Kansas State University Libraries [New Prairie Press](https://newprairiepress.org/)

[Conference on Applied Statistics in Agriculture](https://newprairiepress.org/agstatconference) 1995 - 7th Annual Conference Proceedings

A REVIEW OF ANALYSIS OF EXPERIMENTAL DATA

Roger Mead

Follow this and additional works at: [https://newprairiepress.org/agstatconference](https://newprairiepress.org/agstatconference?utm_source=newprairiepress.org%2Fagstatconference%2F1995%2Fproceedings%2F2&utm_medium=PDF&utm_campaign=PDFCoverPages)

Part of the [Agriculture Commons](http://network.bepress.com/hgg/discipline/1076?utm_source=newprairiepress.org%2Fagstatconference%2F1995%2Fproceedings%2F2&utm_medium=PDF&utm_campaign=PDFCoverPages), and the [Applied Statistics Commons](http://network.bepress.com/hgg/discipline/209?utm_source=newprairiepress.org%2Fagstatconference%2F1995%2Fproceedings%2F2&utm_medium=PDF&utm_campaign=PDFCoverPages) \bigcirc 000

This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License.](https://creativecommons.org/licenses/by-nc-nd/4.0/)

Recommended Citation

Mead, Roger (1995). "A REVIEW OF ANALYSIS OF EXPERIMENTAL DATA," Conference on Applied Statistics in Agriculture.<https://doi.org/10.4148/2475-7772.1328>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

A Review of Analysis of Experimental Data

Roger Mead

Department of Applied Statistics, The University of Reading, . *UK*

1. Initial Assumptions

We assume that it is standard practice to base an initial analysis of experimental data on a linear model including terms for blocks, treatments and covariates. This produces a summary analysis of variance, indicating the major components of variation relative to the fitted model, and tables of means (one- or two-way) as indicated from the analysis of variances, with appropriate standard errors. Treatment contrasts as indicated by the questions motivating the experiment may be calculated, with standard errors. Residuals may be calculated and related to the pattern of experimental units. Multiple comparisons should be avoided because they are rarely appropriate, are based on algebra for situations which hardly ever occur, and they lead to experimenters 'interpreting differences', ignoring patterns and generally failing to think. For details of basic analyses and a discussion of the fundamental concepts of experimental design, including further comments on multiple comparisons, see Mead (I988).

The following sections examine some of the main features of analysing experimental data. These may amplify, extend or replace the initial analysis. They may be needed before the initial analysis or to combine several analyses. They may be required to relate the results of different analyses.

2. Identification of Structure

Unfortunately it is all too easy to input data and a brief description of experimental design structure into a statistical package and obtain an analysis of variance and tables of treatment means with standard errors. Unless the statistician has been closely involved in the detailed planning and progress of the experiment it is not easy to be certain about the structure of the experiment and the appropriate linear model.

The biggest problem is the identification of the experimental units and their interrelationship, and in particular the different levels of variation present in the structure and the levels at which information is available about different treatment comparisons. Examples of problems are many. Detection of the correct structure is often a time-consuming and even a depressing activity requiring highly developed consulting skills.

2a. Example 1

The student said his was a simple problem with 8 treatments and 10 replicates per treatment. His F ratio for treatment mean square/error mean square was 500. Under questioning he explained that his 8 treatments were 2 fungicides \times 2 leaf ages \times top/bottom of leaf for injection; treatments were applied to small discs cut from leaves. But how were the discs arranged within the same leaf, in leaves from the same plant or from different plants? Eventually it emerged that there had been 8 plants, with one leaf taken from each plant, and the analysis was in three sections.

Example 2

To investigate effects on soil warming of different materials for the pots containing the soil and of different materials for covering the frames within which the pots were placed, six frames with four pots per frame were used. Three frames had glass covers, three had polythene; in each frame pots of four materials were allocated randomly to the four comers. After two weeks, with one week of measurement, the experimenter decided to swap glass and polythene covers leaving everything else unchanged and repeated the whole procedure. The analysis has three sections in a criss-cross form of design.

Methods for analysing multiple level data structures are discussed in section 4.

3. Data Reduction and Variable Definition

Experimenters have always made many measurements and there has always been a need to calculate appropriate derived variables. Averages of values for different plants, animals or parts of plants, ratios of variables, differences between initial and final measurements or combinations of all four arithmetical operators have been needed. Recently, however, the capacity for recording data has increased with automated data recording so that measurements are recorded at very many time points, at many spatial sites or at an almost infinite number of wavelengths in spectra.

