Expl Agric. (2005), volume 41, pp. 109–120 © *2005 Cambridge University Press*
DOI: 10.1017/S0014479704002364 *Printed in the United Kingdom*

THE FACTORIAL FIELD EXPERIMENT

By S. C. PEARCE†

Institute of Mathematics and Statistics, University of Kent, Canterbury, Kent CT2 7NF, United Kingdom

(*Accepted 02 August 2004*)

SUMMARY

Factorial experiments are described and the importance of interactions emphasized with suggestions about how they should be interpreted. A distinction is made between factors introduced only to see if they will provoke an interaction and those actually under study. Split-plot designs receive special attention. Factorial experiments often involve a large number of treatments and ordinary block designs may be ineffective in controlling environmental variation. If factors have few levels, as in exploratory experiments, the usual device for reducing block size is confounding, which is explained along with single-replicate experiments, partial replication and hidden replication. Alternatively, non-orthogonal designs and analysis of data by nearest-neighbour and spatial methods might prove useful. The need for randomization and the role of significance are discussed. It is pointed out that interactions can sometimes be avoided by transformation of the variate.

FACTORS AND INTERACTIONS

In factorial experiments the range of treatments under study is constructed from all possible combinations of two or more sets of treatments, known as factors. For example, there might be six 'treatment-combinations' formed from two fungicides and three times of spraying. Often the intention is to study the interactions between the factors; in this instance, the best time for spraying might depend upon the fungicide used. In another experiment, there might be 15 treatment-combinations formed by using five nitrogenous fertilizers, each at three dosages, and it might emerge that some fertilizers are better not used in large amounts. Such experiments have been in use for a long time. The first exposition of the analysis of variance was illustrated by Fisher (1925), though a thorough statistical examination had to wait until the classic paper of Yates (1935).

The opportunity to assess interactions between factors is not always valued by agronomic experimenters, who sometimes regard them as mathematical abstractions that have little to do with practical issues, but really they lie at the heart of a biological approach. Anyone who declared that a certain treatment would have a stated effect on crop weight irrespective of variety, season, soil conditions or weather would be regarded as impossibly naïve. There are always conditions and they are expressed as interactions.

† Email: pearceclifford@aol.com

https://doi.org/10.1017/S0014479704002364 Published online by Cambridge University Press

Table 1. The derivation of the elements of an interaction given the treatment means. The data come from a glasshouse experiment in which T1, T2 and T3 represent successive times for transplanting into the field and H1, H2, H3 and H4 are rooting hormones. The data represent growth during a fixed time after transplanting and are means from two replicates.

	T1	T2	T3	Means
Data and means				
H1	12.0	15.0	17.5	14.83
H2	18.0	21.5	15.5	18.33
H3	18.0	20.5	19.5	19.33
H4	17.5	19.5	25.0	20.67
Means	16.38	19.13	19.38	18.29
Interaction				
H1	− 0.92	− 0.67	1.58	
H2	1.58	2.33	− 3.92	
H3	0.58	0.33	− 0.92	
H4	− 1.26	− 2.01	3.24	

Form of an interaction

Sometimes the form of the interaction is not obvious. Table 1 sets out data from a greenhouse experiment in which there were two replicates of twelve treatments resulting from two factors (Clarke, 1980). Seedlings were transplanted to the field on three occasions, T1, T2 and T3, after treatment with four rooting hormones, H1, H2, H3 and H4. Data represent growth in a fixed period after transplanting. Both main effects were significant as was the interaction.

The form of the interaction is not, perhaps, obvious but it usually becomes so when its elements are isolated. First, it is necessary to evaluate the main effects. That is done by subtracting the general mean from each of the factor means. For the hormones that gives − 3.46, 0.04, 1,04, 2.38 respectively. If the factor had no effect, each of these figures would be zero apart from experimental error. In any case, they should sum to zero. Significance is judged by finding the probability of such values arising from experimental error alone. The error variance was 4.29, which makes the probability about 1 in 40, so the hormone effect must be judged significant. For the times of transplanting the corresponding values are − 1.91, 0.84 and 1.09. (The sum is not exactly zero on account of rounding errors.) Again, the values must be judged significant, the probability being less than 1 in 100.

