Producción Animal Tropical 1983 8:288-307

THE DESIGN AND INTERPRETATION OF ANIMAL FEEDING TRIALS IN THE TROPICS

J R Morton and D Wyllie

Department of Biology, University of Papua New Guinea, Port Moresby PNG, School of Agriculture, West Mains, Edinburgh, Scotland

Various problems which arise in the statistical design and analysis of feeding trials in the tropics are considered. Examples are used from the literature to show that in many cases the design and interpretation of experiments in tropical animal production could be improved. In general, it is apparent that greater clarity is required on the objectives of experiments. A consideration of the analysis that will be applied to the experiment before it is carried out is essential to determine if the design will answer the questions posed. With some prior thought and common sense, a given number of animals and effort can be made to yield far more meaningful results than is often the case.

Key Words: Experimental design, statistics, tropics, animal nutrition, feeding trials.

In recent years the emphasis in tropical animal production has moved away from attempts to ameliorate the tropical environment so as to permit the use of temperate animal systems and towards the exploitation of local sources of feed for animals able to withstand the local environment. This change introduces a number of problems in the design and analysis of experiments. The population of animals under test may be unfamiliar; predictions of response based on temperate experience are often inapplicable to tropical feed sources; and there is a lack of sophisticated modern computational equipment in the tropics where such experiments have to be carried out. A large computer can be, and too often is used to cover the inadequacies of the original experimental design. However, by careful design none of the problems mentioned is unsurmountable.

Problems in the design of animal feeding trials have been discussed in brief and general terms by Morton and Ridgman (1977). The purpose of the present paper is to apply these generalities to the specifics of tropical animal production.

In so doing we shall use as examples the experiments of a number of workers in the field. For convenience, these are taken largely from papers published in Tropical Animal Production, although similar examples can be found in most other journals dealing with tropical animal experiments. It was considered best to treat a few examples in detail rather than give a more general coverage using more examples.

The subject can perhaps best be divided into three topics represented by the questions:

a) Are the objectives of the experiment decided with sufficient clarity and will the design answer the questions that they pose?
b) What is the basic unit of variation in the experiment that will be used in the analysis?

c) Of what more general populations are the animals in the experiment a representative sample?

Specification of Objectives

As contrasts: When statistical analysis is taken beyond the simple F-test, the literature in animal production in the tropics abounds with the use of Least Significant Difference (LSD) or, more properly, multiple range tests such as that of Duncan (1955), either to decide which of the treatment mean differences are contributing to the significant F value, or directly to test for significant differences between means. LSD's based on simple t tests can lead to over-confidence in the significance of results where many treatments are compared. Although multiple range tests protect against this over-confidence, they are commonly not the most efficient way of breaking down the degrees of freedom for treatments, and inefficient statistics tend to produce wrong conclusions. In those cases where a multiple range test is the only proper mean of further analysis, this is often due to the inadequacy of the original experiment design, which itself commonly derives from a lack of clarity about the objectives of the experiment.

The danger in the use of simple LSD's can be illustrated from the work of Ferreiro et al (1977b) in which 8 diets are compared. In the results section of the paper one or two comparisons are pointed out as "significantly different". Yet if these were based on simple "t-tests", (and in the absence of a mention of any more sophisticated method we must assume that they were) it is clear that there are 28 possible two-way comparisons among 8 means and one of these is expected to be "significant at the 5% level" by chance alone.

Multiple range tests like Duncan's overcome this problem of the "expected number of significances", but their proper use is for the case where we have no hypothesis about the treatments, in what is sometimes described as a "look-see" experiment. Of course, occasions do arise in which our ignorance of the treatments that we are using is so complete that all we can do is to apply them, see if they have a significant effect and/or use a multiple range test to discover which of the treatment comparisons are in themselves significant. An example of such a case is the effect of various blood groups on chicken embryonic mortality (Gilmour and Morton 1970) in which paper the use of Duncan's test with and without a significant overall F-test is discussed. But an example from tropical animal production is hard to to find, since there are usually good reasons for trying, for example, certain supplements and whether or not it is expressed, each additional treatment is chosen to test a hypothesis.

It is thus a useful discipline to define a contrast, that is to define which treatments are to be compared, for each degree of freedom available. For the most efficient use of the data obtained, and hence the most efficient use of the facilities available, the contrast chosen should be orthogonal to one another. Thus *orthogonal contrasts*, which are treated in many tests as one of many methods of analysis, provide in fact a most useful design tool (Ridgman 1975 pp 71-80). But to analyse an experiment with properly defined orthogonal contrasts by LSD or multiple range test wastes the design and may lead to erroneous conclusions.

In factorial designs, the factors are orthogonal to one another and there exists a set of orthogonal contrasts for each main and interaction effect. Consider in this connection the work of Salais et al (1977). This consisted of a 22 factorial of ad libitum molasses/urea vs sugar cane stem (A) with coarse vs fine chopped cane top supplement (C), applied to 8 bulls in two Latin Squares. An additional factor of a supplement of rice polishings (R) was applied to one Latin Square.

This is a somewhat unconventional design in that it provides no error term for the interaction of the rice polishings with other factors. There exist basically two ways of analysing these data, as laid out in Table 1. Firstly the two Latin Squares can be analysed as a single experiment as shown in ANOVA (a) and the complete set of contrasts. The latter are set out in the conventional manner (see eg Ridgman 1975) with the means represented in lower case letters "a" representing the presence of the second ad libitum treatment (cane stem); "c" the fine chopped cane top, etc; and the capital letters the main and interaction effects of these treatments. Effects A, C and AC can be tested against the error within squares and R against the variance between animals. The effects AR, CR and ACR can be calculated but not tested. Alternatively the Latin Squares without and with rice polishings can be treated as two separate experiments, each with ANOVA (b) and with the orthogonal contrasts enclosed in the squares in Table 1.

Table 1:
Ways of analysing experiments with Latin squares confounded with one treatment.

