

THE STATISTICAL ANALYSIS OF REPRODUCTIVE DATA FROM GRAZING EXPERIMENTS

INTRODUCTION

P.K. O'ROURKE*

Research designed to improve reproductive efficiency in cattle and sheep is large in scale, covers several years and is expensive. Efficiency in experimental design and data analysis requires the use of sophisticated and often complex statistical methods. This contract reviews these methods, and describes how they can maximise the benefit from research effort and their contribution to practices readily adoptable by the producer.

Breeding systems have reproduction, survival and growth as interacting components. Research projects may look at changes to the commercial system in its environmental context or they may seek detailed understanding and explanation of one component. The former projects tend to be of simple design and oriented towards direct adoption by producers while the latter often use complex designs and measurements to aid interpretation of basic principles. Data for statistical analysis are of two types, continuous, as for liveweight, or categorical, as for conception. Analysis of variance is traditionally used to analyse continuous data but methods of analysis for categorical data range from use of contingency tables to analysis of variance and log-linear modelling. Unequal sub-class numbers complicate each of these forms of analysis through non-orthogonality and problems in model selection.

This contract addresses the statistical analysis of reproductive data from a broad perspective. O'Rourke and Howitt introduce the initial steps of planning and design which set up the analytical methods and scope for interpretation and practical use of results from the research project. Rudder reviews methods for data recording, storage and management, which consume up to three-quarters of analytical time. Mayer introduces log-linear modelling for categorical data and compares this new method with the traditional use of analysis of variance. Shepherd discusses model selection and the pitfalls involved in it. Finally, McCosker considers extrapolation from research results to the complete breeding system and extension of these results for use by producers.

PLANNING AND DESIGN OF REPRODUCTIVE EXPERIMENTS

P.K. O'ROURKE AND C.J. HOWITT*

The planning and design of research projects is the initial step, dictating the form of analysis, limits for interpretation of results and qualifiers for extrapolation. These aspects will be reviewed here.

PLANNING OF REPRODUCTIVE EXPERIMENTS

Reproductive data from grazing experiments have a wide range of potential response variables but it is most efficient to select a variable which gives the best early indication of response. Breeder mortality, weaning rate and growth rate of progeny are of most consequence to the producer. However, they are relatively insensitive response indices for research. Earlier and more sensitive indicators are conception, birth and survival to weaning, with conception being the most critical of these. Lactating animals are known to be capable of conceiving and are most susceptible to nutritional stress; hence, conception rate among lactating animals provides the best early indication of response to a

*QDPI Biometry Branch, Brisbane, Qld 400 1.

treatment.

On extensive properties, cost, size and management may set limits on what and how often measurements may be made. Mustering may not be possible during the wet season, during calving or lambing or at other stressful times. Costs of complete musters may be prohibitive. Fencing may be insecure so that animals are lost from their paddock or conceive outside the scheduled breeding season. Data on birth difficulties, date and weight at birth, date and cause of mortality and abortion are frequently unavailable. Two musters per year with measurement of pregnancy status, lactation status, liveweight and fleece weight (sheep) should be considered minimal for breeders. Mustering efficiency should be adequate to allow estimates of breeder mortality based on absence at two consecutive musters. Numbers and weaning weight of progeny should also be obtained at musters.

Background management options must be compatible with problem definition, specific objectives for the research project and the target population. The selection of continuous or seasonal joining 'must reflect the population to which results will be extrapolated. Similarly, appropriate choices must be made for weaning, culling and replacements. Each of these management options is likely to interact with the response variables and so influence interpretation. Alternatively, basic research may establish principles which need further development and demonstration by extension officers before adoption by industry.

EXPERIMENTAL DESIGN

The first issue in design is identification of the experimental unit as either the paddock, group or individual animal (Blight and Pepper 1982; Blight 1984). The animal may be considered as the experimental unit only when it receives its treatment individually and performs independently of other animals in its group. These conditions are often appropriate for reproductive experiments conducted under commercial conditions. Alternatively, nutritional or management experiments often have paddocks as the experimental unit.

Designs for reproductive experiments are simple ones, such as randomised block or completely randomised designs. Replication rates are required to yield at least ten degrees of freedom for the estimate of experimental error so that tests of significance have adequate power to recognise as significant, differences which are large enough to be of practical importance. This is problematical when paddocks are the experimental unit but easily achieved otherwise. Methods to estimate the number of paddocks and animals per group are given by Blight (1984). When the response variable is binary, as for conception rate or mortality rate, large numbers of animals are needed to identify moderately sized differences as significant. For instance, if the conception rate of a control group is 60%, 135 animals per group will be needed to have a 15 percentage point response significant at $P < 0.05$ with probability 0.80.

