

Invited paper

Scientific research: the planning process

R.M. Gous

Animal and Poultry Science, University of KwaZulu-Natal, Pietermaritzburg 3209, South Africa

Abstract

This paper describes a few important aspects of the process that should be followed before an experiment is conducted. One should start by defining the hypothesis or theory to be tested, which gives a good indication of the way in which the experiment should be designed, and also informs of the way in which the data should be analysed. In order to ensure the success of a trial the number of replications, or animals, should be calculated beforehand, and if too few are possible, or available, then the experimental design should be changed to accommodate this. The approach to the design of response experiments differs from that when two or more independent treatments are being compared, in that fewer replications and more doses would be favoured. Also, it is worth extending the range of inputs beyond the conventionally-applied doses. This is because a response surface must be fitted to the data, and the more points and the wider the range the better for this purpose. Duncan's multiple range test is always inappropriate when analysing a dose/response experiment, and should never be used for this purpose. The optimum dose should be chosen on the basis of the hypothesis being tested, but should preferably include economic data such that an optimum economic dose can be determined, which could be modified as economic circumstances change.

Keywords: Duncan's multiple range test, dose response experiments, interpretation of results, replications
Corresponding author. E-mail: gous@ukzn.ac.za

Introduction

Many papers submitted to the South African Journal of Animal Science are rejected because the experiment has not been correctly designed, and in many cases the statistical analysis used by the authors is inappropriate for the design used. Experiments can generally be divided into two categories: those that make comparisons between treatments, and those that measure responses. It is critically important to decide which of the two is being attempted when designing the experiment as the optimum number of replications required will depend on the design as will the statistical analysis. In this paper, two aspects of the design process are detailed, these being the appropriate number of replications to use in order to achieve significant differences between treatments, and the appropriate statistical analysis to use particularly when dose response experiments have been conducted. Once such experiments have been analysed there is still some controversy regarding the way in which the results should be interpreted, and this aspect is also dealt with.

The interpretation of results from dose response trials was the subject of a paper presented by Prof. Trevor Morris of the University of Reading at a conference in Nottingham (Morris, 1983), where he made use of the results of a trial conducted by Morris & Blackburn (1982) to illustrate the point that the method of analysing results can make a large difference to the conclusions drawn. Morris commented that '... many dose/response trials are interpreted with the aid of nothing more elaborate than a Student's *t*-test or a multiple range test, which is rather like trying to peel an apple with an axe.'

An excellent book on statistical design and analysis for animal scientists was subsequently published by Morris (1999), which contains virtually all the information an animal scientist would need for correctly designing and analysing experiments, from the simple to the most complex. Many of the more important aspects of planning an experiment, dealt with in this paper, are drawn from his book, where further details of all aspects of planning and analysing experiments are available. The book should be compulsory reading for all researchers in animal and poultry science as it contains a wealth of useful information often sought after by researchers.

The planning process

The purpose of an experiment

An experiment should never be conducted in the hope of discovering a theory; rather, it should be conducted to test a theory, to measure the numbers that will make a theory work, or to choose between two theories. Experiments can generally be divided into two categories: those that make comparisons between independent treatments, and those that measure responses to increasing levels of a factor. It is critically important to decide which of the two is being attempted before designing the experiment. Every aspect of the experiment depends on which direction the research is to follow, including the number of treatments and replications, and the way in which the data are analysed.

Comparisons can be made between independent factors such as genotypes, feed additives, vaccination procedures or the feeds from different mills. These factors generally do not have different levels and should therefore be analysed using an analysis of variance, with means being compared using Student's *t*-test, for example. But the purpose of many experiments is to compare different levels of factors such as temperature, dietary lysine content, nutrient density, daily concentrate intake etc. In this case the purpose is not to show, with a high degree of confidence, that there are significant differences in the response between levels of input, but to make use of some form of regression analysis to illustrate the trend resulting from the logical structure to the treatments. The number of replications required for the two approaches are therefore bound to differ.

How many replications?

