ANALYSIS OF DATA FROM ALL POSSIBLE RECIPROCAL CROSSES BETWEEN A SET OF PARENTAL LINES

F. YATES

Rothamsted Experimental Station

I. INTRODUCTION

WHEN assessing the value of different parental lines in plant breeding work on plants which are normally cross-fertilised, the device of making all possible reciprocal crosses and comparing the progeny of these crosses is often of value.

The analysis of the data obtained in such a set of reciprocal crosses can be carried out by an adaptation of the analysis of variance technique. This adaptation, analogous to those appropriate to the

TABLE 1

Fertility (number of seeds per 100 florets pollinated) in reciprocal crosses of 12 sibs of an F_1 family of Trifolium hybridum

♀ Parents

Sib	2	3	7	11	1	5	8	9	10	12	4	6	Total*	Mean*
2	11	24	74	10	167	134	153	144	87	68	320	155	1228	154
3	10	24	75	13	95	178	190	163	84	113	61	158	1042	130
7	9	48	27	33	235	203	158	132	136	252	204	129	1449	181
11	20	30	58	20	108	170	168	199	119	78	139	150	1131	141
1	95	112	124	122	21	35	29	20	20	105	223	148	929	133
5	83	123	123	205	63	50	33	55	21	219	152	148	1053	150
8	80	114	163	156	41	28	6	27	61	212	130	131	986	141
9	73	152	162	210	58	32	39	42	64	248	171	237	1253	179
10	70	125	145	89	61	71	30	69	53	177	143	134	883	126
12	175	105	218	165	78	266	204	181	85	0	3	123	1600	160
4	192	116	202	163	150	195	157	154	133	0	2	155	1617	162
6	120	121	155	178	154	202	163	156	110	170	239	0	1768	161
Total *	888	968	1292	1288	987	1348	1193	1129	754	1642	1782	1668	14939	
Mean *	111	121	162	161	141	193	170	161	108	164	178	152		152

(♂ Parents)

* Excluding incompatible crosses.

more complicated forms of experimental design, such as incomplete randomised blocks, was developed to deal with the set of data shown in table 1, which gives the fertility in reciprocal crosses of 12 sibs

of a certain F_1 family of *Trifolium hybridum* (Alsike). These data were obtained in the course of work at the Welsh Plant Breeding Station, and my thanks are due to Mr Watkin Williams for permission to use them to illustrate the method of analysis.

This particular set of crosses is complicated by the fact that, in addition to nearly complete self-sterility, certain groups of sibs (2, 3, 7, 11), (1, 5, 8, 9, 10), and (12, 4) are incompatible. This incompatibility is clearly shown by the data.

Before discussing the analysis of these results, we will consider the simpler case in which partial or complete self-sterility exists but in which there are no incompatible groups, and also the case in which the selfed plants behave similarly to the crosses.

2. SELF-STERILITY WITHOUT GROUP INCOMPATIBILITY

If the plants are self-sterile, but all crosses are fertile, a table of results of the type given in table 2 will be obtained.

TABLE 2

Data from reciprocal crosses when there is self-sterility

Male parent	Female parent					Total
	1	2	3	4	...	
1	—	y_{12}	y_{13}	y_{14}	...	$Y_{1\cdot}$
2	y_{21}	—	y_{23}	y_{24}	...	$Y_{2\cdot}$
3	y_{31}	y_{32}	—	y_{34}	...	$Y_{3\cdot}$
4	y_{41}	y_{42}	y_{43}	—	...	$Y_{4\cdot}$
...	...	...	...	...	...	...
Total	$Y_{\cdot 1}$	$Y_{\cdot 2}$	$Y_{\cdot 3}$	$Y_{\cdot 4}$	...	$Y_{\cdot\cdot}$

The y's in this table may represent measurements on single plants, or the means or totals of measurements on a number of plants, possibly derived from a number of replicated plots. The Y's denote the marginal totals of the y's.

It is apparent that the direct comparison of $Y_{1\cdot}$, $Y_{2\cdot}$, . . . will not give valid estimates of the differences between the male parents, since one female parent is missing from each total. $Y_{1\cdot}$, for example, does not contain progeny from female parent 1.

Efficient estimates can be obtained by the method of least squares, fitting constants for the effects of the male and female parents. (See, for example, Yates (1933).) Let there be k parents of each sex, and let the constants be :

Mean : m.
Male parents : $a_1, a_2, a_3, \ldots$ $\qquad$ $S(a) = 0$.
Female parents : $b_1, b_2, b_3, \ldots$ $\qquad$ $S(b) = 0$.
We then have

$$y_{12} = m + a_1 + b_2 + \epsilon_{12}, \text{ etc.,}$$

where ϵ_{12}, etc., are residuals the sum of whose squares is to be minimised.

