A METHOD OF INTERPRETING THE RESULTS OF FIELD TRIALS

By BERNARDO G. CAPÓ

Biometrician

In the comparison by means of field trials of a relatively small number of agricultural practices, the complete randomized block and Latin square designs are most commonly employed by the majority of research workers. When a relatively large number of such practices are to be tested, other designs are used, such as the incomplete randomized blocks, including the lattices, and designs which make use of check plots in one form or another. The reason for this attitude is that the effects of the differences in fertility between the different plots within any given block, file or column, as the case may be, are not eliminated from the estimate of the experimental error obtained by interpreting tests performed according to the above-mentioned designs by the simple methods of statistical analysis developed for use with said types of experiments.

It occasionally happens, however, that field tests are performed in places where the fertility of the soil varies a great deal between spots relatively near to one another. Under these conditions, and especially with crops in connection with which relatively large plots must be used, the usual method of statistical analysis yields too high an estimate of error and it is impossible to determine any significant difference between the practices under trial. The possibility of the use of some alternate method of interpreting the results of such experiments might conceivably be of benefit in a number of cases. It is the purpose of this article to suggest the use of a method applicable to cases of this kind which occur when relatively small numbers of practices have been tested.

The method is based on the assumption of a different effect constant for every different pair of adjacent plots in the field when an even number of plots has been used. If an odd number of plots has been used, the previous assumption is made for all but 3 plots lying together in the field, for all three of which the same effect constant is assumed to hold. By fitting then a multiple regression equation to the results of such an experiment, it is possible to obtain an estimate of the effect constant of each of the practices tested. Mathematical statisticians certainly need no more information with respect to the way in which the method works than the one conveyed in the three preceding sentences. The rank and file of field workers, however, are not so fortunate in this respect. Due to this, the explanation of the method will be performed by following a numerical example throughout for a case where an even number of plots was used. The slight modification needed for the cases of odd numbers of plots will be dealt with further on.

DIAGRAM I

Field arrangement of alfalfa experiment

50 A 30	40 H 26	30 G 30	20 F 35	10 E 44
49 C 37	39 J 36	29 I 29	19 H 30	9 G 34
48 E 59	38 B 38	28 A 22	18 J 2[illegible]	8 I 33
47 I 60	37 F 42	27 E 35	17 D 24	7 C 30
46 G 62	36 D 52	26 C 45	16 B 38	6 A 33
45 D 57	35 A 33	25 J 39	15 I 40	5 H 48
44 H 43	34 E 38	24 D 43	14 C 39	4 B 49
43 E 39	33 I 36	23 H 41	13 G 45	3 F 57
42 F 43	32 C 42	22 B 47	12 A 40	2 J 52
41 J 48	31 G 46	21 F 50	11 E 52	1 D 56

Diagram I shows the field arrangement of the plots of an alfalfa experiment performed at the Agricultural Experiment Substation farm at Isabela by Messrs. L. A. Serrano and C. J. Clavell. In each case, the first number is the plot identification number, the letter beneath it is the treatment identification letter, and the number beneath it is the observed yield in tons of green alfalfa roughage per acre obtained from 16 successive cuttings.

In table I the treatments are described, and in addition, the total and mean yields obtained with each treatment are presented.

Although in this experiment ten different treatments were tested, each one of the treatments may be expressed in terms of the amounts of nitrogen, phosphorus and potassium applied per acre. Thus, instead of trying to determine whether the observed difference in mean yields caused under the effects of any two given treatments is significant or not, one may try to determine whether the nitrogen applications have affected the alfalfa yields

TABLE I

Yields in tons of green alfalfa roughage per acre

Letter	Treatments						Yields	
	Units applied			Lbs./*A*.				
	N	P	K	NH_3	P_2O_5	K_2O	Total	Mean
A	0	0	0	0	0	0	158	31.6
B	0	1	3	0	48	216	211	42.2
C	0	1	4	0	48	288	193	38.6
D	1	1	3	36	48	216	232	46.4
E	2	1	3	72	48	216	228	45.6
F	1	1	2	36	48	144	227	45.4
G	1	1	4	36	48	288	217	43.4
H	1	0	3	36	0	216	188	37.6
I	1	2	3	36	96	216	198	39.6
J	2	2	4	72	96	288	204	40.8

significantly or not; and similarly for the other two elements. In order to determine the effects of these elements, one may assume that the yield of any plot is the additive result of a series of terms as follows:

$$N_i b_n + P_i b_p + K_i b_k + B_i = Y_i , \qquad (1)$$

where Y_i is the yield of plot i; B_i is the yield constant of the block of which plot i forms part; N_i , P_i , and K_i are the units of nitrogen, phosphorus and potassium applied to the crop in plot i; and b_n , b_p , and b_k are the respective increases in yield caused by the application of each unit of nitrogen, phosphorus or potassium respectively. b_n , b_p , and b_k are termed partial regression coefficients. The values of these coefficients are then the ones to be tested to determine whether the applications of each respective element have affected the yields significantly or not.