	ANOVA (a)	ANOVA (b)		Orthogonal contrasts							
				(1)	c	a	ac	r	cr	ar	acr
Treatments	7	-	C	-	+	-	+	-	+	-	+
A, C & AC	3	3	A	-	-	+	+	-	-	+	+
R	1	-	AC	+	-	-	+	+	-	-	+
AR, CR, & ACR	3	-	R	-	-	-	-	+	+	+	+
Animals	6	3	CR	+	-	+	-	-	+	-	+
Periods	6	3	AR	+	+	-	-	-	-	+	+
Error within			ACR	-	+	+	-	+	-	-	+
Squares	12	6									

The election between both methods depends on whether the researcher considers the interactions of rice polishings with the other factors are important or not, because it is usual in factorial designs, as we will see below, that the estimation of main effect is not likely to be significant if an interaction involving it is statistically significant. In fact,this is a possible use of AR, CR and ACR constrasts, although it is statistically probable, these could confirm or reject researcher's preconceptions about their utilities.

Salais et al (1977) appear to have used a combination of these approaches since they tested R overall, and A, C and AC separately with and without R (see Table 2 reproduced from their Table 1). Although it would

seem quite legitimate in this case to use the combined error within Squares (with 12 df) for testing each of the two experiments, we shall continue the analyses completely separately as Salais et al presented them in Table 2.

Table 2:

The effect of fineness of chopping of cane tops and supplementation with rice polishings on the digestibility and consumption of diets of cane tops with molasses/urea or cane stem.

Diet	Molasses/urea		Cane Stem		Standard error of mean[3]
Form of cane tops	5-20 cm	Fine chop	5-20 cm	Fine chop	
DM digestibility %					
With rice polishings	72.7[a]	71.0[a]	61.0[b]	56.4[c]	0.50
Without rice polishings	75.9[a]	71.6[b]	60.8[c]	55.5[d]	0.55
Voluntary intake, kg DM					
With rice polishings	3.97[a]	4.12[a]	3.71[ab]	3.30[b]	0.17
Without rice polishings	3.49	3.15	2.94	2.84	0.23
Consumption index[1]					
With rice polishings	2.31[a]	2.34[a]	2.04[b]	1.83[c]	0.02
Without rice polishings	2.05[a]	1.87[b]	1.71[b]	1.47[c]	0.02
Digestible dry matter consumption index[2]					
With rice polishings	1.67[a]	1.51[b]	1.24[c]	1.03[d]	0.03
Without rice polishings	1.58[a]	1.36[b]	1.04[c]	0.82[d]	0.01

[abcd] Means with different superscripts differ ($P < 0.05$)

[1] Food intake, kg DM/100 kg live weight/day

[2] Daily digestible dry matter intake/100 kg live weight

[3] Each value is the mean of 4 determinations

Applying the contrasts in the rectangular boxes of Table 1 to the means with their error structure of Table 2 produces a table of main and interaction effects as shown in Table 3. It seems from this table that the effects on Voluntary Intake are essentially the same as those on Consumption Index, but of much lower significance, presumably simply as a result of the variation in body weight of the bulls used. Voluntary Intake will therefore not be further considered. In the Latin Square with rice polishings supplement, there is a significant interaction effect for DM digestibility and for Consumption Index. Either by consideration of the signs of the contrasts and mean effects or by reverting to the original table of means, we can see that this significant interaction arises because the deleterious effect of fine chopping of the cane tops is virtually confined, in each case, to two treatments with cane stem fed ad libitum. Thus although a significant average effect of chopping cane tops can be demonstrated, in the presence of the interaction this average is not meaningful, being made up of an effect in one half and the lack of an effect in the other half of the experiment. It could be claimed that the average advantage, in both traits of molasses/urea over cane stem, is so large relative to the inter-

Table 3:
Reanalysis of the experiment of Salais et al (1977) - Main and interaction mean effects[a]

Effect	DM Digestibility	Voluntary Intake	Consumption Index (CI)	Digestible DM
With rice polishings				
C	-1.6***	-.07	-.05**	-.09***
A	-6.6***	-.27*	-.20***	-.23***
AC	-0.7*	-.14	-.06***	-.01
Without rice polishings				
C	-2.4***	-.22	-.12***	-.11***
A	-7.8***	-.22	-.20***	-.27***
AC	-0.3	+.06	.01	0

[a]Mean effects in the sense of $\frac{m_1 - m_2}{2}$ where m_1 and m_2 are the means of two treatments

*P < .05

**P < .01

***P < .001

action that it remains real. But,statistically, we cannot say that some other, maybe even coarser, method of chopping the cane top would not bring the value of a cane stem diet up to that of one based on molasses/urea, although biologically we may be able to discount this possibility. Thus,as noted earlier, one needs to be most careful in ascribing meaning to a significant main effect, when there is evidence for an interaction involving it. Before leaving the results for the part of the experiment with rice polishings, we could ask why this interaction does not appear in the analysis of the digestible DM consumption index, which is only the product of the other two traits. It is because there is an error in the Table of Means presented by Salais et al for the treatment molasses/urea + fine chopped cane tops + rice polishings. 2.34 x 0.71 does not equal 1.51 but 1.66. Assuming that this last, derived figure is the one in error, it is clear that an analysis of the corrected figures for digestible DM consumption index would give results very similar to those for digestible DM and consumption index separately. It is a by-product of an appropriate statistical analysis that it will test through its logic those arithmetical errors which inevitably creep into talbes of results.

The results from the half of the experiment without rice polishings are much simpler. There is no evidence for interaction and the size and sign of the main effects are similar to the average effects for the other half of the experiment. Thus the results of the whole experiment could be summarised as follows.

"Molasses/urea is a superior ad libitum food to sugar cane stem as regards both digestibility and consumption. Supplementation with coarse

ground cane tops is superior to that fine ground in both measures, except that in a molasses/urea diet supplemented additionally with rice polishings this difference disappears".

This exemplifies the use of orthogonal contrasts in the main purpose of a statistical analysis to describe as simply as possible what the experiment has demonstrated. But they can also be used in experimental design each contrast representing the use of one of the degrees of freedom for treatment. The contrasts must, of course, be defined before the data are surveyed and it is of great advantage to define them while the experiment is being planned. We shall consider this case in relation to the experiment of Ferreiro et al (1977b), whose design is shown in Table 4. We shall use the results on daily gain (also shown in Table 4), although in this case the standard error appears to have been calculated on the between-animal rather than between-pen variance. We shall use the published standard error, for want of any other, but since there were two pens of 4 bulls per treatment, we shall take only 8 df for error.

