Jour. Ind. Soc. Ag. Statistics
49 (Golden Jub. No.), 1996-97 : 11–20

Temporal or Spatial Repetition

David J. Finney
13 Oswald Court, Edinburgh, UK

SUMMARY

This paper looks systematically at experiments designed according to sound principles in which data are recorded in accordance with a plan of measurements repeated on several occasions. Types of data are described, with a following discussion on types of analysis: split-plot, a separate analysis for each occasion, use of carefully specified compound variates, and multivariate analyses. The dangers of uncritical split-plot analysis are emphasized; an example of specifying compound variates is discussed in detail with an account of how the choice affects error mean squares. The paper ends with brief notes on software and the apparent absence of facilities in standard packages.

Key words : Types of data, Precautions in analysis, Error estimates, Choice of compound variate, Multivariate analysis.

1. Introduction

From the earliest days of involvement of statisticians in agricultural research, important problems in the analysis and interpretation of data from experiments have arisen in connexion with field plots on which measurements of performance have been recorded at regular intervals during a short or long period of continuous treatment. Indeed, the reason for the appointment in 1919 of a statistician at Rothamsted Experimental Station was that data of this character had been accumulating for nearly eighty years without serious attempt to interpret them comprehensively for the benefit of cereal growing practice in England.

The story of R. A. Fisher's appointment and innovatory analyses on yields from the continuous wheat experiment is too well known to need recapitulation. By the standards of today, the experiment was of flawed design, lacking randomization or replication, but careful management had produced potentially important data. In two early papers, Fisher ([6], [7]) presented ingenious approaches that, although now seeming archaic, contained the seed of the technique of analysis of variance[1] that he was later to develop so superlatively.

1 Although the abbreviation ANOVA is used in some circles, to anyone who has been closely associated with analysis of variance since the 1930s it must surely seem a pointless and ugly abomination.

Analysis of variance soon became the major statistical procedure for all analyses of data from planned experiments. Although it originated in research on field crops, it is equally relevant to animal experiments, and to research in many branches of medical science. In all circumstances where interest lies in studying the performance of plots or other experimental units[1] over a period of time in which they are subject to either a continuously applied treatment or a well-planned sequence of changes in treatment, complications arise of a kind commonly known in statistical literature as due to 'repeated measurements'. Difficulties enter because of non-independence of successive observations on the same unit, alternatively described as correlation of components of experimental error within a unit and attributable to the impossibility of randomizing order in time !

Gross mistakes of method in statistical analysis of repeated measurement data are deplorably common. Major faults have arisen through neglect of correlational structure, implicit adoption of an erroneous model, and lack of understanding of the essential randomization features of split-plot designs. These mistakes are not restricted to small details of arithmetical exactness but may lead to total misunderstanding of the lessons that an experiment can teach. In an age when a professional statistician who is employed to analyze experiments, or indeed any agronomist or other biologist who needs to analyze his own data, expects ready access to software that will relieve him of all arithmetical labour, danger lies in unthinking abuse of software that may be perfectly sound for a different purpose. The aim of the present paper is to formulate and illustrate systematic rules and warnings. There is a lack of language adequate for exact description of the possible patterns of design and data records, but a selection of examples will illustrate the main points to be emphasized.

2. *Types of Data*

In a typical modern field experiment, p different treatments (crop varieties, fertilizers, methods of cultivation, etc.) are allocated at random to r different replicate plots. The design may be simply unrestricted randomization over the N = pr plots, but more usually it will have a combinatorial structure. For example, it may have r randomized blocks of p plots each. The treatments may be a simple or complicated factorial set, so that p is a suitable composite integer. Further constraints may be placed on the replicates, introducing such familiar terms as Latin square, lattice, split-plots, confounding, and many more. During a growing season, each plot will generate a single value for a yield variate, y; this may be a simple weight of crop, a measure of pest or disease attack, any other defined and measurable quantity, or even a value of one of these transformed in accordance with statistical recommendations. A standard analysis of variance for the experiment consists in forming, and suitably partitioning, the total sum of squares for y with (pr - 1) degrees of freedom.

1 In what follows, the terms 'experimental unit' and 'plot' will often be used as synonymous.

The further possibility is that, instead of a single y, a series of values is determined at intervals on each plot. For example, disease status might be reassessed weekly throughout the season or the weight of crop might be recorded annually on the same plots for several years. Let there be τ such occasions, so that for each plot the data comprise the τ values of y_t for $t = 1, 2, ..., \tau$, so representing $pr\tau$ values in all for analysis.