If there is no interaction each combination of factors should give a mean equal to:

$$\text{General mean} + \text{the two relevant main effects.}$$

Thus, the mean for the combination of H1 and T1 should be:

$$18.29 - 3.46 - 1.91 = 12.92$$

It is, in fact, 12.00, or 0.92 less than expected. The elements of the interaction are presented in the lower part of Table 1. The probability that they have arisen by chance lies somewhere between 1 in 20 and 1 in 40. They suggest that the relative responses to

https://doi.org/10.1017/S0014479704002364 Published online by Cambridge University Press

the herbicides were different at T3 than at T1 and T2. That conclusion is confirmed by ignoring data from T3 and using only T1 and T2 to perform the calculations again, when it will emerge that all elements of the diminished interaction lie in the range ± 0.38. They could well have arisen by chance and provide no evidence of any interaction.

Some alternatives

There are occasions when interactions are not very meaningful. For example, if the factors represent different amounts of nitrogen and potassium in a fertilizer mixture, there may well be an interaction but the results are better expressed by fitting a response surface and finding its maximum.

Again, when one factor is 'qualitative', e.g. comparison of substances, and the other 'quantitative', interactions can be confusing. For example, there might be two fertilizers, A and B, applied in single and double applications, and an untreated control. There are five treatments, 0, A1, A2, B1 and B2. (If the control is to provide a basis from which the others are to be judged it might be wise to double its replication.) Some writers have recommended estimating the interaction from A1, A2, B1 and B2. There is no need to do so because there must be one. It cannot be supposed that the difference between the two substances will be the same whatever the level of application, however large or small that may be. The best way forward may be to compare the response curves for the two substances.

Classes of factors

Factors are of two kinds. They may represent the subject of enquiry, in which case they can be called 'substantive', or they may have been included only to see if they will cause an interaction if one exists, in which case they are 'provocative'.

For example, in an experiment on the control of a fungal disease, the first factor might be four different concentrations of a fungicide, and the second two varieties, chosen because they have different resistance to the disease under study. The object of the experiment is to find the optimal spraying treatment, so that factor is substantive. The varieties have been included to see if they affect the result, so they are provocative. So far from being under investigation, they were chosen for properties that were well understood.

Relationships between effects

The interaction of two factors, A and B, is written $A \times B$. There is no need to consider their order. If A affects the response to B then B will affect the response to A. Further, if $A \times B$ depends upon the level of C, there is a three-factor interaction, $A \times B \times C$, and the response to C will depend upon the particular combination of levels of A and B. The same can be said of $A \times C$ and B and of $B \times C$ and A. All these statements are equivalent. This progression to higher order interactions can be extended indefinitely.

https://doi.org/10.1017/S0014479704002364 Published online by Cambridge University Press

Hidden replication

To return to the design with four levels of a spray substance in conjunction with two varieties; if the data analysis indicates an interaction it would be necessary to report eight means for each quantity studied. However, if there were no evidence of interaction, it would be possible to merge the results over varieties and present only four. They would have a smaller standard error on account of the doubled replication. In designing the experiment, however, it would be unwise to rely on the 'hidden replication', as the pioneers called it, because an interaction is possible, though not expected. Suppose, however, that there was a third factor, would anyone expect a three-factor interaction? If not, it would safe to use hidden replication in the estimation of the two-factor interactions. The cautious might object, but where there are four or more factors most experimenters would be prepared to ignore the higher order interactions and use them to increase the effective replication. They could also include the sums of squares attributed to these high order interactions in the estimation of experimental error.

The interpretation of three-factor interactions

With high order interactions, the estimation of, for example, sums of squares is a matter of mathematics and presents few problems. The difficulty lies in their interpretation, i.e. their biological implications.

Three-factor interactions do arise, though usually with one or more of their associated two-factor interactions. If there is only one, e.g. if A × B exists as well as A × B × C but not A × C or B × C, the first step is to explore the nature of A × B and then to consider the combination of A and B as a single factor and to enquire as to its interaction with C.

The situation is more difficult if A × C exists also. One approach is to note that A has given rise to a significant interaction whenever it occurs, i.e. with B, C and B × C, so it has brought about a radical change in conditions. The analysis of variance can then be calculated anew. The main effect of A can remain, but it has become necessary to calculate the effects of B, C and B × C for each level of A separately. With a design in randomized complete blocks, this is quite easy. One method is to analyse all data at the first level of A in isolation, then those at the second and so on. The sum of squares and degrees of freedom can be transferred to the original analysis and the rest can be discarded. It will be found that both the degrees of freedom and the sums of squares have the same totals as in the original analysis.