Thus, if one assumes block 1 to consist of plots 1 and 2, and assumes further that the yield constant of the plots in block 1 is B_1, one may write the following equation for plot 1:

$$b_n + b_p + 3b_k + B_1 = 56, \tag{2}$$

since plot 1 received treatment D, which consisted in the application of 1 unit of nitrogen, 1 unit of phosphorus and 3 units of potassium. One may similarly write the following equation for plot 2:

$$2b_n + 2b_p + 4b_k + B_1 = 52. \tag{3}$$

In a similar way, one may write down equations for the other plots, obtaining a total of 50 equations, one for each plot. In these equations there would be the 28 unknowns $b_n, b_p, b_k, B_1, B_2, \cdots, B_{25}$. It is impossible to determine values for these 28 unknowns which will fit the 50 equations. The next best solution, that of finding the most probable values of the constants, must be resorted to. Once these most probable values are determined, one may estimate the yield of any plot by means of equation (1) above. The difference between the actually observed yield on any given plot and the value of said yield estimated by means of equation (1) is the error of estimate of the yield of that plot. Thus, for plot (1) the error of estimate would be:

$$d_1 = 56 - b_n - b_p - 3b_k - B_1, \tag{4}$$

since the actually observed yield was 56 and the estimated yield would be the value of $b_n + b_p + 3b_k + B_1$ when the most probable values of these constants are substituted in equation (4).

According to the principle of Least Squares, the most probable values of constants like the ones under discussion are those which render the sum of the squares of the errors of estimate a minimum. Thus, if d_1 is the error of estimate of plot 1, d_2 is the error of estimate of plot 2, etc., those values of the constants $b_n, b_p, b_k, B_1, B_2, \cdots, B_{25}$ which would make $Sd^2 = d_1^2 + d_2^2 + \cdots + d_{50}^2$ a minimum, would be their most probable values, and therefore, the ones to be used in equation (1) for estimating the yield of any given plot.

Now,

$$d_1^2 = (56 - b_n - b_p - 3b_k - B_1)^2. \tag{5}$$

Similarly,

$$d_2^2 = (52 - 2b_n - 2b_p - 4b_k - B_1)^2. \tag{6}$$

In a similar way one may find the values of $d_3^2, d_4^2, \cdots, d_{50}^2$.

In applying the criteria for a minimum to the expression $Sd^2 = d_1^2 + d_2^2 + \cdots + d_{50}^2 = (56 - b_n - b_p - 3b_k - B_1)^2 + (52 - 2b_n - 2b_p -$

$4b_k - B_1)^2 + \cdots + (30 - B_{25})^2$, one obtains 28 equations, each one of which is a partial derivative of Sd^2 with respect to one of the constants b_n, b_p, b_k, B_1, B_2, $\cdots$, B_{25} equated to zero. The equation obtained on finding the partial derivative of Sd^2 with respect to B_1 and simplifying is

$$56 - b_n - b_p - 3b_k - B_1 + 52 - 2b_n - 2b_p - 4b_k - B_1 = 0. \quad (7)$$

From this equation one gets

$$B_1 = 54 - 3b_n/2 - 3b_p/2 - 7b_k/2. \quad (8)$$

If one now substitutes this value of B_1 in equation (4) above, and simplifies, one gets

$$d_1 = 2 + b_n/2 + b_p/2 + b_k/2 = (d_1 - d_2)/2. \quad (9)$$

Substituting the value of B_1 in the corresponding equation for plot 2, i.e.,

$$d_2 = 52 - 2b_n - 2b_p - 4b_k - B_1,$$

one gets $$d_2 = -2 - b_n/2 - b_p/2 - b_k/2 = -(d_1 - d_2)/2. \quad (10)$$

It will be noticed that by means of these substitutions B_1 has been eliminated from the equations corresponding to the first two plots and, since these are the only two equations where B_1 occurs, from the whole set of 50 equations. In a similar way all the other B's may be eliminated from the other equations resulting finally in a system of 50 equations in but the three unknowns b_n, b_p, and b_k. In this new system of equations, the right-hand side of the equation corresponding to plot 1 will be equal to the right-hand side of that corresponding to plot 2 except that the signs are reversed. Similarly the right-hand sides of the equations corresponding to plots 3 and 4 will be equal except for the signs changed; and similarly for the remaining pairs of equations.

