KM 32(1989)
pag 65 - 72

DEVELOPMENTS IN EXPERIMENTAL DESIGN IN AGRICULTURAL AND HORTICULTURAL RESEARCH

D. A. Preece

The earliest experimental design that I know of was for a trial in 1788 on fattening sheep. The experiment was devised and reported by the Frenchman Cretté de Palluel, and the design was a systematic 4 x 4 Latin square. There were 16 sheep of 4 different breeds, fed on 4 different diets, and killed in 4 successive months, as follows, where the animals are numbered 1,2,...,16:

	Elève du Pays	Beauceron	Champenois	Picard
Pommes de terre	1: Février	2: Mai	3: Avril	4: Mars
Turneps	5: Mars	6: Février	7: Mai	8: Avril
Betteraves	9: Avril	10: Mars	11: Février	12: Mai
Grains &c	13: Mai	14: Avril	15: Mars	16: Février

In modern terminology, this design is a "fractional replicate" of a factorial, and this fraction was used as a "main effect plan". Unlike most Latin squares used as designs for agricultural experimentation, this one was not used as a row-and-column design, as the Breed and Diet factors - like Slaughter-date - were under deliberate investigation. There are no "blocking" factors here; in an analysis produced by the ANOVA directive of the general statistical program Genstat, all three factors would be coded as treatment factors. But they are factors of three conceptually very different types. Breed is inherent to the animals; once you have a sheep, you cannot randomly assign a breed to it. Diet

---

D.A.Preece, Biometrics Department, Institute of Horticultural Research, East Malling, Maidstone, Kent ME19 6BJ. Tekst van een lezing gehouden op de 4e Conferentie voor Landbouwstatistici in Lunteren, 11-12 januari 1989.

and Slaughter-date are factors whose levels are capable of random assignment, but again these factors differ from one another, in that Slaughter-date is a time-factor whereas diet is not, and a time-factor invites special questions about the appropriateness of standard assumptions about error terms in a model. True, different animals were killed on the different occasions, so we are not here dealing with so-called "repeated measurements"; nevertheless, the fact that the "plots" of the experiment did not all co-exist throughout the entire life of the experiment is a warning of possible hazards in the interpretation of results.

I was asked to consider past, present and future in this talk. So I quote Cretté de Palluel's design, not only to indicate that the subject of experimental design is now more than two hundred years old, but also because the future demands an improvement on our simplistic dichotomy into just block factors and treatment factors. (Why, incidentally, are statisticians so partial to dichotomies? We have: Estimation and Hypothesis-testing; Point-estimation and Interval-estimation; Type I and Type II; Fixed and Random; Block and Treatment; Bayesian and non-Bayesian; and many more. We seem to have binary minds!) The design of experiments is going to fall more and more into the hands of computers, less and less into the hands of thinking, trained humans. Whether we like it or not, statisticians concerned with experimental design will be replaced by "expert systems" - some being far from expert. If such a system were to know only two types of factor - block factors for variability that we are not interested in, and treatment factors for which randomisation is needed - what dreadful consequences there could be! But how many types of factor do we need? How are breed-factors and sex-factors to be categorised or handled? I don't know the answer, but I do know that the problem is not as clear-cut as it may seem on first sight. Consider a crop-variety trial. The allocation of varieties to plots of land can of course be randomised, but the allocation of varieties to the seed cannot. So, if the seed of one variety is anomalously poor, whereas that of another variety is unusually good, variety will be confounded with quality. Indeed a barley-variety trial can never be more than a comparison between particular batches of seed. In this sense it is just like an animal-trial that aims to compare different breeds. Perhaps, then, we need a spectrum of types of factor - or do we just need flexibility in interpreting factor-type?

In English, we often say that the basic principles of experimental design are the so-called "3 R's": Replication, Randomisation, and Blocking. (What a pity that no English word means "Blocking" and starts with an "R"!). To the 3 R's we add the concept of factorial experimentation. These four concepts seem to me to be so fundamental that I see no place for "development" of them. However, their implementation still needs thought and attention. Take replication first. Many experiments are now done to study binary responses. In horticultural research, cuttings of plant material may be pushed into rooting medium and the information obtained for each cutting may be merely whether or not it produces at least one root; if an experiment has "plots" each having n cuttings, the recorded observation for each plot may well be the number of cuttings (out of n) that have rooted. Different rooting media may be compared within an experiment, as may different sources of cutting, different "pre-treatments" of the cuttings, different moisture regimes for the rooting-environment, and so on. For some treatment-combinations there may be little or no rooting, for some there may be roots on all or nearly all of the cuttings, and all intermediate extents of rooting are possible. People doing this sort of experiment frequently ask how much replication is needed to "establish" a difference between, say, 40% rooting and 50% rooting; these people are often horrified by the amount of replication suggested by the formula for binomial variance. Education about binomial and extra-binomial variability is much needed. We may not yet be clear how much our biological colleagues should be taught about generalised linear models with binomial errors and the logit link function, but we ought to be clear that researchers in plant-propagation should be more familiar with the implications of the formula SQRT(pq/n).

