ANALYSIS OF VARIANCE OF THE HALF DIALLEL TABLE

R. MORLEY JONES
A.R.C. Unit of Biometrical Genetics, Department of Genetics, University of Birmingham, England

Received 24.x.64

1. INTRODUCTION

An analysis of variance for the complete diallel table was given by Hayman (1954*a*), developing in one direction that of Yates (1947). Frequently, however, reciprocal differences are assumed absent, and only one of each pair of reciprocal crosses is raised. The analysis appropriate to this case is given in the present note.

Using the same model as Hayman, the determination of the sums of squares corresponding to additive effects (a), and on the assumption of no epistasis to mean dominance (b_1), to additional dominance effects that can be accounted for by genes having one allele present in only one line (b_2) (the remaining $n-1$ lines being assumed to carry the same alternative allele), and to residual dominance effects (b_3), is in essence a straightforward application of fitting constants by least squares. The derivations are therefore given here in a rather concise form.

We shall use small heavy type letters to denote column vectors, but we shall write the elements of the observation vector $\mathbf{y}$ and similar vectors in the natural triangular form to make clear their structure. Thus

$$\mathbf{y} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1n} \\ & y_{22} & \cdots & y_{2n} \\ & & \cdots & \cdots \\ & & & y_{nn} \end{bmatrix}$$

The sum of squares for fitting the mean, or mean correction, which will not ordinarily need to be calculated, is $(\mathbf{m}'\mathbf{y})^2/\mathbf{m}'\mathbf{m}$, where

$$\mathbf{m} = \begin{bmatrix} 1, & 1, & \ldots, & 1 \\ & 1, & \ldots, & 1 \\ & & \ldots, & \ldots \\ & & & 1 \end{bmatrix}$$

i.e. $2y_{..}^2/[n(n+1)]$.

The additive effects are obtained by regression on the differences of the $\mathbf{j}_r$, where

$$\mathbf{j}_1 = \begin{bmatrix} 2, & 1, & 1, & \ldots, & 1 \\ & 0, & 0, & \ldots, & 0 \\ & & & \ldots, & \ldots \\ & & & & 0 \end{bmatrix}, \qquad \mathbf{j}_2 = \begin{bmatrix} 0, & 1, & 0, & \ldots, & 0 \\ & 2, & 1, & \ldots, & 1 \\ & & & \ldots, & \ldots \\ & & & & 0 \end{bmatrix},$$

Regression on the $\mathbf{j}_r$'s themselves would give a sum of squares including the mean correction, since $\Sigma\mathbf{j}_r = 2\mathbf{m}$.

If $u_r = \mathbf{j}_r'\mathbf{y} = y_{r\cdot} + y_{rr}$, since $\mathbf{j}_r'\mathbf{j}_r = n+3$ and $\mathbf{j}_r'\mathbf{j}_s = 1$ when $r \neq s$, the sum of squares for regression on the $\mathbf{j}$'s is

$$\mathbf{u}'[(n+2)\mathbf{I}+\mathbf{E}]^{-1}\mathbf{u}$$

where $\mathbf{I}$ is the unit matrix and $\mathbf{E}$ is the square matrix all of whose elements are unity. If we also write $\mathbf{\Delta} = \mathbf{I} - \frac{1}{n}\mathbf{E}$, then $\mathbf{\Delta}^2 = \mathbf{\Delta}$, $\left[\frac{1}{n}\mathbf{E}\right]^2 = \frac{1}{n}\mathbf{E}$ and $\mathbf{\Delta E} = \mathbf{E\Delta} = \mathbf{O}$, so that for any non-zero α and β,

$$(\alpha\mathbf{\Delta}+\beta\mathbf{E}/n)^{-1} = \alpha^{-1}\mathbf{\Delta}+\beta^{-1}\mathbf{E}/n.$$

Hence the sum of squares becomes

$$\mathbf{u}'[(n+2)\mathbf{\Delta}+2(n+1)\mathbf{E}/n]^{-1}\mathbf{u} = \mathbf{u}'\left[\frac{1}{n+2}\mathbf{\Delta}+\frac{1}{2n(n+1)}\mathbf{E}\right]\mathbf{u}$$
$$= \frac{1}{n+2}\,dev^2u_r + \frac{1}{2n(n+1)}(\Sigma u_r)^2.$$

The second term is the mean correction, so that the sum of squares for additive effects is

$$\frac{1}{n+2}\,dev^2u_r \qquad (a)$$

with $n-1$ degrees of freedom.

