely (1969) method, can be used directly to obtain the five genetic components (additive, dominance, additive \times additive, additive \times dominance, and dominance \times dominance) for the Hinkelmann's nonorthogonal triallel model.

Acknowledgements. The authors wish to express their sincere thanks to Prof. A. R. Hallauer, Prof. R. Chakraborty and the referees for their valuable comments and suggestions leading to the improvement of the paper.

References

- Cockerham CC (1961) Implications of genetic hybrid breeding programme. *Crop Sci* 1:47-52
- Hinkelmann K (1965) Partial triallel crosses. *Sankhya* 27:173-196
- Koch GG (1967) A general approach to estimation of variance components. *Technometrics* 9:93-118
- Koch GG (1968) Some further remarks concerning a general approach to estimation of variance components. *Technometrics* 10:551-558
- Ponnuswamy KN (1971) Some contributions to design and analysis of diallel and triallel crosses. PhD thesis, I.A.S.R.I., New Delhi, India
- Ponnuswamy KN, Srinivasan MR (1988) Comparative study of quadratic unbiased estimators for variance components - A simulation study *AMSE Rev.* 8:24-48
- Radhakrishnan G (1981) Some contributions to the estimation of design and genetic components of variance. PhD thesis, University of Madras, Madras, India
- Rawlings JO, Cockerham CC (1962a) Triallel analysis. *Crop Sci* 2:228-231
- Rawlings JO, Cockerham, CC (1962b) Analysis of double cross hybrid populations. *Biometrics* 18:229-244
- Seely J (1969) Estimation in finite dimension vector space with application to the mixed model. PhD thesis, Iowa State University, Ames, USA
- Srinivasan MR (1986) Quadratic least squares estimators for variance components (design and genetic) based on mating design. PhD thesis, University of Madras, Madras, India
- Yuan CH (1977) A least squares approach to quadratic unbiased estimation of variance components in general mixed model. PhD thesis, North Carolina State University, Raleigh, USA