Methods used to allocate animals to paddocks and treatments must be compatible with the objectives of the research project. Mixed ages of breeders relate well to commercial practice, giving results of wide generality with little added cost; but research with a single cohort provides easy interpretation. Allocation to groups is usually by stratified randomisation based on age, lactation status, stage of pregnancy, body weight or fleece weight. These criteria for stratification must be chosen so that either they will not interact with the intended treatments or their influence can be isolated and tested during analysis. Stratification will only be more efficient than complete randomisation when it removes a substantial proportion of the residual variation in the major response variables. Selection of replacements for dead, missing or culled animals should be specified as part of the allocation. In reproductive

experiments the worst treatment is likely to have high mortality and low conception and weaning rates, so that a high level of replacement is required. This contrasts markedly with the needs of the best treatment. Clear specification of replacement policy, in line with the objectives and problem definition, is necessary to cover these extremes.

Reproductive experiments typically extend over at least three breeding cycles to sample a range of seasons. Some treatments only exhibit their superiority under adverse seasonal conditions, while favourable rainfall during the dry season will tend to eliminate or mask other effects. As well as seasonal factors, previous reproductive history, particularly recent lactation, will influence response, so that long term experiments are required. If weaning rate of previously lactating animals is the response variable, the animal's history is required over the previous two years under the conditions of the experiment. Five years are required to collect data on three breeding cycles.

Contrasting examples of major reproductive experiments are given by Holroyd et al. (1983) and McCosker (pers. comm.). Holroyd et al. compared six treatments in four paddock replicates over 900 ha with a single group of 288 heifers. They confounded age with seasonal effects over four years. Their experiment was run at Swan's Lagoon, which made possible close management and supervision of the animals and gave accurate records of births and deaths. Measurements were made at monthly musters. McCosker used ten paddocks and 900 breeders grazing 14000 ha at Mt Bunday. Data were collected at musters twice per year for five years. Although it is difficult to achieve close herd management on commercial properties in northern Australia, seasonal mating and weaning were achieved. Replacement policy for these mixed-age herds was a complex issue. Both experiments related well to their objectives so that results could be extrapolated to relevant sectors of the commercial industry.

RECORDING, EDITING AND STORING REPRODUCTIVE DATA

T.H. RUDDER*

Many research workers and most administrators under-estimate the work and money needed to prepare sets of reproductive data for analyses, and the time taken by biometricians to edit and correct data files that have been poorly designed and edited. Experience has shown that biometricians spend up to three quarters of the time with a set of reproductive data coping with inconsistencies in format and coding, and checking atypical data values with research workers.

Until recently, the biometrician was the only member of the research team with access to computing resources to thoroughly edit data. Therefore, there were limitations to editing at the field level. During the past three to five years microcomputers with word processing and sort-merge capabilities have been distributed to many research centres and this has given these research workers the capacity to edit data thoroughly and present files in a format compatible with analytical needs. This paper discusses procedures designed to reduce the time spent in data management by both the research worker and the biometrician.

REQUIREMENTS FOR RECORDING SYSTEMS

Reproductive projects may involve the accumulation of large quantities of data over many years. Data are collected to assist project decision making and for statistical analysis. Maintaining large data collections requires data to be assembled in a logical and systematic manner.

*QDPI Beef Cattle Husbandry Branch, Brisbane, Qld 4001.

It is most important that the system used is compatible with the skills and resources available at the field level. Generally, research workers have skills and aptitudes biased towards the biological aspects of the experiments and have limited skills or inclination towards use of computers. Therefore, the system must be based on available user-friendly software packages.

The finished file must give data in a form suitable for analysis. Historically, biometricians have been involved in experimental design and subsequently in analysis and interpretation of results, but have made very little contribution to the design of data records. This aspect should be considered before the project commences by the biometrician and research worker.

Finally, whatever system is used it must be well documented and changes in format and coding not made without adequate consultation. This is particularly important in reproductive experiments because large numbers of animals are required to obtain reliable estimates of treatment effects and these numbers are usually accumulated by replication over years. In turn, this means that records may be kept by two or more workers which frequently introduces discontinuity in format and coding of the response variables.