Table 4:

Composition and amounts of supplements fed (g/animal/day) (Experiment of Ferreiro et al 1977b)

	A	B	C	D	E	F	G	H
Rice polishings	-	-	-	-	-	-	1000	1500
Maize grain	-	500	-	-	500	-	-	-
Fishmeal	-	-	125	125	125	187	-	-
Soybean	-	-	375	375	375	563	-	-
Maize oil	-	-	-	30	-	-	-	-
Total	-	500	500	530	1000	750	1000	1500
Mean values for performance characteristics (2 groups of 4 bull/ treatment - 98 days trial)								
Daily gain	.037	,051	,333 $\pm$.060	.517	.555	,669	,728	,651
				P < .001				
Ferreiro et al's contrasts								
1) B vs A	-1	+1	0	0	0	0	0	0
2) B + A vs rest	-3	-3	+1	+1	+1	+1	+1	+1
3) C vs A - B	-1	-1	+2	0	0	0	0	0
4) D & E vs C	0	0	-2	+1	+1	0	0	0
5) F vs D & E	0	0	0	-1	-1	+2	0	0
6) F vs G	0	0	0	0	0	-1	+1	0
7) H vs G	0	0	0	0	0	0	-1	+1
Duncan's test (means with a common letter not significantly different P <0,05).	c	c	b	ab	a	a	a	a

To quote Ferreiro et al (1977b) "The object of this experiment was to test a variety of supplements providing different combinations of essential amino acids, starch and unsaturated lipids, comparing them with rice polishings as the control treatment". It is clear from other parts of the paper that G is the control treatment and that the possibility was entertained that growth on some of the treatments might exceed G. The appropriate test among mean responses is then a Dunnett (1955) 2-tailed test against G, by which treatments A, B and C are significantly inferior to G. But a full reading of Ferreiro et al (1977b) shows that the objectives of the experiment was rather more than is indicated by the summary above. It was to discover whether response to rice polishings as a supplement to sugar cane/urea diets was due to their protein, starch or oil content or to some combination of these. From this standpoint, the Dunnett's Test enables us to conclude that it is neither the starch nor the protein alone which is responsible, a somewhat limited conclusion from a large experiment.

In the first two paragraphs of the Results section of their paper, Ferreiro et al (1977b) makes a number of comparisons among the treatment means, equal to the number of degrees of freedom for treatments and converted here to Ferreiro et al's contrasts in Table 4. Of these, contrasts 2 and 3 are said to be significant, contrast 4 to have an effect, 5 to appear to have an effect and 1, 6 and 7 to show no effect.

Recalling that the test for orthogonality of contrasts is that the product of each pair should sum to zero, we see that 3 is not orthogonal to 4 nor 4 to 5 to 6 nor 6 to 7. It may at times be necessary to design an experiment in which not all contrasts are orthogonal, but the degree of non-orthogonality here suggests that the comparisons were made after viewing the data. There is of course nothing to prevent the experimenter drawing the attention of the readers of his report to any comparisons he thinks of interest, but he cannot apply normal statistical tests to these comparisons. The above treatments 3, 4 and 5, for example, are 3 out of 5 possible comparisons among 8 treatments three at a time, of which nearly 3 would be expected to be "significant" at the 0.05 level even if no real effects existed. Such *a posteriori* comparisons require especial tests such as Duncan's (1955) to protect against such false significancies. The results of applying Duncan's test to these data is also shown in Table 4, confirming the result of the Dunnett's test and most of the conclusions of Ferreiro et al, but still not really answering the problem of the advantage of the rice polishings.

It becomes clear why the experiment is so difficult to analyse satisfactorily if (ignoring the results) one tries to choose a set of orthogonal contrasts for these treatments as applied, which will test hypotheses of interest. For the present authors this proved impossible, which indicates that the treatments applied cannot satisfactorily answer the questions posed. The use of orthogonal contrasts at the plannins stage would have detected the errors in design.

While many possible orthogonal designs could answer this problem, one is presented as an example in Table 5. Contrast I tests whether the effect of rice polishings can be accounted for by its protein, starch and oil content. If it cannot, this contrast being significant, in a sense the experiment is at the end. One might still, however, rescue some information from the other treatments by a Dunnett's test of them against "Protein

Table 5:

Proposed design to evaluate the reasons for the good response to rice polishings as a supplement to sugar cane/urea diets.

Treatments	Starch + oil	Protein + starch	Protein + oil	Protein + starch + oil	Rice Polishings	MDD*
Contrasts						
I	0	0	0	-1	+1	103
II a	-2	0	0	+1	+1	89
b	0	-2	0	+1	+1	89
c	0	0	-2	+1	+1	89

* M.D.D. = Minimun Detectable Difference (as increment only) in daily gain (g/animal/day) at $P < 0.05$

+ starch + oil as control". But if, as expected, contrast I is non-significant, we apply contrasts II a-c. If all are significant, then the effect of rice polishings is due to all the constituents acting together. If any two are significant and the third not, we know which is the non-essential constituent. If none are significant, then, depending on the level of significant difference we can detect (see below), rice polishings are not such a special supplement after all. Only if we achieve just one significant result among these three are we forced to do a further experiment. If say "a" alone were significant we would need to discover whether protein alone could provide the same boost to growth as protein + starch or oil.

By using 15 of the 16 pens of animals available to Ferreiro et al (1977b) 3 replications of this experiment could be achieved, with 10 df.By setting the levels of protein, starch and oil in the compounded supplements the same as those found in rice polishings, we would need to test only for positive effects of the contrasts and so could use a 1-tailed t-test. If the error variance were the same as in Ferreiro et al's (1977b) experiment, differences as small as those in Table 5 would be detected. Note also that the greatest powers of determination are reserved to the 3 contrasts of most importance.

Response curves and surfaces: If a Duncan's multiple range test of LSD is a method of last resort in the kind of experiments discusses so far, it is clearly inappropriate where the experiment is in the nature of a response curve (Mead and Pike 1975), that is where some form of yield is measured against varying levels of an applied commodity as in the experiments of López et al (1978), Ferreiro et al (1977a), Meyreles et al(1977a and 1977b) and Ffoulkes and Preston (1978a).

