Expl Agric. (1992), volume 28, pp. 245–253
Printed in Great Britain

DATA ANALYSIS IN AGRICULTURAL EXPERIMENTATION. I. CONTRASTS OF INTEREST

By S. C. PEARCE

Applied Statistics Research Unit, University of Kent at Canterbury, England

(*Accepted 29 November 1991*)

SUMMARY

An experiment has its origin in the need to find answers to stated questions. From the start great care is given to its correct conduct (e.g. the application of treatments and the method of recording) as well as to statistical design, always with the original questions in mind. The analysis of its data is the climax of a long process and the questions to be answered must dominate all else. It is not enough to feed data into a computer package in the hope that it will provide an automated path to a true interpretation.

Where the treatments have been chosen with care to answer specific questions, the statistical way of designating purpose is to declare 'contrasts of interest', each corresponding to a degree of freedom between treatments. They derive solely from the reasoning behind the selection of treatments. If possible the questions posed should be equal in number to the degrees of freedom and should admit of separate study because no one can give a single answer to several diverse questions.

This paper shows how to define a contrast of interest and how to isolate it in the analysis of variance. Attention is given both to its contribution to the treatment sum of squares and to its variance (i.e. the precision with which it is estimated). Independence of estimation is also considered. Algebraic formulae are given for a restricted though important range of designs, which includes those that are completely randomized, in randomized blocks or in Latin squares, all treatments having the same replication. The methods can, however, be generalized to cover all designs. With these formulae it is possible both to test the existence of an interesting effect and to set confidence limits round its estimated value.

Análisis de datos en la experimentación agrícola. I

RESUMEN

Todo experimento surge de la necesidad de encontrar respuesta a cuestiones planteadas. Desde un principio se presta mucha atención tanto a la correcta forma de conducirlo (por ejemplo, la aplicación de tratamientos y el método de registro), como a su diseño estadístico, siempre teniendo en mente los interrogantes originales. El análisis de los datos constituye el clímax de un largo proceso, y también debe de estar dominado por las cuestiones a responder. No basta con alimentar datos en un sistema de computador esperando que éste pueda ofrecer una vía automatizada para llegar a una verdadera interpretación.

En los casos en que se han elegido tratamientos seleccionados cuidadosamente para responder cuestiones específicas, la forma estadística de denominar el fin consiste en declarar 'contrastes de interés', cada uno de los cuales corresponderá a un grado de libertad entre los tratamientos, los cuales se derivan exclusivamente del razonamiento que subyace la selección de tratamientos. De ser posible, la cantidad de preguntas planteadas debe ser igual al número de grados de libertad, y deberá poder admitir un estudio por separado, dado que nadie puede dar una sola respuesta a tantas preguntas diversas.

Este trabajo muestra cómo definir el contraste de interés y cómo aislarlo en el análisis de variaciones. Se presta atención tanto a su contribución a la suma de cuadros del tratamiento como a su variante (o sea, la precisión con que se la estima). Tamhién se considera la

Con todo, los métodos pueden generalizarse para cubrir todos los [illegible] puede tanto comprobar la existencia de un efecto interesante como también determinar límites de confianza alrededor de su valor estimado.

THE ROLE OF DATA ANALYSIS

The analysis and interpretation of data is the last stage in a long process that began with someone formulating a question. It continued by way of an experiment laid out after careful design, in which treatments were applied, plants measured and records kept until at last all was complete. Every stage of that process was considered and its form determined with reference to the initial question; the final stage alone, namely the data analysis, appears to be exempt. It is not enough that the data should be analysed using an approved program or computer system. That is equivalent to saying that microscopic examination of sections requires a good microscope. Certainly that helps, but it is even more important that the user should know what to look for. Perhaps the object is merely exploration. A section is taken and examined to see if it exhibits any abnormalities; sometimes the search is for something more specific. It is the same with a statistical analysis. There also the approach must depend upon the reason for undertaking the task at all. Where some specific point needs to be clarified or some question answered, it can be defined by a 'contrast' between the treatment means. In this paper the object is to show how such 'contrasts of interest' can be defined and then estimated or tested according to the needs of the investigation.

TYPES OF EXPERIMENTS

There are many different kinds of experiment. Statistically speaking much depends upon the extent to which their initiator has succeeded in thinking through the issues beforehand.

In the most elementary case the experimental objective may simply be exploration. The investigator has no clear lead and is looking for one. Such an enquiry is not highly regarded as a scientific pursuit but there are times when it is the only way forward. A number of treatments are tried in the hope of getting a clue as to underlying causation of a phenomenon. If that is the position, there are usually no clearly defined contrasts in mind at the initiation of the investigation.