Now, the most probable values of the three unknowns b_n, b_p, and b_k are, as stated above, those which would make Sd^2 a minimum. From equations (9) and (10) above, it follows that $d_1 = -d_2$, and therefore, $d_1^2 = d_2^2$ and $-2d_1d_2 = 2d_1^2 = 2d_2^2 = d_1^2 + d_2^2$. Thus, $(d_1 - d_2)^2 = d_1^2 - 2d_1d_2 + d_2^2 = 2d_1^2 + 2d_2^2$. A similar relation may be found in the cases of the corresponding errors of estimate of the plots belonging to the other blocks. Therefore,

$$(d_1 - d_2)^2 + (d_3 - d_4)^2 + \cdots + (d_{49} - d_{50})^2 = 2d_1^2 + 2d_2^2 + 2d_3^2 + 2d_4^2 + \cdots + 2d_{49}^2 + 2d_{50}^2 = 2Sd^2. \quad (11)$$

For simplicity of calculations Sd^2 is calculated in the method under discussion by calculating $2Sd^2$ by the use of the left-hand side of equation (11), and dividing by 2. The numerical values of the coefficients of the squares and products of the unknowns which appear in Sd^2 may thus be conveniently found in the example under study by arranging the calculations as they appear in table II.

TABLE II

Calculation of the numerical values of the coefficients of squares and products

(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Difference															
Plots	Treatments	d_N	d_P	d_K	d_Y	d_N^2	$d_N d_P$	$d_N d_K$	$d_N d_Y$	d_P^2	$d_P d_K$	$d_P d_Y$	d_K^2	$d_K d_Y$	d_Y^2
1–2	D-J	−1	−1	−1	4	1	1	1	−4	1	1	−4	1	−4	16
3–4	F-B	1	0	−1	8	1	0	−1	8	0	0	0	1	−8	64
5–6	H-A	1	0	3	15	1	0	3	15	0	0	0	9	45	225
8–7	I-C	1	1	−1	3	1	1	−1	3	1	−1	3	1	−3	9
10–9	E-G	1	0	−1	10	1	0	−1	10	0	0	0	1	−10	100
11–12	E-A	2	1	3	12	4	2	6	24	1	3	12	9	36	144
13–14	G-C	1	0	0	6	1	0	0	6	0	0	0	0	0	36
15–16	I-B	1	1	0	2	1	1	0	2	1	0	2	0	0	4
18–17	J-D	1	1	1	5	1	1	1	5	1	1	5	1	5	25
20–19	F-H	0	1	−1	5	0	0	0	0	1	−1	5	1	−5	25
21–22	F-B	1	0	−1	3	1	0	−1	3	0	0	0	1	−3	9
24–23	D-H	0	1	0	2	0	0	0	0	1	0	2	0	0	4
26–25	C-J	−2	−1	0	6	4	2	0	−12	1	0	−6	0	0	36
27–28	E-A	2	1	3	13	4	2	6	26	1	3	13	9	39	169
30–29	G-I	0	−1	1	1	0	0	0	0	1	−1	−1	1	1	1
31–32	G-C	1	0	0	4	1	0	0	4	0	0	0	0	0	16
34–33	E-I	1	−1	0	2	1	−1	0	2	1	0	−2	0	0	4
36–35	D-A	1	1	3	19	1	1	3	19	1	3	19	9	57	361
37–38	F-B	1	0	−1	4	1	0	−1	4	0	0	0	1	−4	16
39–40	J-H	1	2	1	10	1	2	1	10	4	2	20	1	10	100
41–42	J-F	1	1	2	5	1	1	2	5	1	2	5	4	10	25
44–43	H-B	1	−1	0	4	1	−1	0	4	1	0	−4	0	0	16
46–45	G-D	0	0	1	5	0	0	0	0	0	0	0	1	5	25
47–48	I-E	−1	1	0	1	1	−1	0	−1	1	0	1	0	0	1
49–50	C-A	0	1	4	7	0	0	0	0	1	4	7	16	28	49
Sum						29	11	18	133	20	16	77	67	199	1480
Coefficients (= Sum/2)						14.5	5.5	9	66.5	10	8	38.5	33.5	99.5	740

Columns (1), (2) and (6) have been filled from the information given in diagram I. In finding the differences corresponding to any given pair of adjacent plots, said differences have been taken in such an order as to yield a positive value of d_Y in every case. Thus, since the yield of plot 2 was smaller than the yield of plot 1, the differences are to be found by subtracting the data of plot 2 from the corresponding data of plot 1. This is indicated by writing "$1 - 2$" in column (1). Since plot 1 received treatment D

and plot 2 received treatment J, the corresponding entry in column (2) is $D - J$. The entry in column (6) is $4(= 56 - 52)$. The rest of columns (1), (2), and (6) is filled in a similar way. Thus, for plots 33 and 34, the order will be $34 - 33$ since the yield of plot 34 exceeded the yield of plot 33. In columns (2) and (6) the corresponding entries are $E - I$ and 2, respectively.

