REPLICATION OF FIELD TRIALS IN SPACE AND TIME.

D.J. Saville,
Agricultural Research Division
Ministry of Agriculture & Fisheries,
Lincoln.

ABSTRACT

The design and analysis of a series of field trials is discussed. It is recommended that field trials should be spread over as many years as possible.

For the analysis, it is suggested that the quantity (or quantities) of interest should be calculated for each year. The mean and standard error of these yearly values can then be computed and a t-test performed. Alternatively, these yearly values can be correlated to other variables.

INTRODUCTION

In agricultural experimentation, statistical analyses are commonly performed on the data from a single experiment. This means that conclusions in scientific papers are often tentative, with the author stressing the fact that the results apply to a particular paddock in a particular year.

Such reservations about the conclusions do not necessarily detract from the value of the experimental results. For example the first experiment in a new field of study will often warrant publication as a case study to give the lead to other researchers who may be interested in working in that field.

This situation aside, the bulk of experimental work is aimed at ascertaining whether or not one treatment is consistently superior to another treatment. For example, does a new wheat cultivar average out to be higher yielding than the standard cultivar, over several seasons and locations? As a second example, what is the average response of lupins to a single irrigation at time of flowering, over several seasons and locations?

In this paper recommendations will be made on the design of this type of experimental work. An example will then be presented to describe the method of statistical analysis, and to point out the dangers of using incorrect methods.

DESIGN OF A SERIES OF FIELD TRIALS

Population(s) of study:

When considering field trial design, the trial locations, and the years in which the trials are conducted, can be thought of as a random sample from the set of all locations and the set of all years in the population being studied.

Often the population under study can meaningfully be broken into smaller populations which are somewhat different to each other and which are less variable than the entire population. For example, to assess the response to nitrogen applied to second year winter sown wheat in Canterbury, one may do four trials each year, with one trial for each of the populations:

(i) shallow soils in low rainfall districts
(ii) deep soils in low rainfall districts
(iii) shallow soils in high rainfall districts
(iv) deep soils in high rainfall districts

Analysis:

If a single trial is carried out each year at a different place, the combined analysis is simple because the variation between years and between places are both included in the error term. If two or more trials are conducted each year, the data from these trials can be averaged, then used in an analysis combining results between years. Therefore the only advantages in having two or more trials for each population each year are:-

(i) insurance against natural and manmade disasters
(ii) an evening out of some of the variation between places within each year.

If data from sufficiently many years and/or sites is available, the understanding of the results can be improved by correlating the trial data with other information such as climatic variables or paddock history. Secondly, the average of the main variable can be improved using the correlations to adjust to the long term average of the other variables.

Yearly variation:

For certain types of trial, variation from year to year can "a priori" be assumed to be important. For example, irrigation responses can be expected to vary dramatically with weather conditions. For other types of trial, the question can be raised: is yearly variation important? For example, does a quantity such as optimum seeding rate of barley vary from year to year?

The difficulty is in distinguishing work in which yearly variation is important from work in which it is not. The safest approach is to allow for the possibility of yearly variation regardless of prior hunches. If there is no such variation this will be demonstrated by the consistency of the trial results from year to year. Such a demonstration will be much more convincing in a scientific publication than a statement that yearly variation was assumed to be unimportant.

Recognition of work:

A researcher who embarks upon research programmes which are spread in space and time may not still be present when the work is completed, and may not receive sufficient credit for having organised the programme. This fact could militate against the establishment of such programmes.

Recommendations:

That field trials of a single design be carried out for several years at a rate of one or two per year. The number of years will depend on the size of the variation from year to year, but three years would seem to be the absolute minimum.

In general each trial need not be heavily replicated. The ratio of number of replicates in each trial to number of trials, depends on the variation within trials, the variation between trials, and the relative costs of within and between trial replication.

TABLE 1: Average grain yield response to nitrogen (kg/ha) by treatment and by trial.