At a higher level are 'pick the winner' situations in which several alternatives are available and the object is to find the best. It is commonly encountered with a breeding programme where a number of new strains have to be tested, sometimes with only limited material for each. If the aim were really to find the best strain in some particular respect and two leading candidates were very similar, the task would be fearsome indeed, but in fact it usually suffices to eliminate the greater number of new strains as inferior, leaving a few with obvious merit. (Actually there are different kinds of merit and the survivors are not always competitors.) There is a variant in which the experimenter is concerned only to establish the relevance of a certain kind of treatment, e.g. variety or nutrition, without at that stage attempting to identify the best. In neither of these cases is it usually necessary to specify contrasts.

The most highly regarded experiments are of the 'critical' sort. The name derives from the Greek *κριτησ*, meaning a judge. That is to say, they exist to decide a contentious issue. That cannot be done without knowing what the issue is and in what respect it is contentious. Specified contrasts of interest now become essential if subsequent data analysis is to be strictly relevant to purpose, and they must be studied intensively. If in addition a lot of diffuse information about related topics emerges, that may be useful in provoking further thought, but it is not what the experiment was primarily intended to achieve.

It is usually best to study each contrast of interest, i.e. each question, in isolation and in logical order. That is important because no one can give a single answer to a composite query. To take an example, an experimenter may propose the comparison of two treatments, e.g. the application or omission of an insecticide. After discussion with colleagues a third treatment is added, namely an application at half strength, so there are now two degrees of freedom between three treatments. Two degrees of freedom imply two questions, but what are they? There are several possibilities. Perhaps the third treatment was introduced to establish the approximate straightness of the response curve. Alternatively, it may have been hoped that half the usual concentration would prove effective enough. (In that case curvature is being assumed.) Other possible reasons can be suggested. The implied questions are different in each case and a statistical approach that is satisfactory with one will not fit another. (Appropriate contrasts of interest for the two cases will be considered in the next section.)

Whatever the reason for the addition of a third treatment and whatever the two questions, it is unlikely that anything will be gained by carrying out a significance test on the two degrees of freedom taken together. If the result is positive, which of the questions is receiving the answer 'Yes'? The first, the second or both? If it is negative, possibly one of the contrasts leads to a positive response that was diluted by a negative response to the other.

The questions that can be asked are immensely varied and defy classification; consequently there are no immutable rules for data analysis. It is true that some approaches will be needed so often as to appear almost standard, while others will be needed only from time to time, but the sole criterion for use is relevance to the problem in hand.

The method of contrasts is old (Tharp *et al.*, 1941) and continues to be recommended (e.g. by Mead, 1988). The computing procedures presented here are those given by Pearce *et al.* (1988). They derive from an earlier and more mathematical publication (Pearce, 1983).

To consider first the example just given, it has been suggested that there are at least two ways in which the data can be approached and only those involved in the investigation can say which is correct.

First, there are those who want the middle level in order to check that the graph of response against levels is approximately straight, i.e. they want to know how far the response to the middle level departs from the mean of the other two. Written as a contrast, that is $(-\frac{1}{2}\ +1\ -\frac{1}{2})$, the coefficients denoting multipliers of the treatment means. (Note that the coefficients must sum to zero.) That leaves a subsidiary question: if the graph is straight, what is its slope? That is best answered by taking the difference between the extreme levels, i.e. by studying the contrast $(-1\ 0\ +1)$. (Note that, if the graph is curved, the second contrast no longer measures the slope, which is not constant.)

The alternative approach is for those who wanted the half application because they thought it might prove to be as effective as the full one. For them, the first contrast of interest will be $(0\ -1\ +1)$. If it shows a difference, they will have to concede that they were mistaken and it will scarcely be necessary to enquire further. If, however, no difference appears, that could result from the preparation being ineffective anyway. To remove that possibility it would be necessary to look at the values of $(-1\ +1\ 0)$ or $(-1\ 0\ +1)$ or perhaps $(-1\ +\frac{1}{2}\ +\frac{1}{2})$.

It is sometimes said that the contrasts should depend upon the treatments. It would be better to say that they should derive from the questions that led to the treatments being chosen.

THE USE OF CONTRASTS

It will here be assumed that the design belongs to a limited though important class, namely, it is completely randomized, in randomized blocks or in Latin squares, and all treatments have the same replication, r. The methods to be given can be generalized to cover all valid designs, but for such an extension the reader is referred to more advanced texts, such as Pearce *et al.* (1988).