After filling columns (1), (2), and (6); columns (3), (4), and (5) must be filled. In filling them use is made of table I. The entries corresponding to plots 1 and 2 in these columns are found as follows: d_N = units of N of treatment D minus the units of N of treatment $J = 1 - 2 = -1$; and likewise, $d_P = 1 - 2 = -1$, and $d_K = 3 - 4 = -1$. For plots 33 and 34, where the corresponding difference between treatments is $E - I$, $d_N = 2 - 1 = 1$, $d_P = 1 - 2 = -1$, and $d_K = 3 - 3 = 0$. In a similar way, the rest of columns (3), (4), and (5) may be filled.

Once columns (1) to (6) are filled, the corresponding entries in columns (7) to (16) may be made by performing in each case the operation indicated by the heading at the top of each column. Thus, for the entries corresponding to plots 1 and 2, $d_N^2 = d_N d_N = (-1)(-1) = 1$; $d_N d_P = (-1)(-1) = 1$; $d_N d_K = (-1)(-1) = 1$; $d_N d_Y = (-1)(4) = -4$; $d_P^2 = d_P d_P = (-1)(-1) = 1$; $d_P d_K = (-1)(-1) = 1$; $d_P d_Y = (-1)\,4 = -4$; etc. The entries corresponding to plots 33 and 34 are, likewise, as follows: $d_N^2 = (1)(1) = 1$; $d_N d_P = (1)(-1) = -1$; $d_N d_K = (1)(0) = 0$; $d_N d_Y = (1)(2) = 2$; etc.

Once these entries are made, it remains but to add the entries of columns (6) to (16) and to divide each sum by 2. Since

$$Sd^2 = b_n^2 Sn^2 + 2b_n b_p Snp + 2b_n b_k Snk - 2b_n Sny + b_p^2 Sp^2 + 2b_p b_k Spk - 2b_p Spy + b_k^2 Sk^2 - 2b_k Sky + Sy^2, \quad (12)$$

one can write the value of Sd^2 for this case as follows:

$$Sd^2 = 14.5b_n^2 + 2(5.5)b_n b_p + 2(9)b_n b_k - 2(66.5)b_n + 10b_p^2 + 2(8)b_p b_k - 2(38.5)b_p + 33.5b_k^2 - 2(99.5)b_k + 740. \quad (13)$$

As previously stated, the most probable values of b_n, b_p, and b_k are those which will make Sd^2 minimum. Among the mathematical requirements for Sd^2 to be a minimum, the requirements stated by three following equations must be fulfilled:

$$b_n Sn^2 + b_p Snp + b_k Snk = Sny, \quad (14)$$

$$b_n Snp + b_p Sp^2 + b_k Spk = Spy, \quad (15)$$

and

$$b_n Snk + b_p Spk + b_k Sk^2 = Sky. \quad (16)$$

Equations (14), (15), and (16) are known as "normal equation in b_n," "normal equation in b_p," and "normal equation in b_k," respectively. In the sample under study, the equations are as follows:

$$14.5b_n + 5.5b_p + 9b_k = 66.5, \tag{17}$$

$$5.5b_n + 10b_p + 8b_k = 38.5, \tag{18}$$

and

$$9b_n + 8b_p + 33.5b_k = 99.5. \tag{19}$$

These three simultaneous linear equations in 3 unknowns may be solved by any of the well-known methods studied in the elementary courses in algebra to obtain the values of the partial regression coefficients b_n, b_p, and b_k. Due, however, to the fact that the knowledge of the standard error of any given statistic is almost as essential as the value of the statistic itself for the proper evaluation of the significance of said statistic and of differences between it and some other statistic, Fisher's modification of Gauss' method of correlatives or indeterminate multipliers is usually resorted to in practice. Snedecor (4, p. 302) and Rider (3, p. 39) present discussions and give examples of solutions of these systems of equations by this method. Doolittle's method of solution of normal equations as discussed by Mills (2, p. 656) may be incorporated to advantage to the method of indeterminate multipliers. The author has found these methods, however, too tedious and intricate for use in practice. He suggests, therefore, the use of the following method which, for a small number of constants to be fitted, is much easier and takes considerably less time to apply than the methods previously referred to, specially if one has a printed or mimeographed form stating the calculations to be performed, so that one has to fill in merely the values asked for in said form. The suggested method of solution is as follows:

Following Fisher, (1, p. 144), the equations to be solved are:

$$C_{aa}a_1 + C_{ab}b_1 + C_{ac}C_1 = 1, \tag{20}$$

$$C_{aa}a_2 + C_{ab}b_2 + C_{ac}C_2 = 0, \tag{21}$$

$$C_{aa}a_3 + C_{ab}b_3 + C_{ac}C_3 = 0, \tag{22}$$

$$C_{ab}a_1 + C_{bb}b_1 + C_{bc}C_1 = 0, \tag{23}$$

$$C_{ab}a_2 + C_{bb}b_2 + C_{bc}C_2 = 1, \tag{24}$$

$$C_{ab}a_3 + C_{bb}b_3 + C_{bc}C_3 = 0, \tag{25}$$

$$C_{ac}a_1 + C_{bc}b_1 + C_{cc}C_1 = 0, \tag{26}$$

$$C_{ac}a_2 + C_{bc}b_2 + C_{cc}C_2 = 0, \tag{27}$$

and

$$C_{ac}a_3 + C_{bc}b_3 + C_{cc}C_3 = 1, \tag{28}$$

where $a_1 = Sn^2$, $a_2 = b_1 = Snp$, $a_3 = c_1 = Snk$, $b_2 = Sp^2$, $b_3 = c_2 = Spk$, and $c_3 = Sk^2$.

Now, if A_{12} represents the determinant of the second order $\begin{vmatrix} a_1 & b_1 \\ a_2 & b_2 \end{vmatrix}$, A_{123} means the determinant of the third order $\begin{vmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{vmatrix}$, and, in general, $A_{123\cdots n}$ means the determinant of the nth order

$$\begin{vmatrix} a_1 & b_1 & \cdots & m_1 \\ a_2 & b_2 & \cdots & m_2 \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ a_n & b_n & \cdots & m_n \end{vmatrix},$$

the following relations may be easily deducted from the fact that any given determinant is equal to the sum of all terms which may be formed by writing down the principal diagonal $a_1b_2c_3 \cdots m_n$ and forming all possible permutations of the subscripts; the sign of each term being positive if an even number of inversions must be performed to bring the subscripts into an ascending series and negative if said number of inversions is odd.

$$A_{12} = a_1b_2 - a_2b_1 \, ; \qquad A_{123} = A_{12}c_3 - A_{13}c_2 + A_{23}c_1 \, ,$$

$$A_{1234} = A_{123}d_4 - A_{124}d_3 + A_{134}d_2 - A_{234}d_1 \, , \text{etc.}$$

If now $A_{103} = a_1c_3 - a_3c_1$; $A_{203} = a_2c_3 - a_3c_2$ and $B_{23} = b_2c_3 - b_3c_2$, the values of C_{aa} , C_{ab} , and C_{ac} may be determined from equations (20), (21), and (22) above to be as follows:

$$C_{aa} = B_{23}A^{-1}_{123} \; ; \quad C_{ab} = -A_{203}A^{-1}_{122} \; ; \quad \text{and} \quad C_{ac} = A_{23}A^{-1}_{123} \, .$$

From equations (23), (24), and (25), the values of C_{bb} and C_{bc} come out to be as follows:

$$C_{bb} = A_{103}A^{-1}_{123} \; ; \quad \text{and} \quad C_{bc} = -A_{13}A^{-1}_{123} \, .$$

Finally, from equations (26), (27), and (28),

$$C_{cc} = A_{12}A^{-1}_{123} \, .$$

The number of operations required to evaluate these statistics is not large and does not take too long, specially if, as already stated, one has some form indicating the operations and has to fill only the numbers in it. Such a form, already filled, appears as Table III.

In this table, A corresponds to b_n , B corresponds to b_p , and C corresponds to b_k which are the unknowns whose values were to be determined. These values of b_n, b_p, and b_k may be checked by substituting them in equations (17), (18), and (19) above.

TABLE III

Solution of simultaneous linear equations

$$Aa_1 + Bb_1 + Cc_1 = Say$$
$$Aa_2 + Bb_2 + Cc_2 = Sby$$
$$Aa_3 + Bb_3 + Cc_2 = Scy$$

$a_1 = 14.5$	$b_1 = 5.5$	$c_1 = 9$	$Say = 66.5$
$a_2 = 5.5$	$b_2 = 10$	$c_2 = 8$	$Sby = 38.5$
$a_3 = 9$	$b_3 = 8$	$c_3 = 33.5$	$Scy = 99.5$
$a_1b_2 = 145$	$a_2b_1 = 30.25$	$a_3b_1 = 49.5$	$b_2c_3 = 335$
$a_1b_3 = 116$	$a_2b_3 = 44$	$a_3b_2 = 90$	$b_3c_2 = 64$
$a_1c_3 = 485.75$	$a_2c_3 = 184.25$	$a_3c_1 = 81$	
		$a_3c_2 = 72$	

