Theor. Appl. Genet. 62, 353–354 (1982)

© Springer-Verlag 1982

Evaluation of the Normal Equations for the Analysis of Diallels

P. A. Baghurst

Waite Agricultural Research Institute, University of Adelaide, South Australia (Australia)

Summary. Analysis of diallel experiments usually requires specialised computer programs. A simple relationship between the normal equations for a diallel and the normal equations for a hypothetical model in which the general combining ability factor is viewed as two separate distinct simple factors may be useful in adapting general statistical analysis packages to the desired analysis.

Key words: Estimation – Diallels

Introduction

A feature distinguishing diallels from other linear models (Yates 1947; Kempthorne 1956) is the inclusion of a factor (often called the general combining ability or gca) which makes two contributions (usually one from each parent) to every experimental observation y, e.g.

$$E(Y_{ij}) = \mu + gca_i + gca_j\,, \quad i = 1, \ldots, p\,. \tag{1}$$

Wilkinson (1970) has pointed out that the model (1) is a constrained form of a model containing only "simple" factors, viz

$$E(Y_{ij}) = \mu + m_i + f_j\,, \tag{2a}$$

$$m_i = f_i\,. \tag{2b}$$

Analysis of such models poses no new theoretical problem but it would appear that the additional features and conveniences of most commonly available statistical packages are denied the analyst because the package was not specifically designed to handle models containing "nonsimple" factors.

One cumbersome way of circumventing this problem is to define a dummy variate (with values 0, 1 or 2) for each level of the gca factor, but for experiments with many parental lines and considerable replication the data space required may become prohibitively large (unless there is a facility for progressive accumulation of the coefficients of the normal equations). It is the purpose of this communication to present an alternate method in which the normal equations for model (1) are obtained by simple row and column operations performed on the normal equations for the unconstrained model (2a). Any reasonably flexible computing package which permits user intervention after accrual of sums and squares and products may then be used for the analysis.

The Normal Equations

If the incidence matrices for {m} and {f} are denoted by $\mathbf{X}_m$ and $\mathbf{X}_f$, the normal equations for model (2a) (ignoring the constraint, 2b) may be written in partitioned matrix form as

$$\begin{pmatrix} n & \mathbf{1}^T\mathbf{X}_m & \mathbf{1}^T\mathbf{X}_f \\ \mathbf{X}_m^T\mathbf{1} & \mathbf{X}_m^T\mathbf{X}_m & \mathbf{X}_m^T\mathbf{X}_f \\ \mathbf{X}_f^T\mathbf{1} & \mathbf{X}_f^T\mathbf{X}_m & \mathbf{X}_f^T\mathbf{X}_f \end{pmatrix} = \begin{pmatrix} \mathbf{1}^T\mathbf{Y} \\ \mathbf{X}_m^T\mathbf{Y} \\ \mathbf{X}_f^T\mathbf{Y} \end{pmatrix} \tag{3}$$

where n is the number observations and $\mathbf{1}$ is an $n \times 1$ vector of units.

Introduction of the constraint $m_i = f_i$ is accomplished simply by noting that the gca factor has incidence matrix $X_{gca} = X_m + X_f$, and hence the normal equations for the diallel model are

$$\begin{pmatrix} n & \mathbf{1}^T\mathbf{X}_m + \mathbf{1}^T\mathbf{X}_f \\ \mathbf{X}_m^T\mathbf{1} + \mathbf{X}_f^T\mathbf{1} & \mathbf{X}_m^T\mathbf{X}_m + \mathbf{X}_f^T\mathbf{X}_m + \mathbf{X}_m^T\mathbf{X}_f + \mathbf{X}_f^T\mathbf{X}_f \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{1}^T\mathbf{Y} \\ \mathbf{X}_m^T\mathbf{Y} \quad \mathbf{X}_f^T\mathbf{Y} \end{pmatrix}. \tag{4}$$

0040-5752/82/0062/0353/$ 1.00

It is readily apparent that (4) is obtainable from (3) by a "condensation procedure" in which those p rows corresponding to $\partial S/\partial f_k = 0$, $k = 1, \ldots, p$ where S is the sum of squares of residuals, are added to those p rows corresponding to $\partial S/\partial m_k = 0$; and similarly for the columns.

Extensions

In a practical context most models contain other covariates, factors and interaction terms. It can be quickly verified that inclusion of these extra terms does not invalidate the condensation procedure described above.

The sums of squares and products (ssp) for an interaction term involving the gca factor are obtained first by considering the corresponding interactions involving simple factors m and f and performing, in a manner exactly analogous to that outlined above, a further condensation operation on the appropriate rows and columns of the expanded ssp matrix.

Analysis of the rather rare triallel can be accomplished by defining three simple factors and adding rows and columns for all three factors.

Practical Comments

Any package which permits access to the ssp matrix and subsequent resubmittal for regression analysis could be used, in principle, for diallel analysis. The result described here has been used in conjunction with the GENSTAT programming language (Alvey et al. 1977) which is intended primarily for analysis of models containing only simple factors. The GENSTAT ssp matrix has an additional row (column) containing the column means of each variate or dummy variate, and in the last element the number of units used in its construction, but this was not a problem since individual elements were found to maintain their correct meaning when "condensed". Other packages which may allow the required modification of the ssp matrix are P-STAT and SAS, although the latter would require the complete analysis to be specified in terms of matrix operations.

It might be noted that the sweep algorithm for Anova (Payne and Wilkinson 1977) could also be adapted to the analysis of diallels (Wilkinson 1970) provided there was a high degree of balance, but there is no generally available implementation of the adapted form. The technique described here is well suited to the analysis of the various partial diallels (Griffing 1956; Curnow 1963; Bray 1971; Gordon 1980) which have been proposed for situations in which it is impractical to consider all p^2 possible crosses. It may also be used in the analysis of data sets which contain unequal replication within crosses, such sets commonly arising in practice as a result of intentionally increased replication of more important crosses, by misadventure or because certain potential crosses are infertile.

Literature

Alvey, N.G.; Banfield, C.F.; Baxter, R.I.; Gower, J.C.; Krzanowski, W.J.; Lane, P.W.; Leech, P.K.; Nelder, J.A.; Payne, R.W.; Phelps, K.M.; Rogers, C.E.; Ross, G.J.S.; Simpson, H.R.; Todd, A.D.; Wedderburn, R.W.M.; Wilkinson, G.N. (1977): GENSTAT, A General Statistics Program. Rothamsted Experimental Station, Harpenden, Hertfordshire

Bray, R.A. (1971): Quantitative evaluation of the circulant partial diallel cross. Heredity **27,** 189–202

Curnow, R.N. (1963): Sampling the diallel cross. Biometrics **19,** 287–306

Gordon, G.H. (1980): A method of parental selection and hybrid prediction using incomplete diallels. Theor. Appl. Genet. **56,** 225–232

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

Kempthorne, O. (1956): The theory of the diallel cross. Genetics **41,** 451–459

Payne, R.W.; Wilkinson, G.N. (1977): A general algorithm for analysis of variance. Appl. Statistics. **26,** 251–260

Wilkinson, G.N. (1970): A general recursive procedure for analysis of variance. Biometrika **57,** 19–46

Yates, F. (1947): Analysis of data from all possible reciprocal crosses between a set of parental lines. Heredity **1,** 287–301

Received February 19, 1982
Communicated by R. W. Allard

Dr. P. A. Baghurst
CSIRO Division of Human Nutrition
Kintore Avenue
Adelaide, South Australia 5000 (Australia)