The first task must be to estimate the value, V, of the contrast under study. That is a simple matter. If the contrast is the difference between two means, M_A and M_B, then $V = M_A - M_B$. If the next contrast concerns the extent to which their joint mean differs from that of a third treatment, M_C, then the value of that contrast is $\frac{1}{2}(M_A + M_B) - M_C$. In fact, the operation is performed by multiplying out the coefficients of the contrast and the corresponding treatment means and adding the products.

With the designs here considered, in order to find the component of the treatment sum of squares corresponding to a particular contrast, it is next necessary to find the constant U, equal to the sum of squares of the coefficients of the contrast. Then the sum of squares contributed by that contrast has one degree of freedom and is equal to rV^2/U, where V is the value of the contrast. Once thc sum of squares has been found, an F-test can easily be carried out if anyone asks about the significance of the contrast.

ORTHOGONALITY OF CONTRASTS

Continuing with the example of the three levels of insecticide, several standard methods exist for finding the treatment sum of squares with two degrees of freedom. Suppose now that someone is interested in the contrasts $(-\frac{1}{2}+1-\frac{1}{2})$ and $(-1\ 0+1)$. If the treatment means are written respectively as M_1, M_2 and M_3, the first contrast equals $V_1 = (-\frac{1}{2}M_1 + M_2 - \frac{1}{2}M_3)$ and the second equals $V_2 = (-M_1 + M_3)$.

The next part of the calculation is to find U for each contrast:

$$U_1 = (-\tfrac{1}{2})^2 + (+1)^2 + (-\tfrac{1}{2})^2 = 3/2 \text{ and } U_2 = (-1)^2 + 0^2 + (+1)^2 = 2.$$

Then the sums of squares for the two contrasts are:

$$S_1 = rV_1^2/U_1 \qquad \text{and} \qquad S_2 = rV_2^2/U_2$$

each with one degree of freedom. In this instance it will be found that $S_1 + S_2 =$ the treatment sum of squares. That is to say, the two components add up correctly, but that will not always happen. It has done so in this case because the two contrasts are estimated independently of one another, i.e. knowing the value of one provides no evidence as to the value of the other. They are said to be 'orthogonal', a property that can be established by multiplying out the coefficients, thus:

$$(-\tfrac{1}{2})(-1) + (+1)(0) + (+\tfrac{1}{2})(+1) = 0.$$

The sum of products is zero. Only when that happens will the components, each with one degree of freedom, add up to the total sum of squares for treatments.

The contrary case is exemplified when the contrasts of interest are $(0+1-1)$ and $(-1\ 0+1)$, making the sum of products of coefficients equal to

$$(0)(-1) + (+1)(0) + (-1)(+1) = -1.$$

Here it is not surprising that the two are related. If by chance M_3 has a high or low value, it will affect both contrasts equally.

Although the mathematicians are quite right in wishing contrasts to be estimated independently of one another, orthogonality is not always feasible. Where the treatments have been chosen on some logical plan there is usually no conflict, all pairs of contrasts of interest proving to be mutually orthogonal, but if a choice has to be made there can be no doubt which property is to be preferred. As Gill (1973) has said, 'meaning must take precedence over orthogonality'.

SPECIFICATION OF CONTRASTS

It may be noted that the same contrast can be written in many ways, the choice between alternatives depending on convenience and intelligibility. The important points are first, that the coefficients should sum to zero and second, that however a contrast is written, the ratio of its components must be preserved. To take the example of $(-\frac{1}{2}\ +1\ -\frac{1}{2})$, that is effectively the same as $(-1\ +2\ -1)$ or indeed as $(+5\ -10\ +5)$. Changing it to $(-1\ +2\ -1)$ doubles the value of V and multiplies those of V^2 and U by four, leaving V^2/U unaltered. Similarly in the next section, where the standard error of V is considered, it will be found that using $(-1\ +2\ -1)$ instead of $(-\frac{1}{2}\ +1\ -\frac{1}{2})$ doubles both V and its standard error. Consequently the conclusions given by a significance test, if one is called for, will be the same either way. (A change of sign throughout a contrast will alter its direction, but that will correspond to a change in the way the question has been posed.)