Finally, there is the case when all three of the associated two-factor interactions are significant. This is indeed difficult and admits of no general solution. The best course in writing a report is simply to present the means for all the treatment-combinations and to draw attention to anything that the author can discern. Some readers may be more percipient.

Split-plot designs

Some combinations of factors raise difficulties in the field. Some, such as spraying, are more easily applied to compact areas, while others, like sowing, are more easily

https://doi.org/10.1017/S0014479704002364 Published online by Cambridge University Press

carried out on strips. Some are difficult to apply to small plots but there is no problem with others. Indeed, some, such as depth of ploughing, may require large plots to be meaningful. Also, interpollinating varieties have to be dispersed evenly. To minimize such problems, recourse is sometimes made to designs with split-plots. Blocks are formed and divided into 'main plots' for the application of factors suited to such dimensions. Then the main plots are divided into 'subplots' to which the other factors are applied. The analysis of variance has to be divided into two parts, one for the study of main plots within blocks; the other of subplots within main plots. Convenient though this scheme is, it requires caution. In general, the main plot treatments will be estimated less precisely than those on subplots. For one thing, they are compared within blocks whereas the subplot treatments are compared within main plots, which are smaller and so can be more effective in the control of environmental variation. Also, in the main plot analysis there will usually be markedly fewer degrees of freedom for the estimation of error.

If factor M is applied to main plots and S to subplots, their interaction, M × S, will be found in the subplot analysis. Its interpretation requires care. If the responses to S are compared for each level of M, all the relevant differences lie between subplots in the same main plots and their standard errors are derived solely from the subplot error. If, however, comparison is made between responses to M within each level of S, both errors are involved and larger standard errors are to be expected. For that reason, there is an advantage if the factors on main plots are provocative and those on subplots substantive, but this consideration can conflict with allocating factors to main plots or subplots for reasons of practicality. The solution might be to have plots of only one size even if that does call for more land than was originally expected.

Where both analyses will be required, it is possible to affect their relative sensitivities by the shape of the main plots. Their role is ambiguous. In the main plot analysis, they are used as plots, which means that within each block they should be as similar as possible. That is to say that they should be long and narrow and formed across fertility contours so as include as many kinds of soil as possible. In the subplot analysis, on the other hand, they act as blocks. Consequently they should be compact, so that soil differences lie between them and not within them. These considerations are clearly incompatible and the designer has to choose between using long narrow subplots, which will favour the main plot analysis, or compact so as to favour the other.

Block size

The main difficulty with factorial designs comes from the large number of treatment-combinations. If a factor has four levels, a simple design can cope. If there are two such factors, there are 16 treatment-combinations, which call for blocks too large to be completely successful. Three such factors give 64 treatment-combinations; if each were represented by only a single plot that single replicate would be larger than many complete experiments. This situation has no parallel except in the testing of new strains

https://doi.org/10.1017/S0014479704002364 Published online by Cambridge University Press

from a breeding programme, but statistically that is quite different because there is no structure between the new strains whereas here there is the structure essential to a useful interpretation of results. The problem is to avoid blocks so large that they are ineffective in controlling environmental variation within the experimental area.

When nearest-neighbour methods were first attracting attention, Bartlett (1938) suggested that they presented a possible solution. The present time is, perhaps, the occasion to reconsider his suggestion (Pearce, 1998). Also, the introduction of spatial methods of controlling variation, i.e. those that assume a pattern of environmental variation, offer another way forward (Dyke *et al*., 1982; Cullis and Gleeson, 1991). A further possibility is to use non-orthogonal designs, i.e. ones in which each block contains only a selection of the treatments. A lot of effort has been expended on evolving such designs in which all comparisons are estimated with about the same efficiency, though it has been suggested (Pearce, 1963,1983) that they might also be useful when the contrasts differ in importance. There are several kinds of non-orthogonal design that fit factorial experiments very well, e.g. those that are group-divisible. However, the main thrust of study to reduce block size has been in confounding, which will be described next.

