Theoretical and Applied Genetics 44, 378–380 (1974)

A General Approximate Method for Fitting Additive and Specific Combining Abilities to the Diallel Cross with Unequal Numbers of Observations in the Cells

F. ENGLAND

Scottish Plant Breeding Station, Pentlandfield, Roslin, Midlothian (Scotland)

Summary. A method is given and illustrated by examples, for the analysis of incomplete diallel tables including the calculation of the coefficients in the expectations of mean squares. The method is applicable to any of the four basic diallel mating designs.

For many plant and animal breeding purposes, approximate estimates of additive combining ability (*GCA*) and specific combining ability (*SCA*) from a diallel cross are sufficient. Many sets of data are, for some reason incomplete, and a general method for recovering estimates of *GCA* and *SCA* variance seems desirable. Hartley (1967) and Rao (1968) have given a method for "synthesising" the numerical values for the coefficients in the expectations of mean squares for any random model, balanced or unbalanced. Rao (*loc. cit.*) points out however that it is convenient to document the appropriate formulas for specific unbalanced designs which are frequently encountered.

The method to be described is completely general in that it provides an analysis of any of the four possible basic diallel cross designs as given by Griffing (1956) and in the balanced case leads to exact estimates. The method is essentially that given by Gilbert (1967) with the addition of the calculation of expectations of mean squares for the random model. Jones (1965) has used a similar analysis for the case of extra replication of the selfs in a diallel which included them.

The basic statistical model is

$$y_{ijk} = \mu + g_i + g_j + s_{ij} + r_{ij} + e_{ijk}$$

where

y_{ijk} = the mean yield of the k^{th} individual of the cross between the i^{th} and j^{th} parents
μ = the grand mean of all observations
g_i = the additive effect of the i^{th} parent
s_{ij} = the specific effect due to the i^{th} with the j^{th} parent
r_{ij} = the reciprocal or maternal effect
e_{ijk} = the usual error term.

For our purposes it is possible to simplify the model. The r_{ij} may or may not be estimable, depending on the design; if they are they will be estimated by a separate analysis of the reciprocal differences. The e_{ijk} will usually be separately estimated from a random block or similar analysis and can for the meantime be omitted from consideration (although not of course strictly from the basic model). Let y_{ij} be the mean yield of all progeny of the cross between the i^{th} and j^{th} parents and let $y'_{ij} = y_{ij} - \mu$ giving the new model,

$$y'_{ij} = g_i + g_j + s_{ij} + (\text{error})$$

let

n_{ij} = the number of plants in the cross between the i^{th} male and j^{th} female
$n_{..} = \sum_i \sum_j n_{ij}$ = total number of observations
$m_{ij} (i \neq j) = n_{ij} + n_{ji}$ = total number of plants in the cross between the i^{th} and j^{th} parents ($m_{ij} = m_{ji}$)
$m_{ii} = \sum_{j \neq i} m_{ij} + 4n_{ii}$
$r_{i.} = \sum_{j \neq i} m_{ij} + 2n_{ii}$ (equals m_{ii} if selfs are not included)
$c_{ij} (j \neq i) = m_{ij}^2$ ($c_{ij} = c_{ji}$)
$c_{ii} = \sum_{j \neq i} c_{ij} + 4n_{ii}^2$
$w_{i.} = \sum_{j \neq i} c_{ij} + 2n_{ii}^2$ (equals c_{ii} if selfs are not included)
$\boldsymbol{M} = [m_{ij}]$
$\boldsymbol{C} = [c_{ij}]$
$x_{ij} (j \neq i) = n_{ij} y'_{ij} + n_{ji} y'_{ji}$ = Sum of all deviations from the grand mean of all progeny of the cross of the i^{th} and j^{th} parents
$x_{ii} = 2n_{ii} y'_{ii}$
$t_{i.} = \sum_j x_{ij}$
$\boldsymbol{X} = [x_{ij}]$
$\boldsymbol{t} = [t_{i.}]$ (i.e. the column vector of the $t_{i.}$)
k_1 = the coefficient of σ_s^2 in the *GCA* mean square
k_2 = the coefficient of σ_g^2 in the *GCA* mean square
k_3 = the coefficient of σ_s^2 in the *SCA* mean square.

For the method to work $\boldsymbol{M}$ must have an inverse; in effect this means that $\boldsymbol{M}$ must have no blank rows, i.e. that each parent, either male or female, must have at least one offspring.

We may calculate the following raw sums of squares and their expectations, ignoring error terms.

Between crosses: $E\left(\sum_{i}\sum_{j>i} x_{ij}^2/m_{ij} + n_{ii}y_{ii}'^2\right) = n_{..}\sigma_s^2 + Tr(\boldsymbol{M})\,\sigma_g^2$

where $Tr(\boldsymbol{M}) = \text{Trace}\,(\boldsymbol{M}) = \sum m_{ii}$

General combining ability $E(\boldsymbol{t}'\boldsymbol{M}^{-1}\boldsymbol{t}) = Tr(\boldsymbol{M}^{-1}\boldsymbol{C})\sigma_s^2 + Tr(\boldsymbol{M})\,\sigma_g^2$

Correction factor: $E\left(\left(\sum_i\sum_j x_{ij}\right)^2/4n_{..}\right) = \left(\sum_i w_{i.}/2n_{..}\right)\sigma_s^2 + \left(\sum_i r_{i.}^2/n_{..}\right)\sigma_g^2$.