RECORDING DATA

Historically, data have been recorded in a row and column format in books or loose sheets of paper. Most research workers can relate to this approach, therefore it seems preferable to base computer files on it. It is a mistake to try to fit either too much data, or unrelated data on one record and preferable to maintain a number of smaller records. These smaller records can be merged depending on the set of data being analysed.

There is a number of measurements that can be included in reproductive studies and typical data include:

- (i) Pedigree records which contain all the unalterable information at birth, plus details of the animal's survival to weaning, e.g. animal identification, dam and sire identification, breed, date of birth, sex, birth weight, multiple or single birth, natural or assisted birth.
- (ii) Liveweight records.
- (iii) Joining records which contain starting date of joining, period joined, sire joined, joining method, pregnancy status and date tested, lactation status, mating outcome.
- (iv) Miscellaneous records, e.g. tick, worm e.p.g., blight, rectal temperature measurements, fleece weight, wool growth rate, fibre diameter.
- (v) Disposal records that contain live weight, carcase weight, fat thickness, reason for disposal.

The same categories can be successfully adopted for extensive field projects. Uncontrolled mating situations and incomplete musters can be encountered resulting in missing data, and recording procedures for this and for including replacement animals are required.

After the format and content of the records have been determined the next step is to plan data entry. Data collection sheets can be generated automatically after the animal identification is recorded. Ideally, automatic data capture units that would interface with a microcomputer would be used, but their general use seems to be some time away at present.

Records should be updated immediately new information is collected because errors, omissions and inconsistencies are more easily detected and resolved than

at a later date. Inspection of records shows that many entries are common to groups of animals. Therefore, with a little planning, data entry can be streamlined. For example, starting date, period, sire and method of joining are common for groups of breeders. Most word processing packages have adequate find and replace functions to allow accurate entry of repetitive data.

EDITING

Errors and inconsistencies in format and coding will always occur irrespective of who enters data. The importance of thorough editing cannot be over-emphasised and is the responsibility of whoever collects the data. The biometrician will make numerous checks for apparently spurious values. However, this way of finding errors is costly in terms of labour and computing.

There is no general recipe for editing data that will substitute for observation and commonsense. Strategic sorting by treatment groups or other unique categories can be used to aid detection of errors. Constraints can be set in most sort-merge packages to isolate biologically improbable occurrences, e.g. liveweights outside predetermined ranges, non-pregnant breeders being credited with an offspring the following year.

While every effort should be made to minimise missing values, they will occur. If the value can be logically and unambiguously estimated and only a small proportion of values is involved, the biologist could estimate the value, e.g. a cow missing at pregnancy diagnosis and is at the next muster with a calf. Missing liveweights should be recorded as missing and dealt with by the biometrician. This subject should be discussed by the biologist and biometrician at the start of the research project.

STORING DATA

Research data are expensive to collect; therefore security and complete use of the information these data contain are important. Experiments involving reproductive data usually extend over a number of years and contain information not necessarily directly related to the immediate objective. In some cases subsequent experiments may have a common link with previous ones leading to the opportunity to obtain extra information, e.g. heritability estimates, data to compare analytical options, seasonal influences on productivity. Microcomputer disks are suitable for only short to mid term storage, and microcomputers have limitations for sorting and merging large data sets to produce analytical files. Storage of data on a main-frame computer is highly desirable for ease of data manipulation to produce analytical files and for security of data.

There is limited information to assist the decision for the best means of storing and retrieving biological data. The use of early database packages was expensive in terms of computer and operator time to update and retrieve files for analyses and required a relatively high degree of programming skill. Complete testing of the current generation database packages is still needed to assess their value for animal data. While data base packages have the potential to streamline manipulation of large data sets, careful planning is needed with regard to the choice of package and the staff needed to operate and maintain it.

Experience with reproductive and liveweight data stored in a fixed format has shown that this system can be used effectively. The main advantages are that updating and retrieval of subsets for analyses can be implemented by novice computer operators. The major disadvantage is that it uses storage space less efficiently than free format.

ANALYTICAL METHODS FOR BINARY RESPONSE DATA

D.G. MAYER*

Reviews of available analytical methods and comparisons between these are not common, although Cox (1970) gives a theoretical overview. For simpler experimental designs, chi-square and other non-parametric tests, and binomial and Poisson models have been used, although results tend to be similar to those from the more widely used analytical methods (Haseman and Kupper 1979). For more complex experimental designs, such as factorials, statistical models based on the above distributions become difficult, and analytical techniques need to be more generalised.