	Year 1		Year 2		Year 3		
No phosphate:	Farm 1	Farm 2	Farm 1	Farm 2	Farm 1	Farm 2	MEAN
N in July	1240	510	-520	190	870	270	430
Aug	880	450	-120	580	810	640	540
Sept	910	670	-270	410	990	860	600
Oct	890	740	40	270	860	1050	640
Phosphate:							
N in July	830	420	-490	350	1150	370	440
Aug	820	790	-530	730	700	550	510
Sept	790	690	-170	490	1230	1190	710
Oct	1110	810	20	520	810	1250	750
Average:	940	640	-260	450	930	770	

ANALYSIS OF A SERIES OF FIELD TRIALS

Some data from the Timaru district is now presented to allow discussion of the method of statistical analysis of a series of trials. The objective of the work was to determine for the Timaru district the best time of nitrogen application for a second successive winter sown wheat crop, in the presence and absence of superphosphate.

Design:

Trials of identical design were carried out on two farms in the years 1973, 1974 and 1975. Within each farm, trials were on different paddocks each year.

Each trial was laid out in a randomised block design with three replicates. There were eight treatments in a 4 x 2 factorial structure, with factors time of nitrogen application (July, August, September and October) and phosphate (none, 250 kg/ha superphosphate).

Each of the trial plots consisted of three drill strips, with each strip receiving nitrolime at 0, 250 or 500 kg/ha (assigned at random). For each plot the "average response to nitrogen" was calculated by subtracting the yield of the no-nitrogen treatment from the average of the two nitrogen treatments. This quantity was averaged over the three replicates in each trial to produce the data given in Table 1.

TABLE 2: Average grain yield response to nitrogen (kg/ha), by treatment and by year.

	Year 1	Year 2	Year 3
No phosphate:			
N in July	870	-160	570
Aug	660	230	730
Sept	790	70	930
Oct	820	160	960
Phosphate:			
N in July	620	-70	760
Aug	800	100	620
Sept	740	160	1210
Oct	960	270	1030
Average:	790	100	850
(S + 0 − J — A):	180	280	720

Analysis using an incorrect approach:

Treatment means were averaged over the two farms for each year (Table 2). If these values are statistically analysed as three replicates (= years) of a randomised block design, the results are as given in Table 3.

In Table 3, three independent comparisons have been specified for the times of nitrogen factor. These correspond to the contrasts (S + O − J − A), (O − S) and (A − J), where J, A, S and O refer to the main effect means corresponding to July, August, September and October times of application. Notice that the (S + O − J − A) contrast is significantly greater than zero (P = .01), so the September/October time of application is concluded to be better than the July/August time of application.

TABLE 3: Analysis of variance table using years as replicates.

Source of variation	d.f.	Mean square	F ratio
Replicates (=years)	2	1,398,000	
Times of N applicn.	(3)	(88,000)	
(S + O − J − A)	1	232,000	12.2**
(A − J)	1	25,000	1.3
(O − S)	1	7,000	0.4
Phosphate	1	15,000	0.8
P x Times	3	8,000	0.4
Error (= Trts x years)	14	19,000	
Total	23		

The thing which is incorrect about this approach is that the variations from year to year in (S + O − J − A), (O − S), (A − J), P, P × (S + O − J − A), P x (0 - S) and P × (A − J) have all been pooled to produce the error mean square of 19,000. This means that the larger variation of the more substantial quantities (e.g. (S + O − J − A)) is diluted by averaging with the lesser variation of the less substantial quantities (e.g. P). Also the degrees of freedom for error are spuriously high (14 d.f.). The nett result is that, for the larger effects, the significance of the F test is increased beyond what it should be.

Table 4 shows the error term broken into components. As one would expect in general, the largest effect, the (S + O − J - A) contrast, has the largest variation from year to year (42,000).

Analysis using the correct approach:

The significance of an effect is properly tested by comparing its mean square with the mean square for the interaction of the effect with the years. For the (S + O − J − A) contrast, the $F_{1,2}$ value is 232,000/42,000 = 5.5, which is nonsignificant. The resulting conclusion is the opposite of that reached in the last section. It is that the September/October time of application has not been proven to be better than the July/August time of application.