The corrected sums of squares and their expectations may be obtained in the usual way

$E\ (GCA \text{ corrected}) = E\ (GCA \text{ raw}) - E\ (CF)$
$E\ (SCA \text{ corrected}) = E\ (\text{Between crosses raw}) - E\ (GCA \text{ raw})$.

It would obviously be unwise to apply the analysis to very unbalanced data; sensible estimates of the components of variance can only be expected when each parent is used roughly the same number of times and has a roughly equal chance of having been crossed with each of the others.

Example 1

The analysis given can most easily be illustrated by applying it to a balanced design using published data. There is little point in illustrating the calculation of the numerical values of the between cross and correction factor sums of squares; only their expectations will be given. The *GCA* sum of squares will however be calculated since the estimates of individual gca's are conveniently derived during its calculation.

The data are taken from a half diallel including selfs published by Whitehouse *et al.* (1958) and are quoted and analysed by Jones (*loc. cit.*) in his analysis of the half-diallel table. The wheat varieties used were sown in four replications and each plot contained nine plants. Since only four lines were used the random model does not strictly apply but the data are convenient for demonstration.

$$
\begin{aligned}
n_{..} &= 360\\
\text{all } m_{ij}\ (j \neq i) &= 36\\
\text{all } m_{ii} &= 252\\
\text{all } r_{i.} &= 180\\
\text{all } c_{ij}\ (j \neq i) &= 1296\\
\text{all } c_{ii} &= 9072\\
\text{all } w_{i.} &= 6480\\
\boldsymbol{M} &= [252, 36]^*\\
\boldsymbol{M}^{-1} &= \frac{1}{2160}\,[9, -1]^*\\
\boldsymbol{C} &= [9072, 1296]^*
\end{aligned}
$$

* where the first term is the value of all the diagonal terms and the second of all the off-diagonal terms

$$Tr(\boldsymbol{M}^{-1}\boldsymbol{C}) = 144$$
$$\boldsymbol{t}_i' = 36\,[6.95, 2.65, -10.35, 0.75]$$

E (between crosses raw) $= n_{..}\sigma_s^2 + Tr(\boldsymbol{M})\,\sigma_g^2 = 360\,\sigma_s^2 + 1008\,\sigma_g^2$.

The individuals gca's may now be calculated as the elements of the column vector $\boldsymbol{M}^{-1}\boldsymbol{t}$

$$(\boldsymbol{M}^{-1}\boldsymbol{t})' = [1.1583, 0.4417, -1.7250, 0.1250]$$

and the raw gca sum of squares as $(\boldsymbol{M}^{-1}\boldsymbol{t})'\,\boldsymbol{t} = 977.98$ which has an expectation of $Tr(\boldsymbol{M}^{-1}\boldsymbol{C})\,\sigma_s^2 + Tr(\boldsymbol{M})\,\sigma_g^2 = 144\,\sigma_s^2 + 1008\,\sigma_g^2$.

The expectation of the correction term is

$$\left(\sum_i w_{i.}/2n_{..}\right)\sigma_s^2 + \left(\sum r_i^2/n_{..}\right)\sigma_g^2.$$

The "corrected" sums of squares therefore are

$$GCA = 108\,\sigma_s^2 + 648\,\sigma_g^2$$
$$SCA = 216\,\sigma_s^2.$$

So that dividing by the appropriate degrees of freedom we get

$$k_1 = \frac{108}{3} = 36,\ k_2 = \frac{648}{3} = 216 \text{ and } k_3 = \frac{216}{6} = 36.$$

Note that the coefficients in the expectations of mean squares are obtained from those in the corresponding sums of squares by dividing by the degrees of freedom.

Example 2

Kempthorne and Curnow (1961) have described incomplete diallel designs having some degree of symmetry and give the expectations of mean squares for them. Fyfe and Gilbert (1963) give worked examples of two incomplete diallels. All the designs given in these two papers have the same number of crosses per parent. The calculation of the coefficients of the expectations of sums of squares and mean squares will be illustrated for Design (1) of Fyfe and Gilbert. The analysis of variance is given in the original paper and there is no point in reproducing it here. The expectations of mean squares are not given in the paper although the least squares matrix $\boldsymbol{M}$ and its inverse are reproduced and for convenience are given below.