Analysis of variance (ANOVA) is a commonly applied analytical method when the experimental units are taken as cell proportions. The arc-sine transformation (ASINE) can be used if some cell proportions tend towards 0 or 1. For unequal cell numbers, a weighted ANOVA is required, with weights usually being the number of observations in each cell. This method can encounter problems in obtaining adequate degrees of freedom for the error term.

If the individual animal is taken as the experimental unit, ANOVA of the raw binary data (RAW.AN) has been proposed as a practical method of analysis, provided certain precautions are observed (Lunney 1970, Harvey 1982). The long-accepted rule is that where n and p are the number of observations and proportion of success, respectively, in each cell both np and $n(1-p)$ should be greater than or equal to 5. More recently, Lunney (1970) has demonstrated the importance of error degrees of freedom, suggesting that if p lies between 0.2 and 0.8, 20 degrees of freedom for the error term are sufficient, but if p lies outside this range there should be at least 40 degrees of freedom. Subjecting raw data to an ANOVA has a number of advantages, such as ease of analysis and interpretation, automatic weighting and high error degrees of freedom.

The validity of this method of analysis is somewhat questionable, however, as two of the assumptions of the ANOVA are violated when binary data are considered. The first is that the data are distributed normally. Binary data have a two-point distribution, which is decidedly skewed when p is close to 0 or 1. The second is that treatments have equal variances. Treatment variance is $p(1-p)$, which gives a maximum of 0.25 at $p = 0.5$, with variance tending to 0 as p tends to 0 or 1. Generally, violation of the first assumption results in the method under-estimating the true level of significance, i.e., giving too many significant results. Violation of the second assumption is more important leading to loss of efficiency in estimation and significance tests of treatment differences. Statistical methods are available for partitioning the error variance into homogeneous components. However as this is not readily available on most ANOVA packages, it is seldom used. When data which violate one or more of the underlying assumptions are subjected to ANOVA, it is general practice to regard the significance levels as approximate only. However, as ANOVA is quite robust with respect to departures from the assumptions of normality and equal variance (Boneau 1960; Lunney 1970), this approximation may be quite good.

More recently, generalised linear models using logits (LOGIT) have emerged as a method of analysing binary response data (Fienberg 1980). This approach involves an iterative weighted regression technique using binomial error and the logit link, as found in GENSTAT and GLIM (Baker and Nelder 1978). A series of main effects and interactions is fitted, up to the saturated model where the highest-order interaction is involved. This produces an analysis of deviance

*QDPI Biometry Branch, Brisbane, Qld 4001.

table, where the deviance of each term is distributed approximately as a chi-square variable. The main advantages of logit models are that the means are constrained to the range 0-1, and partitioning of treatment and interaction effects in factorial experiments is easy. One feature of this successive model fitting method is that, as with non-orthogonal ANOVA, the significance levels of the treatments alter with the order in which they are fitted.

COMPARISONS OF ANALYTICAL TECHNIQUES

Two measures can be used to compare analytical methods, namely power and significance levels. The power of a test is the probability of rejection of the null hypothesis, and can only be calculated when the extent of departure from the null hypothesis is known, such as in Monte Carlo simulations with defined treatment effects. The level of significance of each test can be compared for either real or simulated data. Several Monte Carlo simulation studies have demonstrated the superiority of LOGIT and RAW.AN over other methods, with little difference between the former two (Butcher and Kemp 1974; Levy and Narula 1977).

The data used in this comparison came from four factorial experiments with reproductive data. These contributed 16 main effect and interaction terms, and were analysed by LOGIT, RAW.AN, ASINE, and ANOVA of cell proportions and Freeman-Tukey arc-sine transformed cell proportions. The last two methods gave results which were similar, but slightly inferior, to ASINE, and these methods are omitted from further discussion. As LOGIT is theoretically the best method, it was chosen as the 'standard'. It identified nine terms with significance levels less than 0.20. These are listed in Table 1, which also includes the mean absolute departure (MAD) from the logit regression significance levels, and the correlation coefficient (r) against the logit significance levels.