REDUCTION OF BLOCK SIZE

A simple example

The reduction of block size is simple when all factors have two levels, e.g., presence or absence of some feature. Such experiments are of special value in exploration where there have been several suggestions why plants are growing abnormally, and the need is to try them out singly and in combination to suggest a way forward. To take the example of only two factors, A and B, there will be four treatments-combinations, which will be called *1*, *a*, *b* and *ab*. Then, using the method of contrasts the 'main effect' of A is given by the contrast $(-1, 1, -1, 1)$. That is to say, a difference is taken between those treatments in which A is present and those in which it is not. Similarly the main effect of B is given by $(-1, -1, 1, 1)$. The interaction of two effects is found by multiplying them out, which gives

$$(-1 \times -1, 1 \times -1, -1 \times 1, 1 \times 1) = (1, -1, -1, 1).$$

These three effects represent all the ways in which four things can be divided into two pairs. The third is clearly the interaction because it calls for the comparison of treatments *1* and *ab* with *a* and *b*. If there is no interaction it would not matter whether the factors were applied together or separately, so the two pairs should give the same total result within the limits of experimental error.

If the interaction proves to be large the two main effects are of little use. They give the mean effect of each factor averaged over two different conditions. What is needed now are the 'particular effects'. That of factor A in the absence of B is given by $(-1, 1, 0, 0)$ and in its presence by $(0, 0, -1, 1)$. Similarly the particular effects of factor B are $(-1, 0, 1, 0)$ and $(0, -1,0,1)$.

https://doi.org/10.1017/S0014479704002364 Published online by Cambridge University Press

More than two factors

There is no reason to stop at two factors. Let there be three, A, B and C, leading to eight treatments *1*, *a*, *b*, *ab*, *c*, *ac*, *bc* and *abc*. The three main effects are represented by the contrasts:

A $(-1, 1, -1, 1, -1, 1, -1, 1)$
B $(-1, -1, 1, 1, -1, -1, 1, 1)$
C $(-1. -1, -1, -1, 1, 1, 1, 1)$

The three two-factor interactions are given by:

B × C $(1, 1, -1, -1, -1. -1. 1. 1)$
A × C $(1, -1. 1, -1, -1, 1, -1, 1)$
A × B $(1, -1, -1, 1, 1, -1, -1, 1)$

It will be seen that any of these is the interaction of the other two and that is for a good reason. Consider, for example, the interaction of A × B and B × C. Since all elements of the contrast for the main effect of B are either + 1 or − 1, all elements of B × B equal + 1, so A × B × B × C = A × C. This happens whenever a main effect with two levels occurs twice in an interaction.

The contrast for the three-factor interaction is found by multiplying out those for all the main effects, i.e A × B × C. It is $(-1, 1, 1, -1, 1, -1, -1, 1)$.

Confounding

As more factors are introduced, the blocks increase in size to the point where they would be of little use in reducing the effect of local variation. If an interaction is thought most unlikely to exist, the device of confounding becomes available and will be exemplified using A × B × C, described above, to reduce size of blocks from eight plots to four but doubling their number, thus leaving the total number of plots unchanged. Half the new blocks will be assigned the treatments *a*, *b*, *c*, *abc*, which will be allocated at random within it, and the other half, *1*, *bc*, *ac*, *ab*. It is clear that the difference between the two sorts of block corresponds to the three-factor interaction, which will no longer appear in the analysis of variance, which is 'intrablock', i.e. concerned with plots within blocks, though it can be 'recovered' by comparing the two kinds of new block. However, the 'interblock analysis', as it is called, is likely to be rather uninformative, partly because there will be few degrees of freedom for the estimation of experimental error. One the other hand, the intra-block analysis contains the other contrasts and remains effective.

Confounding with four or more factors

If the confounding of an interaction with three factors is thought unwise, it is surely acceptable to confound one with four. So a larger experiment will be considered in which there are four factors, A, B, C and D, each with two levels, giving rise to 16 treatment-combinations, viz,

1, *a*, *b*, *ab*, *c*, *ac*, *bc*, *abc*, *d*, *ad*, *bd*, *abd*, *cd*, *acd*, *bcd*, *abcd*.

https://doi.org/10.1017/S0014479704002364 Published online by Cambridge University Press