The least square equations are as follows :—

$k(k-1)m+(k-1)\ (a_1+a_2+\ \ldots)+(k-1)\ (b_1+b_2+\ \ldots)=Y..$

$(k-1)\ (m+a_1)+b_2+b_3+\ \ldots \qquad\qquad\qquad =Y_1.$

$(k-1)\ (m+b_1)+a_2+a_3+\ \ldots \qquad\qquad\qquad =Y._1$

From these equations, using the identities $S(a)=0$, $S(b)=0$, we have :

$$m = \bar{y}$$

$$a_1 = \frac{(k-1)\ Y_1.+Y._1-Y..}{k(k-2)}\ \text{etc.}$$

$$b_1 = \frac{(k-1)\ Y._1+Y_1.-Y..}{k(k-2)}\ \text{etc.}$$

The analysis of variance is shown in table 3, $\text{dev}^2 y$ being used to denote the sum of squares of deviations of all the y's from their mean.

TABLE 3

Analysis of variance of the data of table 2

	Degrees of freedom	Sum of squares
Differences between parents (male and female)	$2(k-1)$	$a_1Y_1.+a_2Y_2.+\ldots+b_1Y._1+b_2Y._2+\ldots$
Remainder	k^2-3k+1	By difference
Total	$k(k-1)-1$	$\text{dev}^2 y$

This form of analysis is based on the assumption that the effects of the parents, male and female independently, are additive, and that any departures from this additive law are random and independent. If the y's are themselves each derived from a number of replicated plots the remainder term may be compared with the experimental error obtained from an ordinary analysis of these replicates. This comparison will indicate whether there is any departure from the additive law.

The above analysis may be extended in two ways. In the first place, since the effects of the male and female parents of the same line are likely to be similar, we may recast the analysis so as to estimate the average effects of a parent of a particular line (male or female), and the differential effects of the male and female parents of that line. There may, for example, be evidence that the progeny of line 1 are superior to those of other lines, but no evidence that this superiority is more marked for the male or female parent of this line.

In the second place, there may well be some similarity between both reciprocal crosses of a pair of parents, either due to some degree of incompatibility or to some specially favourable conjunction of genes. In this event the y's which are diagonally opposite one another, y_{12} and y_{21}, etc., will be more closely correlated than the y's not so related. This point may be investigated by subdividing the remainder degrees of freedom in the above analysis into those derived from contrasts of the sums of pairs of diagonally opposite values, and those derived from contrasts of the differences of such pairs.

Since the estimates of the combined effects of male and female parents of each line are derived solely from the sums of diagonal pairs, and the differences between male and female parents solely from the differences of diagonal pairs, the whole analysis splits into two parts, one based on the sums, and the other on the differences.

The sums and differences of the diagonal pairs can be set out in two tables of the form given in table 4.

TABLE 4

Sums and differences of diagonal pairs of the data of table 2

(a) Diagonal sums, $y_{12}+y_{21}$, etc.						(b) Diagonal differences, $y_{12}-y_{21}$, etc.					
Male parent	Female parent				Total	Male parent	Female parent				Total
	1	2	3	4 ...			1	2	3	4 ...	
1	—	u_{12}	u_{13}	u_{14} ...	U_1	1	—	v_{12}	v_{13}	v_{14} ...	V_1
2		—	u_{23}	u_{24} ...	U_2	2		—	v_{23}	v_{24} ...	V_2
3			—	u_{34} ...	U_3	3			—	v_{34} ...	V_3
4				— ...	U_4	4				— ...	V_4
...				...	...	...				...	...
Total					S (U)	Total					0

The totals U_1, etc., shown in the table of diagonal sums are obtained by summing both the row and column of the parent concerned. Thus $U_3=u_{13}+u_{23}+u_{34}+$ . . . The totals V_1, etc., in the table of diagonal differences are obtained similarly, except that the values in the column concerned are subtracted. Thus $V_3=-v_{13}-v_{23}+v_{34}+$ . . . We see that $U_1=Y_1.+Y_{.1}$, etc., $V_1=Y_1.-Y_{.1}$, etc., and $S(U)=2Y..$

As before, the effects of the parents may be represented by constants, whose values can be estimated by the method of least squares. The two tables can be treated independently, thus providing separate estimates of the residual variances. It will be most convenient to use constants which represent the values of table 4 directly. Let these constants be :

