32

STUDIES IN CROP VARIATION.

II. THE MANURIAL RESPONSE OF DIFFERENT POTATO VARIETIES.

BY R. A. FISHER, M.A. AND W. A. MACKENZIE, B.Sc.
Rothamsted Experimental Station, Harpenden.

(With Two Charts.)

1. INTRODUCTORY.

IT is not infrequently assumed that varieties of cultivated plants differ not only in their suitability to different climatic and soil conditions, but in their response to different manures. Since the experimental error of field experiments is often underestimated, this supposition affords a means of explaining discrepancies between the results of manurial experiments conducted with different varieties; in the absence of experimental evidence adequate to prove or disprove the supposed differences between varieties in their response to manures such explanations cannot be definitely set aside, although we very often suspect that the discrepancies are in reality due to the normal errors of field experiments.

On the other hand, if important differences exist in the manurial response of varieties a great complication is introduced into both variety and manurial tests; and the practical application of the results of past tests becomes attended with considerable hazard. Only if such differences are non-existent, or quite unimportant, can variety tests conducted with a single manurial treatment give conclusive evidence as to the relative value of different varieties, or manurial tests conducted with a single variety give conclusive evidence as to the relative value of different manures.

In a recent experiment at Rothamsted twelve potato varieties were tested with six manurial treatments, and since the tests were carried out in triplicate the normal error may be evaluated with some accuracy. There is thus afforded a basis for comparing the discrepancies between the different varieties with those to be expected if all varieties had responded alike to the differences in manurial treatment. Both the general response to manurial treatment, and the general differences in the yield of varieties were well marked, so that the data are well suited to the present enquiry.

The general question of the existence of differences in the response of varieties to manurial treatment cannot, of course, be answered from data respecting a single species. The principal purpose of the present note is to call attention to (i) the kind of data which is adequate to supply an answer in any particular case, and (ii) the statistical treatment by which such an answer can be elicited.

2. Arrangement of Experiment.

The experimental work was carried out by Mr T. Eden, of this Station, and formed part of a larger investigation into the effects of fertilisers on potatoes.

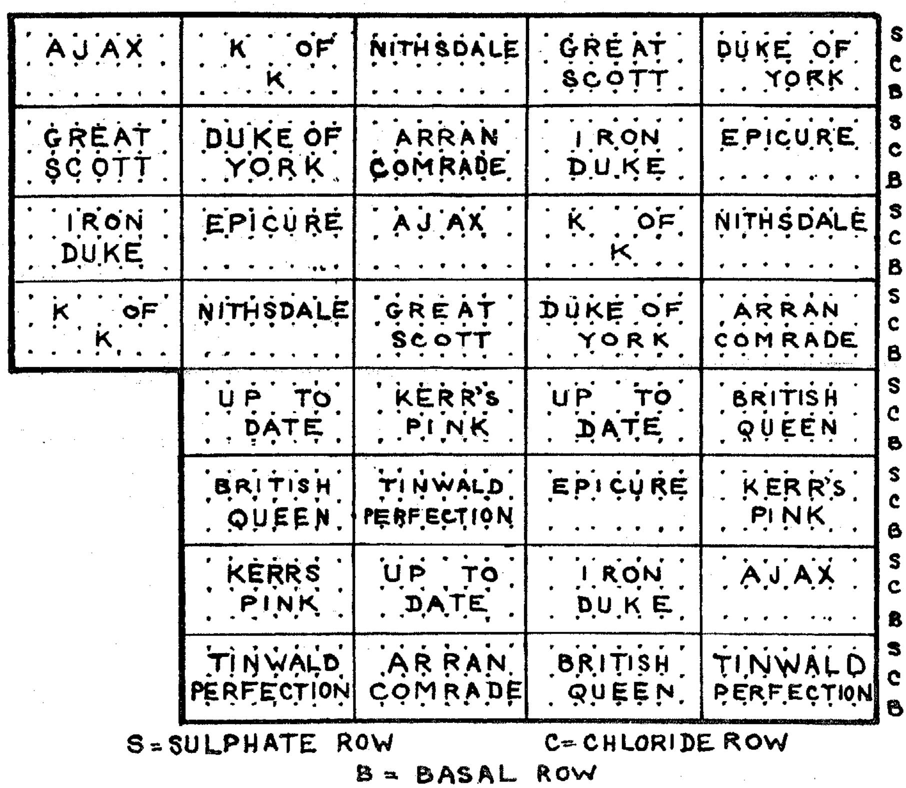

Diagram 1. Plan of experiment. Farmyard manure series.