Table 1 Significance levels (P) of the test statistics, mean absolute departure and correlation with logit significance levels, for $P(\text{logit}) < 0.20$

Method	Term									MAD	r
	1	2	3	4	5	6	7	8	9		
Logit model	.035	.074	.000	.001	.000	.000	.017	.000	.000	-	-
ANOVA (raw data)	.053	.076	.000	.000	.000	.000	.000	.077	.026	.016	.611
ANOVA (ASINE)	.184	.071	.000	.137	.044	.023	.023	.045	.098	.056	.272

As is evident in Table 1, RAW.AN had far better correspondence with LOGIT than did ASINE in the region of most interest ($P < 0.20$). Performance in the region $P > 0.20$ was similar for LOGIT, RAW.AN and ASINE, with average significance levels of 0.643, 0.668 and 0.692, respectively. Compared with LOGIT, RAW.AN had a marginally lower MAD than ASINE (0.117 vs 0.138, respectively) and similar correlation coefficient (0.775 vs 0.738).

Whilst the 16 terms presented above may seem a relatively small sample, a similar study on binary data from 7 horticultural trials investigated a further 26 terms. These showed similar results for the overall treatment test statistic, namely close agreement between LOGIT and RAW.AN. Hence, the preliminary ANOVA often used to determine the correct order of inclusion of terms in the logit model may, under most conditions, be an acceptable approximate analysis.

SELECTION OF STATISTICAL MODELS FOR REPRODUCTIVE DATA

R.K. SHEPHERD*

Selecting models for reproductive data is not straightforward when the number of animals in each subclass is unequal. The unbalanced data cause lack of independence between the terms in any statistical model. This lack of orthogonality is discussed here, with the companion issues of parsimony, number of possible models, hierarchical models, marginality and model selection methods. The issues are illustrated in an example for both least squares and logit models. Computer programs for fitting these models are discussed.

PROPERTIES OF A STATISTICAL MODEL

Statistical analysis of reproductive data aims to construct a model of the response variable (e.g. conception) in terms of explanatory variables. The explanatory variables may be design variables, such as treatment, or observational variables, such as sex of calf.

The model provides fitted values for each observed value of the response variable by minimising some criterion like the residual sum of squares, which is the goodness of fit criterion for least squares models. For logit models the goodness of fit criterion is called the deviance. A complicated model involving large numbers of parameters always fits the data more closely than a simpler model containing a subset of those parameters. However the simpler model is often preferred as a better summary if it excludes unnecessary parameters but still fits the data reasonably well. Such a model is called parsimonious and is more useful as it allows the researcher to think more clearly about the data and provides better predictions.

The number of possible models increases as the number of explanatory variables increases. For example, with three factors there are nine hierarchical models in which all main effect terms are included while for four factors there are 113 such models. Hierarchical models are constructed using the marginality principle (Nelder 1982). For four factors A,B,C and D, the principle requires the main effects A and B to be included in any model containing the AB interaction term, and it also requires terms for A,B,C, AB, AC and BC to be included in any model containing the three factor interaction ABC.

Selection of terms to be included in least squares models is not a problem when the number of animals in a subclass is equal as the resulting statistical analysis has the property of orthogonality. This guarantees the independence of the various terms of the analysis of variance and provides a simple arithmetic method for the calculation of the sum of squares for the effect of each term. With unequal sub-class numbers, these properties do not hold. For example, the effect of age group on conception rate in a non-orthogonal analysis of variance would depend on the presence or absence of terms like year and genotype in the model. Hence in the analysis it should be stated clearly whether the effect of each term is adjusted for all, some or none of the other terms in the model. Two common methods used in computer programs are the "eliminating all other terms" method and the "ordered sequential fitting" method. The first method adjusts each term for all other terms in the model. In fitting ordered sequences of terms, a term is only adjusted for preceding terms in the sequence. There is some controversy in the literature (Nelder 1982) regarding which adjusted terms should be used in a non-orthogonal analysis of variance.

*QDPI Biometry Branch, Townsville, Qld 4810.

As a result of the non-orthogonality, the non-significant terms inflate the variances of other terms, and so should not be selected in the final model. However marginality must be taken into account when deleting non-significant terms. For example, if the AB interaction is significant then the A and B main effects must remain in the model regardless of their significance.

MODEL SELECTION METHODS

The "best" model will contain a balance between goodness of fit and parsimony. One method of finding the "best" model is to fit all possible models and on the basis of goodness of fit and parsimony, choose the "best" one. This method is usually not used as it requires too much computer time. There is no alternative method for selecting the "best" model. McCullagh and Nelder (1983) discuss a stepwise regression method which combines forward selection and backward elimination of terms in the model. With forward selection, the best unselected term satisfying a criterion is included next until no more such terms remain. Backward elimination begins with all the terms and deletes the worst ones from the model until all remaining terms are necessary. Two other methods are Brown's method of calculating two test statistics for each term, and the method of standard parameter estimates for the full model including all terms (Fienberg 1980). All these methods will determine a model but there is no guarantee that it will be the "best". In practice, I use a stepwise method which uses forward selection followed by backward elimination while adhering to the principle of marginality. The procedure is not precise but allows for perception and imagination during the modelling process. In fact, McCullagh and Nelder (1983) state that modelling remains, partly at least, an art.