$A_{12} = a_1b_2 - a_2b_1 = 114.75$	$A_{12}c_3 = 3844.125$
$A_{13} = a_1b_3 - a_3b_1 = 66.50$	$-A_{13}c_2 = -532$
$A_{103} = a_1c_3 - a_3c_1 = 404.75$	$A_{23}c_1 = -414$
$A_{23} = a_2b_3 - a_3b_2 = -46$	$A_{123} = 2898.125$
$A_{203} = a_2c_3 - a_3c_2 = 112.25$	$A^{-1}_{123} = 0.000345051$
$B_{23} = b_2c_3 - b_3c_2 = 271$	
$C_{aa} = B_{23}A^{-1}_{123} = 0.0935088$	$C_{bb} = A_{103}A^{-1}_{123} = 0.139659$
$C_{ab} = -A_{203}A^{-1}_{123} = -0.0387320$	$C_{bc} = -A_{13}A^{-1}_{123} = -0.0229459$
$C_{ac} = A_{23}A^{-1}_{123} = -0.0158723$	$C_{cc} = A_{12}A^{-1}_{123} = 0.0395946$

Checks:

$C_{aa}a_1 = 1.35588$	$C_{ab}a_2 = -0.213026$	$C_{ac}a_3 = -0.142851$
$C_{ab}b_1 = -0.213026$	$C_{bb}b_2 = 1.39659$	$C_{bc}b_3 = -0.183567$
$C_{ac}c_1 = -0.142851$	$C_{bc}c_2 = -0.183567$	$C_{cc}c_3 = 1.32642$
1.0000	1.0000	1.0000

$C_{aa}Say = 6.218335$	$C_{ab}Say = -2.575678$	$C_{ac}Say = -1.055508$
$C_{ab}Sby = -1.491182$	$C_{bb}Sby = 5.376872$	$C_{bc}Sby = -0.883417$
$C_{ac}Scy = -1.579294$	$C_{bc}Scy = -2.283117$	$C_{cc}Scy = 3.939663$
$A = 3.147859$	$B = 0.518077$	$C = 2.000738$

The significance of these coefficients of regression may be checked by calculating their standard errors and using the "t-test." The total sum of squared deviations corrected for variations in fertility from block to block was found in table II to be 740, subject to 25 degrees of freedom, since from the total original number of 49 degrees of freedom, 24 (= 25 − 1) belong to the block statistics assumed to exist. The reduction in this sum of squares caused by fitting the statistics b_n, b_p, and b_k is found, as usual, by means of the relation

$$Sy'^2 = b_nSn^2 + b_pSp^2 + b_kSk^2 = 209.33 + 19.95 + 199.07 = 428.35, \quad (29)$$

corresponding to 3 degrees of freedom, since it corresponds to the three statistics so fitted. The reduced sum of squared deviations is, therefore, $Sd^2 = S(Y - Y')^2 = 740 - 428.35 = 311.65$, corresponding to 22df.

The estimate of the variance, V, is, therefore, $V = 311.65/22 = 14.1659$. The variance of b_n is then $C_{aa}V = (0.0935088)14.1659 = 1.3246$. The standard error of $b_n = \text{S.E.}_{b_n} = (1.3246)^{1/2} = 1.151$, and the corresponding value of $t = 3.1479/1.151 = 2.73$, which is significant at the 5% point; since the value of t at the 5% for 22df. is 2.074.

TABLE IV

Analysis of the total sum of squared deviations

Source of the deviations	Degrees of freedom	Sum of squared deviations	Variance estimate	F values		
				Experimental	5%	1%
Total	49	4,651				
Blocks	4	802				
Treatments	9	922	102.4	1.26	2.15	2.94
Error	36	2,927	81.31			

Similarly, $V_{b_p} = 1.9784$, $\text{S.E.}_{b_p} = 1.407$, and $t_{b_p} = 0.37$, which is not significant. For b_k, the corresponding figures are: $V_{b_k} = 0.5609$, $\text{S.E.}_{b_k} = 0.7489$, and $t_{b_k} = 2.67$, which is also significant at the 5% point.

The performed analysis indicates that the crop responded significantly to the applications of nitrogen and potash, whereas it did not do so with respect to those of phosphorus.

The usual analysis of variance of the yield data of the experiment under study yielded table IV, the first block consisting of plots 1 to 10, the second block of plots 11 to 20, etc.