In the contrast for A × B × C × D, the treatments *1*, *ab*, *ac*, *bc*, *ad*, *bd*, *cd* and *abcd* will have elements equal to + 1 but *a*, *b*, *c*, *abc*, *d*, *abd*, *acd* and *bcd* will have − 1. In order to confound this effect the number of blocks will be doubled and the block size reduced to eight plots. Half the blocks are chosen to be of Type I and are assigned the treatments with + 1. Treatments with − 1 are assigned to the other blocks, which are of Type II. (The treatments in each block must be allocated at random to the plots.) Some may question whether the other effects are validly estimated but they are. To take an example, consider A × B, which requires the comparison of treatments *1*, *ab*, *c*, *abc*, *d*, *abd*, *cd* and *abcd* with *a*, *b*, *ac*, *bc*, *ad*, *bd*, *acd* and *bcd*. On both sides of the comparison four treatments come from blocks of Type I and four from those of Type II, so the result is unaffected by block differences. The same applies to all the other unconfounded effects.

The method can be taken further. Would it not be possible to reduce block size further by confounding another interaction? Could not B × C × D also be sacrificed so that block size can be reduced to four plots? In that contrast *1*, *a*, *bc*, *abc*, *bd*, *abd*, *cd* and *acd* have the element + 1 but *b*, *ab*, *c*, *ac*, *d*, *ad*, *abc* and *abcd* have − 1. The method is to use four kinds of block. Those of Type I will contain treatments with positive elements in both the contrasts to be confounded, i.e. *1*, *bc*, *bd* and *cd*. Blocks of Type II will have those with a positive element in the first contrast but a negative one in the second, i.e. *ab*, *ac*, *ad* and *abcd*; those of Type III having a negative element in the first but a positive one in the second, i.e. *a*, *abc*, *abd* and *acd*. That leaves Type IV to receive those that have a negative element in each, i.e. *b, c*, *d* and *bcd*. Then A × B × C × D has been confounded between blocks of Types I and II as compared with those of Types III and IV and B × C × D has been confounded between those of Types I and III as compared with those of II and IV. It emerges that A is confounded between blocks of Types I and IV as compared with those of Types II and III; this always happens. If two contrasts are confounded, their interaction is confounded also.

Conventional wisdom says that only interactions of three or more factors should be confounded, but it is here suggested that conventional wisdom may be wrong. If A is a provocative factor, introduced only to see if it gives rise to interactions, its main effect does not require investigation. Its interactions are retained except for A × B × C × D.

When the number of factors exceeds three there will be interactions that are poorly estimated and difficult to interpret and, except for special reasons, can be disregarded. Any sum of squares associated with them can be merged with error. They do however provide hidden replication. If it can be assumed that A × B × C × D does not exist, then A × B × C will be estimated twice, once for each level of D. The same is true of any other main effects and interactions involving A, B and C. Effectively, the replication has been doubled.

Single replicate designs

Because of this hidden replication, it is sometimes sufficient to use no actual replication at all. Usually at least six factors are needed and the method will be illustrated with that number. A single replicate will require 64 plots, which leaves no

https://doi.org/10.1017/S0014479704002364 Published online by Cambridge University Press

degrees of freedom for error, but if it is agreed that no interactions with more than three factors need to be considered that leaves 15 interactions with four factors, the six with five and the single six-factor interaction, which can be merged to give an error sum of squares with 22 degrees of freedom. Further, the three-factor interactions can be estimated with hidden eight-fold replication. Thus, the interaction of A × B × C can be estimated by the comparison of *1*, *ab*, *ac* and *bc* with *a*, *b*, *c*, *abc*, but if there is a fourth factor, D, and if A × B × C × D can be ignored, then it can be estimated equally well by comparing *d*, *abd*, *acd* and *bcd* with *ad*, *bd*, *bcd* and *abcd* and in six other ways. There is however a difficulty. An experiment with 64 plots will almost certainly need blocks to control local variation, but they can be provided by confounding the six-factor interaction to obtain two blocks each of 32 plots. If that is not enough, four blocks, each of 16 plots, could be used by confounding A × B × C × D and C × D × E × F together with their interaction A × B × E × F. That would reduce the degrees of freedom for error to 19, which would still be enough.