Moving to the second of the 3 R's, randomisation, we come to a topic where theoretical work is still needed. Randomisation theory for designs with orthogonal block structures is well established and well known; and, in practice, randomisations for designs with standard blocking of this sort can now be produced very easily by computer, particularly by use of Genstat5. Just one rare type of orthogonal blocking structure has been causing trouble related to randomisation. This is the blocking structure of designs such as a Latin cube when the cube is used as a three-dimensional design with three systems of blocks, these being the design's three sets of "layers" parallel to the faces of the cube. If the layers of the cube are randomised separately in each of the three systems, then, as we can see from Professor J.A.Nelder's randomisation theory, the analysis of variance for the design must have seven strata (not four, as some people have supposed); this is because strata arise not only for the main effect of each blocking system, but also for each of the two-factor interactions between blocking systems, and finally for the three-factor interaction.

Notwithstanding the well established place of randomisation in our subject, we still lack satisfactory theoretical work on randomisation for change-over designs where residual effects of treatments are to be estimated, and for the superimposition of one set of treatments on another - as when a set of orchard trees is to be used for a new experiment whilst still carrying residual effects of the treatments of a previous experiment. For a few very special types of superimposition, randomisation procedures are known that are valid in the usual senses, but otherwise we don't have such procedures and - worse - we seem not to know the consequences of not having them. This lack is intellectually unsatisfying to those of us who are aware of the theoretical background to our statistical practice, and unsatisfactory for our teaching: we regard randomisation as so important that we make it one of our 3 R's, we stress it for block designs and row-and-column designs, and then suddenly we go quiet about it when we don't know how to do it! For change-over designs, Professor H.D.Patterson of Edinburgh University has suggested to me that our need is for randomisation procedures having approximate validity, and I imagine that the same is true for superimposition of one set of treatments on another. But how approximate? And we must be prepared for some unpleasant theoretical surprises. For example, for certain situations where we need a balanced superimposition of one Youden square on another, I know of a valid randomisation procedure for the superimposition, but it's a strange one: it requires the second Youden square to be selected at random from a certain set of possibilities, but selected with unequal probabilities, and different possible outcomes are of different statistical efficiencies. Which of us would be prepared to let the efficiency of a design depend on the outcome of a randomisation procedure, and which of us would be prepared to risk losing a useful amount of efficiency in order to achieve strict validity in randomisation?

Just one more randomisation topic deserves mention in this talk: "restricted randomisation" or - in the American usage - "constrained randomisation". Many of us have known experimenters who worry about what they regard as "bad" randomisations: outcomes of the randomisation process that seem to have some undesirable property - such as having treatment H at the left-hand end of every block. These experimenters say: "When we get a randomisation like that, please may we throw it away and have another shot by obtaining a fresh randomisation?" Indeed, unloved randomisations have frequently been thrown away and replaced - without, I suspect, any great harm being done. But this invites two comments. Firstly, a dislike of having treatment H at the left-hand end of every block suggests a belief that there may be an end-effect; perhaps the ends of the blocks should have been omitted from the experiment, or perhaps the blocking was in the wrong direction and should have been parallel to the "ends", or perhaps a row-and-column design (with its 2 systems of blocks) should have been considered in place of a design with just the one set of blocks. So an objection to particular outcomes of a randomisation process should perhaps be seen as an objection to the block-structure that was being considered. But another possibility is to use a procedure of restricted randomisation, and my second comment is that there should be more awareness of what restricted randomisation is about and what the possibilities are. The basic idea is this. You don't like some of the possible outcomes of the standard randomisation procedure, so you try to find a subset of the outcomes, this subset consisting only of acceptable outcomes and such that random selection from the subset satisfies the same criteria of validity as the standard randomisation procedure does. Such subsets can indeed be found for particular examples, and are related to groups from mathematical Group Theory. Such a subset can be expected to exclude not only the outcomes that we wanted to reject as being very "patterned", but also a balancing set of outcomes that might be described as very "unpatterned". Anyone wanting to follow this topic up can consult some of the many papers on the subject by Professor R. A. Bailey of Rothamsted Experimental Station, who is the expert on the subject.

Moving to the third of our 3 R's, we come to Blocking, so often described merely as a way of controlling nuisance-variability in experimental material or in the experimental environment. Today, however, I stress the second role of blocks: the control of variability introduced by what human beings do between the setting-up of the plots and the recording of the observations. I'm talking here about what Mr.G.V.Dyke and Professor S.C.Pearce have called the "management role" or "administrative role" of blocks - but I suggest that the name "operational role" might be less misleading. I'm saying that field operations such as weeding of the plots can themselves introduce variability: one man may weed more efficiently than another, and the differences may cause differences in the crop. Prudence then has different men doing different blocks, so that any simple difference in efficiency is confounded with blocks and so removable from experimental error. Likewise, the harvesting of a field experiment may be impossible to complete in a day, with the result that time-differences may intrude into the data. Prudence here suggests that no block should be left only partly harvested at the end of the day; then any simple temporal difference in the results is picked up as, again, a difference between blocks. This operational role of blocks seems to be recognised by very few experimenters, yet it should be borne in mind not only during the running of field experiments but also during their planning; operational considerations may or may not determine the configuration of the blocks to be used. The operational role can be summed up in the slogan "DO IT BY BLOCKS", but some people have misunderstood this to mean "Have blocks in your experiment", so I now prefer the slogan "DO IT BLOCK BY BLOCK".