The correctness of this approach can be seen intuitively when the computations are expressed in terms of the t-test. Using a t-test, the method of analysis for any particular quantity is to calculate the quantity for each year and then do a t-test to decide whether the underlying true mean is greater than zero. For example, if we are interested in whether the late application times (September and October) are better than the early times (July and August), we could calculate for each year the quantity (S + O − J − A). The answers from Table 2 are 180, 280 and 720 kg/ha for the years 1973, 1974 and 1975 respectively.

For our t-test we regard these three years as a random sample from the population of all possible years, and test the hypothesis that the quantity (S + 0 − J − A) has a true mean of zero. Our numbers have a mean of 393, a standard error of 288 and a standard error of the mean of 166. Hence t_2 = 393/166 =2.4, which is nonsignificant.

Note that (t_2) = 5.5 is precisely the $F_{1,2}$ statistic of the last section, and gives precisely the same statistical result.

TABLE 4: Table 3 with error term broken down into components and testing done against the appropriate component.

Source of variation	d.f.	Mean square	F ratio
Replicates (=years)	2	1,398,000	
Times of N applicn.	(3)	(88,000)	
(S + O − J − A)	1	232,000	5.5ns
(A − J)	1	25,000	0.9
(O − S)	1	7,000	0.6
Phosphate	1	15,000	2.1
P x Times	3	8,000	0.5
Error (= Trts x years)	(14)	(19,000)	
(S + O − J − A) x yrs	2	42,000	
(A − J) x yrs	2	27,000	
(O − S) x yrs	2	11,000	
P x yrs	2	7,000	
P x Times x yrs	6	15,000	
Total	23		

Correlation with other variables:

The usual reason for pooling mean squares is "to increase the error degrees of freedom", which means "to increase the accuracy of the variance estimate". However, pooling mean squares which are likely to be different will not increase the accuracy of the variance estimate — it will just fool us into thinking that we have! The only ways to increase accuracy are to increase the number of years in which trials are conducted, or to correlate results with other information.

For example, one may expect a greater advantage from delaying the nitrogen application in years of heavy winter rainfall. That is, one may expect a correlation between the quantity (S + 0 − J − A) and the winter rainfall.

In 1973, 1974 and 1975 the winter (June, July, August) rainfalls were 142, 90 and 207 mm respectively at the Timaru meteorological station, corresponding to (S + O − J − A) quantities of 180, 280, and 720. The resulting correlation coefficient of 0.81 was not statistically significant.

This illustrates the point that three years' data is often not sufficient to attempt correlations with other variables.

DISCUSSION

In this paper it is pointed out that the analysis combining data from several years can be carried out for each quantity of interest simply by calculating the quantity for each year, and performing a t-test. Equivalently an F-test can be carried out using the correct component of error (Table 4). If the quantities of interest are chosen to be "orthogonal contrasts" the appropriate mean squares can be obtained using computing packages available at most research centres.

Researchers should be warned against pooling the variabilities of quantities for fear of being misled as to the accuracy of their results. Mistakes of this nature could serve to discredit researchers in the eyes of the farming community. For example a researcher who said a certain quantity fell within the limits 100 to 300 in 19 out of 20 years would become a laughing stock if farmers regularly observed values outside the range 100 to 300. This point has also been made by Cochran and Cox (1957) and Colwell (1978).

Correlation of the quantity of interest with other variables can lead to a better estimation of the long term average of the quantity. However, for this correlation to be carried out data over several years is essential.

The series of trials described in the second half of this paper is the largest series with which the author has come into contact, and was chosen for presentation for this reason. However, the design of this series should not be taken as a model. Rather, the model procedure suggested in the first half of this paper is to lay down one trial each year, on a different farm each year.

REFERENCES

Cochran, W.G., Cox, G.M. 1957. "Experimental Designs", Second Edition. J. Wiley, New York pp. 550-553.

Colwell, J.D. 1978. "Computations for Studies of Soil Fertility and Fertiliser Requirements". Commonwealth Agricultural Bureaux pp. 283-288.

ACKNOWLEDGEMENTS

Many thanks to Mr. C.C.S. McLeod, MAF, Timaru, for the use of his data, and to Miss I. Gravett, Lincoln College, for helpful suggestions. Thanks also to Biometrics and other MAF staff for helpful discussion and to Mr. J.R. Proffitt for technical assistance.