Fractional replication

The matter can be taken further. Finney (1946) explored the use of only some of the treatments generated from a large number of factors. Suppose that there are seven. That gives rise to 128 treatments. He proposed the use of a defining contrast, e.g. the seven-factor interaction, to divide them into two groups, one of treatments with an element of + 1 in the defining factor and the others with − 1. If only one group is used the experiment needs only half a replicate, i.e. 64 plots. Its data can still be interpreted, but there is a complication. With only half the treatments present, it is not possible to separate an effect and its interaction with the defining contrast; the two are said to be aliases, two names for the same thing. Consequently, if the defining contrast is the interaction between seven factors, each main effect has as its alias the six-factor interaction of the other treatments. That raises no difficulty. Any difference found must almost certainly be due to the main effect. An interaction with two factors is indistinguishable from its alias with five. Again, there is unlikely to be much doubt as to which is the one responsible for any difference found. Three-factor interactions however are more difficult because each has a four-factor interaction as its alias and there may be less certainty as to which is the active one. Perhaps both are.

Adding a second defining contrast brings into existence a third, namely, the interaction of the two already introduced. There is now a quarter replication and each effect has three aliases. Fortunately, such complexity is rarely called for.

Confounding factors with four levels

The methods described above are readily extended to the confounding of effects with four levels. If A has four levels, 0, 1, 2 and 3, they can be regarded respectively as 00, 01, 10 and 11, where the two digits show the levels of two factors, U and V, each with two levels. The effect of A with three degrees of freedom can now be considered as the aggregate of U, V and U × V, each with one.

https://doi.org/10.1017/S0014479704002364 Published online by Cambridge University Press

The approach can be illustrated by considering a single replicate of 64 treatments, made up of three factors, A, B and C. Writing A as above, B as W, X and W × X taken together and C as Y, Z, Y × Z, it is necessary only to confound both U × W × Y and V × X × Z, which causes the six-factor interaction to be confounded also. The single block of 64 plots has now become four blocks each of 16. Each of the main effects of A, B and C can be found by summing three components and all the two-factor interactions by summing nine, e.g. A × B with 9 degrees of freedom is made up of U × W, U × X, U × W × X, V × W, V × X, V × W × X, U × V × W, U × V × X and U × V × W × X. As to the 27 degrees of freedom of the three-factor interaction, three have been confounded, leaving 24 for estimating the experimental error. There are many similar possibilities.

Use of non-orthogonal designs

In most experiments each block contains the same set of treatments but that is not essential. Such designs are termed 'orthogonal'. Sometimes however difficulties with the site and the formation of blocks makes non-orthogonality desirable, if not essential. Another reason is the need to minimize block size. If blocks differ in their content, some contrasts, if not all, will suffer a loss of information in the intrablock analysis, which can sometimes be recovered by an interblock analysis. Non-orthogonal designs exist in which the loss fits the factorial structure of the treatments, notably those that have factorial balance or are group-divisible (Pearce, 1963; 1983).

An example is afforded by the well-known method for confounding of the interaction in an experiment with two factors, each with three levels. (Yates, 1935). Designating each treatment by two digits to represent the levels of the two factors (i.e. 0, 1 or 2), he suggested replacing two blocks with nine plots by six blocks with three. In one block, chosen at random, the treatments would be 00, 12, 21; in a second 01, 12, 20; in a third 02; 11; 20; in a fourth 00, 11, 22; in a fifth 01, 12, 20; and in the last 02, 10, 21. (It is, of course, possible to use further groups of six blocks to gain further replication.) The resulting design has factorial balance and provides for partial confounding of the interaction; information about it is divided equally between the two analyses.

It is unlikely that anyone would introduce two factors in an experiment without being interested in their interaction, but the scheme can be modified so that the main effects are partially confounded, leaving information about the interaction entirely in the more sensitive intrablock analysis. To do that, the treatments should be allocated to blocks thus: I 00, 01, 02; II 10, 11, 12; III 20, 21, 22; IV 00, 10, 20; V 10, 11, 12; and VI 02, 12, 22. This design also has factorial balance. There are many other possibilities.

DISCUSSION

Availability of software

Before embarking on an unusual design, the experimenter should enquire about the availability of computer software to deal with the data that will arise. Most modern

https://doi.org/10.1017/S0014479704002364 Published online by Cambridge University Press

programs can cope with confounding and many with non-orthogonality but it is wise to check.