Table 4 (a) : Mean : m′
                      Parents : c_1, c_2, c_3 . . .   S(c) = 0
Table 4 (b) : Differences (male-female) : $d_1, d_2, d_3,$ . . . S(d) = 0

Thus

$$u_{12} = m' + c_1 + c_2 + \epsilon'_{12}$$
$$v_{12} = d_1 - d_2 + \epsilon''_{12}$$

The least square equations are :

$$\tfrac{1}{2}k(k-1)m' + (k-1)(c_1 + c_2 + \ldots) = \tfrac{1}{2}S(U)$$
$$(k-1)(m' + c_1) + c_2 + c_3 + \ldots = U_1$$

$$\cdot \qquad \cdot \qquad \cdot \qquad \cdot \qquad \cdot \qquad \cdot$$

$$(k-1)d_1 - d_2 - d_3 \ldots = V_1$$

Whence

$$\cdot \qquad \cdot \qquad \cdot \qquad \cdot \qquad \cdot \qquad \cdot$$

$$m' = 2\bar{y}$$

$$c_1 = \frac{kU_1 - 2Y_{..}}{k(k-2)} \text{ etc.}$$

$$d_1 = \frac{1}{k}V_1 \text{ etc.}$$

Thus, as we should expect, $c_1 = a_1 + b_1$ and $d_1 = a_1 - b_1$, etc.

It will be noted that the difference between the effects of different male parents is given by the differences of the quantities $\tfrac{1}{2}(c+d)$, and of female parents by the differences of $\tfrac{1}{2}(c-d)$. If the differences between male and female parents can be neglected the parental effects (male or female) are measured by the quantities $\tfrac{1}{2}c$. Thus, under these circumstances, the expected difference between the cross with parents 1 and 2, and the cross with parents 1 and 3, is $\tfrac{1}{2}(c_2 - c_3)$; that between the crosses with parents 1 and 2, and with parents 3 and 4, is $\tfrac{1}{2}(c_1 + c_2) - \tfrac{1}{2}(c_3 + c_4)$.

The analysis of variance now splits into two parts, as shown in table 5. The extra factor of $\tfrac{1}{2}$ has been introduced into the sums of squares to bring them to units of a single entry of table 2.

TABLE 5

Analysis of variance of sums and differences of diagonal pairs

		Degrees of freedom	Sum of squares
Diagonal sums (a)	Parents	$k-1$	$\dfrac{1}{2(k-2)}$ dev²U
	Remainder	$\tfrac{1}{2}(k-2)(k-1)-1$	By difference
	Total .	$\tfrac{1}{2}k(k-1)-1$	$\tfrac{1}{2}$dev²u
Diagonal differences (b)	Parents	$k-1$	$\dfrac{1}{2k}S(V^2)$
	Remainder	$\tfrac{1}{2}(k-2)(k-1)$	By difference
	Total . .	$\tfrac{1}{2}k(k-1)$	$\tfrac{1}{2}S(v^2)$

The contrast between the mean squares for remainder (a) and remainder (b) will indicate whether there is any significantly greater

T

similarity between reciprocal crosses of the same parents than between crosses of different parents, after allowance for the average effects of parents, male and female independently, has been made.   In order to test if there are any differences between parents which are consistent for all crosses the mean square for parents (*a*) should be compared with remainder (*a*).   Comparison between parents (*b*) and remainder (*b*) gives the similar test for consistency of differences between male and female parents.   If an estimate of experimental error is available the remainder mean squares may be compared with the error mean square.

The standard errors of the *c*'s and *d*'s are given by the formulæ

$$V(c) = \frac{1}{(k-2)} (2\sigma_1^2)$$

$$V(d) = \frac{1}{k} (2\sigma_2^2)$$

the appropriate estimates of σ_1^2 and σ_2^2 being obtained from the analysis of variance, and the factor 2 being introduced because each value of table 4 is the sum or difference of two values in table 2.   In general the mean square for remainder (*a*) will provide the appropriate estimate for σ_1^2, and that for remainder (*b*) for σ_2^2, though if experimental errors only are under consideration the mean square for experimental error, if available, will be appropriate.

3. NO SELF-STERILITY

In the absence of self-sterility the data can be arranged in a $k \times k$ table similar to table 2, but with the diagonal cells occupied.   The analysis of variance of this table presents no difficulties, since the two sets of marginal totals are orthogonal, and provide estimates of the differences between parents, male and female independently.   The corresponding totals may be added to provide estimates of the differences between parents averaged over both sexes, and subtracted to provide estimates of the differential effects of the two sexes.