The first task of analysis is now frequently the reduction of an enormous mass of data to a manageable number of relevant variables. Ideally this is mainly achieved by consideration of the experimental objectives and subjective decisions about the definition of variables. Data reduction techniques, notably principal components can also be helpful. Data reduction must precede, and not interact with analysis. There are enormous problems about inference if the choice of variables is influenced by optimising some aspect of the analysis.

4. Multiple Levels of Variation

There are some simple structures where each treatment effect is identified with a single level of variation in which all the information about that effect is contained. Such structures are orthogonal and are mostly variants of the split unit design. However there are many other structures where information on individual effects is divided between levels. Examples are incomplete block designs including multiple blocking, and multiple location trials with varying treatment sets, as found in on-farm research.

The general technique for analysis structures with multiple levels of variation is REML (1987) (Residual Maximum Likelihood) for which an iterative system estimates alternately (a) the several variance components required to describe the variation structure of the experimental units, and (b) the parameters (treatment effects) of the expected values for the observations.

Examples of the use ofREML can be found for various row and column structures for field plot experiments, for animal breeding data structures and in spatial models (Gleeson & Cullis; 1987). A simple example concerns the repeated use of control treatments in a series of experiments, often with just one non-control treatment.

A common procedure in laboratory trials is to run a sequence of experiments each with a common control and with one or more test treatments. The inclusion of the control has two purposes:

- (i) for adjusting for changes between experiments, and
- (ii) because comparison with control is necessary for each test treatment.

If there is little difference between experiments we should pool the information on the control from different experiments, ignoring differences between experiments. However if there are differences between experiments we would wish to adjust for them. We can envisage three models - adjustment, pooling, or random experiment effects fitted by REML.

For data from four experiments, shown in Table 1, the model specification and results are shown in Table 2. The random experiment effects model gives means between the values for the other two models and give the smallest SE for comparing control with any treatment and an intermediate SE for comparing treatments (note: the corresponding SE for comparing two treatments from the pooling analysis is clearly too optimistic in view of the clear need for some adjustment).

4.1 Another aspect of multiple levels of variation is involved in experiments at multiple sites and/or in different years. The analysis of variance of the combined data typically includes three types of terms:

There is much argument for this (and for equivalent situations) about the 'right' mean square against which to assess the variation between treatments. I believe that there is no single right answer but that all three mean squares must be interpreted together in an "analysis of inconsistency". That is we have to examine and interpret the pattern of variation of treatment differences both within experiments (between blocks) and between sites, with possible explanatory variables measured at the different sites.

A related problem is when, and how, a combined analysis of variance for all the data for the separate experiments should be calculated. The first question is whether it is appropriate to calculate a combined analysis of variance. For this we have to decide that the pattern of random variation is sufficiently similar in the separate analysis; the prevailing wisdom seems to be that a useful combined analysis is obtained if the individual experiment error mean squares vary apparently randomly and by a factor of not more than 10. The vague similarity of EMS is a better criterion than CVs (Coefficients of Variation) or Treatment Factors each of which is a ratio and can lead to omitting experiments for the wrong reason.

If a combined analysis is performed and we examine and interpret the mean squares for effects and their interactions with sites then the overall pattern of sizes of mean squares is important. Also important is to allow for the inevitable variation when splitting site interaction variation into many components. An example for a set of 11 barley experiments in Syria, for which the treatments were a 4 Nitrogen \times 4 Phosphorus factorial set, is shown in Table 3. In this case the mean squares are remarkably wellbehaved with main effects always considerably larger than interactions involving those main effects, linear effects similarly larger than the remaining treatment factor variation, and the interaction of sites \times residual factor variation being very close to the pooled error mean square. Such good behaviour is unusual; if the treatment \times site interaction is split into more and more components with fewer d.f. it is increasingly likely that unusually high or low values will occur for some components and give rise to F ratios for (Treatment MS/Treatment \times Site MS) which are inappropriately low or high.

5. Modelling Variation

There have been considerable advances recently in attempting to model the error variation in experimental data. Particular attention has been given to modelling probability distributions for errors and to spatial and temporal models. One area where a very simple approach is often needed is for data in which a substantial proportion of the observations are zero. A realistic model of this situation can often include two distinct kinds of zero; either one less than 1.0, or impossible. I believe we must not allow one model to be used for both. Such data require two variables representing a qualitative and a quantitative effect. In Table 4 data from a $2⁴$ germination experiment with four replicate dishes for 50 seeds for each treatment combination are shown. It is clear that several treatments lead to virtually zero or total germination with virtually no variation to analyse. I believe the data in rows 1, 3, 5, 8 and 9 should be treated as qualitatively different and separated from the remaining data before analysis.