Similarly the contribution of mean dominance is found from the regression on

$$\mathbf{l} = \begin{bmatrix} -n+1, & 2, & 2, & \ldots, & 2 \\ & -n+1, & 2, & \ldots, & 2 \\ & & & \ldots, & \ldots \\ & & & & -n+1 \end{bmatrix}$$

giving for the contribution to the sum of squares for this single degree of freedom

$$\frac{[2y_{\cdot\cdot}-(n+1)y_{\cdot}]^2}{n(n^2-1)} \qquad (b_1).$$

The sum of squares for (b_2) can be obtained by regression on

$$\mathbf{l}_1 = \begin{bmatrix} -n+1, & 2, & 2, & \ldots, & 2 \\ & -1, & 0, & \ldots, & 0 \\ & & -1, & \ldots, & 0 \\ & & & \ldots, & \ldots \\ & & & & -1 \end{bmatrix}, \quad \mathbf{l}_2 = \begin{bmatrix} -1, & 2, & 0, & \ldots, & 0 \\ & -n+1, & 2, & \ldots, & 2 \\ & & -1, & \ldots, & 0 \\ & & & \ldots, & \ldots \\ & & & & -1 \end{bmatrix}, \ldots$$

which are orthogonal to the $\mathbf{j}_r$, but as before, since $\Sigma\mathbf{l}_r = 2\mathbf{l}$, this includes (b_1).

If

$$z_r = \mathbf{1}_r'\mathbf{y} = 2y_{r.} - ny_{rr} - y.$$

so that

$$\Sigma z_r = 2[2y_{..} - (n+1)y_.],$$

the sum of squares is

$$z'[(n^2-4)\mathbf{I}+3n\mathbf{E}]^{-1}\mathbf{z} = \mathbf{z}'[(n^2-4)\mathbf{\Delta}+4(n^2-1)\mathbf{E}/n]^{-1}\mathbf{z}$$
$$= \frac{1}{n^2-4}\,dev^2 z_r + \frac{1}{4n(n^2-1)}\cdot 4[2y_{..} - (n+1)y_.]^2$$

in which the second term is (b_1), the first term representing the (b_2) sum of squares. The latter is more conveniently taken in the form

$$\frac{1}{n^2-4}\,dev^2 t_r \qquad (b_2)$$

with $n-1$ degrees of freedom, where

$$t_r = 2y_{r.} - ny_{rr} = 2u_r - (n+2)y_{rr}.$$

The contribution of (b_3) is obtained by subtraction from the total.

2. EXAMPLE

To illustrate the application of these results, we shall consider the F_1 diallel of Whitehouse *et al.* (1958) for yield of spring wheat. On most desk calculators, it is easy to compute the sum of squares of deviations from their mean (dev^2) of a set of quantities in a single operation. The analysis is given in table 1, which represents all that need be written down.

It appears from the analysis that the F_1 yield data, taken alone, can be adequately accounted for by a general advantage of crosses over selfs.

It is perhaps worth observing here that, if we assume no epistasis and only two alleles per locus, the components (b_2) and (b_3) in a 4×4 diallel with homozygous parents have a rather simple interpretation. For in this case, all genes which contribute to the variation belong to one of two classes. Those of the first class exist in the form of one allele in one of the parents, and the other allele in the other three parents. Those of the second class have one allele in two of the parents and the other allele in the other two parents. Genes of either class, if balanced so that their dominance contributions to all crosses are the same, affect only (b_1). In so far as they are not so balanced, genes of the first class contribute to (b_2) but not (b_3), while genes of the second class contribute to (b_3) but not (b_2).