Before any experiment is attempted the number of replications (or animals, where these are housed individually) required to ensure a reliable outcome need to be determined. The required number depends on the variability of the experimental material and the size of the difference needed and can be estimated by inverting the equation for calculating the least significant difference (LSD):

$$\text{LSD} = t \cdot \sqrt{2} \cdot \sqrt{(s^2/n)}$$

where *t* = Student's *t* value for a chosen probability and d.f. appropriate to the error variance

s = the error variance

n = the number of replicates of each treatment

If we define *d* as the difference needed for significance in the planned trial, we need to find the value of *n* that will result in *d* being just equal to the LSD.

$$d = \text{LSD} = t \cdot \sqrt{2} \cdot \sqrt{(s^2/n)}$$

Squaring both sides of the equation gives

$$d^2 = t^2 \cdot 2 \cdot s^2/n$$

$$\text{and therefore } n = 2 t^2 s^2 / d^2$$

To avoid problems with differences in the units of the parameters *s* and *d*, this equation can be written as:

$$n = 2 t^2 (\text{CV})^2 / d\%^2 \quad (\text{Morris, 1999})$$

The CV (%) can be estimated by making use of the results from previous trials in your facility or in some previously published research, by measuring the trait in some animals, using an analogy or making an intelligent guess (the inverse of heritability would suffice). For example, body weight and growth rate have a CV around 10 % whereas the CV of reproductive traits, which have a low heritability, is around 20 – 30 %. Milk yield and rate of laying have a CV around 20 – 25 %.

Estimating d , the difference to be expected, can be done by surveying the literature for similar experiments which, as Morris (1999) suggests, might even help to develop a theory to be tested: will the difference be greater under different circumstances, for example? Alternatively, one could ask the question, what would be an economically worthwhile response? In this case one would need to take account of the cost of the treatment and determine what improvement would be necessary to cover that cost. For the more practical case, the answer could be found by asking how big the difference needs to be before farmers would be prepared to adopt the new strategy.

If the d.f. for error in the planned experiment is in the range 20 to 60, t for $P = 0.05$ will be about 2, and the equation above can be rewritten as:

$$n = 8 (CV)^2 / d\%^2$$

As an example, consider a pig experiment, with pigs being fed individually and with six independent feeds to be tested. If the CV of liveweight gain = 12% and an 8% difference is needed for significance at $P = 0.05$, $n = 18$ pigs per treatment, so 108 pigs would be required.

Morris (1999) suggests three alternatives if there are not enough animals. In the first case, one should search for ways of reducing CV% and this could be accomplished for example by measuring protein gain rather than body weight gain or measuring ovulation rate in sheep or pigs instead of litter size. Secondly, one could widen the range of treatments, which may allow larger $d\%$ values to be used; and thirdly, one could go to the library and try to come up with a new or different hypothesis.

Where response trials are being planned the number of replications required will differ from the case where independent treatments are being compared, unless interacting factors are being tested in the response trial, in which case the rules outlined above still apply. But in cases where the responses of two or more factors are not being tested simultaneously it is often more valuable to increase the number of levels of a factor rather than the number of replications. A visit to a statistician would be a valuable exercise in this case.

Designing and analysing response trials

In this case the objective is not to prove that one treatment is significantly better than another; instead, the objective is to find the optimum dose (which may be that which produces the maximum response, the minimum response or the optimum economic response). So the number of doses (treatments) becomes more of an issue than the number of replications, as describing the response surface is the major objective. An example is given in Figure 1 where an incorrect interpretation of the response of animals to increasing environmental temperatures would result if only levels 1, 3 and 5 had been applied. By increasing the number of 'doses' to six and halving the number of replications a more accurate estimate of the response would be obtained, and a more accurate decision about the optimum temperature would be possible even though the response at each level is not as accurately determined.