The usual hypothesis thus fails to show any significant effect of the treatments on the yields. However, since in the previous analysis but 3 statistics were fitted to explain the effects of the fertilizer applications whereas in this analysis 9 such statistics were fitted, a new analysis was made in an attempt to explain the sum of squares due to treatments in this last analysis by the use of only 3 statistics. Such an attempt indicated that of the sum of squares due to treatments, 922, a total of 334 could be explained by fitting the 3 statistics, there remaining 588 to be assigned to interactions of one sort or another. If this remainder is pooled with the sum of squares due to error, 2,927, a total of 3,515 is obtained, subject to

36 + 6 = 42df. The new estimate of the error variance would be, therefore, 83.69, of the same order of magnitude as the value formerly obtained of 81.31. On testing the fitted statistics, whose values came out to be: $b_n = 2.167$, $b_p = 0.529$ and $b_k = 1.188$, it is found that none of them is significant.

On comparing the variance estimate obtained by the last method of analysis, 83.69, with that obtained by the proposed method, 14.1659, one obtains a relative precision of 5.9078, (= 83.69/14.1659), in favor of the 2-plot blocks hypothesis as against the usual hypothesis, in this case the 10-plot blocks hypothesis. There has been, in this case, therefore, an increase in precision of 491 per cent due to the change in hypothesis.

The results of this experiment might have been also interpreted by assuming 5-plot incomplete blocks consisting of either the plots receiving treatments *A*, *C*, *E*, *G* and *I* or treatments *B*, *D*, *E*, *H*, and *J*. The results of such an analysis appear in table V.

TABLE V

Analysis of the total sum of squared deviations

Source of the deviations	Degrees of freedom	Sum of squared deviations	Variance estimate	*F* values		
				Experimental	5%	1%
Total	49	4,651				
Blocks	9	2,317				
Treatments	8	830	103.75	2.21	2.25	3.12
Error	32	1,504	47.00			

The observed differences between the treatment means are again not significant, although there has been an increase in precision of 81.31/47 − 100% = 73%, by the use of the smaller 5-plot blocks.

In case that an experiment consists of an odd number of plots, either by design or accident, the contribution of all but 3 plots to the numerical coefficients of the squares and products which appear in Sd^2 will be calculated as described above in the case of table II, by finding the respective differences between the adjacent plots, finding all possible squares and products of these differences, adding these squares and products, and dividing each of these sums by 2.

The contribution of the 3 plots left out in the previous calculations may be computed as follows: The set of differences corresponding to the first of these plots will be formed by subtracting the sum of the corresponding figures of the second and third plot from twice the corresponding figures of the first plot. The set of differences corresponding to the second plot will be likewise formed by subtracting the sum of the corresponding figures of the first and third plot from twice the corresponding figures of the second plot. In a similar way are found the differences corresponding to the third plot.

Again the sums of all possible squares and products of these sets of differences are found, but in this case the sums must be divided by 9 in order to determine the contributions of these last 3 plots to the numerical coefficients. The súm of the contributions of both sets of plots will give the required numerical coefficients of the squares and products which appear in Sd^2.

To illustrate the way of finding the contribution of these last 3 plots in such an analysis, one may assume that plots 48, 49, and 50 of diagram I are the remaining 3 plots in the case of an experiment with an uneven number of plots. The value of d_Y for plot 48 would be found as follows: Since plot 48 received treatment E, with 2 units of nitrogen; plot 49 received treatment C, with no nitrogen; and plot 50 received treatment A, with no nitrogen, the respective difference, d_N, would be, therefore, $2(2) - 0 - 0 = 4$. The corresponding d_P would be $2(1) - 1 - 0 = 1$; the corresponding d_K would be $2(3) - 4 - 0 = 2$; and the corresponding d_Y would be $2(59) - 37 - 30 = 118 - 67 = 51$. In a similar way, the differences corresponding to plots 49 and 50 are found.

In the above discussion, the question of the significance of the difference between any pair of statistics fitted to data of the kind mentioned has been dispensed with. In the cases where several varieties or field practices have been tested, the interest centers, not on whether a given variety or practice produces a significant departure of the measured effect from the mean effect of all the varieties or practices tested, but on whether the given variety or practice causes a significantly greater or smaller effect than some other variety or practice included in the test. In such cases, the number of statistics fitted to the data is one less than the number of treatments under study. Thus, if, say, varieties A, B, C, D, and E were tested in one of such experiments, statistics for A, B, C, and D only would be fitted, it being understood that the statistic for E is equal to the negative sum of the statistics corresponding to the other four varieties, i.e., $E = -(A + B + C + D)$. The corresponding C_{ei}'s would be found by means of the relations:

$$C_{ea} = -(C_{aa} + C_{ab} + C_{ac} + C_{ad}), \tag{30}$$

$$C_{eb} = -(C_{ab} + C_{bb} + C_{bc} + C_{bd}), \tag{31}$$

$$C_{ec} = -(C_{ac} + C_{bc} + C_{cc} + C_{cd}), \tag{32}$$

$$C_{ed} = -(C_{ad} + C_{bd} + C_{cd} + C_{dd}), \tag{33}$$

and

$$C_{ee} = -(C_{ae} + C_{be} + C_{ce} + C_{de}). \tag{34}$$

The variance estimate of a difference between any two such statistics, say A and B, would be found by means of the relation $(C_{aa} - 2C_{ab} + C_{bb})V$, from which the corresponding values of the standard error and t would be found accordingly.