The possibility of the existence of greater similarity between reciprocal crosses can best be tested by the procedure of section 2, omitting the progeny of selfed parents.   The inclusion of the selfed matings in this analysis introduces additional complications, and would add little to the information provided by the partial analysis.

4. SELF-STERILITY WITH MUTUALLY INCOMPATIBLE GROUPS

If certain groups of parents are mutually incompatible the general procedure of section 2 can be followed, though the solution is somewhat more complicated, owing to the fact that all comparisons will not be of the same accuracy.   The least square equations, however, are still directly solvable.

Let the k parents be divided into groups of p, q, r, ... mutually incompatible parents, so that $p+q+r+$ . . . $=k$, and denote the constants, etc., of the first group of parents by the suffixes $p1$, $p2$, . . . with the convention that $c_{p1}+c_{p2}+$ . . . $=S(c_p)$, etc.

The least square equations are :

Diagonal sums

$$\tfrac{1}{2}\{p(k-p)+q(k-q)+...\}m'+(k-p)S(c_p)+(k-q)S(c_q)+...=\tfrac{1}{2}S(U)$$

Group p $\begin{cases}(k-p)(m'+c_{p1})+S(c_q)+S(c_r)+... &=U_{p1}\\ . \quad . \quad . \quad . \quad . \quad . \quad .\end{cases}$
(p equations)

Group q $\begin{cases}(k-q)(m'+c_{q1})+S(c_p)+S(c_r)+... &=U_{q1}\\ . \quad . \quad . \quad . \quad . \quad . \quad .\end{cases}$
(q equations)

Diagonal differences

Group p $\begin{cases}(k-p)d_{p1}-S(d_q)-S(d_r)-... &=V_{p1}\\ . \quad . \quad . \quad . \quad . \quad .\end{cases}$
(p equations)

Group q $\begin{cases}(k-q)d_{q1}-S(d_p)-S(d_r)-... &=V_{q1}\\ . \quad . \quad . \quad . \quad . \quad .\end{cases}$
(q equations)

The c and d equations may be rewritten

$$\left.\begin{aligned}&(k-p)(m'+c_{p1})-S(c_p)=U_{p1}\\ &\quad . \quad . \quad . \quad . \quad . \quad .\\ &(k-q)(m'+c_{q1})-S(c_q)=U_{q1}\\ &\quad . \quad . \quad . \quad . \quad . \quad .\\ &\qquad\quad (k-p)d_{p1}+S(d_p)=V_{p1}\\ &\quad . \quad . \quad . \quad . \quad . \quad .\\ &\qquad\quad (k-q)d_{q1}+S(d_q)=V_{q1}\\ &\quad . \quad . \quad . \quad . \quad . \quad .\end{aligned}\right\} \quad\text{(A)}$$

Summing the equations of each group, we have :

$$p(k-p)m'+(k-2p)S(c_p) \quad =S(U_p)$$
$$q(k-q)m'+(k-2q)S(c_q) \quad =S(U_q)$$
$$\quad . \quad . \quad . \quad . \quad . \quad .$$
$$kS(d_p)=S(V_p)$$
$$kS(d_q)=S(V_q)$$
$$\quad . \quad . \quad . \quad .$$

Hence

$$S(c_p)=\frac{1}{k-2p}\,S(U_p)-\frac{p(k-p)}{k-2p}\,m'$$

etc.

Since $S(c_p)+S(c_q)+\ldots = 0$ we have

$$\mathrm{K}m' = \frac{1}{k-2p}\,S(U_p) + \frac{1}{k-2q}\,S(U_q) + \ldots$$

where

$$\mathrm{K} = \frac{p(k-p)}{k-2p} + \frac{q(k-q)}{k-2q} + \ldots$$

It may be noted that since $p+q+\ldots = k$

$$\mathrm{K} = \tfrac{1}{2}k\left\{1 + \frac{p}{k-2p} + \frac{q}{k-2q} + \ldots\right\}$$

The above equation gives the numerical value of m' (which will not in this case be exactly equal to twice the mean of the original values). By substitution for $S(c_p)$ and $S(d_p)$ in equations (A) we also find

$$(k-p)c_{p1} = U_{p1} + \frac{1}{k-2p}\{S(U_p)-(k-p)^2 m'\}$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

$$(k-p)d_{p1} = V_{p1} - \frac{1}{k}\,S(V_p)$$

$$\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot$$

It will be noted that the first of the original least square equations has not been used. This equation is redundant in virtue of the identity $S(c) = 0$.