In choosing the range of levels to apply in a response trial it is often worthwhile going beyond the conventional levels that have been applied previously or conventionally, as this affords the researcher the opportunity of obtaining a more accurate picture of the response to the input. Two examples are given here, one of which is illustrated below, where more information has been added to our knowledge by going outside the conventional range of treatments. In Figure 2 the response of broilers, in 35 d body weight, to increasing daily photoperiods is illustrated. This is from a trial reported by Lewis *et al.* (2009). Conventionally, only photoperiods greater than 12 h have been used in broiler trials. By going outside the 'accepted' range it was possible to show that body weight is unaffected by photoperiods > 7 h. In applying very short photoperiods it was discovered that broilers eat successfully in the dark when given less than 16 of light, and that a 12 h photoperiod results in the highest efficiency, calculated using the European Efficiency Factor, and bones with the greatest breaking strength.

The second example relates to the series of experiments reported by Lewis *et al.* (2007), which involved transferring broiler breeder pullets from an 8 to a 16 h photoperiod over a range of ages from 84 to 225 d, whilst ensuring a wide range of 140-d body weights within each treatment. Conventionally, light stimulation in broiler breeders is not applied before 126 d, and body weights much smaller or larger than 2.1 kg at 20 weeks are not used. Given these restricted ranges it would not have been possible to develop an empirical model of these effects on age at sexual maturity

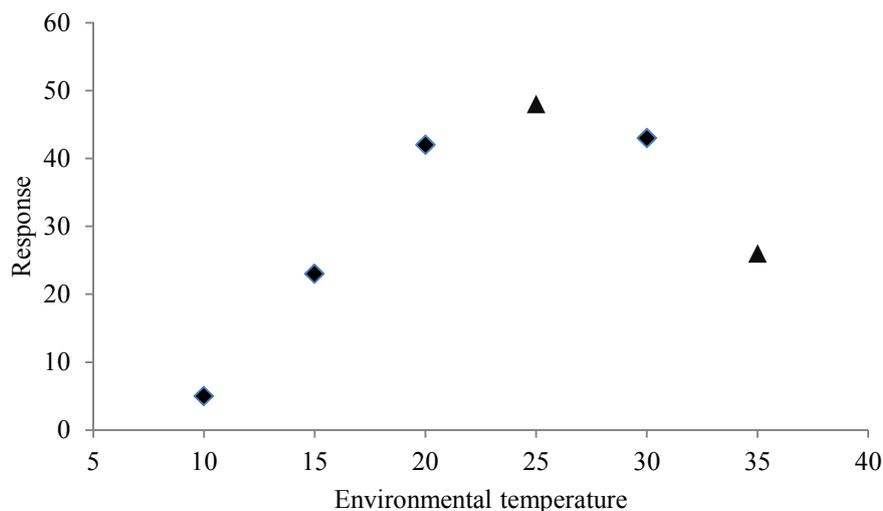


Figure 1 The results of an hypothetical response trial in which animals were subjected to increasing environmental temperatures. The interpretation of the result would differ depending on the number of levels of temperature used in the trial.