****

By letting Cretté de Palluel's experiment lead me into my talk, I have bypassed any serious consideration of why we do experiments anyway. We may say: Oh we do them in a spirit of scientific enquiry, we all see comparative experimentation as part of the scientific process. But Cretté de Palluel's purpose was hardly within what we nowadays call fundamental science. Professors M.J.R.Healy and J.A.Nelder have drawn our attention to the distinction that can be made between "science", in a restricted sense, and "technology". If we make that distinction, many of us are more concerned with agrotechnology (or hortotechnology - if there is such a word!) than with fundamental science. So we should perhaps ask whether the criteria for good experimentation should be the same for our more technological work as for our more scientific. Indeed many of us have had pressure put on us to design field trials to have a demonstrational role as well as a scientific one. That pressure is common in connection with experimentation in developing countries. Another distinctive role for experiments is a legalistic one, as when experiments are used to decide whether a new drug or new crop-variety is to be released. Scientific open-mindedness is not then required; instead, a firm decision must be made, which an aggrieved manufacturer or customer could challenge in a court of law. So; scientific, demonstrational, and legal: do we here have three possibly conflicting roles for comparative experiments?; should we have three separate sets of criteria for good experimentation? Certainly we have a conflict if we are asked to have unrandomised treatments in order to provide a better demonstration for the benefit of visiting farmers or growers. My usual response then is to say that compromise isn't possible; if you want a demonstration, fine, have one - but don't try to claim it as a scientific experiment. To deal with legal needs too, there will have to be scientific sacrifice: a set of pre-specified rules will have to be followed every time, and seen to be followed, even if the spirit of scientific enquiry would have suggested changes in particular instances. The need for more thought on these matters would perhaps have been clear were it not for the recipe-book approach that many poor experimenters have adopted towards experimental design.

****

So far, in this talk, I have dealt with general principles and not with specific examples of new or newish types of design. Nor have I yet said anything about combinatorial aspects of the mathematics of experimental design. However, combinatorial work continues. For example, I have recently been working with Mr.

P.J.Owens of the University of Surrey on enumerating the distinct complete sets of mutually orthogonal 9 x 9 latin squares, where the word "distinct" has a fairly straightforward meaning; since these complete sets come from four different mathematical structures, and the different structures have different amounts and kinds of inherent symmetry, the job is not easy. I think we have finished it, but I thought it had too little general statistical interest for more than passing mention in this talk.

No, the emphasis of my talk has reflected, in part, my belief that it is all too easy to slip into bad statistical practice by neglect of fundamentals and by taking general principles too much for granted - and it is all too easy to get carried away by the sort of mathematical enthusiasm that has led so much time to be spent on constructing more and more examples of various different types of incomplete block design for which no practical need was clearly apparent. Also, though, my emphasis has reflected my lack of personal involvement with the development of - for example - the alpha-designs that are now so important for cereal-variety trials, or the nearest-neighbour designs that are being examined in connection with neighbour-methods of analysis for agricultural field experiments. The development of alpha-designs and related designs does, however, tie in well with some of what I have been saying. These designs were developed in connection with statutory variety trials, so the link with legal thinking is obvious; also, good alpha-designs are now generated by an algorithm that is available on an IBM micro, so the bond between design and modern computing is manifest. In a recent review-paper dealing with the development of the designs for the variety-trials, Dr.L.J.Paterson of Heriot-Watt University said that "fruitful research in design seems likely to continue to be towards developing flexible mathematics that can exploit computers."

As for the nearest-neighbour work, of which there has been so much, I have not been able to keep up with it, and I wait until the dust has settled. I am clear that some wild claims have been made for some of the methodology, and that any claims that are made should be questioned carefully. Also, we must remember that agricultural experiments are done on various different types of land: sometimes research-station land, sometimes farmers' fields or growers' orchards; sometimes land that has been in cultivation for years or decades or even centuries, sometimes land newly claimed from bush or forest or perhaps sea. Refinements of long-established methods of design and analysis may well prove to be beneficial for some some types of land and environment, not for others.

So, I cannot see far into the future of experimental design in agricultural and horticultural research. If so-called civilised man again resorts to large-scale war, perhaps there may be no such future. Equally, if man were to decide that little further agricultural and horticultural research was justified, there would be little future for the subject - which would then have served its purpose and be set aside. If there is indeed a rosy future for the subject, let us do what we can to make it an intellectually active future.

****
****

Dr. Preece gaf een uitstekende voorzet voor het informele deel van de Conferentie voor Landbouwstatistici, door het op de volgende bladzijde lied ten gehore te brengen. Menig statisticus zal zich hierin herkennen ....

Ontvangen: 08-05-1989
Geaccepteerd: 08-05-1989