The portion of land selected for the purpose had previously received uniform manurial treatment, thus avoiding the possibility of variation in yield due to the varying residual effect of different manurial systems. The total area used was quite small, being only ·162 acres. This minimised the differences which might arise from variations in the natural fertility of the soil, and from the position of the plots in the field. The smallness of the area used, and the uniform treatment of the land prior to the experiment, secured initial homogeneous conditions for all the rows as far as this was possible.

The total area was divided into two equal parts, one of which was used for the farmyard manure series, and the other for the series without farmyard manure. Each half was divided into 36 small plots and the twelve varieties of potatoes selected for the experiment were planted in triplicate on the "chess-board" system on both series. The accompanying plans show how the varieties were distributed over the experimental area in both the dunged and undunged series.

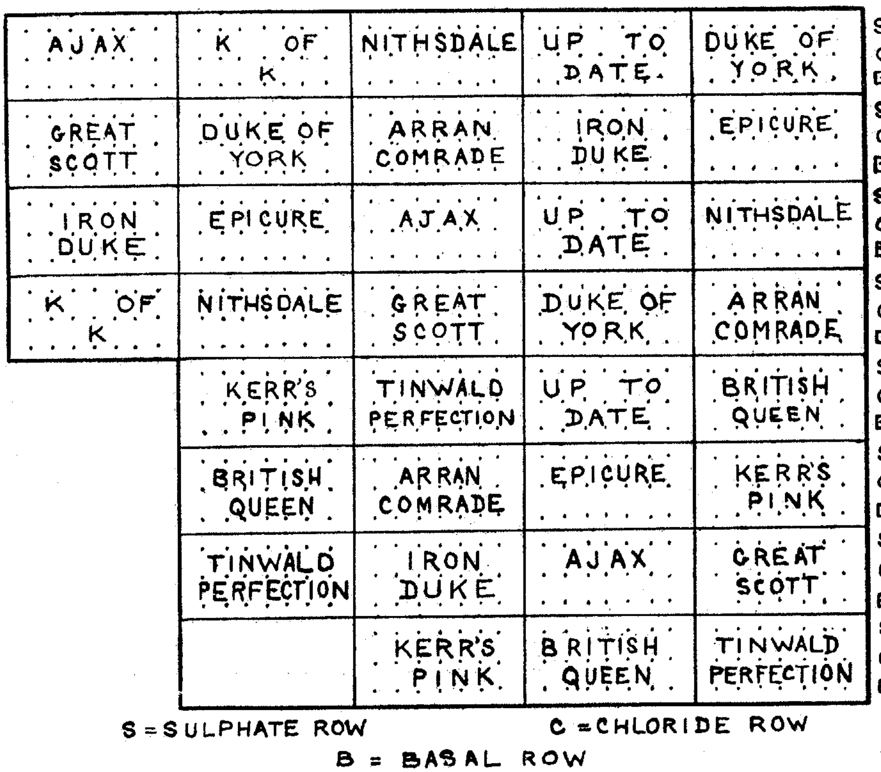

Diagram 2. Plan of experiment. Series without farmyard manure.

It will be noticed from Diagram 2 that, in the series without farmyard manure, the variety K. of K. was only planted in duplicate instead of triplicate and allowance has had to be made for this irregularity in the data in applying the Statistical Method explained below.

Three rows of seven plants each were set on each plot: each row received the basal manuring of the series to which it belonged; and in addition, one row received a dressing of sulphate of potash, and another a dressing of muriate of potash.

The actual quantities of manure applied to the different types of rows are shown in the following table.

Table I.

