XIII. The Technique of Field Experiments

Foreword

Sir E. J. Russell

FOREWORD

By Sir E. JOHN RUSSELL

Director of the Rothamsted Experimental Station, Harpenden

THE papers presented herewith were read at a Conference at Rothamsted on 7th May, when the technical experts concerned in the carrying out of field experiments assembled to describe the methods they are actually adopting.

The old method, involving the use of large single plots placed side by side, is simple and effective for the purpose of demonstrating known facts so long as the differences to be observed are large and do not require to be expressed numerically. But it breaks down as soon as accurate measurements are needed, because it takes no account of variations in the soil, which are now known to be considerable.

In the classical fields at Rothamsted Lawes and Gilbert had two unmanured plots, one at each end of the series : the yields from these usually differed by about 10 per cent. The differences they were demonstrating were, however, much larger, so that they felt safe in attributing their results to the treatments. In 1846-47 they split the Broadbalk plots lengthwise into two halves, which from that time onwards were harvested separately : this was the first duplication of field plots of which we can find any record. In 1847-48 and occasionally afterwards one half of each plot was treated differently from the other, so that they ceased to be strict duplicates.

Better duplication appears to have been practised by P. Nielsen, the founder of the Danish experimental stations about 1870, in his experiments on grass mixtures for pastures. Some of the Norfolk chamber of agricultural experiments carried out in the later 1880's were systematically repeated thus : A B C D D C B A.

Nothing more was done in this country till 1909. In that year A. D. Hall, and somewhat later T. B. Wood, both urged the need for estimating experimental errors in field work and gave approximate methods for doing this. Somewhat later Dr. E. S. Beaven of Warminster designed his well-known strip method of replication

5

specially suitable for variety trials and adopted by the National Institute of Agricultural Botany ; it is described by Mr. S. F. Armstrong on page 30.

The subject was taken up seriously at Rothamsted in 1919 and has been much developed since. In 1919 Dr. R. A. Fisher applied to the study of variation an arithmetical analysis known as the analysis of variance, which had the advantage over the ordinary calculus of correlations of avoiding both the calculation of a large number of irrelevant values and also the numerous corrections to which correlations are liable, especially with small samples. He applied the method to the Broadbalk wheat yields and showed its value for measuring the effect of distinct groups of causes. This investigation, however, showed the need for more exact methods for treating the small number of cases, or samples, generally available in agricultural investigations. The first example of an analysis of variance in its modern form was the examination of the results of T. Eden's experiment in 1922 on the response of different potato varieties to manures (Fisher and Mackenzie, *Journ. Agric. Sci.*, 1923). Somewhat later, "Student" gave alternative proofs by himself and by Fisher of formulæ appropriate to cereal variety experiments. Thus rigorous methods of statistical examination were elaborated.

The next step was to develop a correspondingly rigorous field technique, and this was done by Dr. Fisher in co-operation with T. Eden and E. J. Maskell. The chief difficulty was to overcome the effects of the irregularities in the soil which had long been a serious stumbling-block to field experimenters.

Part of the irregularity or heterogeneity could be eliminated by suitable arrangements of the plots, but there was always an unknown remainder of residual errors. It was shown that the statistical analysis previously developed could eliminate the effects of soil irregularities and at the same time afford a valid estimate of the remaining errors, provided that the plots were sufficiently replicated and deliberately randomised.

Dr. Fisher then devised various types of experiments to meet the requirements of the statistical analysis, and tested these on the results of uniformity trials so as to discover which were the most accurate and convenient in actual working. Two types stood out as satisfactory ; randomised blocks and the Latin square. The randomised block is the simpler and the more easily adjusted to suit the pecularities of the field and the crop. The experimental area is divided into several strips or blocks, each of which contains one plot of each treatment, the arrangement being deliberately at random and determined not by selection, but by writing the possible arrangements on separate cards, shuffling them, and drawing one out. Since

one block is not directly compared with another, the differences in soil fertility between them are eliminated ; and since the arrangement within the blocks has been entirely at random, the significance of the results can be estimated.

The Latin square is the more accurate but less widely applicable in fertiliser experiments. The plots are arranged with as many rows and columns as there are treatments. Each treatment appears once, and only once, in each row and each column. A surprisingly large number of arrangements are possible, but the selection is again deliberately at random and, as before, is effected by the shuffling and drawing of cards. From the figures for yield, a standard error is worked out which shows the degree of trustworthiness of the result. A difference in yield equal to the standard error of this difference can be obtained about once in three trials, even when the experimenter is convinced that he has given exactly the same manuring and cultivation to each of the plots, but a difference twice this size would be obtained by chance only once in twenty-two times : it is therefore much more likely to be true. The chances against the difference in yield being due to causes other than the difference in treatment are :—

For difference equal to its standard error .	3 to 1
For difference double its standard error .	22 to 1
For difference three times its standard error .	370 to 1
For difference four times its standard error .	15,780 to 1

For most agricultural purposes a chance of about 30 to 1 is good enough. The " standard errors " for the yield values have to be multiplied by 1·414 (*i.e.* $\sqrt{2}$) in order to give the standard error of the difference between treated and untreated plots—the figure one usually wants. To attain a probability of 30 to 1, a difference must be roughly three times the standard error of the yield.

Dr. Wishart (p. 15) shows how the results are to be worked out. Our experience proves that the methods are quite practicable not only on the fields of the Experimental Station but also on those of ordinary farms : Mr. Garner (p. 49) gives particulars of the methods, and Messrs. Lewis, Manson and Proctor (p. 37) show how to extend them to cover a series of trials. Large numbers of these experiments are now made, the numbers of plots in each ranging from 16 to 144. Usual standard errors per plot on our present methods of good working are :—

USUAL STANDARD ERRORS PER PLOT FOR GOOD WORKING

	Randomised Block.	Weight per Acre Latin Square	Per Cent. of Yield Latin Square	Random-ised Block
Potatoes	0·7 tons	0·4 tons	4·4	8·4
Sugar Beet : Roots	0·8 ,,	0·6 ,,	5·7	10·2
Tops		0·7 ,,	5·7	10·9
Barley : Grain	1·5 cwt.	1·3 cwt.	5·6	9·1
Straw	1·9 ,,	1·9 ,,	6·0	7·2
Oats : Grain		2 ,,	8	
Straw		2 ,,	6	

The standard error precisely measures the accuracy of the experiment and it includes errors of working, inequalities due to variable natural agencies, such as weather, birds, insects, diseases, and also soil variations within the individual plots, but not the large variations between plot and plot, which are eliminated by the method of arranging the experiment. It is not, however, an absolute measure, since it depends to some extent on the size and arrangement of the plots. Thus a standard error of 0·4 tons per acre of potatoes in a Latin square experiment is not strictly comparable with a standard error of 0·4 tons in a randomised block experiment having more plots. Nevertheless, it is a useful guide to the experimenter, as showing the standard of performance he is attaining in his work. The standard error is much the same whether the crop is large or small, so that a heavy crop has a lower percentage error than a light one.

There are several plots of each treatment, and the standard error of the final result is much less than the figures of errors per plot ; it is usually now at Rothamsted about 2 to 4 per cent. of the mean yield.

In addition detailed observations are made on various growth factors : these are discussed by Mr. Watson on page 54.

These principles have been applied to horticultural experiments by Mr. Hoblyn, who deals with the problem on page 42.