Where "yield" is estimated against various levels of a treatment, there is, at least subconsciously, an hypothesis that the two are related. If we have some theory as to the nature of the relationship, we can of course test this. More commonly in tropical animal proudction, no theory exists as to how the animals should respond, and rather we seek an empirical rela

tionship which will give productive responses within the range of levels of treatment used. One approach that has frequently been used is to fit a curvilinear response curve of the form $a + bx + cx^2$ by least squares. This generalised least squares method is essential in many cases, where, for example, the data are collected by survey of published data as in the review of NPN supplementation for dairy cattle by Satter and Roffler (1977) because the points on the abscissa are unequally spaced and each point has a different weight.

But in designed experiments, such as those of Silvestre et al (1977a, 1977b), Ravelo et al (1977) and Ferreiro and Preston (1977) such an approach is clumsy and requires care in testing whether the model is appropriate. For example in Silvestre et al (1977a), it is clear from inspection of the published graph that the relationship between live weight gain of Zebu bulls and urea concentration in molasses from 60-120 days after the start of the treatment might be fitted by a linear or a cubic response curve, and that the quadratic fitted is inappropriate.

In such cases, the fitting of an orthogonal polynomial is simpler and less liable to error. Orthogonal polynomials also have the advantage that response surfaces to various levels of two or more treatments can be simply fitted, because as we saw for orthogonal contrasts, the interaction polynomials are the product of the main response curve. The fitting of a response surface by generalised least squares may require greater computing facility than is available.

But, as an example, we will consider the fitting of a response curve to the data of López et al (1978). This concerns the effect of space per pig on growing performance in the Cuban winter. There were 20 pigs assigned to each pen and 3 pens to each treatment, allocated at random at each level, and the four treatments were achieved by using 4 different positions to the end barrier of the pen to provide 0.5, 0.6, 0.7 and 0.8 m^2 per pig. The data and analysis are shown in Table 6. Since the treatment levels are equally spaced, the orthogonal polynomials can be copied directly from Fisher & Yates Tables (eg 1957), the introduction to which provides as good instructions as any to their use. López et al (1978) provide a standard error for the mean daily gains over the whole experiment, but as will be explained later, there is doubt as to whether the unit of variation used in its calculation was pens or pigs. The mean squares have therefore been expressed in multiples of r, (the replication), which does not however affect the variance ratios, which are clearly significant for linear $(\frac{1}{1})$ quadratic $(\frac{1}{1})$ and cubic $(\frac{1}{1})$ effects.

The response curve for overall daily gain shows a maximum at .608 m^2/pig and a minimum of .766 m^2/pig, and the curve is shown in Figure 1. It is much more difficult to think of a biological reason for the minimum than for the maximum. For the partial stages of fattening, López et al (1978) give no measures of error, so that no analysis of variance is possible. So for consistency the same model is fitted to these stages and the maximum growth rate of each calculated, with the results in Figure 1 and Table 6. The overall conclusion, perhaps banal, is that larger pigs need more space. But it is easy to calculate from this analysis that if the pigs were provided with the ideal amount of space at each stage they would be expected to average 582 g/d, while if given the best space for the overall

Table 6:
Growth rate (g/d) of pigs at different densities (after López et al)

Area per pig	.5	.6	.7	.8	m^2	
Growth period						
33–65 kg	548	524	498	502		
65–100 kg	524	608	569	584		
100–110kg	417	610	615	602		
33–110 kg	516	573	556	550	± 7.0 (SEM)	
Orthogonal Polynomials					33–100 kg MS	VR
$\}_1^1$	-3	-1	+1	+3	361r	7.35
$\}_2^1$	+1	-1	-1	+1	992r	7.35
$\}_3^1$	-1	+3	-3	+1	361r	7.35
				Error	49r	

Response Curve

33–110 kg $Y = 548.75 + 4.25\}_1^1 - 15.75\}_2^1 + 4.25\}_3^1$

or $Y = 568.44 + 205.4\ (x - \bar{x}) - 1{,}575\ (x - \bar{x})^2 + 14{,}167\ (x - \bar{x})^3$

Ymax = 573 at x = .608, Ymin = 546 at x = .766

33–65 kg $Y = 518.00 - 8.20\}_1^1 + 7.00\}_2^1 + 1.60\}_3^1$

or $Y = 509.25 - 2743.31\ (x - \bar{x}) + 700\ (x - \bar{x})^2 + 5{,}333\ (x - \bar{x})$

Ymax = 550 at x = .468

65–100 kg $Y = 571.25 + 7.05\}_1^1 - 17.25\}_2^1 + 8.85\}_3^1$

or $Y = 592.80 - 463.8\ (x - \bar{x}) - 1{,}725\ (x - \bar{x})^2 + 29{,}500\ (x - \bar{x})^3$