If all genes with dominance effects are of the second class, then Jinks' independent gene distribution condition is satisfied. Hence, if (b_2) is not significant, any apparent departure of the W/V graph

(Jinks, 1954; Hayman, 1954*b*) from a straight line of unit slope is certainly not significant. The converse is not strictly true, but (b_2) goes a long way towards providing an exact test of " upsets " in the W/V graph for the 4×4 diallel.

TABLE 1

Analysis of a 4×4 half-diallel table

y_{rs}				u_r	t_r
33·9	42·0	35·6	38·7	184·1	164·8
	31·0	36·7	39·1	179·8	173·6
		30·0	34·5	166·8	153·6
			32·8	177·9	159·0
				708·6	651·0

$y_{..} = 354·3$
$y_{.} = 127·7$

	D.F.	S.S.	M.S.
a . .	3	27·17	9·06
b_1 . .	1	81·90	81·90***
b_2 . .	3	18·31	6·10
b_3 . .	2	0·82	0·41
Total . .	9	128·20	
Error . .	18		3·73 (derived from Whitehouse *et al.*)

3. DIALLEL WITH REPLICATED SELFS

Diallel crosses have sometimes been carried out with each self included more than once. Suppose a complete diallel table is available with reciprocal crosses included and the diagonal terms replicated k times. Then, by a similar technique to that illustrated above, the sums of squares for the (a), (b_1) and (b_2) items are found to be:

$$(a)\quad \frac{dev^2[y_{r.}+y_{.r}+2(k-1)y_{rr}]}{2(n+2k-2)}$$

$$(b_1)\quad \frac{k(y_{..}-ny_{.})^2}{n(n-1)(n+k-1)}$$

$$(b_2)\quad \frac{k.dev^2(y_{r.}+y_{.r}-ny_{rr})}{(n-2)(n+2k-2)}$$

where y_{rr} is the mean of the k observations on the rth self, and then $y_{r.}$, $y_{.r}$ are the row and column sums, respectively, $y_{..} = \Sigma y_{r.}$, and $y_{.} = \Sigma y_{rr}$ as before.

As special cases, when $k = 1$ we recover Hayman's analysis, and when $k = 2$ we obtain the analysis for the half-diallel, except for a factor of 2 which appears because the half-diallel now consists of sums of two observations on reciprocal crosses, or two selfs, as the case may be.

The analysis of reciprocal differences is exactly as given by Hayman, being unaffected by the replication of the diagonals.

By giving an appropriate value to k, not necessarily an integer, the analysis may be deduced for the case in which the error variance differs between the diagonal and off-diagonal terms.

This analysis is included as part of a Fortran programme for the analysis of diallel crosses, written by Mr B. E. Cooper of the Atlas Computer Laboratory, N.I.R.N.S., Chilton, Berks. The programme is intended for use on the I.B.M. Stretch, the I.B.M. 7090 Computer, and the I.C.T. Atlas.

4. REFERENCES

HAYMAN, B. I. 1954*a*. The analysis of variance of diallel tables. *Biometrics*, *10*, 235-244.

HAYMAN, B. I. 1954*b*. The theory and analysis of diallel crosses. *Genetics*, *39*, 789-809.

JINKS, J. L. 1954. The analysis of continuous variation in a diallel cross of *Nicotiana rustica* varieties. *Genetics*, *39*, 767-788.

WHITEHOUSE, R. N. H., THOMPSON, J. B., AND DO VALLE RIBIERO, M. A. M. 1958. Studies on the breeding of self-pollinating cereals. 2. The use of a diallel cross analysis in yield prediction. *Euphytica*, *7*, 147-169.

YATES, F. 1947. Analysis of data from all possible reciprocal crosses between a set of parental lines. *Heredity*, *1*, 287-301.