	Farmyard manure tons per acre	Super. cwts. per acre	Sulphate of ammonia cwts. per acre	Sulphate of potash lbs. per acre	Muriate of potash lbs. per acre
Dunged Series:					
Sulphate row	15	4	1·5	184	—
Chloride row	15	4	1·5	—	147
Basal row	15	4	1·5	—	—
Undunged series:					
Sulphate row	—	6	2	244	—
Chloride row	—	6	2	—	197
Basal row	—	6	2	—	—

The actual weight of produce lifted from each of the 213 rows was recorded, and may be found in the *Report* of the Rothamsted Experimental Station for 1921 and 1922. From this data, two series of means were calculated: one giving the mean yield of each of the twelve varieties, irrespective of the manuring applied; and the other giving the mean yield of each system of manurial treatment irrespective of the variety grown. Table II shows the mean yield for each combination of manure and variety, together with the mean yield of each variety and each manurial system.

Table II.

Manurial treatment	Up to Date	K of K	Kerr's Pink	Tinwald Perfection	Iron Duke	Great Scott	Ajax	British Queen	Nithsdale	Epicure	Arran Comrade	Duke of York	Mean yield
Dunged Series:	lbs.	lbs.	lbs.	lbs.	lbs.	lbs.	lbs.	lbs.	lbs.	lbs.	lbs.	lbs.	lbs.
Sulphate row	25·3	28·0	23·3	20·0	22·9	20·8	22·3	21·9	18·3	14·7	13·8	10·0	20·12
Chloride row	26·0	27·0	24·4	19·0	20·6	24·4	16·8	20·9	20·3	15·6	11·0	11·8	19·82
Basal row	26·5	23·8	14·2	20·0	20·1	21·8	21·7	20·6	16·0	14·3	11·1	13·3	18·62
Undunged Series:													
Sulphate row	23·0	[20·4]	18·2	20·2	15·8	15·8	12·7	12·8	11·8	12·5	12·5	8·2	15·33
Chloride row	18·5	[17·0]	20·8	18·1	17·5	14·4	19·6	13·7	13·0	12·0	12·7	8·3	15·47
Basal row	9·5	[6·5]	4·9	7·7	4·4	2·3	4·2	6·6	1·6	2·2	2·2	1·6	4·47
Mean yield	21·47	20·45	17·63	17·49	16·89	16·58	16·21	16·08	13·49	11·88	10·54	8·86	15·635

The figures enclosed within the square brackets are based upon two records only, but in calculating the marginal means, they have been given full weight. This eliminated the bias in the means concerned, due to the fact that the three missing quantities belonged to the series of one of the *highest* yielding varieties, and to the three *lowest* yielding manurial treatments.

3. Analysis of Variation on the Sum Basis.

In order to make clear the method of the analysis of variation, we will first carry through the process on the assumption that the yield to be expected from a given variety grown with a given manurial treatment is the sum of two quantities, one depending on the variety and the other on the manure. This assumption is evidently an unsatisfactory one, but for a clear understanding of the method it is desirable to separate the difficulties of a more adequate theory, from the method adopted to analyse the variation.

The theory of this analysis depends on a well-known mathematical identity. If ξ, η, ζ stand for the yields of three parallel plots, x for the mean yield of the three, and $\bar{x}$ the mean value of all the plots, then it is well known that

$$S\{(\xi-\bar{x})^2+(\eta-\bar{x})^2+(\zeta-\bar{x})^2\}$$
$$=S\{(\xi-x)^2+(\eta-x)^2+(\zeta-x)^2\}+3S(x-\bar{x})^2.$$

Thus the sum of all the squares of deviations from the general mean may be divided up into two parts: one measures the variation within the triplicates, while the other measures the variation between triplicates differently treated. Further, if all the plots are undifferentiated, as if the numbers had been mixed up and written down in random order, the average value of each of the two parts is proportional to the number of degrees of freedom in the variation of which it is compared. Had the record been complete, 216 plots would have given 215 degrees of freedom; each triplicate contains two degrees of freedom in which its numbers may mutually differ, giving 144 for variation within the triplicates; while the 72 different combinations of variety and manure account for the remaining 71 degrees of freedom. Actually we have 69 triplicates and three duplicates, so that the total degrees of freedom are 212, of which 141 are within, and 71 between triplicate sets.

