Heredity (1977), **38** (3), 401-407

NOTES AND COMMENTS

A NOTE ON THE HAYMAN ANALYSIS OF VARIANCE FOR A FULL DIALLEL TABLE

D. E. WALTERS
A.R.C. Statistics Group, Department of Applied Biology, Cambridge CB2 3DX
and
J. S. GALE
Department of Genetics, University of Birmingham, Birmingham B15 2TT

Received 7.ix.76

SUMMARY

Hayman's (1954) analysis of variance is derived, both for the reciprocal effects and the maternal effects models, without using progressive fitting of parameters. It is shown that, apart from a minor modification, Hayman's analysis is appropriate for detecting the various effects for which the analysis was devised.

1. INTRODUCTION

HAYMAN (1954) set out his analysis of variance of the full $n \times n$ diallel table very briefly, with the main emphasis on applications. The analysis has been criticised on two grounds. Firstly, the analysis appears to have been calculated using a progressive fitting of the unknown parameters; in this way, a truly non-orthogonal analysis of variance is made orthogonal, with the component sums of squares summing up to the total sum of squares (Kempthorne 1956). Secondly, Hayman's representation of the maternal effects in the model on which his analysis is based has been questioned (Wearden, 1964; Topham, 1966). It has been pointed out that Hayman's model is a " reciprocal effects " model rather than a " maternal effects " model in the strict sense.

In this paper, we shall consider both these models and derive sums of squares directly from the least squares estimates of the various parameters; progressive fitting will not be used. This approach should serve to clarify the points at issue.

2. THE RECIPROCAL EFFECTS MODEL

For simplicity, we suppose in the first instance that a single observation has been made for each cell of the diallel table. As usual, we write y_{rs} for the datum on an individual resulting from a cross between the rth line as mother and sth line as father. The rth leading diagonal observation has expectation

$$m + 2j_r$$

and the $(r \times s)$ off-diagonal observation has expectation

$$m + j_r + j_s + j_{rs} + k_r - k_s + k_{rs}.$$

In this model

m = the mean of the inbred lines
j_r = the genic effect of the rth line
$2j_{rs}$ = the dominance deviation for the $(r \times s)$ reciprocal sum (given that an additive-dominance model is adequate)
$2k_r$ = the difference between the effects of the rth parental line used as female parent and as male parent
$2k_{rs}$ = remaining discrepancy in the $(r \times s)$ reciprocal difference (*i.e.* the k_{rs} represent interactions between parental effects and own-genotype of the progeny effects).

These parameters are subject to the restraints

$$\sum_r j_r = 0, \sum_r k_r = 0, \sum_r k_{rs} = 0, \sum_s k_{rs} = 0, k_{sr} = -k_{rs}.$$

On the other hand, no restraints are imposed on the j_{rs} which are defined to be dominance deviations and *not* specific combining abilities.

The j_{rs} are further subdivided as

$$j_{rs} = l + l_r + l_s + l_{rs}$$

where l = the mean dominance deviation
l_r = the further dominance deviation due to the rth parent
l_{rs} = residual dominance deviation in the $(r \times s)$ reciprocal sum

with

$$\sum_r l_r = 0, \sum_r l_{rs} = 0, \sum_s l_{rs} = 0$$

and one further restraint on the l_{rs}.

When introducing this subdivision of the j_{rs}, Hayman rewrites the expectation of the rth diagonal as

$$m + 2j_r - (n-1)l - (n-2)l_r$$

instead of

$$m + 2j_r.$$

This revision of the model is a convenient computational device, easing the derivation of the sums of squares corresponding to the l_r and l_{rs} effects respectively; these sums of squares are not affected by the change in model. We shall not, however, use this device here.