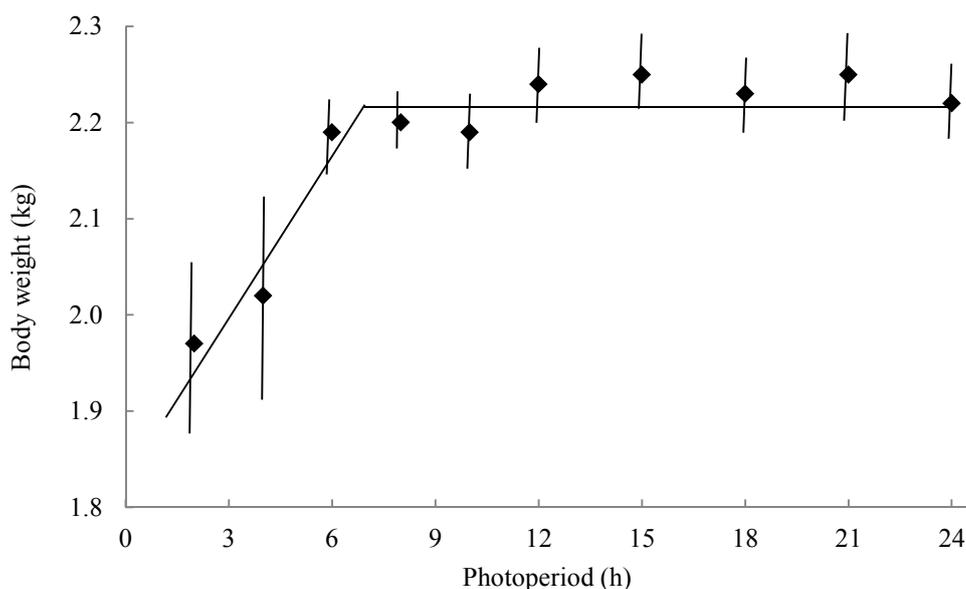


Figure 2 Mean body weight at 35 d of broilers reared on a 2-, 4-, 6-, 8-, 10-, 12-, 15-, 18-, 21-h photoperiod from 2 d or given continuous illumination (After Lewis *et al.*, 2009).

Preparing and reporting the feeds used

An important principle that should be applied when conducting response trials involving feed nutrient levels is that the intermediate feeds can be manufactured by blending the 'outer' feeds. When measuring the response to dietary protein, for example, where the levels of all other nutrients and energy are to remain constant in all the feeds, two basal feeds can be formulated and blended to produce all the levels of protein required. Where the responses to protein and energy are to be measured simultaneously in a trial, four basal, or 'corner', feeds are formulated and all intermediate protein and energy levels can be achieved by appropriately blending these four feeds.

There is then no need to present details of the ingredient and nutrient composition of each of the feeds used in the experiment, as anyone knowing a little arithmetic could work out their contents from the composition of the basal feeds. Only the composition of the basal feeds should be given, as well as either a table or a schedule showing the mixing proportions.

It is not necessary to have equal spacings between doses. It would be more prudent to have more around the doses of interest, and fewer towards the extremes. Such an example is in Morris & Blackburn (1982).

Introducing factors into a response trial

It is possible to investigate the response to two or more variables simultaneously (e.g. environmental temperature and feed energy content), and if interactions are expected between these factors then replication becomes more important than if only one factor were being investigated, as mentioned above. Such a design would be used, for example, to answer the question whether males and females, or three genotypes, would respond in the same way to a range of feeds differing in protein content, or to a range of environmental temperatures. Each 'group' i.e. males and females, might receive the same levels of dietary protein, or, if it is known that the requirement of the females for protein is less than that for males, the highest protein level may not be fed to the females, and the lowest may not be fed to the males. In most cases, however, the groups would receive the same levels of the input being applied.

The responses by the different groups then need to be compared, and this can be done using simple (or multiple) linear regression with groups in Genstat (2008). A regression is first fitted to all the data combined; the constant terms are then fitted separately using the same slope for each group; and finally the constant term and slope are fitted separately. In this way the responses can be compared statistically to determine whether they respond in the same way, but with one always higher or lower than the other, or whether their responses are totally different. It is not correct simply to apply a test of significance to the regression coefficients to determine whether they differ significantly.

An example is given in Figure 3 of the results of a trial (Danisman & Gous, 2008) in which the allometric relationship between thigh weight and body protein weight of four broiler strains were compared using simple linear regression with groups. In this case the response was the same in all four strains so a common constant term and slope represented all strains.

