Problems in the statistical design and analysis of feeding trials

By J. R. MORTON and W. J. RIDGMAN, *Department of Applied Biology, University of Cambridge, Pembroke Street, Cambridge CB2 3DX*

The problems that arise in the statistical analysis of feeding trials are more often due to mistaken philosophy than to particular technical difficulties related to the analysis of the trials. For example, the title originally proposed for this paper was 'Problems of applying statistical designs and analysis to feeding trials', which suggests that the nutritionist would happily conduct his trials in isolation, were it not for the need to adapt them to satisfy the peculiar demands of the statistician. The problem that in fact we face is to design feeding trials in such a way that they will produce truly meaningful results which are of use to the rest of the world.

At its simplest, an experiment in applied science asks whether treatment A differs from treatment B in its effect on some measurable character. The experimenter then tests A and B on representative samples of the population to which the result is to be applied. By statistical analysis he can compare the difference obtained with the variation between items making up the samples and, after making some assumptions about the distribution of this variation, may be able to claim that if the effects of A and B are truly the same he has witnessed an unlikely event, and he can say how unlikely. It is then up to the experimenter and his readers to decide whether they believe that an unlikely event has occurred or that there is a true difference.

The theoretical concepts of truth and representative sample lead to great difficulty in interpreting and hence designing feeding trials. The nearest that statisticians can get to defining truth is to call the true (or expected) value the average over the whole population, and for them the ideal material is an infinite population of random variates, which will obviously differ from a real population of animals with which we have to deal. Similarly, a truly representative sample is a random sample which is drawn rigorously so that every item has the same chance of being selected, but it is much simpler conceptually and practically to generate a set of random numbers from a particular distribution using a computer than it is to select a random sample of pigs intended to represent all pigs now present or likely to be bred in the foreseeable future. The nature of the sample of animals probably accounts for many of the discrepancies between results from different centres. The best that can be said of feeding trials is that the results refer to the animals used and any other population of which they could be considered a random sample, but the latter is difficult to assess.

This problem becomes particularly acute where, for purposes of economy, experiments are carried out on an institute's own herd of animals. This practice

could lead to the erroneous assumption that exact comparisons can be made between experiments on the same stock. More importantly, the institute's own herd may be atypical genetically or in its status in respect of, for example, immunity or vitamin reserves. At least one odd herd was recently found by F. J. Dieguez (personal communication 1976, Morton, 1976) when the strain of Durocs from one Cuban pig breeding centre was shown, atypically, to be leaner than their Yorkshires. Even where the effect of breed is taken into account, there is a tendency to treat breeds as homogeneous entities, so that, for example, few comparisons of the effects of dietary treatments on different breeds attempt to obtain random samples of the breeds.

Where experimenters have been aware of the problem their solutions have been less than ideal. The trial of Cole, Brooks, English, Livingstone & Luscombe (1975) used diverse centres and genotypes, but they did not suggest that these were a random sample of any population. Similarly, Fowler & Livingstone (1971) seemed to be attempting to allow for future as well as present pig populations by using minimal disease, specific pathogen-free and conventional pigs in their experiments. Even the large co-ordinated trials of Braude (e.g. Braude & Rowell, 1957; Braude & Hosking, 1975) were conducted on experimental farms where pen sizes are smaller than usual, and the management presumably better. The latter may represent future conditions of pig fattening, but the former surely does not. It may be that co-ordinated trials on more commercial units might be worth while, even at the expense of the degree of control and accuracy and completeness of results seen in Braude's trials.

The importance of random sampling depends on the importance of interactions between diet and type of animal or accommodation. There have been a few attempts to assess this in pigs, with variable results. Braude, Townsend, Harrington & Rowell (1962) found treatment×centre interactions to be unimportant and C. Rico (personal communication 1976, Morton, 1976) found little evidence of changes of ranking of genotypes, within or between breeds, when fed on cereal, molasses or swill-based diets. On the other hand, Braude & Hosking (1974) found significant interaction between protein levels in the feed and co-ordinating centres, which accounted for approximately the same amount of variation as the replicates×protein levels within centres. Also Cole *et al.* (1975) reported variations between co-ordinating centres in the degree of interaction between storage and processing methods of barley, which was a major effect on feed utilization and growth. (This might have been more easily seen if the cases where there was storage×processing interaction had been published as 2×2 contingency tables.) Braude & Hosking (1974) discussed whether their findings should encourage an increase in the number of centres to provide a wider range of genotypes and environments over which the effects of the diets are averaged, or a more regular use of replication within centres to detect such interactions, and on balance favoured the former. Yet knowledge of the presence or absence of diet×centre interaction is surely of prime importance in knowing the degree of general applicability of the results.