To take an example, a factorial set of treatments, (1), A, B, AB, is used in which a basic procedure, (1), is modified by two changes, A and B, and by using both together, AB. It might be asked whether A interacts with B. One person might interpret that as meaning 'Is the effect of B the same in the presence and absence of A?', that is, asking whether the value of the contrast $(0\ -1\ 0\ +1)$ is the same as that of $(-1\ 0\ +1\ 0)$, leading to a study of their difference, the contrast $(+1\ -1\ -1\ +1)$. Another might look at it differently and ask an equivalent question, 'If A and B are applied to a pair of plots, will the mean effect be the same whether they are applied to different plots or to the same one?' that is, are the values of $(0\ +\frac{1}{2}\ +\frac{1}{2}\ 0)$ and $(+\frac{1}{2}\ 0\ 0\ +\frac{1}{2})$ the same? The difference, $(-\frac{1}{2}\ +\frac{1}{2}\ +\frac{1}{2}\ -\frac{1}{2})$, is the contrast obtained previously apart from the constant multiplier of -0.5, which alters nothing. (It may be noted that neither $(0\ +\frac{1}{2}\ +\frac{1}{2}\ 0)$ nor $(+\frac{1}{2}\ 0\ 0\ +\frac{1}{2})$ counts as a contrast because their coefficients do not sum to zero, but their difference does qualify.)

Given a choice in how to express a contrast, intelligibility is most important. If the aim of statistical analysis is to reply to questions, the answers must as far as possible be readily understandable.

VARIANCE OF ESTIMATION OF A CONTRAST

Another question concerns the precision with which the value, V, of the contrast is known. With the same restricted range of designs as before it is evaluated with a variance equal to:

$$\textit{Error mean square} \times U/r.$$

(There is a small difficulty with nomenclature here. In this context some talk about the 'error' mean square; others prefer the word 'residual'. No difference of meaning is intended.) The square root of the variance is of course the standard error. If that quantity is multiplied by the appropriate value of t, it can be used to set confidence limits round V. To take an example, let the value, V, of the contrast $(+1\ -1\ +1\ -1)$ be found as 12.0 on the basis of six $(r = 6)$ replicates. It will be noted that:

$$U = (+1)^2 + (-1)^2 + (+1)^2 + (-1)^2 = 4.$$

Further, it will be supposed that the error mean square was 24.0 with 15 degrees of freedom. Then the contrast has been estimated with variance $24.0 \times 4/6 = 16.0 = 4.0^2$, so the standard error is 4.0. For 15 degrees of freedom and a significance level of 0.01, t equals 2.947. Hence, confidence limits should be set at a distance of 11.8 ($=2.947 \times 4.0$) from the estimated value of 12.0, i.e. at 0.2 and 23.8. That is to say, the value of the contrast can be declared to lie within those limits, but with the recognition that there is one chance in a hundred that it could in fact lie outside.

If anyone enquires about significance, the sum of squares for the contrast is $6 \times 12.0^2/4 = 216.0$. It has one degree of freedom, so the mean square likewise equals 216.0. Since the error mean square is 24.0, F equals 9.00 ($=216.0/24.0$) with 1 and 15 degrees of freedom and is just significant at the level $P = 0.01$. The calculation shows the relationship between the two approaches. Asserting that the observed value of the contrast is significantly different from zero is equivalent to asserting that zero lies outside the confidence limits.

TESTING AND ESTIMATING

Statistical theory makes an important distinction between testing and estimating. In the first, the question concerns the existence of an effect. Does deeper planting affect the date of maturity? Would a spray containing magnesium reduce the incidence of leaf scorch? The only acceptable answer to such a question is 'Yes' or 'No' and a significance test is required to provide it. The second sort assumes that the effect exists and requires a quantitative answer, for example, if applications of nitrogen were increased by half, what effect would that have on yield? The distinction is logically important but frequently ignored. Modern statistical techniques were developed in a context of testing and seem never to have outgrown their origins. Many experimenters are taught that every assertion must be the subject of a test and sometimes the outcome is absurd. To take a common example, an experimenter wishes to study some cultural treatments and decides to do so on a range of widely different cultivars. The report contains a significance test to establish that there were indeed differences between them, though no one would believe it if it said that there were not. It might be helpful to know how each did behave in the conditions of the experiment, but what is the test testing? A non-significant result, if one were found, would establish only that the experiment was insensitive.

If the supporters of significance tests overstate their case, their opponents also are open to criticism. It is true that any differential treatment must do something and an effect on any organ must have effects elsewhere. It is not possible to cultivate differently or to apply a fertilizer or a spray and for the plants to continue as if nothing had happened. Nevertheless, treatment differences can be so small as to be unimportant and better ignored.