Ymax = 608 at x = .595

100–110 kg $Y = 561.00 + 28.0\}_1^1 - 51.5\}_2^1 + 8.5\}_3^1$

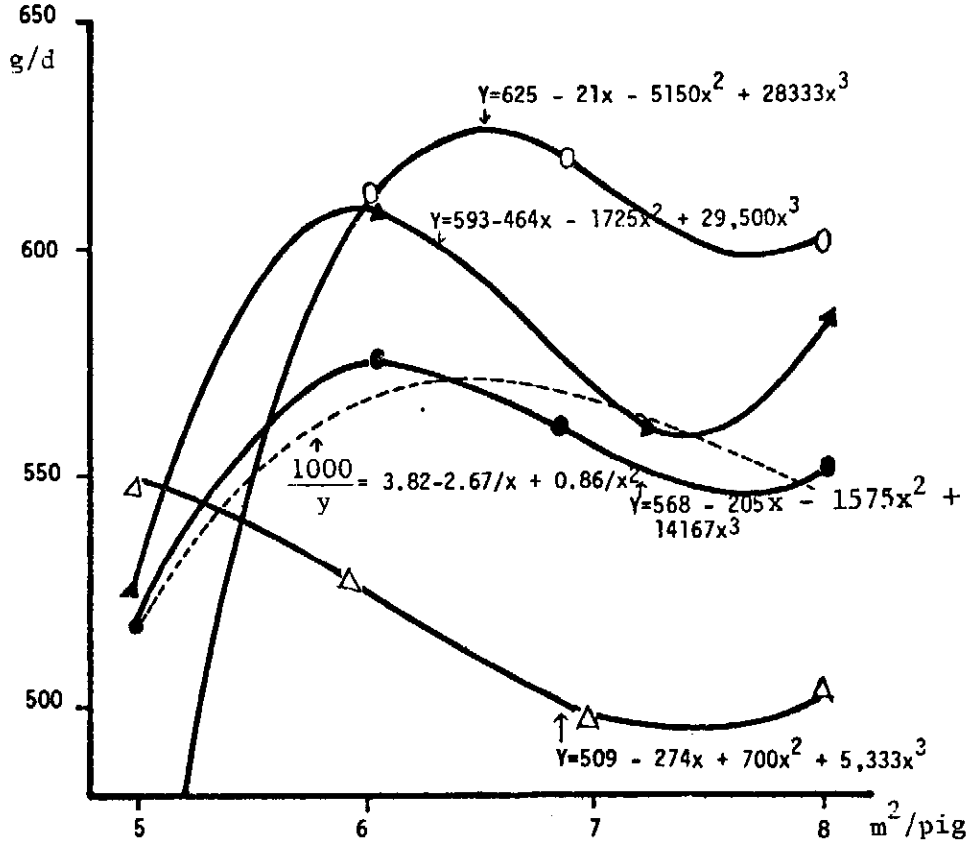
or $Y = 625.38 - 20.8\ (x - \bar{x}) - 5{,}150\ (x - \bar{x})^2 + 28{,}333\ (x - \bar{x})^3$

Ymax = 625 at x = .652

fattening period they would be expected to gain 573 g/d, a comparison that would be of value in considering different ways of designing pig accommodation.

The minimum value of Y produced may be an artifact of the analysis. This is because of the way in which fitting orthogonal polynomials works. First we fit mean values and a linear response. If this is unsatisfactory we superimpose a parabola (quadratic response), and so on for cubic (sigmoid form) and maybe quartic and quintic responses, each additional power of x giving an additional stationary point to the response curve.The problem is that reasonable biological mechanisms seldom produce more than one maximum or minimum and some authors (eg Ridgman 1975 p 101) suggest using all terms beyond the quadratic as a test of goodness of fit of the quadratic model. On this basis, the significant cubic effect in the data

Figure 1:
Fitted curves to the growth rate of pigs relative to the area available to each pig from 33-65 kg (△), 65-100 kg (▲), 100-110 kg (○) and 33-110 kg (●) (After López et al 1978)



Note: Y = expected value of y (growth rate in g/d)

of López et al (1978) merely tells us that quadratic (space per pig) does not provide a good model of the biology involved. Similarly the fitting of the cubic implies either that the growth rate of pigs increases with increased space for all distance, or that there exists a second maximum at larger areas/pig and that a quartic (at best) response curve is needed. Although extrapolation of a polynomial fit outside the experimental points is most dangerous, the behaviour of the model beyond these points should be reasonable (Nelder 1966) and here it is not. It is biologically reasonable to suppose that beyond some given area/pig provision of further space has negligible effects, that is, the response curve approaches some asymtotic values with increasing x, perhaps the most common form of curve in biology which is not well fitted by orthogonal polynomials.

Where the response to a treatment reaches an asymtote at high levels of the treatment, some form of inverse polynomial is required. The simplest of these is the linear regression of the reciprocal of y against the reciprocal of x. Unless a specifically exponential form of response, such as that of Mitscherlich (1909); is expected, for which special methods of fitting exist (see eg Snedecor y Cochran 1967), the family of curves produced from the various possible linear regressions of y^{-1} against x^{-1} provide the simplest method of fitting an asymptotic response. But in the case of the data of López et al (1978) we have clearly a maximum before the asymptote, so that an additional term is needed. The best known inverse polynomial of this kind is Nelder (1966), which takes the form

$$y^{-1} + \beta_o x^{-1} + \beta_1 + \beta_2 x$$

However, this response form is for the case, quite common in applied biology, where the treatment, while benefitial at low levels, is toxic and finally lethal at higher levels, so that the asymptote where x is very large is $y \cong 0$. This is not the case in López et al's experiment, and we must obviously try the simple inverse polynomial.

$$y^{-1} = \beta_o + \beta_1 x^{-1} + \beta_2 x^{-2} + \text{etc}$$

Although this can be fitted by the usual method for orthogonal polynomials, the levels of treatment expressed as 1/x are no longer equally spaced and the $\}^1$ need to be calculated, which is arithmetically tedious. (The corollary of this is that where an asymptotic response curve is expected the treatment should be applied at equal spacings of 1/x; even in the case of the simple reciprocal linear regression the facilities will thus have been more efficiently used). Therefore the inverse polynomial was fitted by generalised least squares. The relationship found

$$1000/Y = 3.82 - 2.67/x + 0.86/x^2$$

has a residual mean square in Y of 176 r and hence a non-significant variance ratio of 3.41, so that it can be considered a satisfactory fit. It should be noted that in an inverse polynomial the invariance term, here 3.82, is the asymptotic value fitted for x very large. Since, in the fitted curve, the approach to the asymptote is very gradual, it is not well estimated. If we needed to know whether a second degree inverse polynomial was a good fit and to have a good estimate of the asymptote, we should need to perform another experiment with equally spaced values of 1/x ranging perhaps from 2 to 1/3.

However, practically, we are much more interested in finding the value of x for which Y is a maximum. The inverse polynomial gives this as Y = 573 g/d at .608 m^2/pig). An experiment to locate this maximum accurately or perhaps better, a series of experiments to detect the maximum for different ages or weights of pig, might be more useful.