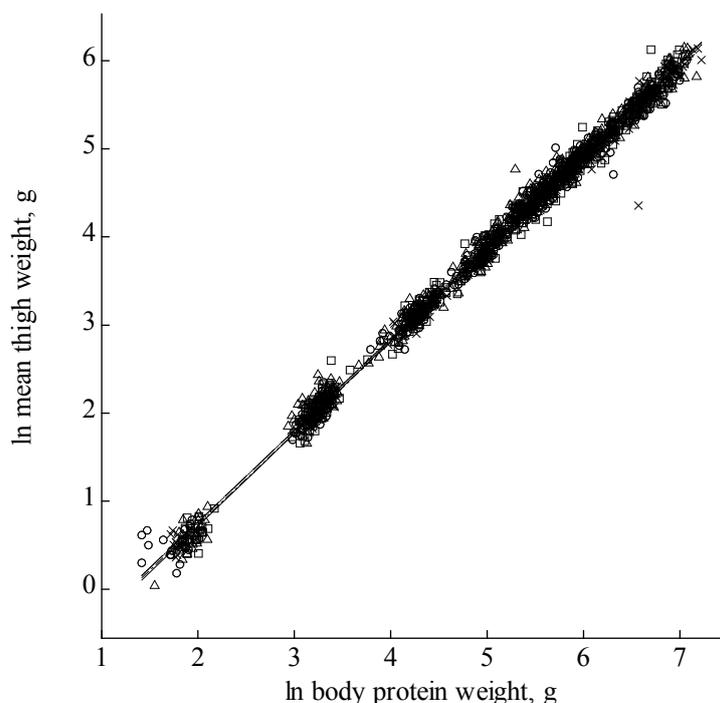


Figure 3 Allometric relationship between mean thigh weight and body protein weight for four strains. Cobb, ... Δ ; Ross 308, - - - \circ ; Ross 788, ___ \square ; Hybro, _._._ \times . (after Danisman & Gous, 2008).

Statistical analysis of a response trial

Duncan's multiple range test (Duncan, 1955) is very often inappropriately used to compare treatments that are factorial in nature or that correspond to several levels of a quantitative or continuous variable (Chew, 1976). It is therefore incorrect to determine whether there are statistically significant differences between treatment levels when using a dose/response experiment. The comparison between treatments by means of a multiple range test is inappropriate when there is a logical structure to the set of treatments, and the use of a conventional 5 % level of significance is inappropriate when trying to obtain the best estimate of some end point, 'as opposed to requiring a high degree of confidence that we have not gone too far along some input scale' (Morris, 1983).

A response surface must be fitted to the data. A good procedure to follow is to start by calculating the means of each dose, plotting these points and then drawing the resultant response surface by eye. In this way it is possible to determine whether the response is linear, curvilinear, asymptotic etc, whereafter the correct regression curve can be fitted to the data. An important question to be asked at this stage is whether the response conformed to the original hypothesis.

Presentation of results

The means of all variables measured should be presented in a table, together with the standard error (SE) of the mean or residual mean square, but with no super- or subscripts indicating statistical differences between means. The coefficients of the curve fitted to the data need to be displayed together with their SE's. When graphing the results, the actual means for each level of the factor should be displayed and not the fitted means, and the continuous function fitted to the data should be drawn through these means. It is incorrect to use a bar chart when illustrating a response experiment, as this implies that the factor levels were independent treatments, which they are not.

Interpretation of the response

It is assumed that the purpose of a response trial is to determine the optimum dose to apply in practice, or to develop a model that can be used subsequently for this purpose. The solution is not to find a curve that fits the data with the minimum statistical error but instead to apply biological and preferably economic reasoning when choosing the correct curvilinear analysis. Many different curves have been fitted to response data, and the interpretation of the response is dependent on the method used. Morris (1999) gives examples of these methods: a broken stick method will always underestimate the optimum dose, as this reflects the response of the average individual; a parabolic curve often fits the data well, but is unrealistic in many cases as the predicted response diminishes at higher levels of input, and also because it is unduly sensitive to the range of treatments selected; the inverse polynomial and exponential functions give asymptotic curves which also fit the data well, but they predict continuing responses at high inputs when the real response has ceased, so some subjective judgement has to be applied to determine at what percentage of the maximum response the optimum should be.

The favoured approach when interpreting the response of animals or birds to increasing doses of a nutrient is to make use of the Reading Model (Fisher *et al.*, 1973) as this 'has a curvature largely independent of the choice of treatments and therefore gives realistic estimates of the optimum dose even with few data points, results from different trials can be combined even when mean performances differ between trials, and it is suitable for extrapolation to levels of performance which lie outside of the range of experimental data' (Morris, 1999). It is also possible to account for the marginal cost of the input and the marginal revenue derived from the output, thereby making decisions about the optimum dose on the basis of the economic value of the relationship.