Finally, there are times when a writer has to choose between two courses of action. For example, a graph has been obtained relating yield to the quantity of fertilizer. Can it be regarded as straight? The next step may well depend upon the answer to that question. The decision how to proceed should not be made capriciously and a significance test can help to justify a course of action. There are those who dislike testing but who would nonetheless recognise its usefulness in such a situation.

CONFIDENCE LIMITS OF A CONTRAST

There is much sense in the argument put forward by Chew (1980), which goes a long way towards resolving the conflict between testing and estimating. In the course of an excellent account of the role of significance testing, he suggests that a good way of presenting results is to cite the values of relevant treatment contrasts and to set confidence intervals round each. As explained in the section on the standard error of contrasts, if zero lies within the range indicated, then the difference is non-significant, that is, the test is subsumed within the estimate. To take an example, following convention a paper might state that deeper ploughing had led to a 5% increase in crop yield, significant at the level $P = 0.05$. Such a report might encourage some to adopt the new method immediately, but Chew's suggestion would lead to better information. Using confidence intervals the paper might state that the improvement lay between 1 and 9% and that there was one chance in 20 of its lying outside those limits. In that case there could well be a call for research to evaluate the improvement more precisely, but no more. Many would hesitate to go to the expense of ploughing more deeply to obtain an improvement in yield that might turn out to be no more than 1%. On the other hand, if the limits were 4 and 6%, the response might be more positive.

Tukey (1991) reaches much the same conclusion as Chew though by a different path. He first agrees with the proponents of estimation that the statistician's null-hypothesis, that the treatments have had no effect, is absurd. There must be an effect and the first question concerns its direction. A conventional significance test may leave the matter open (non-significance) or it may indicate a change in one direction or the other, but that is not enough. Like Chew, he recommends estimating the magnitude of the effect and setting limits round the estimate.

PRESENTATION OF RESULTS

There is a conflict here that needs to be faced. When presenting the results of an experiment, is one answering only the initial questions or informing others whose questions may be different? If the writer, having explained the reasoning that gave rise to the experiment, deals only with the contrasts of interest so indicated, the report will be barren for those who approach the subject from a different viewpoint and would like to look at other contrasts. It also inhibits browsing and searching for clues. On the other hand, a report fails in the opposite direction if it presents only the treatment means with a standard error for each, because then the writer does not really come to terms with his or her own questions and the purpose behind the investigation. There is need for both approaches even if that does lead to some duplication. It is here suggested that tables and text call for different priorities.

In a table space is short and uniformity of presentation desirable. It is therefore convenient to follow convention and just give treatment means and the standard error of each, like this:

A	B	C	D	SE
14.6	15.1	13.5	19.1	0.88

It is important that sufficient information is provided in the text for any reader who wants to study contrasts of personal interest. (Many papers fail in that respect, especially if the design is complex.) For example, the replication should be clearly stated, as should the design. (Here it will be assumed that there were five randomized complete blocks, so r equals 5 and there will be 12 degrees of freedom for error.) For simple designs, such as those considered here, the error mean square equals r times the square of the standard error of a treatment mean, that is, here it equals $5 \times 0.88^2 = 3.872$, but that determination suffers from rounding error and it would be better for the value to be stated explicitly.

In the text the author can be more specific and consider the contrasts relevant to the investigation. Suppose, for example, that the point being made concerns the way in which Treatment D, which perhaps has some special feature, stands away from the rest, that is, attention is concentrated on $(-\frac{1}{3} -\frac{1}{3} -\frac{1}{3} +1)$, it follows that the contrast had a value of +4.7 with confidence limits at +2.5 and +6.9 (P = 0.05). It is not enough to leave that to the reader.

Acknowledgement. The writer thanks Dr D. J. Pike for his helpful comments on an earlier draft.

REFERENCES

Chew, V. (1980). Testing differences among means: correct interpretation and some alternatives. *HortScience* 15:467–470.

Gill, J. L. (1973). Current status of multiple comparisons of means in designed experiments. *Journal of Dairy Science* 56:973–977.

Mead, R. (1988). *The Design of Experiments*. Cambridge: Cambridge University Press.

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

Pearce, S. C., Clarke, G. M., Dyke, G. V. & Kempson, R. E. (1988). *A Manual of Crop Experimentation*. London: Charles Griffin, and New York: Oxford University Press.

Tharp, W. H., Wadleigh, C. H. & Barker, H. D. (1941). Some problems in handling and interpreting plant disease data in complex factorial designs. *Phytopathology* 31:26–48.

Tukey, J. W. (1991). The philosophy of multiple comparisons. *Statistical Science* 6:100–116.