NO. 11—THE USE OF THE METHOD OF PAIRED DIFFERENCES FOR ESTIMATING THE SIGNIFICANCE OF FIELD TRIALS.

BY

P. C. MAHALANOBIS,

Statistical Laboratory, Presidency College, Calcutta.

Very recently we received a paper for criticism in which the author used the method of paired differences for estimating the significance of field trials. The use of this method is justified only in special cases, and a discussion of the principles involved may prove useful to field workers.

Consider a hypothetical experimental field which is completely free from systematic variation in fertility. Suppose we have "$2n$" plots of which "n" are sown with each of the treatments (or variates) A and B. Let x_1, x_2, x_3............ x_n be the yield of the "n" plots under A, and y_1, y_2, y_3.............y_n for the "n" plots under B. Let $d_1=(x_1-y_1)$, $d_2=(x_2-y_2)$...............etc., be the difference of yield for paired plots of A and B. The observed mean value of the difference $\bar{d}=(\bar{x}-\bar{y})$, where $\bar{x}$ and $\bar{y}$ are the mean yields of A and B. The observed value $\bar{d}$ is clearly equal to the real difference in yield between A and B *plus* the experimental errors, since by hypothesis there are no systematic differences in fertility between different plots. We may, therefore, proceed to calculate in the usual way s_d, the variance of the difference, and use it to judge whether $\bar{d}$ is significantly different from zero or not. Fisher's t-test with

$$t=\frac{\bar{d}\sqrt{n}}{s_d} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

and $(n-1)$ degrees of freedom can then be applied in the usual way.

It is clear that in this method we can pair x and y in any way we like. For example, we can take $d_1=x_1-y_4$, or $d_2=x_{12}-y_5$, etc. The mean value of d and the standard deviation s_d would not be sensibly affected by the method of pairing (since x and y are supposed to be quite independent), and hence the value of "t" will remain the same.

2. The conditions under which the method of analysis of variance is used are entirely different. Systematic changes in fertility cannot be assumed to be absent (and, in fact, are usually known to be present), and hence the differences in yield between any two plots will definitely include the effect of differences in soil fertility.

In this case we first try to allow for the effects of soil heterogeneity by the elimination of the variance " between blocks", and then judge the significance of the effect of treatments (or varieties) in the usual way with the help of Fisher's z-test.

3. Where systematic differences in soil fertility are present (or cannot be assumed to be absent) it is clear that we cannot legitimately use the method of paired differences. Soil heterogeneity will almost never be entirely absent, and hence the necessary conditions for the use of paired differences will never be strictly fulfilled in practice. But under certain circumstances we may reasonably assume the effect of extraneous factors to be either absent or to be appreciably constant in magnitude. For example, consider pairs of adjacent plots sown with two different varieties. In ordinary circumstance (that is, unless changes in fertility are very sharp) we may. as a first approximation. assume that the soil fertility will remain nearly the same for each pair of adjacent plots. Hence if we take the difference in yields from two adjacent plots, we may be reasonably certain that the effect due to the soil factors will be eliminated in the process of differencing. Under these circumstances, that is, when the differences in yield refer to adjacent plots. the use of paired differences may possibly be justified.

But the method cannot obviously be extended to the case of plots which are not adjacent. Consider x_1 the yield of A from Plot No. 1, and y_{16} the yield of B from Plot No. 16 which is situated at a considerable distance from Plot No. 1. The difference in yield ($d=x_1-y_{16}$) must obviously include not only the effect due to the varietal difference between A and B and the residual errors, but also the effect due to differences in soil fertility from one part of the field to another. The use of the method of paired differences in these circumstances has, therefore, absolutely no justification.

4. A numerical example may make the position clear. In a manurial experiment on wheat, two treatments A and B were laid out in a Randomized Block in 8 replications, and the yields obtained (in lbs. per 1/40th acre) were as shown in Table I. The serial number of the plot is shown against each plot.

TABLE I.

1	2	3	4	5	6	7	8
A 47·0	B 68·5	B 62·0	A 55·0	A 62·5	B 73·5	A 56·0	B 66·0
B 63·5	B 61·0	A 44·5	B 57·0	A 66·0	A 52·0	B 34·0	A 38·0
9	10	11	12	13	14	15	16

Using the usual method of analysis we obtain the mean value of yield under A=52·64, and the mean value of the yield under B=60·69, so that the difference in mean yield is 8·05. The standard error of the difference is 5·23, so that t=1·54. From Fisher's Table IV we find that with n=7, the probability of occurrence lies between 0·1 and 0·2. So that even if the difference between A and B is *nil*, the observed difference would occur in from 10 per cent. to 20 per cent. of cases. The difference between the two treatments cannot therefore be considered significant.

Now let us try to use the method of paired difference. We choose the pairs as shown in Table II.

TABLE II.

Treatment B		Treatment A		Difference in yield
Plot No.	Yield	Plot No.	Yield	
6	73·5	13	66·0	7·5
2	68·5	5	62·5	6·5
8	66·0	7	56·0	10·0
9	63·5	4	55·0	8·5
3	62·0	14	52·0	10·0
10	61·0	1	47·0	14·0
12	57·0	11	44·5	12·5
15	34·0	16	38·0	—4·0

Mean difference = 8·05

Standard error of difference = 2·018

$$\therefore t = \frac{8 \cdot 05}{2 \cdot 18}$$

$$= 3 \cdot 994$$

$$P < 0 \cdot 01$$

The probability of occurrence is now less than 0·01, so that on the results of the present mode of analysis the difference between treatments A and B would be considered to be definitely established. But there is no justification for choosing the pairs in this special way, and hence the estimate of the standard error is wholly invalid.

In fact the value of the standard error depends entirely on the particular manner in which the plots are paired. Consider for example the following system of pairing (Table III).

TABLE III.

Treatment B		Treatment A		Difference in yield
Plot No.	Yield	Plot No.	Yield	
6	73·5	16	38·0	35·5
2	68·5	11	44·5	24·0
8	66·0	1	47·0	19·0
9	63·5	14	55·0	8·5
3	62·0	4	52·0	10·0
10	61·0	7	56·0	5·0
12	57·0	5	62·5	–5·5
15	34·0	13	66·0	–32·0

Mean difference = 8·05

Standard error of difference = 7·23

$$\therefore t = \frac{8 \cdot 05}{7 \cdot 23}$$

$$= 1 \cdot 11$$

$$P > 0 \cdot 3$$

The value of t is now only 1·11, and the probability of occurrence is greater than 0·3. On the result of this analysis the effect of both treatments would appear to be statistically indistinguishable.

We can pair the plots in a large number of different ways, and each particular way of pairing would lead to a different value of the "standard error", and hence to a different value of t. But all such values are equally invalid, and no legitimate inference can be drawn from any arbitrarily-paired system of differences.

REFERENCES.

Fisher, R. A. (1930). Statistical Methods for Research Workers, p. 139.

Pearson, Karl (1930). Tables for Statisticians and Biometricians, Part I, 26.