The degrees of freedom and sums of squares in the analysis of variance are given in table 6, in which $N = \tfrac{1}{2}\{p(k-p)+q(k-q)+\ldots\}$, the number of entries in the table of diagonal sums and differences.

TABLE 6

Analysis of variance with self-sterility and mutually incompatible groups

		Degrees of freedom	Sum of squares
Diagonal sums (a)	Parents	$k-1$	$\tfrac{1}{2}\{\tfrac{1}{2}m'S(U)+c_{p1}U_{p1}+\ldots+c_{q1}U_{q1}+\ldots\}-\dfrac{\{\tfrac{1}{2}S(U)\}^2}{2N}$
	Remainder	$N-k$	By difference
	Total	$N-1$	$\tfrac{1}{2}\mathrm{dev}^2 u$
Diagonal differences (b)	Parents	$k-1$	$\tfrac{1}{2}(d_{p1}V_{p1}+\ldots+d_{q1}V_{q1}+\ldots)$
	Remainder	$N-k+1$	By difference
	Total . .	N	$\tfrac{1}{2}S(v^2)$

It will be noted that, since m' is not equal to the mean, it must be introduced explicitly in the sum of squares for the constants, and the ordinary correction for the mean deducted.

The formulæ for the standard errors of the differences of the c's and d's may be derived from the solution of auxiliary sets of equations in the manner followed in partial regression analysis. There are certain points of difference, however, which are of general recurrence in least square solutions of this type, and which will therefore be worth describing. For this description we will use the notation customarily adopted in regression analysis (with the exception that, to avoid confusion, the c's of this notation will be replaced by c''s). In this notation the regression coefficients b_1, b_2, ..., satisfy the equations

$$b_1 S(x_1^2) + b_2 S(x_1 x_2) + \ldots = S(x_1 y)$$
$$b_1 S(x_1 x_2) + b_2 S(x_2^2) + \ldots = S(x_2 y)$$

$$\cdot \qquad \cdot \qquad \cdot \qquad \cdot \qquad \cdot \qquad \cdot \qquad \cdot$$

c'_{11}, c'_{12}, c'_{13}, ... satisfy a similar set of equations with the numerical terms $S(x_1 y)$, $S(x_2 y)$, $S(x_3 y)$, ... replaced by 1, 0, 0, ..., c'_{21}, c'_{22}, c'_{23} a similar set with numerical terms 0, 1, 0, ..., etc., and $c'_{12} = c'_{21}$ etc. Also

$$V(b_1) = c'_{11}\sigma^2,\ V(b_2) = c'_{22}\sigma^2,\ \text{etc.},$$
$$\mathrm{cov}(b_1 b_2) = c'_{12}\sigma^2,\ \text{etc.},$$

so that

$$V(b_1 - b_2) = (c'_{11} + c'_{22} - 2c'_{12})\sigma^2.$$

In regression analysis the numerical solution of the equations for the b's is usually effected by first determining the matrix of c''s, and thence the b's from the relations

$$b_1 = c'_{11} S(x_1 y) + c'_{12} S(x_2 y) + \ldots$$

$$\cdot \qquad \cdot \qquad \cdot \qquad \cdot \qquad \cdot \qquad \cdot \qquad \cdot$$

In cases such as the present, in which the equations are such that a simple algebraic solution is possible, the procedure may be reversed, and the c''s determined from this solution.

If there are redundant regression coefficients or constants such that

$$\mu_1 b_1 + \mu_2 b_2 + \mu_3 b_3 + \ldots = \mu_0$$
$$\mu'_1 b_1 + \mu'_2 b_2 + \mu'_3 b_3 + \ldots = \mu'_0$$

$$\cdot \qquad \cdot \qquad \cdot \qquad \cdot \qquad \cdot \qquad \cdot \qquad \cdot$$

it has been shown (Yates and Hale, 1939) that the numerical terms 1, 0, 0, . . . of the set of equations for c'_{11}, c'_{12}, c'_{13}, . . . must be replaced by

$$1 - \mu_1\, d - \mu'_1\, d' - \ldots,\quad -\mu_2\, d - \mu'_2\, d' - \ldots,\quad -\mu_3\, d - \mu'_3\, d' - \ldots,\ \ldots$$

where d, d', . . . are so chosen that the relations to which the redundant constants give rise between the coefficients of the normal equations still persist.