The generalised linear models of McCullagh and Neider (1990), most clearly specified in the use of the GUM program, allow for a wide range of distribution modelling for the random variable to describe individual observations from experimental units. The use of deviance in the form of $a - 2$ log likelihood to describe departures from a sequence of fitted models is now becoming widespread, even in some agricultural journals. It is also becoming widely recognised that in many situations the relatively simple distributional model is not sufficient to describe all the deviations and overdispersion is a well discussed concept.

An analysis of the remaining data in Table 4 using binomial error structure for a logit link gives an analysis of deviance table, (the equivalent of an analysis of variance table for ordinary linear models) shown in Table 5, which appears to show a simple pattern of three main effects and their three two-factor interaction with the residual deviance close to its d.f.. Note that all effects involving the chill factor make negligible deviance contributions. However, the satisfaction with the model is diminished by the prediction for constant temperature, dark and $H_2 0$ for which a germination proportion of 0.2 (10/50) is predicted when 7 out of 8 replicates produced zeros. Clearly the model needs further investigation.

Spatial models have become important in field crop experiments (Bartlett (1978), Wilkinson *et al* (1983), Besag and Kempton (1986), Cullis and Gleeson (1989, 1991)); the methods are also more widely applicable for spatially structured data for example in laboratory assays. These models are mainly based on first differences from a sequence of plots with an autoregressive model possibly combined with seconddifferencing to describe the variance structures. Temporal models (Diggle (1988)) developed mainly for repeated measures data have utilised semi-variogram infonnation and modelling of the semi-variogram.

The two approaches are very similar in practice, involving the usual iteration between fitting a model for the variance structure and for the expected value parameters. An example of this form of modelling is taken from data on intercropping safflower and cowpea in a systematic design (Willey and Rao : 1981). Safflower and cowpea are planted in alternate rows, the cowpea density remaining constant from row to row, but the safflower density changing gradually through 14 rows to produce an overall density change from 4.4 plants/m² to 16.7 plants/m² in steps of 10%.

A variance model

$$
E_i = U_i + \eta_{ij} + W_i(j)
$$

where U_i is the whole plot effect $(\text{var}(U_i) = \nu^2)$, η_{ii} is the row j in plot i effect $var(\eta_{ij}) = \tau^2$ and $W_{i(j)}$ is a stationary random process

$$
Cov(W_i(j), W_i(k)) = \sigma^2 \exp(-\phi(j-k)^2).
$$

The mean model for yield per unit area (y_i) in terms of plant density (p_i) is

$$
E\big(\log(y_i)\big) = \log\bigl(\rho_j\big/\bigl(\alpha + \beta \rho_j\bigr)\bigr)
$$

where α and β are parameters depending on other factor levels.

Using an iterative scheme to minimise -2 log likelihood (Castro : 1992) we can examine the invariance of parameters over four main plot treatments, which are four overall chickpea densities. The results are summarised in Table 6 from which the model with a single common value for v^2 , σ^2 , ϕ and α , and separate values of τ^2 and β appear best.

One possible heretical use for spatial models is in large-plot validation trials or ecological experiments when true replication is very limited or non-existent. If a credible spatial model can be fitted to within plot samples or to an adjacent replicated trial then extrapolation of the variance model could permit comparison of results from unreplicated large plots.

6. Interaction

The basic concept of an interaction between two, or more, factors is as important to analysis as factorial structure is to design. The idea of representing variation between experimental treatments by a linear combination of main effects and interactions, which effects are defined as combination of yields from the treatment combinations, and then choosing which interactions to include in the prediction model provides great flexibility in the interpretation of results.

One approach to interpreting the effects of interactions is to treat two factors symmetrically and to examine either the interaction effects or the two-way table of fitted means. In the table of interaction effects we can pick out large individual effects or trends. In table 7(a) we have a typical diagonal trend pattern for two 4-level quantitative factors: in $7(b)$ we have, also for two 4-level quantitative factors, a single large value. In the table of means the maximum fitted value will be given by the combination of levels with the highest main effects unless an interaction effect is greater than the smaller of the main effects.

Alternatively we can approach interactions primarily as modifiers of the response of the factor with the largest main effect. Usually this factor will be included in the larger two-factor interaction. This leads to analysis of contrasts which can be a very powerful interpretive analysis.