This analysis may be carried further if we wish to distinguish the effects of variety and manure. There are 5 independent differences among the manures, 11 among the varieties, while the remaining 55 show the differences between the varieties in their response to manurial treatment. To each of these classes can be assigned a definite fraction of the sum of the squares of deviations; thus, if $a_1, a_2, \ldots, a_{12}$ are the mean yields of the 12 varieties, $b_1, b_2, \ldots, b_6$ the mean yields of the 6 manures,

$$3S(x-\bar{x})^2=18S(a-\bar{x})^2+36S(b-\bar{x})^2+3S(x-a-b+\bar{x})^2.$$

In Table III is shown the analysis of the variation into these four classes; the mean square deviation is found by dividing the sum of squares in each class by the number of degrees of freedom, while the standard deviation is shown in the last column. When this value is significantly greater than the standard deviation of the differences between parallel plots, we may conclude that the corresponding effect is not due to chance.

Table III.

Variation due to	Degrees of freedom	Sum of squares	Mean square	Standard deviation
Manuring	5	6,158	1231·6	35·09
Variety	11	2,843	258·5	16·07
Deviations from summation formula	55	981	17·84	4·22
Variation between parallel plots ...	141	1,758	12·47	3·53
Total	212	11,740	—	—

In comparing the standard deviations in the last column we may use the fact that 3·53, for example, has the same accuracy as if it had been determined from a sample of 142; the variance of its natural logarithm is therefore $\frac{1}{2 \times 141}$. Thus, to test if the deviations from the summation formula are significantly greater than would occur by chance, we compare the difference of the logarithms with its standard error, namely $\sqrt{\frac{1}{282} + \frac{1}{110}}$:

$\frac{1}{282}$ =	·003546	$\log_e 4{\cdot}22$ =	1·4398
$\frac{1}{110}$ =	·009091	$\log_e 3{\cdot}53$ =	1·2613
Sum	·012637	Difference	·1785 ± ·1124
Standard Error	·1124		

The difference in the logarithm thus exceeds its standard error, but not sufficiently to be significant; while the effects both of manuring and of variety are very strongly marked, and clearly significant.

4. Trial of Product Formula.

The above test is only given as an illustration of the method; the summation formula for combining the effects of variety and manurial treatment is evidently quite unsuitable for the purpose. No one would expect to obtain from a low yielding variety the same actual increase in yield which a high yielding variety would give; the falsity of such an assumption is emphasised by the fact that the expected values $(a + b - \bar{x})$ calculated on such an assumption, are often negative in the unmanured series. A far more natural assumption is that the yield should be the product of two factors, one depending on the variety and the other on the manure. As a preliminary test of this assumption we may take the formula $\frac{ab}{\bar{x}}$; the computed yields from each combination of variety and manure are shown in Table IV.

Table IV.

	a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}
b_1	27·6	26·3	22·7	22·5	21·7	21·3	20·9	20·7	17·4	15·3	13·6	11·4
b_2	27·2	25·9	22·3	22·2	21·4	21·0	20·5	20·4	17·1	15·1	13·4	11·2
b_3	25·6	24·4	21·0	20·8	20·1	19·7	19·3	19·1	16·1	14·1	12·5	10·6
b_4	21·0	20·0	17·3	17·2	16·5	16·3	15·9	15·8	13·2	11·6	10·3	8·7
b_5	21·2	20·2	17·4	17·3	16·7	16·4	16·0	15·9	13·4	11·8	10·4	8·8
b_6	6·1	5·9	5·0	5·0	4·8	4·7	4·6	4·6	3·9	3·4	3·0	2·5

The computed values are now more variable than before, but agree more closely with the observations. Since the formula $\frac{ab}{\bar{x}}$, where a and b are the variety and manurial means, is not the best fitting product formula, this circumstance speaks strongly in favour of the product as against the summation formula. If we now make out a table showing the subdivision of the variation, it will be seen that the total no longer tallies with the sum of the squares of the 213 deviations from the general mean, thus indicating that we are not using a least square solution, and that the fit can be improved.

Table V.

Variation of	Degrees of freedom	Sum of squares	Mean square	Standard deviation
Computed values (manures and varieties)	16	9,348	—	—
Deviations from product formula ...	55	873	15·87	3·98
Parallel plots	141	1,758	12·47	3·53
Total	212	11,979	—	—

Table V shows that with the product formula there is even less indication of varietal differences in manurial response. In the present material evidently the varieties show no difference in their reaction to different manurial conditions. The method, however, is at this stage unsatisfactory, for, in dividing up the variation according to the different groups of degrees of freedom, we assume that adequate methods of fitting have been employed. In the present case an improved fit can only strengthen our conclusion, but it is necessary to indicate how an adequate fit can be obtained when the conclusions are less clear.