By simultaneous estimation of all unknown parameters, taking due account of the restraints, the following least squares estimates are obtained:

$$\hat{m} = \frac{y.}{n}$$

$$\hat{j}_r = \frac{y_{rr}}{2} - \frac{y.}{2n}$$

$$\hat{l} = \frac{y.. - ny.}{n(n-1)}$$

$$\hat{l}_r = \frac{C_r + R_r - ny_{rr} + y.}{2(n-2)} - \frac{y..}{n(n-2)}$$

$$\hat{l}_{rs} = \tfrac{1}{2}(y_{rs}+y_{sr}-y_{rr}-y_{ss}-2\hat{l}_r-2\hat{l}_s-2\hat{l})$$

$$\hat{k}_r = \frac{R_r-C_r}{2n}$$

$$\hat{k}_{rs} = \tfrac{1}{2}(y_{rs}-y_{sr})-\hat{k}_r+\hat{k}_s$$

where

$$y_{\cdot} = \sum_{r=1}^{n} y_{rr},\ y_{\cdot\cdot} = \sum_{r,s} y_{rs},\ C_r = y_{\cdot r} = \sum_{s=1}^{n} y_{sr},\ R_r = y_{r\cdot} = \sum_{s=1}^{n} y_{rs}$$

(note that Hayman writes y_{rr} as y_r).

The variances and covariances of these estimates are as follows, σ^2 being the error variance, assumed to be homogeneous over cells.

$$\text{var}(\hat{j}_r) = \frac{n-1}{4n}\sigma^2, \qquad \text{cov}(\hat{j}_r, \hat{j}_s) = \frac{-1}{4n}\sigma^2$$

$$\text{var}(\hat{l}) = \frac{1}{n-1}\sigma^2$$

$$\text{var}(\hat{l}_r) = \frac{n-1}{4(n-2)}\sigma^2, \quad \text{cov}(\hat{l}_r, \hat{l}_s) = \frac{-1}{4(n-2)}\sigma^2$$

$$\text{var}(\hat{k}_r) = \frac{n-1}{2n^2}\sigma^2, \qquad \text{cov}(\hat{k}_r, \hat{k}_s) = \frac{-1}{2n^2}\sigma^2$$

$$\text{var}(\hat{l}_{rs}) = \frac{n-3}{2(n-1)}\sigma^2, \quad \text{cov}(\hat{l}_{rs}, \hat{l}_{rt}) = \frac{-(n-3)}{2(n-1)(n-2)}\sigma^2,$$

$$\text{cov}(\hat{l}_{rs}, \hat{l}_{tu}) = \frac{1}{(n-1)(n-2)}\sigma^2$$

$$\text{var}(\hat{k}_{rs}) = \frac{n-2}{2n}\sigma^2, \qquad \text{cov}(\hat{k}_{rs}, \hat{k}_{rt}) = \frac{-1}{2n}\sigma^2, \ \text{cov}(\hat{k}_{rs}, \hat{k}_{tu}) = 0.$$

The sums of squares may now be calculated as follows. Firstly,

$$\text{S.S.}(\hat{l}) = \hat{l}^2(n\text{-}1)$$

testing the significance of the average dominance effect, a significant result implying that dominance is directional. The other sums of squares are more difficult to calculate but may be found from general theory, as follows.

Suppose we have a set of n parameters $\theta_1, \theta_2, \ldots, \theta_n$ constrained to add to zero. We conveniently take the "first" $(n\text{-}1)$ parameters as free parameters, with least squares estimates $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_{n-1}$, these estimates being obtained subject to the restraint. On inverting the variance-covariance matrix, with the σ^2 omitted, of these estimates we obtain a matrix with typical element $\mathcal{J}_{rs}$ (say). The sum of squares testing the significance of the departure of the $\hat{\theta}$'s from zero is then

$$\sum_{r=1}^{n-1} J_{rr}\hat{\theta}_r^2 + \sum_{r=1}^{n-1}\sum_{\substack{s=1 \\ s\neq r}}^{n-1} J_{rs}\hat{\theta}_r\hat{\theta}_s.$$

The $(n-1)\times(n-1)$ matrix with $(n-1)$ on the leading diagonal and (-1)

elsewhere inverts to give a matrix with $2/n$ on the leading diagonal and $1/n$ elsewhere. Hence the sum of squares testing the significance of the consistent reciprocal effects is

$$\begin{aligned}\text{S.S. } (\hat{k}_r) &= 2n^2\left\{\frac{2}{n}\sum_{r=1}^{n-1}\hat{k}_r^2+\frac{1}{n}\sum_{r=1}^{n-1}\sum_{\substack{s=1\\ s\neq r}}^{n-1}\hat{k}_r\hat{k}_s\right\}\\ &= 2n\left\{2\sum_{r=1}^{n-1}\hat{k}_r^2+\sum_{r=1}^{n-1}\hat{k}_r(-\hat{k}_r-\hat{k}_n)\right\}\\ &= 2n\sum_{r=1}^{n}\hat{k}_r^2.\end{aligned}$$

The S.S.$(\hat{\jmath}_r)$ and S.S. $(\hat{l}_r)$ follow at once on replacing the $(2n^2)$ in the above expression by $(4n)$ and by $4(n-2)$ respectively.