One important point to consider is whether the input has an effect on an associated variable. Where responses to nutrients are measured, if the level of the nutrient has an effect on food intake, which is likely, then the scale used to describe the input should not be the dietary concentration of the test nutrient but its intake. For example, where the response to an amino acid has been measured, the appropriate input should be the daily intake of the amino acid, not the concentration of the amino acid in the feed, as food intake is influenced by amino acid content (Gous *et al.*, 1987). Once the optimum daily intake has been determined then it is necessary to calculate the dietary content that will guarantee the bird will consume the optimum dose, which requires the prediction of food intake (Gous, 1986).

The optimum dose, when based on biologically meaningful characteristics such as the mean body weight and egg output of a flock of hens, as well as the prevailing economic circumstances, can be modified as the biological and economic circumstances change. Such a dynamic approach is inherent in the Reading Model, which is why this model was such an important step forward in our understanding of the way in which nutrient responses could be interpreted and used in practice.

Conclusions

In planning an experiment one should start by defining the hypothesis or theory to be tested. This gives a good indication of the way in which the experiment should be designed, and also informs of the way in which the data should be analysed. In order to ensure the success of a trial the number of replications, or animals, should be calculated beforehand, and if too few are possible, or available, then the experimental design should be changed to accommodate this. The approach to the design of response experiments differs from that when two or more independent treatments are being compared, in that fewer replications and more doses would be favoured. Also, it is worth extending the range of inputs beyond the conventionally-applied doses. This is because a response surface must be fitted to the data, and the more points and the wider the range the better for this purpose. The optimum dose should be chosen on the basis of the hypothesis being tested, but should preferably include economic data such that an optimum economic dose can be determined, which could be modified as economic circumstances change.

References

- Chew, V., 1976. Uses and abuses of Duncan's multiple range test. Proc. Fla. Slate Hort. Soc. 89, 251-253.
- Danisman, R. & Gous, R.M., 2008. Predicting the physical parts of broilers. XXIII World's Poultry Congress, Brisbane, Australia.
- Duncan, D.B., 1955. Multiple-range and multiple-F tests. Biom. 11, 1-42.
- Fisher, C., Morris, T.R. & Jennings, R.C., 1973. A model for the description and prediction of the response of laying hens to amino acid intake. Br. Poult. Sci. 14, 469-484.
- GenStat 2008. GenStat 11th Edition, VSN International. Hemel Hempstead U.K.
- Gous, R.M., 1986. Measurement of response in nutritional experiments. In: Nutrient requirements of Poultry and Nutritional Research. Eds Fisher, C. & Boorman, K.N., British Poultry Science Symposium Number Nineteen, Butterworths. pp. 41-57.
- Gous, R.M., Griessel, M. & Morris, T.R., 1987. Effect of dietary energy concentration on the response of laying hens to amino acids. Br. Poult. Sci. 28, 427-436.
- Lewis, P.D., Gous, R.M. & Morris, T.R., 2007. A model to predict sexual maturity in broiler breeders given a single increment in photoperiod. Br. Poult. Sci. 48, 625-634.
- Lewis, P.D., Danisman, R. & Gous, R.M., 2009. Photoperiodic responses of broilers: I. Growth, feeding behaviour, breast yield, breast yield, and testicular growth. Br. Poult. Sci. 50, 657-666.
- Morris, T.R., 1983. The interpretation of response data from animal feeding trials. In: Recent Advances in Animal Nutrition. Ed. W. Haresign, Butterworths, London. pp. 12-23.
- Morris, T.R., 1999. Experimental Design and Analysis in Animal Sciences. CABI Publishing, Wallingford, U.K.
- Morris, T.R. & Blackburn, H.A., 1982. The shape of the response curve relating protein intake to egg output for flocks of laying hens. Br. Poult. Sci. 23, 405-424.