5. Theory of Fitting a Product Formula.

Given any set of positive values $x_{pp'}$, where p takes integral values from 1 to n, and p' from 1 to n', we require to find values $a_1, a_2, \ldots, a_n$, $b_1, b_2, \ldots, b_{n'}$, such that

$$\overset{n}{\underset{1}{S}}\ \overset{n'}{\underset{1}{S}}\ (x_{pp'} - a_p b_p)^2$$

is a minimum. Differentiating with respect to a_p, we have

$$\overset{n'}{\underset{1}{S}}\ (b_{p'} x_{pp'}) = a_p \overset{n}{\underset{1}{S}}\ (b_{p'}^2) \quad n \text{ equations} \qquad \ldots\ldots(\mathrm{I}),$$

$$\overset{n}{\underset{1}{S}}\ (a_p x_{pp'}) = b_{p'} \overset{n'}{\underset{1}{S}}\ (a_p^2) \quad n' \text{ equations} \qquad \ldots\ldots(\mathrm{II}).$$

From either of these sets of equations we can deduce

$$\overset{n}{\underset{1}{S}}\ \overset{n'}{\underset{1}{S}}\ (a_p b_{p'} x_{pp'}) = \overset{n}{\underset{1}{S}}\ (a_p^2) \,.\, \overset{n'}{\underset{1}{S}}\ (b_{p'}^2) = \lambda,$$

showing that we have only $n + n' - 1$ independent equations; as is obvious if we consider that in any solution we may multiply all the values of a, and divide all the values of b by any factor, without affecting the products.

Knowing λ, the minimum value sought may be directly calculated, for

$$\overset{n}{\underset{1}{S}}\ \overset{n'}{\underset{1}{S}}\ (x_{pp'} - a_p b_{p'})^2 = \overset{n}{\underset{1}{S}}\ \overset{n'}{\underset{1}{S}}\ (x_{pp'}^2) - \lambda \qquad \ldots\ldots(\mathrm{III}).$$

At this point two courses are open: (i) we may obtain an approximate solution from equations (I) and (II); by inserting in equation (I) the values of b previously used, we obtain a new series a_p, using in fact approximately the right weights, instead of equal weights as in the previous calculation; then from this new series, a_p, we recalculate b_p by equation (II); (ii) we may eliminate a_p and b_p, and obtain an equation for λ. This latter process is exceedingly laborious and we shall only use it so far as to show that an adequate approximation is obtained by the first method.

By the first method the following values of a_p, b_p, and their products were obtained:

Table VI.

		a_1	a_2	a_3	a_4	a_5	a_6	a_7	a_8	a_9	a_{10}	a_{11}	a_{12}
		20·98	20·48	17·55	16·94	16·94	17·02	16·24	15·99	13·91	12·01	10·50	9·01
b_1	1·290	27·1	26·4	22·6	21·9	21·9	22·0	20·9	20·6	17·9	15·5	13·5	11·6
b_2	1·270	26·6	26·0	22·3	21·5	21·5	21·6	20·6	20·3	17·7	15·3	13·3	11·4
b_3	1·187	24·9	24·3	20·8	20·1	20·1	20·2	19·3	19·0	16·5	14·3	12·5	10·7
b_4	·983	20·6	20·1	17·3	16·7	16·7	16·7	16·0	15·7	13·7	11·8	10·3	8·9
b_5	·981	20·6	20·1	17·2	16·6	16·6	16·7	15·9	15·7	13·6	11·8	10·3	8·8
b_6	·300	6·3	6·1	5·3	5·1	5·1	5·1	4·9	4·8	4·2	3·6	3·2	2·7

The variation of these computed values, and of the deviations of the observed values from them are shown in the following table.

Table VII.