Randomization

In general, treatments should be allocated to plots at random to validate significance tests, but there are occasions with factorial designs when it might be better to disperse them over the area. One such occasion arises with split-plot designs. If all factors on the main plots are provocative, there will be no need to carry out any tests and it might be better to arrange that each main plot treatment had its share of fertile and less fertile land. (Of course, that depends on the experimenter being able to discern beforehand which were the better and worse areas.) A similar situation arises in confounding. It is true that an interblock analysis can provide a test of confounded effects, but probably it will be insensitive and the effects small. Some may think it better to disperse the types of block rather than to randomize them.

The role of significance

When significance tests were first introduced, there was a tendency to report all effects significant at the 0.05 level as if they were proven and to regard all that failed to reach that level as non-existent. That was, of course, a mistake. The word 'significant' was chosen originally to imply only that the effect in question merited serious consideration. In practice a plausible effect may carry conviction even if it fails a significance test, while one that looks absurd may be rejected even though its significance level is high. It should be reported nevertheless; someone else may be able to make sense of it, even if the author cannot do so.

This rather loose approach causes difficulties in presenting results from a factorial experiment, because the readers will not necessarily agree with the author as to what is plausible and what is not. If an interaction is significant at some stated level it would be wise not only to report it, despite any scepticism on the part of the author, but to present the particular effects that would be required to interpret it in case a reader should wish to do so. If the particular effects are known, it is easy to derive the corresponding main effects, but the reverse process is not possible.

Estimation of error

Some may think it strange to see high order interactions merged into the error sum of squares, yet in a design like randomized complete blocks the error is formed from the interaction of blocks and treatments and that is generally accepted. Since experimenters are urged to associate blocks with features of the land, e.g. depth of top soil or exposure to wind, and since it is wise to assume that any two-factor interaction may well exist, there is nothing new in using interactions to form the error. If the interaction of treatments with blocks (Tr × Bl) may be regarded as error, there should be no objection to A × Bl and even less to A × B × Bl or A × B × C. In fact, standard errors are only guides and are rarely found with great precision. Modern writers tend to avoid the word 'error' and prefer 'residual', i.e. they see the sum of squares in

https://doi.org/10.1017/S0014479704002364 Published online by Cambridge University Press

question simply as what is left after the treatment effects have been removed. Whatever name is chosen, it is used to estimate the experimental error.

Interactions and choice of variate

It sometimes happens that interactions arise because of a bad choice of variate. For example, fertilizers affect the size of plant by changing its rate of growth. If then plants are graded before planting and the final size (e.g weight, height) is measured it could well be found that two fertilizer elements (e.g. N and K) had interacting effects. If, however, the size measurements are transformed to their logarithms to give growth rates, the interaction may disappear. In any experiment, especially those with factorial design, it is sensible to measure the underlying biological process rather than its result.

REFERENCES

Bartlett, M. S. (1938). The approximate recovery of information from field experiments with large blocks. *Journal of Agricultural Science* Cambridge 28:418–27

Clarke, G. M. (1980). *Statistics and Experimental Design*. 2nd edn. London: Edward Arnold Limited.

Cullis, B. R. and Gleeson, A. C. (1991). Spatial analysis of field experiments – an extension to two dimensions. *Biometrics* 47:1449–1460.

Dyke, G. V., Smith, G. L. and Yeoman, D. P. (1982). Fourier series and response curves. *Journal of Agricultural Science* Cambridge 98:119–122.

Finney, D. J. (1946). Recent developments in the design of field experiments, III. Fractional replication. *Journal of Agricultural Science* Cambridge 36:184–191.

Fisher, R. A. (1925). *Statistical Methods for Research Workers*. Cambridge: Cambridge University Press.

Pearce, S. C. (1963). The use and classification of non-orthogonal designs (with discussion). *Journal of the Royal Statistical Society, B* 126:353–377.

Pearce, S. C. (1983). *The Agricultural Field Experiment. A Statistical Examination of Theory and Practice*. John Wiley & Sons, Chichester.

Pearce, S. C. (1998). Field experimentation on rough land: the method of Papadakis reconsidered. *Journal of Agricultural Science* Cambridge 131:1–11.

Yates, F. (1935). Complex experiments (with discussion). *Journal of the Royal Statisical Society*, Suppl. 2:181–247

https://doi.org/10.1017/S0014479704002364 Published online by Cambridge University Press