It now remains to present a resumé of the results obtained to date in the use of the 2-plot blocks method in the interpretation of the results of field trials. It may be stated that, in general, and in full accord with the theoretical foundations of the method, it is most useful as compared with

TABLE VI

Comparison of efficiency of the 2-plot blocks method with the usual analysis of "variance" method of interpreting the results of complete randomized blocks experiments

Crop	Nature of test	Factor studied	Coefficient of variability obtained by usual analysis of "variance" method	Increase in efficiency due to the use of the 2-plot blocks method
			%	%
Corn, 1st crop	Fertilizer	Yield	11.04	13.15
Corn, 2nd crop	"	"	17.13	−6.43
Corn, 3rd crop	"	"	9.97	−11.00
Alfalfa, 16 cuttings	" unlimed	"	21.93	490.81
" , " "	" limed	"	7.13	32.60
Beans	"	"	16.79	27.86
Sweet potato, 1st crop	"	"	43.71	248.75
" " , 2nd "	"	"	26.92	51.29
Cucumbers, 1932–33	"	"	18.38	11.31
Cucumbers, 1933–34	"	"	18.20	109.32
" , 1935–36	"	"	27.67	74.88
" , 1936–37	"	"	21.02	75.15
" , 1941–42	"	"	25.25	24.81
Cotton	"	"	16.27	−7.61
Sugar cane, Aguirre 1942–43	Varietal	Yield of cane	17.27	53.48
" " "	"	Tons sugar/A	17.31	56.93
" " "	"	Sugar % cane	4.62	−13.18
Sugar cane, Río Piedras 1941–42	"	Tons sugar/A	27.60	190.16

the usual method of analysis in cases where soil heterogeneity affects markedly the effect under study, the more so the more variable the soil. In table VI, the results of comparisons of the 2-plot blocks method with the usual method of analysis are presented. In every case the comparisons between the two methods have been made by using the same number of statistics to explain the effects of the treatments.

As seen above, in by far the majority of the cases, the 2-plot blocks method has been more precise than the complete randomized blocks method. The method, however, takes more time and work to apply to the results of trials not specially designed for its application. By using some sort of balanced design, however, the work of calculation of the results of experiments by means of the proposed method may be considerably shortened.

Thus, in testing 6 treatments, each of the treatments might be included with each of the other ones an equal number of times in the 2-plot blocks. This would require replicating each treatment either 5 times or some multiple of 5. In general, this would require $m(n-1)$ replications where n is the number of treatments tested and m is any positive integer. Such a layout for 6 treatments replicated 5 times could be something as illustrated in diagram II.

DIAGRAM II

Layout for a plot distribution to test 6 treatments (A, B, C, D, E, and F) each replicated 5 times

AB	CD	EF
EC	FA	BD
FD	BC	AE
CA	DE	FB
BE	CF	DA

The advantage of such a balanced design would be that all C_{ii}'s would be equal, i.e., $C_{aa} = C_{bb} = C_{cc} = C_{dd} = C_{ee} = C_{ff}$, and, similarly, all C_{ij}'s would be equal, i.e., $C_{ab} = C_{ac} = \cdots = C_{ef}$. Therefore, the same variance estimate would be used to test each effect constant and, likewise, there would be but one variance estimate to test the difference between any two such effect constants. More information relative to the possibilities of the use of these balanced designs will be given in another article to be published shortly after this one in this same Journal. In said article, the manuscript of which has been already prepared, the modified calculational technique is discussed and a numerical example is presented.

REFERENCES

1. Fisher, R. A., Statistical Methods for Research Workers, 4th edition, revised and enlarged. Oliver and Boyd, Edinburgh. 1932.
2. Mills, F. C., Statistical Methods, Revised. Henry Holt and Co., New York. 1939.
3. Rider, P. R., Modern Statistical Methods. John Wiley and Sons, Inc., New York. 1939.
4. Snedecor, G. W., Statistical Methods, 3rd edition. The Iowa State College Press, Ames, Iowa. 1940.