COMPUTER PROGRAMS

Most computer programs for analysis of data with unequal sub-class numbers do not incorporate automatic procedures for model selection. Rather they will fit the sequence of models requested by the user. The program GLIM (Baker and Nelder 1978) is ideal for evaluating the fit of various least squares or logit models. The Queensland Department of Primary Industries has a modified version of program HARVEY (Harvey 1968) which fits least squares models, evaluates adjusted means and performs protected t-tests. Model selection for least squares analyses is easier using GLIM than HARVEY. Other computer packages which can be used include GENSTAT, SAS, BMDP and SPSS. When interpreting the results, it is important to know how the computer package constructs the analysis of variance or the analysis of deviance table.

EXAMPLE

A study was conducted between 1977 and 1983 on the conception rate of 458 Bos indicus cross cows on Swan's Lagoon. All cows were at least four years old and were lactating at the start of mating. Conception rate was analysed by coding established pregnancy as 1 and non-pregnant as 0 for each cow. The four explanatory factors were year of mating, genotype, time of calving pre-mating (month) and whether a calf was weaned last season. A stepwise method was used to select terms for inclusion in both the logit model and the least squares model of the binary response variable.

The chosen models are shown in Table 1 with the uninteresting main effects of year, genotype and weaning omitted. The mean squares for month of calving and the two interactions were adjusted for all other terms in each model. The three factor interaction, year by weaning by month, was significant ($P = 0.045$) in the least squares model. It was excluded from the chosen model as the significance was marginal and no two factor interaction involving these factors was

significant. The significance of each term in the logit model is given by its deviance which is distributed approximately as a chi-square. The three factor interaction, year by weaning by month, was not significant ($P = 0.25$).

Table 1 Least squares and logit analysis of conception rate

<u>Source</u>	<u>d.f.</u>	<u>M.S. (least squares)</u>	<u>Deviance (logit)</u>
Month	2	1.8 **	20.6 **
Year x Genotype	6	0.68 ** ($P = .001$)	22.5 ** ($P = .001$)
Genotype x Weaning	1	0.77 * ($P = .04$)	4.2 * ($P = .04$)
Residual	440	0.18	481

The two models gave similar predicted conception rates. For the genotype by weaning table these were:

	Brahman		Sahiwal	
	Weaned	Not weaned	Weaned	Not weaned
least squares	72%	53%	43%	43%
logit	80%	57%	41%	39%

A problem with the least squares model was the finite (0-1) range of the probability of conception. Of the 458 cows, 25 had an estimated probability of conception greater than one using the chosen least squares model. Fitting a logit model is an iterative process which requires considerably more computer time than fitting a least squares model. For example, seven times more computer time on a CYBER 76 was required to select the logit model than was required for the least squares model. This cost needs to be taken into account when choosing a selection method and balanced against the cost of collecting the data.

INTERPRETATION AND EXTENSION OF RESULTS

T.H. McCOSKER*

Experiments are designed so that their results apply to a target population, but conclusions are generally extended in a wider context. Interpretation of results is dependent upon several factors which include the experimental design, frequency and precision of data collection, and amount of data available for analysis. Statistical principles set confidence limits on results and help ensure that biologically important and practical sized differences reach reasonable levels of statistical significance.

Effective extension is dependent upon the differences between the research property and the recipient property in management, climate, soils, phenotypes used and nutrition and also on the attitude of the extension officer. Results from experiments conducted in a commercial environment with management options aligned with those of industry are more readily adopted by producers.

*Resource Consulting Services Pty Ltd, Darwin, N.T. 5794.

INTERPRETATION

Grazing experiments are usually large, which leads to the problem of paddock variability. In the dry tropics, a low stocking rate increases the probability of paddock variability, whereas, in temperate regions, very high stocking rates have a similar impact. The effect of paddock variability may be reduced by stock rotation, by increasing the number of replicates, or by running the paddocks without treatments for a period. These remedies, however, are not always possible due to the trial design or treatments, physical, financial and time constraints. A large data base combining data from several years for a large number of animals leads to more confidence in the results, especially if the scale of operation and management are compatible between research and producer.