To detect a maximum, at least three levels of treatment are required. The closer they are to the truer maximum, the more nearly will they approximate to a parabola, which is defined by a quadratic term of the orthogonal

polynomial. We therefore require three points equally spaced in x (m^2/pig) far enough apart to give a significant quadratic term, but not so far apart that the approach to the asymptote distorts the parabola. A possible experimental plan is shown in Table 7. The final phase of this plan, from 16 - 18 weeks post-weaning, is very similar to the final phase of the experiment of López et al (1978) from 100-110 kg,which took about 16 days. From that experiment we could predict that the growth rate of pigs at the three spacings would be 610, 625 and 615 g/d respectively. From the variance between pens of that final stage (not published) the degree of replication necessary for a significant quadratic response could be calculated. The pens of pigs would be randomly redistributed to treatments at each stage and finally the relationship between space per pig for maximum growth versus age of pig could be established. Such a relationship would contain all the biological information necessary for the optimisation of the design of accommodation for fattening pigs in the Cuban winter.

Table 7:

Metre ^{2}/pig at various ages to determine point of maximum growth.

Weeks past weaning	Space		
	Small	Medium	Large
1 - 3	.40	.45	.50
4 - 6	.44	.49	.54
7 - 9	.48	.53	.58
10 - 12	.52	.57	.62
13 - 15	.56	.61	.66
16 - 18	.60	.65	.70

Unit of Variation

In designs such as the split plot, the experimenter is unable to avoid consideration of what constitutes the basic unit of variation in his experiment. But in trials with pens of animals this is often overlooked.

Ffoulkes and Preston (1978b) report on the effects of certain feed supplements. Twelve Zebu and four Holstein x Zebu bulls were allocated according to liveweight to give the two replications of two animals per treatment. In the design of the experiment, pairs of bulls represented the basic unit of variation, but in the analysis daily gains were calculated by the linear regression of individual weights against time (emphasis added). Not only were the pairs of bulls, rather than the individuals, randomised over the experiment, so that they became the proper basic unit for the extinction of natural error, but the pairs were balanced as to weight, increasing the variation between bulls within pairs. Indeed the coefficient of variation of daily gain was more than twice that of other measures, based on bull pairs, in the experiment. As it happens the only difference between treatments that reached conventional statistical signif

icance were those for daily gain, despite the inflation of the error term, because the replication by pairs of bulls was so inadequate. This can be better appreciated by considering what was presumably the skeleton analysis of variance. The provision of some indication of the breakdown of the ANOVA in the published results even of by way of a sample ANOVA is always of value in understanding the experiment. This was

Source of variation	Degrees of freedom
Replication (R)	1
Soy bean level (S)	1
Forage source (F)	1
S x F	1
R x treatments	3
Between bulls within pairs	8

The effect of balancing pairs of bulls by liveweight is to increase the variance in the term "between bulls within pairs" at the expense of the term "R x treatments". This is an entirely proper statistical trick provided that "R x treatments", representing the natural variation between pairs of bulls is used as Error. In this particular case the inflated invalid error term "Between bulls within pairs" gave the more significant result because the valid error term, which had been reduced by balancing, had only 3 degrees of freedom, which cursory perusal of a table of F will show to be too few.

A second example, in some ways more extreme, in others more subtle, is provided by the experiment of López et al (1978) concerning the growth of pigs with varying amounts of area per animal (0.6, 0.6, 0.7 and 0.8 m^2 per pig). This last way of expressing the treatments makes it appear that they are applied to the individulal pigs. But in fact the treatments applied are of 10, 12, 14 and 16 m^2 to groups (pens) of 20 pigs, so these pens are the basic unit of variation. Aside from this caveat, it is unwise in experiments involving animal behaviour to treat numbers and density as alternative and equivalent measures, that is 2 pigs in 1 m^2 may not be at all the same thing from the pig's point of view as 20 pigs in 10 m^2. The skeleton anovar of this experiment can be set out as in Table 8.

Table 8:

Source of Variation	df	Components of variance
Treatments	3	$\sigma^2 + 20\,\sigma^2_p + 60\,\sigma^2_t$
Pen within treatments	8	$\sigma^2 + 20\,\sigma^2_p$
Pig within pens	228	$\sigma^2\ (+\,\sigma^2_c)$

López et al state (p 18) "Mean daily gain for each animal and feed consumption and feed conversion efficiency for each pen were calculated" and it seems clear that their analysis was worked on these basis. Working

back from their published standard error of means, it can be shown that on the basis of a pen error variance the coefficients of variation of final weight, daily gain and food conversion are 2%, 2% and 6% respectively, whereas calculated on the basis or individual pig error variance, the coefficients of variation of final weight and daily gain are 7% and 10% respectively, which is very much more likely. A glance at the components of variance in Table 2 shows that the correct F-test for significant treatment variation (σ_T^2) is of the treatment MS vs the pen within treatment MS, and that for any significance test the use of the between pig within-treatment MS as error will be invalid unless the variance due to pens (σ_p^2) is negligible.

An apparently obvious way of rescuing the replication within pens is to test for the pen effects (σ_p^2) by an F-test of the MS for pens within treatments vs the MS for animals within pens. If this proves clearly non-significant, (σ_p^2) disappears from the components of variance and the two MS can be combined by summing their SS and deriving a joint MS to estimate the basic error variance (σ^2) The treatments can now be tested with greater precision against this, with a result almost identical to using the variance between animals within treatments as error in the first place.

Statisticians raise two objections to this practice. In the first place the absence of a significant F-test for pen effects may not indicate a negligible σ_p^2 but rather a chance equivalence of $20\ \sigma_p^2 = \sigma_C^2$, where σ_C^2 is a source of variance inflating the MS within pens, such as that due to the effect on the growth of each procedure is only valid if we can be sure on biological grounds that such competition effects are not important, which is not an easy assumption in group fed animal experiments. Secondly, it can be objected that this combining of errors will more often combine a smaller than a larger between-pens variance with the within-pens variance, so that this combined error variance will be truncated, lacking part of the upper tail of its distribution. Although true, this criticism seems to us to lack force. Combining a smaller between-pen variance will increase the average error variance and will thus be conservative as regards the error mean square. It will protect against the use of fortuitously small between-pen MSs based on few degrees of freedom, but will be helpful in finding differences because of the increased replication and increased degrees of freedom for error. Nontheless it may be as dangerous to combine a between-pen variance which is very much smaller than the within-pen than one that is much larger, since the former could indicate a real competition effect (σ_C^2).