In the present case we have, in the solution for the diagonal sums,

$$c_{p1}+c_{p2}+ \ldots +c_{q1}+c_{q2}+ \ldots +c_{r1}+c_{r2}+ \ldots = 0$$

and the μ's for all the constants except m are therefore unity. Also the sum of the numerical terms of all the equations except the first, k in number, is equal to twice the numerical term of the first equation. Consequently the numerical terms for the set of equations giving $c'_{p1 \cdot p1}$, etc., are

$$0, \; 1 - \frac{1}{k}, \; -\frac{1}{k}, \; -\frac{1}{k}, \; \ldots$$

Hence the values for $c'_{p1 \cdot p1}$, etc., are given by the substitutions

$$U_{p1} = 1 - \frac{1}{k}, \; U_{p2} = U_{p3} = \ldots = U_{q1} = \ldots = -\frac{1}{k}$$

Substituting these values in the solutions already obtained and denoting the resultant value of m' by m'_{p1}, we find

$$m'_{p1} = \frac{1}{kK}\left\{ \frac{k-p}{k-2p} - \frac{q}{k-2q} - \frac{r}{k-2r} - \ldots \right\}$$

$$= \frac{2}{kK} \frac{k-p}{k-2p} - \frac{2}{k^2}$$

$$c'_{p1 \cdot p1} = \frac{1}{k-p}\left[1 - \frac{1}{k} + \frac{1}{k-2p}\left\{ \frac{k-p}{k} - (k-p)^2 \, m'_{p1} \right\} \right]$$

$$c'_{p1 \cdot p2} = \frac{1}{k-p}\left[-\frac{1}{k} + \frac{1}{k-2p}\left\{ \frac{k-p}{k} - (k-p)^2 \, m'_{p1} \right\} \right]$$

$$c'_{p1 \cdot q1} = \frac{1}{k-q}\left[-\frac{1}{k} + \frac{1}{k-2q}\left\{ -\frac{q}{k} - (k-q)^2 \, m'_{p1} \right\} \right]$$

The other c''s can be written down immediately from these expressions by symmetry, and it can be verified that $c'_{p1 \cdot q1} = c'_{q1 \cdot p1}$.

We now have

$$V(c_{p1}-c_{p2}) = (c'_{p1 \cdot p1}+c'_{p2 \cdot p2}-2c'_{p1 \cdot p2})\sigma^2$$

$$= \frac{2}{k-p}(2\sigma_1^2)$$

$$V(c_{p1}-c_{q1}) = (c'_{p1 \cdot p1}+c'_{q1 \cdot q1}-c'_{p1 \cdot q1}-c'_{q1 \cdot p1})\sigma^2$$

$$= (2\sigma_1^2)\left\{ \frac{1}{k-p} + \frac{1}{k-q} + \frac{1}{(k-p)(k-2p)} + \frac{1}{(k-q)(k-2q)} \right.$$
$$\left. - \frac{2k(p-q)^2}{K(k-2p)^2(k-2q)^2} \right\}$$

where, as before, estimates of σ_1^2 are given by the remainder mean squares of the analysis of variance (table 6). These two expressions give the variances of the difference of any pair of parents in the same or in different groups.

The same procedure may be followed for the diagonal differences. It will be found that

$$V(d_{p1}-d_{p2}) = \frac{2}{k-p}\,(2\sigma_2^2)$$

$$V(d_{p1}-d_{q1}) = \left(1-\frac{1}{k}\right)\!\left(\frac{1}{k-p}+\frac{1}{k-q}\right)\!(2\sigma_2^2)$$

It can be verified that when $p = q = 1$ the expressions for $V(c_{p1}-c_{q1})$ and $V(d_{p1}-d_{q1})$ reduce to the expressions already found for the case in which there is no group incompatibility.

5. ANALYSIS OF NUMERICAL DATA

We may now consider the analysis of the data of table 1. In this case $k = 12$, $p = 4$, $q = 5$, $r = 2$, $s = 1$, the parents in the four groups being sibs (2, 3, 7, 11), (1, 5, 8, 9, 10), (12, 4), and (6) respectively.

The sums and differences of the diagonal pairs are shown in table 7a and table 7b, the incompatible crosses being omitted. The U and V totals, and $S(U_p)$, and $S(V_p)$, etc., are also shown in these tables.