The sum of squares for a parameter group is equal to the overall sum of squares when the parameter group is included in the full model less the overall sum of squares when the parameter group is excluded from the full model; this provides the easiest method for finding the S.S. $(\hat{k}_{rs})$ and S.S. $(\hat{l}_{rs})$. Inspection of the normal equations reveals that the estimates of reciprocal effects are independent of those for own-genotype effects, so that the latter may be ignored when obtaining S.S. $(\hat{k}_{rs})$. Moreover, estimates $\hat{k}_r$ are uncorrelated with estimates $\hat{k}_{rs}$, so that S.S. $(\hat{k}_r)$ is the same irrespective of whether the $\hat{k}_{rs}$ are included or excluded. Hence

$$\begin{aligned}\text{S.S. } (\hat{k}_{rs}) &= \text{S.S. (all reciprocal effects)} - \text{S.S. } (\hat{k}_r)\\ &= \sum_{r<s}\frac{(y_{rs}-y_{sr})^2}{2}-\text{S.S. } (\hat{k}_r)\\ &= 2\sum_{r<s}\hat{k}_{rs}^2.\end{aligned}$$

The S.S. $(\hat{l}_{rs})$ are obtained in a rather similar way, using the all own-effects sum of squares less the all own-effects sum of squares when the $\hat{l}_{rs}$ are excluded.

TABLE 1

Analysis of variance for reciprocal effects model

Parameter group	S.S.	S.S. in Hayman's notation	d.f.
j_r	$4\sum_r \hat{\jmath}_r^2$	—	$(n-1)$
l	$(n-1)\hat{l}^2$	b_1	1
l_r	$\frac{4(n-2)}{n}\sum_r \hat{l}_r^2$	b_2	$(n-1)$
l_{rs}	$2\sum_{r<s}\hat{l}_{rs}^2$	b_3	$\frac{1}{2}n(n-3)$
k_r	$2n\sum_r \hat{k}_r^2$	c	$n-1$
k_{rs}	$2\sum_{r<s}\hat{k}_{rs}^2$	d	$\frac{1}{2}(n-1)(n-2)$

Results are summarised in table 1. They are given in a form which indicates the particular component tested by every sum of squares. However, when account is taken of the difference in notation, it will be seen that, apart from S.S. $(\hat{j}_r)$, all the sums of squares are precisely those given by Hayman.

Moreover, since estimates $\hat{l}$ are uncorrelated with estimates $\hat{l}_r$, and each of these is uncorrelated with estimates $\hat{l}_{rs}$, S.S. $(\hat{l})$, S.S. $(\hat{l}_r)$ and S.S. $(\hat{l}_{rs})$ may be added to give the sum of squares measuring all dominance effects, identical with Hayman's (b) sum of squares.

Thus only S.S. $(\hat{j}_r)$ differs from the sums of squares given by Hayman; his (a) sum of squares being the sum of squares for genic effects only in the special case of no dominance. It has indeed long been recognised that on Hayman's model the correct sum of squares for detecting genic effects is

$$\sum_r y_{rr}^2 - (\sum_r y_{rr})^2/n$$

(identical with our $4\sum \hat{j}_r^2$), the " usual " sum of squares of the diagonal entries. This sum of squares is, however, sometimes calculated but not reported; it would be unusual to make up a diallel set of matings without having first shown that the inbred lines differ significantly, so that the test of significance on the diagonal entries is normally confirmatory only.

For simplicity we have assumed so far only a single observation per cell. With z observations per cell, we may write y_{rs} for the *total* in cell $r \times s$; our formulae for the estimates must now, of course, be divided by z and if these revised formulae are used all sums of squares in table 1 must be multiplied by z.