Therefore, before we can consider the use of within-pen replication in our estimate of error, we need to be convinced that competition effects are unimportant and to demonstrate by rough equality of the between and within pen mean squares that neither within (σ_p^2) nor (σ_C^2) is important. There are statisticians who will claim that these provisos are so difficult to satisfy, particularly in experiments concerned with grouped animals, that the only safe procedure is always to use the between-pen variance as error. But the biometrician must make the best use of the data available, which may not coincide with such a strict interpretation. By analogy we might note that Table 8 is constructed on the premise that all 240 pigs completed the experiment, whereas in fact 6 died (López et al). An effect of these losses is that the coefficient of (σ_p^2) is no longer exactly 20, nor will it

any longer be completely identical in the two lines for treatments and for pens. So that strictly no F test exists at all for the treatment effects (cf Kendall and Stuart 1968). But no practical statistician would hesitate in applying it because the inaccuracy is so minute.

If he hopes to use within pen variance as a part of his estimate of error, there are a number of things that the experimenter can do. Clearly The less variability there is between pens physically, the less likely are pen effects to be important, and it may even, in extreme cases, be worthwhile eliminating end pens from an experiment by placing non-experimental animals in them.

The effects of competition must be reduced to the point where the experimenter can honestly convince himself that they are likely to be negligible. Pigs should have generous space and the feeding facilities for cattle in groups even of 2 and 3 should be designed so as to minimise competition (the ideal presumably being individual mangers and group housing). It is especially important to minimise competition in trials where a small volume of protein or mineral/vitamin supplement is given separately from the main bulk feed, and is consumed in a short space of time. Of course, the reduction of competition, like any move away from normal agricultural practice, reduces the general applicability of the results. In the case of feeding trials, the interaction of diet with competition would be lost.

A diet designed to increase growth through increased intake might have much less and more variable effect in the farm than in an experiment with competition minimised. But the alternative may entail such a lack of replication that the effect is never detected at all.

Having done his best to minimise pen and competition effects, the experimenter will need to check that the data are at least consistent with their negligibility. Before carrying out the analysis, he should choose a value for the F-test of pens vs within-pens, below which he will consider pen effects negligible, having regard to the degrees of freedom that will be available for the test and choosing a fairly high level of probability (10 or 20%) since we are concerned to demonstrate, so far as possible, the absence of an effect. Further, because of our arguments earlier, it is probably wise to set a lower value of F below which combination of error will also be eschewed. For many experiments of the size used in animal nutrition, $1/2 < F < 2$ provides a suitable range within which the within-pen MS can be combined with the between for an overall error, but in the case of a large experiment such as Table 8, the upper limits were best reduced to 1.7.

It will be appreciated that there is a great difference between deciding to combine error terms having considered the validity of such a procedure, and using the between-animal within-treatment MS as error just because it is available, as appears to have been done in many experiments in tropical animal production. Undoubtedly in some of these experiments there would have been real pen effects and inevitably among these there will be cases where a supposed treatment effect is merely the random effect of the pens chosen for that treatment.

What is a representative sample?

The same experiment of Ffoulkes and Preston (1978b) can be used to discuss the question of what more general population are the animals in the experiment a representative sample? With such a small sample of animals (12 Holstein and 4 Holstein x Zebu bulls), there is a clear danger that the randomisation of balanced pairs of bulls to treatments and replicates will produce spurious differences in the treatment results, if, as has often been shown, the growth of Zebu and Holstein x Zebu cattle differs. Nor can this problem be overcome by ensuring that all the Holstein x Zebu bulls represent one of each pair of each replicate, because the replicates then become, like the treatments, fixed effects and the treatment x replicate interaction mean square no longer provide an error variance for testing the treatment effects. This error variance is an estimate, and in this case, based on 3 degrees of freedom, not a very good estimate of population of cattle to which the treatments are applied, and the results of the experiment refer to this population.

But what is this population? As a sample of, say, cattle in the Caribbean, or even in the Dominican Republic, it is hardly representative. If it is unrepresentative of any wider cattle population, then the results of this experiment cease to have bearing as soon as these particular bulls complete their growth phase. This is not to say that experiments on unrepresentative samples should never be done. For reasons of cost and labour experiments using rumen cannulation, for example, are frequently carried out on very small unrepresentative samples. But once any treatments nutritional or otherwise, are proposed for widescale use, they must be tested on a random sample of the wider population.

General

Feeding trials in the tropics frequently encounter unique circumstances and problems, including lack of facilities, animals, money and personnel Given this, the need for well-planned experiments designed to disprove the null hypothesis at all likely levels of management and production becomes all the more important. As Wilson (1975) points out, "too many cattle experiments use the argument that if there is no significant difference between A and B then A is equal to B. There should be a much concentrated effort on the part of those charged with supervising research to ensure that research effort is not wasted in this fashion".

Animal numbers need to be calculated with care. Wyllie and Ferreiro (1979) calculated the replication required to detect various levels of difference among various parameters in cattle feeding trials in tropics using data from experiments carried out in the Dominican Republic and Mexico. They showed that sufficient replication was frequently lacking, making interpretation of many experiments extremely difficult. When interpreting such trials it should be remembered that no experiment can prove two treatments to be equal; it can only give the probabilities that they are not the same. It is surprising, too, that the concept of what constitutes a replication is still not fully understood by many researchers. Too many trials in the tropics, very few of which are published, still place unequal numbers of animals per treatment, or vary the proportions of the sexes, or

in the case of pigs, place castrates and gilts in different pens calling each a replication. Frequently blocking according to initial weight is inadequate. It is often assumed in cattle trials that using linear regression of individual weights against time for the calculation of average daily gain removes the need for careful initial balancing. However, given the correlation between initial weight and subsequent gain the effect of initial weight can only be reduced by careful blocking or by the use of covariance.

The question of what constitutes a representative sample applies not only to the feed used. Feed composition varies widely; for example the composition of sugar cane and molasses differs greatly with location, as does the quality of by-products such as oil-seed meals, meat and bone meal and fishmeal. Thus the results from one particular situation may well be difficult to reproduce elsewhere and care should be taken in the interpretation of results in relation to the level of quality of the particular feeds used. This problem is particularly acute with grazing or supplementation trials where the quantity and quality of pasture varies from year to year. Seasonal effects are important and trials need to be repeated over several years.