Interpretation is constrained by both experimental design and management. Interpretation of results from experiments using atypical management practices is very narrow, and in a commercial sense may be invalid. Management factors such as mating policy, weaning, culling, replacement policy, mustering techniques and nutrition need to be carefully considered in an experiment to balance the conflict between the requirements of an experiment or design and the industry practice. For example, it is simpler to interpret reproductive data from a controlled mated system and therefore most experiments are conducted on that basis, but extensive livestock producers rarely have mating control. Extrapolation from experiments with 60-90 day mating periods is of particular concern here. This type of conflict leads to problems with extension of experimental data into a different set of management circumstances.

Data such as conception, survival and weaning weight and rate need to be interpreted collectively and economically. A high conception rate means little if there are heavy losses of fetuses or offspring between conception and point of sale. An improved weaning rate is of little use if coupled with higher dam loss, low weaning weight and higher weaner mortality. A high survival rate does little for profitability if associated with low reproductive rates. In the extensive livestock industries, where poor nutrition and poor husbandry are frequently the norm, a gain in one parameter usually means loss in another. For instance, weaning in north Australia has increased branding rate but it also increased mortality and lowered liveweight gain of the calf (McCosker et al. 1984). A productivity index, such as weaner weight per breeder mated, could be appropriate to integrate these competing components of reproductive response. It must always be borne in mind that the purpose of improving reproductive performance is to sell more meat, or wool and improve a property's gross margin.

EXTENSION

Extension of experimental results falls into two conceptual areas. The first is to recommend a researched practice in such a way that the producer expects to achieve results similar to those on the research station. The second approach is to recommend the principle of a practice with the expectation that it will improve a producer's situation to a degree unknown. Neither approach provides a producer with objective biological and financial decision-making data. Both approaches are limited by the individual and interactive effects that different management and nutritional regimes have. Stage of property development in terms of both range and herd management imposes a further constraint upon the application of new techniques (McCosker et al. 1982).

There are several examples in the extensive beef industry which illustrate the importance of interactions which occur in the practical setting, but which are not always quantified by research. On Mt Bunday Station, results showed that with controlled mating, non-weaning gave a 50 kg/head advantage to two year old

offspring mass per breeder over weaning on the same plane of nutrition. Uncontrolled mating plus weaning gave higher weaning weights and rates than controlled mating plus weaning (20 kg two-year offspring mass per breeder). The point here is that if one were doing a nutrition experiment, then the results would differ with management regimes, certainly in absolute if not relative terms. A second example of extension pitfalls arises from combining uncontrolled mating, a single weaning time and Bos indicus breeders. Weaning has led to cycling in the dry season, which has led to an out of season calf, resulting in smaller unweanable calves at weaning time and higher breeder mortalities a year later (D.G. Robertson, pers. comm.).

Taylor and Rudder (1982: 1984) investigated, through herd modelling, the long term effects of technological change on breeder herds, and increased drought susceptibility from yearling heifer mating. This has shown the value of herd modelling as a tool to better understand the biology and interactions which occur within a breeding unit. Dynamic herd modelling can act as the missing link between an experiment and an extension situation. However, this approach is limited by the scarcity of suitable data from uncontrolled mating.

The use of models in reproductive research and extension can occur at two levels. Firstly, there is a need for researchers and extension personnel to model experimental results with various assumptions to firmly establish the principles involved. This approach has been illustrated by the work of Baker et al. (1983). Secondly, there is need for extension officers to use models to develop packaged recommendations in conjunction with producers at an individual property level (Lishman et al. 1984).

Modelling alternatives at the property level can overcome several of the problems which occur at that level. Firstly, differences in management systems and competency can be accounted for. For example, it is quite probable that there are practices which would not be recommended to an incompetent manager. Secondly, allowance could be made for changes in scale of operation. Research groups may comprise from 4 to 100 individuals, while commercial groups could range from 50 to 5,000. Larger herds or flocks introduce biological inefficiencies which are generally not measured in experiments, e.g. mis-mothering due to handling large mobs, higher stress, longer periods without feed and water, and greater walking distances. Thirdly, while an extension officer is working with a manager, the problem of different definitions can be resolved. Finally, it would be useful to the producer if the researcher or his extension colleague published results and recommendations in both the scientific and popular rural media, with the latter using industry language and definitions.