TABLE 7a

Diagonal sums of data of table 1

♂ Parent	♀ Parent								U	S(U)
	1	5	8	9	10	12	4	6		
2	262	217	233	217	157	243	512	275	2116	
3	207	301	304	315	209	218	177	279	2010	
7	359	326	321	294	281	470	406	284	2741	9286
11	230	375	324	409	208	243	302	328	2419	
1						183	373	302	1916	
5						485	347	350	2401	
8						416	287	294	2179	10515
9						429	325	393	2382	
10						262	276	244	1637	
12								293	3242	6641
4								394	3399	
6									3436	3436
										29878

TABLE 7b

Diagonal differences of data of table 1

♂ Parent	♀ Parent								V S(V)
	1	5	8	9	10	12	4	6	
2	+72	+51	+73	+71	+17	−107	+128	+35	+340 ⎫
3	−17	+55	+76	+11	−41	+8	−55	+37	+74 ⎪ +414
7	+111	+80	−5	−30	−9	+34	+2	−26	+157 ⎬
11	−14	−35	+12	−11	+30	−87	−24	−28	−157 ⎭
1						+27	+73	−6	−58 ⎫
5						−47	−43	−54	−295 ⎪
8						+8	−27	−32	−207 ⎬ −307
9						+67	+17	+81	+124 ⎪
10						+92	+10	+24	+129 ⎭
12								−47	−42 ⎫ −207
4								−84	−165 ⎭
6									+100 +100
									0

The equations for K, m', the c's and the d's are as follows :—

$$K = 6\left\{1 + \frac{4}{4} + \frac{5}{2} + \frac{2}{8} + \frac{1}{10}\right\} = 29 \cdot 1$$

$$m' = \frac{1}{29 \cdot 1}\left\{\frac{1}{4}9286 + \frac{1}{2}10515 + \frac{1}{8}6641 + \frac{1}{10}3436\right\} = 300 \cdot 7809$$

$$c_{p1} = \frac{1}{8}\left[2116 + \frac{1}{4}\left\{9286 - 8^2 \times 300 \cdot 7809\right\}\right] = \frac{1}{8}2116 - 311 \cdot 3743 = -46 \cdot 8743$$

$$c_{p2} = \frac{1}{8}\,2010 - 311 \cdot 3743 = -60 \cdot 1243$$

.   .   .   .

$$c_{q1} = \frac{1}{7}\left[1916 + \frac{1}{2}\left\{10515 - 7^2 \times 300 \cdot 7809\right\}\right] = \frac{1}{7}\,1916 - 301 \cdot 6617 = -27 \cdot 9474$$

.   .   .   .

$$d_{p1} = \frac{1}{8}\left[+340 - \frac{1}{12}(+414)\right] = +38 \cdot 1875$$

.   .   .   .

$$d_{q1} = \frac{1}{7}\left[-58 - \frac{1}{12}(-307)\right] = -4 \cdot 6310$$

.   .   .   .

The complete set of values for the c's and the d's are given in table 8.

TABLE 8

Values of c's and d's

Group	Parent	c	d
p	2	−46·8743	+38·1875
	3	−60·1243	+4·9375
	7	+31·2507	+15·3125
	11	−8·9993	−23·9375
q	1	−27·9474	−4·6310
	5	+41·3383	−38·4881
	8	+9·6240	−25·9167
	9	+38·6240	+21·3690
	10	−67·8046	+22·0833
r	12	+31·2364	−2·4750
	4	+46·9364	−14·7750
s	6	+12·7410	+8·3333
		+0·0009	−0·0002

The sums of squares for parents in the analysis of variance are calculated as follows :—

$$S(cU) = -46\cdot8743 \times 2116 + \ldots \quad = \quad +196119$$

$$\tfrac{1}{2}m'S(U) = \frac{1}{2} \times 300\cdot7809 \times 29878 \quad = \quad +4493366$$

$$-\frac{1}{4N}\Big\{S(U)\Big\}^2 = -\frac{1}{4\times49} \times 29878^2 = -4554566$$

$$\overline{134919}$$

$$S(dV) = (+38\cdot1875) \times (+340) + \ldots = \quad 45372$$

The sum of squares of deviations of the values in the body of table 7a, dev²u, is 319,721, and the sum of squares of the values of table 7b, $S(v^2)$, is 140,559. Introducing the further factor of $\tfrac{1}{2}$ we obtain the analysis of variance shown in table 9.

The main features of the data are now apparent. The mean square for sibs (a) is significantly greater than the remainder (a) at the 5 per cent. level, and there is also some indication of a difference between sibs (b) and remainder (b), though this is not fully significant. Furthermore, remainder (a) is significantly greater than remainder (b) at the 5 per cent. level ($e^{2z} = 1\cdot99$, 5 per cent. value = $1\cdot70$). There is thus clear indication of differences between sibs, which have a certain degree of consistency over all compatible matings. There is some indication of differences in effect between pollen and ova of the same sib, though these differences, if they exist, are not large.