Various other problems arise in the design and interpretation of tropical feeding experiments. Frequently one is dealing with animal production systems involving local breeds and by-products utilisation or fibrous feeds Feeding regimes, for example with low cereal inputs, may be designed to give a performance below the biological maximum possible but highly economic and practical within the local situation. Experimental designs incorporating production functions or response surfaces are the most useful here for exploring a range of likely economic situations. Experiments at peasant level or imposing the nutritional constraints likely to be met in the field should be considered. Wyllie and Lekule (1980), for example, tested pig diets using nutrient intakes equal to those found at the peasant level rather than those recommended by standard tables of nutrient requirements.

In any difficult experimental situation such as is frequently encountered in the tropics one can only do the best one can. But too often experiments are unnecessarily flawed because of a failure to apply common sense to the problems discussed above. While in all situations in the tropics it is obviously desirable to do something rather than nothing, care should be taken taht the experimental results are such that they can be applied with confidence in the field.

Conclusion

From the papers that have been reviewed here, it is clear that the design and analysis of experiments in tropical animal production could be improved. Such an improvement does not require particularly sophisticated kinds of statistical design or computational facilities. In general it is apparent that greater clarity is required on the objectives of the experiment. Consideration of the analysis that will be applied to the experiment before it is carried out is probably the simplest method of determining whether the design is suitable both in form and replication to answer the questions posed. The application of proper experimental design and

analysis is not the imposition of alien methods of working by the statisticians or the experimental scientist: rather it is the means to increase the efficiency of experimentation by attempting to maximise the useful information obtainable from a given experimental effort.

Acknowledgements

It is a pleasure to acknowledge the many useful discussions with Mr W J Ridgman on the subjects covered in this paper and also to the helpful comments of Dr W Thorpe. We are also grateful to Ing H M Ferreiro and D Ffoulkes for much helpful discussion and for allowing us to use their data so freely.

References

Duncan D B 1955 Multiple range and multiple F-test. Biometrics 11:1-42

Dunnett C W 1955 A multiple comparison procedure for comparing several tretaments withs a control. Journal of the American Statistical Association 50:1096-1121

Ferreiro H M & Preston T R 1977 Digestibility and voluntary intake of derinded sugar cane stalk with and without of cane tops. Tropical Animal Production 2:90-99

Ferreiro H M, Preston T R & Sutherland T M 1977a Investigations of dietary limitations on sugar cane based diets. Tropical Animal Production 2:56-61

Ferreiro H M, Sutherland T M, Wilson A & Preston T R 1977b Fattening cattle with sugar cane: A comparison of different supplements. Tropical Animal Production 2:309-314

Ffoulkes D & Preston T R 1978a The banana plant as cattle feed: Digestibility and voluntary intake of different proportions of leaf and pseudostem. Tropical Animal Production 3:114-117

Ffoulkes D & Preston T R 1978b Cassava or sweet potato forage as combined sources of protein and roughage in molasses based diets: Effects of supplementation with soybean meal. Tropical Animal Production 3:186-192

de soya. Producción Animal Tropical 3:234-236

Fisher R A & Yates F 1957 Statistical tables for biological, agricultural and medical research. 5th Edition Oliver and Boyd, London

Gilmour D G & Morton J R 1970 Association of genetic polymorphisms with embryonic mortality in the chicken II.Teh B blood-group system and the pure and crossbred progeny of two populations. Genetic Research, Cambridge 15:265-285

Kendall M G & Stuart S 1968 The Advances Theory of Statistics Vol III Design and Analysis and Time-Series 2nd Edition Chapter 36, Griffin and Co., London

López O, González J & Otaño T 1978 Efectos del espacio de alojamiento en la ceba de cerdos, Ciencia y Técnica en la Agricultura: Ganado Porcino 1:15-24

Mead R & Pike D J 1975 A review of response surface methodology from a biometric viewpoint. Biometrics 31:803-851

Meyreles Luz, MacLeod N A & Preston T R 1977a Cassava forage as a protein supplement in sugar cane diets for cattle: Effect of different levels on growth and rumen fermentation. Tropical Animal Production 2:73:80

Meyreles Luz, MacLeod N A & Preston T R 1977b Cassava forage as a protein supplement in sugar cane diets for cattle: Effects of different levels without urea on growth and on rumen fermentation. Tropical Animal Production 2:200-205

2:203-208

Mitscherlich E A 1909 Das Gesetz dis Minimums und das Gesetz dis abnehmenden Bodernertrages. Landw Jahrb 38:537-552

Morton R & Ridgman W J 1977 Problems in the statistical design and analysis of feeding trials. Proceedings of the Nutrition Society 36:173:177

Nelder J A 1966 Inverse polynomials, a useful group of multifactor response functions. Biometrics 22:128:141

Ravello G, MacLeod N A & Preston T R 1977 Ensiling mixtures of sugar cane, cassava forage and urea. Tropical Animal Production 2:34-39

Ridgman W J 1975 Experimentation in Biology and introduction to design and analysis Blackie, London

Salais F J, Wilson A & Elliott R 1977 Determination of the apparent digestibility of diets containing coarsely or finely chopped cane tops. Tropical Animal Production 2:306-308

Satter L D & Roffler R R 1977 Protein requirements and non-protein nitrogen utilization. Tropical Animal Production 2:238-259

Silvestre R, MacLeod N A & Preston T R 1977a Voluntary intake and liveweight gain of cattle given chopped sugar cane and solutions of molasses containing different concentrations of urea. Tropical Animal Production 2:1-12

Silvestre R, MacLeod N A & Preston T R 1977b Fattening cattle with molasses/urea: Effects of different urea levels. Tropical Animal Production 2:315-318

Snedecor G W & Cochran W G 1967 Statistical Methods 6th Edition, Iowa State University Press, Ames, Iowa

Wilson P N 1975 In considerations for the Design and Interpretarion of Cattle experiments (Ed P D P Wood) British Society of Animal Production

Wyllie D & Ferreiro H M 1979 Observations on the number of replications required per treatment in cattle experiments. Tropical Animal Production 4:107

Wyllie D & Lekule F P 1980 Cassava and molasses for fattening pigs under village conditions in Tanzania. Tropical Agriculture 57:267-276

Received 19 July 1983