6.1 Genotype - Environment Interaction

A lot of attention has been paid to the interpretation and modelling of genotypeenvironment interactions. Some of the previous approaches, such as examining where the larger interactions occur are useful but most of the modelling is concerned with relating the vector of interaction effects for each genotype to an environment vector. The central problem is what environment vector? Ideally we would like a vector which summarises those characteristics of the environments which cause genotypes to behave differently but we do not usually have physical measurements of the environments. Two approaches, each of which is flawed, are:

- (i) Use the environment main effects vector (the joint regression, Finlay-Wilkinson method). The flaws are (a) why should it work and (b) it gives regression slopes which depend on the sum of squares for the genotype vector and the sum of products with each other genotype vector. Hence the regression slopes are measuring similarity to the other genotypes in the particular study (perhaps this is inevitable).
- (ii) Use the eigenvector of the HH' matrix where H is the matrix of interaction effects (the AMMI model). Unfortunately this method tends to explain a large percentage of the total interaction variation (Mandel (1971)), even when simulated data generated with zero interaction terms are analysed. This method is overoptimistic because of the correlations in the matrix of interaction effects, and because the eigenvector is the vector giving the maximum regression sums of squares and we are essentially overfitting. Studies for methods of correcting the 'optimism' of the AMMI method are being pursued. The philosophic approach of Gauch (1988) still produces overfitting.

7. Modelling Relationships between Variables

In most experiments several variables are recorded. In intercropping experiments with two (or more) crops grown simultaneously on the same area there are two crop yields recorded. In this latter case it is natural to perform a joint (bivariate) analysis on the pairs of yields. In other experiments it may also be reasonable to calculate a bivariate or multivariate analysis for different variables. Experimenters often wish to follow a series of univariate analyses by examining the regression of one variable on another. For all these situations we should consider an initial bivariate analysis which in its simplest form is

The data structure is, of course that of a covariance analysis, adjusting the variate y for its dependence on the covariate x. In a bivariate analysis for two yield variables the two variables are treated symmetrically and the bivariate analysis is used to transform the two original variables to new variables which are uncorrelated relative to the error variance-covariance matrix. This produces a skew axis representation of the original variables, with precision of comparisons independent of direction. This skew axis representation provides a powerful interpretive tool (Mead (1986».

To illustrate the use of bivariate analysis and the skew axes interpretation of data, Table 8 displays the bivariate analysis of variance for an experiment on maize and cowpea with three treatment factors. The residual correlation of -0.46 shows a substantial competitive situation. There are clearly big effects of maize variety (principally on maize yield) and of nitrogen (even more dominantly on maize yield and clear evidence of an interaction of cowpea variety and nitrogen. The skew axes plots for maize variable and cowpea variety \times nitrogen are shown in Figure 1. We can distinguish four directions of effects in a skew axes diagram: (a) parallel to the cowpea yield axis (change of cowpea yield); (b) parallel to the maize yield axis; (c) vertical (both yields increase); (d) horizontal (changing proportions of crop yields). In l(a) the differences between maize varieties are mainly a change of the proportions of maize and cowpea yield. In l(b) the interaction is clearly displayed qualitatively. For variety A (dots) increasing nitrogen gives increasing maize yield (trend parallel to maize axis); for variety $B(+)$ the effect of increasing nitrogen is mainly to change the proportion of maize: cowpea (horizontal) with a tendency towards increasing maize as well.

One of the reasons for the effectiveness of covariance analysis or of a bivariate analysis is that the relationship between x and y is different in the treatment line of the analysis from that in the error line. However when regression analysis is used to see how far y depends on x, it is common to use the total line to calculate the regression. This seems to be incorrect because it ignores the structure of the data. An alternative of examining the regression for each treatment and comparing the regression is practically sensible, though it may be corrupted by block differences, but often the number of replicates per treatment is too low for much power in the comparison of regression. Fitting separate regressions for the treatment line and the error line and comparing them could give relevant information.

7.1 Modelling response relationships

Another form of analysis is modelling the response of the yield variable to one or more input factors. Examples have already been considered in section 5. The full range of linear regression and non-linear regression models are available but it is important that the details of the experimental structure are allowed for in the modelling. In terms of the bivariate analysis structure, if x is an input variable, and if a complete block design is used then variation of x occurs only in the treatment line and this is the only set of information which should be used in fitting a response. For more complex design structures the situation may be more subtle but care is still required.

8. Sequential Information

Too frequently each experimental design and data analysis is treated in isolation. In practice both design and analysis take place within an overall programme of research activity. In terms of analysis, particularly, this means that there is usually some previous information about the parameters, for which the current experiment is intended to provide estimates. The standard statistical approach to such situations is to represent prior information in the form of prior distributions and to use a Bayesian system to combine prior information and present data. This philosophy has been developed formally for sequences of industrial experiments. It also is used, rather vaguely and imprecisely, in combining results from stages of a selection procedure in breeding programmes.