3. The maternal effects model

In the reciprocal effects model, no maternal effect appears in expectations for the leading diagonal. Hence it has been suggested that expectations be written

$$E(y_{rr}) = m + 2j_r + k_r$$

$$E(y_{rs}) = m + j_r + j_s + l + l_r + l_s + l_{rs} + k_r \quad (r \neq s).$$

This gain in biological realism is achieved at a cost, since in this model terms corresponding to the k_{rs} do not appear, that is we are assuming no interaction between maternal effects and own-genotype effects.

If we carry out the same procedure as for the reciprocal effects model we find that the estimates of $\hat{l}$, $\hat{l}_r$ and $\hat{l}_{rs}$ are precisely as before, so that the corresponding sums of squares are unaffected by the change of model. The $\hat{k}_r$ are affected only in a trivial way; with a single observation per cell, we now find

$$\hat{k}_r = \frac{R_r - C_r}{n}$$

and the corresponding sum of squares is now written

$$\frac{n}{2}\sum_r \hat{k}_r^2$$

but this is identical with the sum of squares in table 1, both being

$$\sum_r \frac{(R_r - C_r)^2}{2n}.$$

We can also calculate Hayman's (d) sum of squares, in the same form as before

$$\sum_{r<s} \frac{(y_{rs}+y_{sr})^2}{2} - \sum_r \frac{(R_r - C_r)^2}{2n}.$$

Should the model fit, the corresponding mean square estimates σ^2; thus the (d) mean square, when calculated for the case of z replicates per cell, provides a test of the adequacy of the maternal effects model.

The only sum of squares affected by the change of model is S.S. $(\hat{\jmath}_r)$. We now have, for one observation per cell,

$$\hat{\jmath}_r = \frac{y_{rr}}{2} - \frac{y_{\cdot}}{2n} - \frac{R_r - C_r}{2n}$$

$$\text{var}\,(\hat{\jmath}_r) = \frac{(n-1)(n+2)}{4n^2}\sigma^2, \quad \text{cov}\,(\hat{\jmath}_r, \hat{\jmath}_s) = \frac{-(n+2)}{4n^2}\sigma^2$$

$$\text{S.S.}\,(\hat{\jmath}_r) = \frac{4n}{n+2}\sum_r \hat{\jmath}_r^2, \text{ d.f.} = (n-1).$$

This test for genic effects, free of possible maternal effects, is attractive but conditional on the model; it must not be used if the (d) mean square comes out significant.

A more realistic model would involve, in addition to maternal effects, interactions between maternal and own-genotype effects; these interactions being specified by parameters k_{rr} on the leading diagonal as well as parameters k_{rs}. Here, however, we should be attempting to go beyond the limits imposed by the diallel design; only $\frac{1}{2}(n\text{-}1)(n\text{-}2)$ such interaction parameters are estimable, so that the introduction of k_{rr} parameters would require biologically meaningless restraints on the k_{rr} and k_{rs}. This can be seen purely on genetical grounds. Suppose, for example, the $(r \times s)$ cell showed a particularly striking dominance deviation, as estimated by

$$y_{rs} + y_{sr} - y_{rr} - y_{ss}.$$

With the diallel design, we could not be certain whether the striking effect was due strictly to dominance rather than, say, some strong interaction between genotype of F_1 and maternal effect.

4. Conclusions

A completely realistic biological model for the diallel is attainable only if interactions between maternal and own-genotype effects are absent. Thus if the (d) sum of squares, which tests for the presence of such interactions, proves to be significant, considerable caution must be exercised in interpreting the results of the analysis. If, however, (d) is not significant, the maternal effects model is biologically more appropriate than the reciprocal effects

model. Hence genic effects should be tested using the procedure given in section 3; other sums of squares will be the same on either model.

Acknowledgments.—We are indebted to Professor J. L. Jinks, Dr M. J. Kearsey, Dr P. D. S. Caligari and Mr R. Morley Jones for helpful discussions, at various times, on the diallel design and analysis.

5. References

HAYMAN, B. I. 1954. The analysis of variance of diallel tables. *Biometrics*, *10*, 235-244.
KEMPTHORNE, O. 1956. The theory of the diallel cross. *Genetics*, *41*, 451-459.
TOPHAM, P. B. 1966. Diallel analysis involving maternal and maternal interaction effects. *Heredity*, *21*, 665-674.
WEARDEN, S. 1964. Alternative analyses of the diallel cross. *Heredity*, *19*, 669-680.