SUMMARY AND CONCLUSIONS

P.K. O'ROURKE

Researchers contemplating the planning, design, analysis or interpretation of large scale reproductive research projects should seek assistance from an experienced biometrician at each of these stages. Measurement of conception rate for lactating animals provides the most efficient, early indication of response to a treatment. Issues in experimental design are complex and include specification of the experimental unit, replication rate, method of allocation and policy on replacements.

Methods of data measurement, recording, storage and management can now be largely automated through computers, but the researcher retains responsibility for their accuracy and efficiency. Data base packages are available for large scale reproductive experiments to produce timely summaries and to retrieve data

for statistical analysis.

Logit analysis is a newer, more rigorous analytical method than analysis of variance for the binary response variables from reproductive experiments. The similarity of probability levels from these two forms of analysis demonstrates once again the robustness of analysis of variance even when its major assumptions are violated with binary data. There is no best method for model selection, so guidance is required from an experienced analyst to take into account issues such as non-orthogonality, model parsimony, marginality and features of the computer program. The analyst often must show some perception and imagination in fitting models. Preliminary model reduction for large sets of binary reproductive data using least squares analysis of variance through GLIM is recommended. Further refinements can then be made, if considered necessary, by fitting the selected model using logit analysis. This conclusion and recommendation was also reached by Nicholls and Tyrrell (1980).

Researchers and their collaborating extension officers must be mindful of the use to which their results can be put by a whole range of producers whose management regimes and competency vary, whose property development varies and whose physical resources and objectives vary. Reproductive experiments should be designed to provide basic data for modelling and to validate models as well as for the traditional hypothesis testing objectives. Extension of results from reproductive experiments is dangerous due to varying management regimes, nutritional levels, physical resources and definitions. Use of dynamic herd models is a way of honing extension recommendations to individual property needs.

REFERENCES

- BAKER, R.J., and NELDER, J.A. (1978). The Glim system, Release 3, Generalized linear interactive modelling. Oxford: Numerical Algorithms Group.
- BAKER, J.F., SMITH, G.M. and CARTWRIGHT, T.C. (1983). J. Anim. Sci. 57: 144.
- BLIGHT, G.W. (1984). Proc. Aust. Soc. Anim. Prod. 15: 52.
- BLIGHT, G.W. and PEPPER, P.M. (1982). Proc. Aust. Soc. Anim. Prod. 14: 297.
- BONEAU, C.A. (1960). Psychol. Bull. 57: 49.
- BUTCHER, D.F. and KEMP, K.E. (1974). Appl. Stat. 23: 190.
- COX, D.R. (1970). In "Analysis of Binary Data". (Methuen: London).
- FIENBERG, S.E. (1980). In "Analysis of Cross-classified Categorical Data". 2nd ed. (MIT PRESS: Cambridge, Massachusetts).
- HARVEY, W.R. (1968). In "Instructions for Use of Least Squares and Maximum Likelihood General Purpose Program". (Ohio State University: Columbus).
- HARVEY, W.R. (1982). J. Anim. Sci. 54: 1067.
- HASEMAN, J.K. and KUPPER, L.L. (1979). Biometrics 35: 281.
- HOLROYD, R.G., O'ROURKE, P.K., CLARKE, M.R. and LOXTON, I.D. (1983). Aust. J. Exp. Agric. Anim. Husb. 23: 4.
- LEVY, K.J. and NARULA, S.C. (1977). J. Statist. Comput. Simul. 5: 239.
- LISHMAN, A.W., PATERSON, A.G. and BEGHIN, S.M. (1984). s. Afr. J. Anim. Sci. 14: 164.
- LUNNEY, G.H. (1970). J. Ed. Meas. 7: 263.
- MCCOSKER, T.H., EMERSON, C.A. and GARRARD, A.E. (1982). Proc. Aust. Soc. Anim. Prod. 14: 333.
- MCCOSKER, T.H., EGGINGTON, A.R. and DOYLE, F.W. (1984). Proc. Aust. Soc. Anim. Prod. 15: 452.
- MCCULLAGH, P. and NELDER, J.A. (1983). In "Generalized linear models" (Chapman and Hall: London).
- NELDER, J.A. (1982). Utilitas Mathematica 21B: 141.
- NICHOLLS, P.J. and TYRRELL, R.N. (1980). Proc. Aust. Soc. Anim. Prod. 13: 337.
- TAYLOR, W.J. and RUDDER, T.H. (1982). Proc. Aust. Soc. Anim. Prod. 14: 261.
- TAYLOR, W.J. and RUDDER, T.H. (1984). Proc. Aust. Soc. Anim. Prod. 15: 620.