At a very simple level, if we have three stages of a variety selection programme and we consider those varieties present in all three stages we would expect to get better comparisons by using not only the final stage but also the two earlier stages. However when we consider how to combine the results, e.g. what weighting procedure to use, we have to make assumptions about the extent of stage \times variety interaction. Absence of interaction would suggest weighting by replicate numbers. Large interaction would suggest equal weighting.

9. Questions and Decisions

It is important that, like the experimental design, the choice of the method of analysis is motivated by the need to answer the questions being posed. Many questions are in the form of decisions. Which varieties to select for the next stages of testing? Which optimal combination of factor levels? There are also the questions about size of differences or effects or rates of change. Precision of estimates and probabilities of decisions being correct are desirable adjuncts to the estimates and decisions. However, the concept of significance seems to be almost irrelevant to most answers to questions.

References

- Bartlett, M.S. (1978). Nearest neighbour models in the analysis of field experiments. *J.R. Statist. Soc. B,* 40, 147-174.
- Besag, 1. and R. Kempton (1986). Statistical analysis of field experiments using neighbouring plots. *Biometrics,* 42, 231-251.
- Castro, J.AJ. (1992). Analysis of data from field plot experiments using models for spatial covariance and yield response. Ph.D. thesis University of Reading.
- Cullis, B.R. and Gleeson, AC. (1989). The efficiency of neighbour analyses for replicated variety trials in Australia. *Journal of Agricultural Science,* 113,233-239.
- Cullis, B.R. and Gleeson, A.C. (1991). Spatial analysis of field experiments an extension of two dimensions. *Biometrics,* 47, 1449-1469.
- Diggle, PJ. (1988). An approach to the analysis of repeated measurements. *Biometrics, 44,* 959-971.
- Finlay, K.W. and Wilkinson, G.N. (1963). The analysis of adaptation in a plant-breeding programme. *Austral. J. Agric. Res.* 14, 742-754.
- Gauch Jr, H.G. (1988). Model selection and validation for yield trials with interaction. *Biometries,* 44, 705-715.
- Gleeson, A.C. and Cullis, B.R. (1987). Residual maximum likelihood (REML) estimation of a neighbour model for field experiments. *Biometrics,* 43,277-288.
- McCullagh, P. and Neider, J.A (1990). *Generalized Linear Models,* 3rd edn. Chapman and Hall, London.
- Mandel, J. (1971). A new analysis of variance model for non-additive data. *Technometrics,* 13, 1-18.
- Mead, R. (1986). Statistical methods for multiple cropping. In Multiple cropping systems. C.A Francis (editor). London, Collier Macmillan. 371-350.
- Mead, R. (1988). *The Design of Experiments* : *Statistical Principles for Practical Application, Cambridge University Press, Cambridge, U.K.*
- REML (1987). *Program and Manual Scottish Agricultural Statistics Service*, Edinburgh.
- Wilkinson, G.N., Eckhert, S.R., Hancock, T.W. and Mayo, O. (1983). Nearest neighbour (NN) analysis offield experiments (with discussion). *J.R. Stat. Soc. B.,* 45, 151-211.
- Willey, RW. and Rao, M.R. (1981). A systematic design to examine effects of plant population and spatial arrangement in intercropping, illustrated by an experiment on chickpea/safflower. *Expl. Agric.* 17,63-73.

Results from four completely randomised experiments with control and test structures

Table 2

Model assumptions and results for data from Table 1

Mean Squares and d.f. for a detailed combined analysis of variation for 11 barley trials

14
4

Table 4

Germination counts (out of 50 seeds per dish) for 16 experimental conditions

Changes in deviances for a logit model for part of the data of Table 4

Summary of invariance fits for four variance parameters and two mean parameters, over four main plots with different chickpea densities

 $(C = common parameter value; I = four individual parameter values)$

Two-way table of intraction effects from a $\frac{1}{4}$ replicate of a 4⁴ industrial experiment; (a) Moisture x Temp, showing opposite trends in the first and fourth rows, (b) $pH \times Time$, showing a large negative value in (row 4, column 4), with associated positive values

Bivariate Analysis of Variance for Maize/Cowpea Yield Data

Figure 1

Skew axis representation of maize-cowpea yields as affected by (a) maize variety and (b) cowpea variety \times nitrogen