It could be argued that when animals are grouped together in pens a small degree of replication of treatments over pens can be tested against the variation between animals within pens; then if the former is clearly no greater than the latter, the two can be combined to provide a much improved general error term for the treatment×centre interactions. The counter argument to this would suggest that the equivalence of these two mean squares was fortuitous, for example due to the variance of pen effects happening to equal the variance due to competition between pigs within pens; under these circumstances the combined error would be invalid rather than improved. One needs to be convinced that the test of replicated pens against variance within pens is a valid test of the absence of pen effects before making such error combinations.

Of course, there are experiments in which the experimenter will claim that he is not proposing to tell the world what will happen to other animals but is concerned to detect small changes due to treatments, in order to build up general hypotheses. In this case he will want simply to reduce the variability between observations. There are quite satisfactory ways of doing this. With pigs, for instance, variation within litters may be much less than that between litters, so that, by allocating to each treatment one piglet from each of several litters, the effect of litters is removed from both treatments and the error term and the treatments are tested very precisely. The experimental piggery built on the advice of a statistician at Cambridge in the 1930s was capable of detecting very small differences (about 2·5%) in the many experiments on feeding bacon pigs by Woodman, Evans and collaborators (e.g. Woodman, Evans, Callow & Wishart, 1936; Evans, 1955). It must be stressed that such balancing must be part of the design and the effects removed in the analysis. Balancing by itself and still taking the variation between pigs receiving the same diet as the error term is worse than useless because the error term may be grossly inflated. Similar arrangements may be made when animals of different initial weight or breeding are included in the same experiment, but where a variety of animals is used in a feeding trial, one must beware of simplifying assumptions in the analysis of the effects. Thus the equation, maintenance requirement = body-weight$^{0·73}$ is used very widely and sometimes thoughtlessly. Yet the equation was originally proposed specifically for mature animals and is far from the truth for growing animals (Brody, 1945). Since different breeds and strains, and even animals within strains, mature at different rates and lay down protein and fat at different rates, designation of the energy of feeds to these different purposes requires great care and may not be well represented by simple linear addition of the components (Monteiro, 1975).

A similar problem arises with diets of ostensibly similar value but made up of different ingredients. At the simplest level there is the assumption that the ingredients are constant. Simple chemical analysis shows that this is untrue (Lerman & Bie, 1975), and it may be that differences exist between the nutritive value of, say, barley from different harvests that simple chemical analyses will not reveal. Thus it may be just as dangerous to extrapolate results from one small parcel of barley to barley in general as it is to extrapolate from a few of the

institute's pigs to pigs in general. More sophisticated assumptions are often made such as, that the digestibility of a diet may be estimated by the proportionally weighted linear combinations of the digestibilities of each ingredient, or that the amount of sulphur-containing amino acid needed in a diet is the same regardless of the source of the remaining protein. The first of these has been tested by Frape, Tuck & Boxall (1976) and found to be true for the diets they used; the second by Braude & Esnaola (1973) and found to be doubtful. As Lerman & Bie (1975) showed, uncritical modelling can lead to spurious dietary recommendations.

Frape *et al.* (1976) used a latin square design which appears particularly attractive to nutritionists because many observations can be obtained from few animals. In essence each animal receives each treatment in sequence and it is possible to remove from the error term the variation between animals and that between periods of time when the diets are fed. But for the ordinary analysis to be valid there are hard assumptions to meet. The first is that the true effect must be the same in each period; if the periods extend over much of the growth of an animal, this is clearly untrue. Secondly, there is the possibility of carry-over (often called residual) effects. Frape *et al.* (1976) were convinced that these were negligible in their experiment, and methods exist for estimating them in cases where they do not exceed the length of the subsequent treatment (Cochran & Cox, 1957). But where, for example, they may affect the microflora of the rumen, or the whole of the subsequent course of a lactation, they must obviously not be used.