Variation of	Degrees of freedom	Sum of squares	Mean square	Standard deviation
Computed values (manures and varieties)	16	8,966	—	—
Correction for mean	—	169	—	—
Deviations from product formula ...	55	847	15·40	3·92
Parallel plots	141	1,758	12·47	3·53
Total	212	11,740	—	—

The deviations from the product formula are diminished; the solution is certainly very near the best possible, since the total of the sums of squares now agrees with the true value. The mean of the computed values exceeds that of the observed value by ·025, and the correction inserted above is 72 times the difference of the squares of these two values. In order further to demonstrate that the solution obtained cannot be sensibly improved it will be necessary to consider the exact method of solution.

If we use equation (I) to eliminate the a_p from equation (II), there result equations of the form

$$\overset{n'}{\underset{1}{S}}\{b_{p'}\overset{n}{\underset{1}{S}}(x_{p1}x_{pp'})\} = \lambda b_1,$$

or, writing

$$c_{rs} = \overset{n}{\underset{1}{S}}(x_{pr}x_{ps}),$$

the equations are

$$\begin{gathered}(c_{11}-\lambda)\,b_1 + c_{12}b_2 + \dots + c_{16}b_6 = 0,\\ c_{12}b_1 + (c_{22}-\lambda)\,b_2 + \dots + c_{26}b_6 = 0,\\ \dots\dots\dots\dots\dots\dots\dots\dots\\ c_{16}b_1 + c_{26}b_2 + \dots + (c_{66}-\lambda)\,b_6 = 0,\end{gathered}$$

which, by elimination of the $b_{p'}$, give as an equation for λ

$$\begin{vmatrix} (c_{11}-\lambda) & c_{12} & c_{13} & \dots & c_{16}\\ c_{12} & (c_{22}-\lambda) & c_{23} & \dots & c_{26}\\ \dots & \dots & \dots & \dots & \dots\\ c_{16} & c_{26} & c_{36} & \dots & (c_{66}-\lambda)\end{vmatrix} = 0,$$

an equation of the 6th degree, which may be expanded in the form

$$\lambda^6 - S(c_{pp})\,\lambda^5 + S\left\{\begin{vmatrix} c_{pp} & c_{pq}\\ c_{pq} & c_{qq}\end{vmatrix}\right\}\lambda^4 - S\left\{\begin{vmatrix} c_{pp} & c_{pq} & c_{pr}\\ c_{pq} & c_{qq} & c_{qr}\\ c_{pr} & c_{qr} & c_{rr}\end{vmatrix}\right\}\lambda^3 + \dots.$$

It is fortunately not necessary to evaluate the whole of this equation, since the terms are of diminishing importance. Since

$$S(c_{pp}) = \overset{n}{\underset{1}{S}}\ \overset{n'}{\underset{1}{S}}(x^2_{pp'}),$$

and λ is slightly less than this quantity, we may obtain a succession of approximations for λ with alternately positive and negative errors, by taking account of the successive terms of the equation. The first terms are

$$\lambda^6 - 20{,}928{\cdot}01\lambda^5 + 5{,}853{,}407{\cdot}6454\lambda^4 - 597{,}335{,}955\lambda^3 + \ldots,$$

from the first two terms we have

$$\lambda = 20{,}928{\cdot}01 \qquad 20{,}928{\cdot}01 - \lambda = 0;$$

the error being positive; taking in the 3rd term we find

$$\lambda = 20{,}644{\cdot}48 \qquad 20{,}928{\cdot}01 - \lambda = 283{\cdot}53;$$

with a negative error; with the 4th term

$$\lambda = 20{,}645{\cdot}90 \qquad 20{,}928{\cdot}01 - \lambda = 282{\cdot}11;$$

showing how rapidly an approximation is attained. Since the error of this last value is positive, the sum of the squares of the deviations cannot be less than 282·11, or the value in Table VII less than 846.

Since the value obtained from the approximate method was 847, this must be substantially the true value. The true subdivision of the variation must therefore be that given in Table VII.

Testing the significance of the deviation

log 3·92	1·3661
log 3·53	1·2163
	·1048 ± ·1124.

Conclusions.

(1) The data show clearly significant variation in yield due to variety and to manurial treatment.

(2) There is no significant variation in the response of different varieties to manure.

(3) The yields of different varieties under different manurial treatment are better fitted by a product formula than by a sum formula.

(4) For the purposes of analysing the variation the product formula may be obtained by successive approximation from equations (I) and (II); the exact method of solution should not be necessary in any ordinary case.

(*Received March 20th*, 1923.)