Source	*df*	sum of squares	mean squares	EMS
Total = Between crosses	$(p^2 + p - 2)/2 = 9$	4615.24		
GCA	$p - 1 = 3$	977.98	325.99	$36\sigma_s^2 + 216\sigma_g^2$
SCA	$p(p-1)/2 = 6$	3637.26	606.21	$36\sigma_s^2$

$$M = \begin{bmatrix} 3&0&0&0&0&0&0&1&1&1 \\ 0&3&0&0&0&1&1&0&0&1 \\ 0&0&3&0&1&0&1&0&1&0 \\ 0&0&0&3&1&1&0&1&0&0 \\ 0&0&1&1&3&0&0&0&0&1 \\ 0&1&0&1&0&3&0&0&1&0 \\ 0&1&1&0&0&0&3&1&0&0 \\ 1&0&0&1&0&0&1&3&0&0 \\ 1&0&1&0&0&1&0&0&3&0 \\ 1&1&0&0&1&0&0&0&0&3 \end{bmatrix} \qquad M^{-1} = 1/24 \begin{bmatrix} 13&1&1&1&1&1&1&-5&-5&-5 \\ 1&13&1&1&1&-5&-5&1&1&-5 \\ 1&1&13&1&-5&1&-5&1&-5&1 \\ 1&1&1&13&-5&-5&1&-5&1&1 \\ 1&1&-5&-5&13&1&1&1&1&-5 \\ 1&-5&1&-5&1&13&1&1&-5&1 \\ 1&-5&-5&1&1&1&13&-5&1&1 \\ -5&1&1&-5&1&1&-5&13&1&1 \\ -5&1&-5&1&1&-5&1&1&13&1 \\ -5&-5&1&1&-5&1&1&1&1&13 \end{bmatrix}$$

Matrix $C = M$ in this example since there are no selfs included (all $n_{ii} = 0$) and all the $m_{ij}(i = j) = 0$ or 1.

$n_{..} = 15$, all $m_{ii} = 3$, all $r_i = 3$, all $c_{ii} = 3$ all $w_i = 3$

All the diagonal elements of $M^{-1}C = ((3 \times 13) - 3\,(1 \times -5))/24 = 1$ so that $Tr(M^{-1}C) = 10$. The expectations of sums of squares are:

E (between crosses raw) $= n_{..}\,\sigma_s^2 + Tr(M)\,\sigma_g^2$
$= 15\,\sigma_s^2 + 30\,\sigma_g^2$

E (gca raw) $= Tr(M^{-1}C)\,\sigma_s^2 + Tr(M)\,\sigma_g^2 = 10\,\sigma_s^2 + 30\,\sigma_g^2$

E (correction term) $= \left(\sum_i w_{i.}/2n_{..}\right)\sigma_s^2 + \left(\sum_i r_{i.}^2/n_{..}\right)\sigma_g^2 + 6\sigma_g^2$.

The expectations of corrected sums of squares and mean squares are therefore

	df	Sums of Squares	Mean Squares
GCA	9	$9\sigma_s^2 + 24\sigma_g^2$	$\sigma_s^2 + 2.67\,\sigma_g^2$
SCA	5	$5\sigma_s^2$	σ_s^2

The coefficients in the mean squares are the same as those obtained from Kempthorne and Curnow (*loc. cit.*). The coefficient of σ_s^2 in both the GCA and SCA mean square is equal to the number of replications, which in our example is one, and that of σ_g^2 is equal to (reps) $\times$ (crosses per parent) $\times (p-2)/(p-1) = 3 \times 8/9 = 2.67$.

The example is, of course, trivial and the coefficients are more easily obtained from Kempthorne and Curnow but this is only so because each parent appears in the same number of crosses.

Acknowledgements

It is a pleasure to thank Mr. G. K. Shukla for working out the calculation of the expectations for the diallel table without selfs, Mr. J. L. Fyfe for much helpful advice and Mr. R. Morley Jones and Mr. Robin Thompson for comments on the typescript.

Literature

Fyfe, J. L., Gilbert, N.: Partial diallel crosses. Biometrics **19**, 278–286 (1963).

Gilbert, N.: Additive combining abilities fitted to plant breeding data. Biometrics **23**, 45–49 (1967).

Griffing, B.: Concept of general and specific combining ability in relation to diallel crossing systems. Austr. J. biol. Sci. **9**, 463–493 (1956).

Hartley, H. O.: Expectations, variances and covariances of Anova mean squares by 'synthesis'. Biometrics **23**, 105–114 (1967).

Jones, R. M.: Analysis of variance of the half diallel table. Heredity **20**, 117–121 (1965).

Kempthorne, O., Curnow, R. N.: The partial diallel cross. Biometrics **17**, 229–250 (1961).

Rao, J. K. N.: On expectations, variances and covariances of Anova mean squares by 'synthesis'. Biometrics **24**, 963–978 (1968).

Whitehouse, R. N. H., Thompson, J. B., Do Valle Ribeiro, M. A. M.: Studies on the breeding of self-pollinating cereals. 2. The use of a diallel cross analysis in yield prediction. Euphytica **7**, 147–169 (1958).

Received September 3, 1973
Communicated by A. Robertson

Dr. F. England
Scottish Plant Breeding Station Pentlandfield
Roslin, Midlothian EH25 9RF (Scotland)