Pearce (1969) has pointed out that there are fundamentally three methods of designing and analysing experiments. First the experimenter may think up an hypothesis and design an experiment to test it, either by the classic F-test of the null hypothesis, or by seeing the extent to which the model created by the hypothesis accounts for the variation (as in the experiment of Frape *et al.,* 1976). Secondly, the hypothesis may already have been shown to be well founded, or may itself be trivial, such as that two different diets are different, and interest lies in the value of the parameters which are produced by the model of the hypothesis. These experiments are best designed by concentrating on the specific contrasts whose difference one wishes to measure. In the case of the trivial hypothesis no problems arise, but in more complex hypotheses the dangers discussed earlier with regard to the assumptions made in the hypothesis must be remembered. Thirdly, one may use a form of analysis to generate an hypothesis, which is the particular use of multivariate analysis. What is very tempting, but clearly invalid, is to attempt to use the same data to generate, test and estimate the parameters deriving from the hypothesis.

A nice example of the third method of analysis is provided by Fenlon & Beever (1976). They took an apparently incomprehensible matrix of thirty-three determinations of the loss or gain of sixteen amino acids across the rumen of cannulated mature sheep and used multivariate methods to organize these data into patterns. Fenlon & Beever are careful to say that the point of such analysis is to suggest experiments that might usefully be performed on grazing sheep. One obvious conclusion from their analysis is that lambs fattened on a barley-based diet

may require more general protein but less lysine supplementation than lambs fattened on other diets. But a further experiment would need to be designed to test this hypothesis, and probably a third experiment to estimate the level of supplementation needed.

The other main use of multivariate analysis is more of a counsel of despair where it appears that no other method of analysis is valid. But the interpretation of such analysis is often difficult compared with a univariate method. A case in point is the analysis of data in which repeated measurements are made on the animals subjected to various treatments which was well explained by Cole & Grizzle (1966). Rowell & Walters (1976) recently showed that such experiments can be analysed by univariate methods in which each contrast of interest is tested against its own particular error term. It is seldom any longer sufficient to know that some activity is not the same at all times or that the effects of several treatments over time are not the same. We are more interested in how it varies and which treatments cause what change, and this can be obtained only by asking our question in unambiguous terms.

Perhaps the general conclusion from this brief survey is that many of the problems of the design and analysis of feeding trials disappear when the experimenter is sufficiently clear about what he wishes to compare and to what population of animals he wishes this comparison to refer, and that the design is simple enough and the replication adequate to give unambiguous answers.

REFERENCES

Braude, R. & Esnaola, M. A. (1973). *Br. J. Nutr.* **30**, 437.
Braude, R. & Hosking, Z. D. (1974). *J. agric. Sci., Camb.* **83**, 385.
Braude, R. & Hosking, Z. D. (1975). *J. agric. Sci., Camb.* **85**, 263.
Braude, R. & Rowell, J. G. (1957). *J. agric. Sci., Camb.* **48**, 322.
Braude, R., Townsend, M. J., Harrington, G. & Rowell, J. G. (1962). *J. agric. Sci., Camb.* **58**, 251.
Brody, S. (1945). *Bioenergetics and Growth.* pp. 404–5. Baltimore, Md: Reinhold.
Cochran, W. G. & Cox, G. M. (1957). *Experimental Design* 2nd Edn. pp. 133–139. New York: Wiley.
Cole, D. J. A., Brooks, P. H., English, P. R., Livingstone, R. M. & Luscombe, J. R. (1975). *Anim. Prod.* **21**, 205.
Cole, J. W. L. & Grizzle, J. E. (1966). *Biometrics* **22**, 810.
Evans, R. E. (1955). *J. agric. Sci., Camb.* **46**, 329.
Fenlon, J. S. & Beever, D. E. (1976). *J. agric. Sci., Camb.* **87**, 255.
Fowler, V. R. & Livingstone, R. M. (1971). *Anim. Prod.* **13**, 59.
Frape, D. L., Tuck, M. G. & Boxall, R. C. (1976). *J. agric. Sci., Camb.* **86**, 325.
Lerman, P. M. & Bie, S. W. (1975). *J. agric. Sci., Camb.* **84**, 459.
Monteiro, L. S. (1975). *Anim. Prod.* **20**, 315.
Morton, J. R. (1976). *Bull. Brit. Cuba sci. liaison Cttee* No. 7. p. 5.
Pearce, S. C. (1969). *Exptl Agric.* **5**, 67.
Rowell, J. G. & Walters, D. E. (1976). *J. agric. Sci., Camb.* **87**, 423.
Woodman, H. E., Evans, R. E., Callow, E. H. & Wishart, J. (1936). *J. agric. Sci., Camb.* **26**, 546.

Printed in Great Britain