Finally, crosses between individual pairs of sibs deviate somewhat from the additive law based on the average performance of either sib separately in all compatible crosses.

<div align="center">TABLE 9</div>

<div align="center">Analysis of variance of data of table 1</div>

		Degrees of freedom	Sum of squares	Mean square
Diagonal sums (a)	Sibs	11	67460	6133
	Remainder	37	92400	2497
	Total	48	159860	
Diagonal differences (b)	Sibs	11	22686	2062
	Remainder	38	47594	1252
	Total	49	70280	

The final results may be presented in a table of values of $\frac{1}{2}(c+d)$, $\frac{1}{2}(c-d)$, $\frac{1}{2}c$ and d, giving respectively the estimates of the average effects of the male parents, the female parents, the mean of male and female parents, and the difference of male and female parents. The standard errors of individual comparisons may be calculated, if required, from the formulæ already given. Allowing for irregularities in the performance of the different crosses, *i.e.* basing the error variance on remainder (a) of table 9, the smallest of the standard errors for c comparisons, that between sib 12 and sib 4, for example, is given by

$$V(c_{r1}-c_{r2}) = \frac{2}{10}\,(2497 \times 2) = 31 \cdot 6^2$$

Similarly the largest of these standard errors, that between any member of group p and any member of group q, is given by

$$V(c_{p1}-c_{q1}) = \left\{ \frac{1}{8} + \frac{1}{7} + \frac{1}{8\times4} + \frac{1}{7\times2} - \frac{2\times12(8-7)^2}{29\cdot1\times4^2\times2^2} \right\} (2497 \times 2) = 42 \cdot 2^2$$

The smallest and largest standard errors for the d comparisons are similarly given by

$$V(d_{r1}-d_{s1}) = \frac{11}{12}\left(\frac{1}{10} + \frac{1}{11} \right) \times 1252 \times 2 = 20 \cdot 9^2$$

$$V(d_{q1}-d_{q2}) = \frac{2}{7}\times 1252 \times 2 \qquad\qquad = 26 \cdot 7^2$$

Since the c and d comparisons are independent the standard errors of the comparisons of $\frac{1}{2}(c+d)$ and $\frac{1}{2}(c-d)$ can be obtained from the above standard errors by the ordinary rules for the combination of variances of independent variates.

For many purposes average standard errors which are approximately applicable to all comparisons will be adequate. These can be calculated directly from the mean square deviations of the values in table 8 and the corresponding variance ratios in table 9. The variance ratio for the c comparisons is $6133/2497 = 2·46$, and the mean square deviation of the c values is $18883/11 = 1717$. The average standard error of a single c value is therefore approximately $\sqrt{1717/2·46} = \pm 26·4$. The average standard error of the difference of two values is therefore $\sqrt{2}$ times this or $\pm 37·3$. Similarly the average standard error of a single d value is $\sqrt{518/1·65} = \pm 17·7$ and of the difference of two values is $\pm 25·0$.

Examining the results of table 8 in detail, we see that five sibs, 7, 5, 9, 12 and 4, give decidedly better performance than average. The crosses between these sibs have picked out a good proportion of the high values of table 7a, but there are some remaining, in particular those between 4 and 2, between 4 and 6, between 9 and 11, between 9 and 6, and between 12 and 8 of which one parent is a sib not in this group. Again, some of the crosses between the sibs of the top group, in particular those between 7 and 5, 7 and 9, and 9 and 4, have given values decidedly below expectation. These inconsistencies are, in part at least, a reflection of departures from the additive law. If it is desired to pick pairs of parents which may be expected to give high fertility, we should in these circumstances give some weight to the performance of the individual crosses as well as the performance of either parent separately. We should also perform the cross in the direction which the d value indicates is most favourable, e.g. in a mating between 5 and 12 we should use sib 5 as the female parent.

6. SUMMARY

The paper describes the analysis of data obtained in plant breeding work when all possible reciprocal crosses between different lines are made. The cases discussed are : self-sterility, no self-sterility, self-sterility with incompatibility within groups of lines. The last case is illustrated by a numerical example.

REFERENCES

YATES, F.   1933.
The principles of orthogonality and confounding in replicated experiments.
J. Agric. Sci. 23, 108-145.

YATES, F., and HALE, R. W.   1939.
The analysis of Latin squares when two or more rows, columns, or treatments are missing.
Suppl. J. R. Statist. Soc. 6, 67-79.