Commonly, the treatments of the p plots will remain constant throughout, or will be renewed at regular intervals as with annual fertilizer or cultivation treatment. Alternatively, there may be a complex pattern of balanced changes in the treatments (changes in fertilizer applications or even in the variety or species of crop plant). At one time, such experiments on cropping or management rotations received much attention in agricultural research, and the principles established for statistical interpretation were in full accord with the general outlook of the present paper.

The same possibilities apply to animal experiments or even in human clinical research, there usually being termed cross-over or change-over experimental designs. The unit plot of land may be replaced by a single animal - perhaps a dairy cow whose production of milk is to be recorded in τ successive lactations or other defined portions of her productive life. In other branches of biology, the animal may be a standard laboratory mammal that is to be subjected to a complex treatment regime, the effects of which will be assessed by analyses of blood or other samples taken on τ occasions. Recently, a pharmacologist asked my help with data from an experiment in which each of three doses of a new drug had been administered to three male and three female animals; on 4 occasions during the year after administering the first dose, he had made a biochemical measurement of a blood property on a sample from each of the 18 animals.

A permissible but uncommon variant in the type of data is that encountered when the experimenter chooses to measure different properties of the experimental unit on different occasions; on occasion τ, all measurements will relate to the same characteristic of the animal or plot, but on different occasions the experimenter may prefer to measure something different. For example, the y_1 may all represent blood-sugar determinations for every animals, but the y_2 might be based on a new method of estimating sugar; the y_3 might be a measure of haemoglobin and y_4 any one of these measurements but now made by a different technician.

Exactly the same logical issues can arise in experiments where time is replaced by physical position! For example, a replicated experiment on varieties of a crop might be conducted on long plots running at right angles to an irrigation channel. Effects of irrigation might be studied, with $\tau = 4$, by measuring some property of the growing crop on each plot at 2, 5, 10, 15 metres from the water source.

3. *Types of Analysis*

Statistical analysis of repeated measurement data can be discussed under four main heads, as described below. All make use of analysis of variance. As always for such analysis, past experience allied to current evidence may need to be considered in relation to additivity and any need for transformation as a precursor of inferences resting upon assumptions of distributional Normality.

A. The analysis of experiments designed in split-plots has been familiar to statisticians and agricultural scientists for many years. Its very simplicity has tended to encourage excessive use of a design that is often far from optimal for a multi-factor study, but popularity has ensured that software for the analysis is widely available. One arithmetically possible analysis for the typical repeated measurement experiment is to expand whatever analysis is appropriate to the pr plots into that which has each of these plots split into τ sub-plots. The possibility of performing such an analysis does not guarantee that it is an optimal choice or even that it is logically defensible.

B. Another possibility is to compute τ separate analyses of variance, each on the pr values of y_τ for a particular occasion τ. Each analysis is structured appropriately to the design, exactly as if there were no repeated measurements.

C. Each experimental unit has provided τ observations. Clearly new variates can be compounded from these as indices of the performance of the plot. They may be chosen arbitrarily in a manner that represents the main interests of the experimenter. It is usually preferable that not more than τ compound variates be defined, in order to avoid redundancy of information; often, one or two will suffice.

D. A final possibility is to regard the data as providing τ-variate information on each of the pr plots of the main design. Any form of multivariate analysis of variance may then be considered as a route to interpretation and conclusions.

4. *The Flaws in 'A'*

The split-plot analysis, A, can be a great temptation. Before, raising questions with me, the pharmacologist mentioned above had used software readily available on his Macintosh PC, namely STATVIEW (Anon [1]). This software was new to me. I wasted much time before discovering that the package, specially written for factorial experiments, had regarded time uncritically as the splitting factor in a standard split-plot analysis. Naturally, it produced extensive tables of analyses of variance, with tests of significance for complicated interactions of dose, sex, and time, which could later be seen as largely irrelevant.

It should be intuitively obvious that a plot or animal may have a component of experimental error, a deviation from an additive parametric representation, that persists from one occasion to the next. If this is present, any residual mean square based upon intra-plot variation will have a bias relative to the true plot error. The root of the trouble is the impossibility of randomizing in time; the consequential serial correlation between successive measurements on the same animal contravenes the assumptions of independence that underlie a split-plot analysis. In a classic paper, Rowell and Walters [9] presented abundant empirical evidence of this phenomenon. Fisher [6] must have had the point in mind when he carefully restricted his attention to study of intra-plot variation, and absence of replication protected him from any attempt to use other information on experimental error. Almost certainly under Yate's influence, the same point was implicit in Turner's pioneering paper [10], and Yates [11] developed the theme further. I myself [3] contributed further warnings of the manner in which a split-plot analysis can lead to serious bias in estimation of error mean squares; biases in tests of significance and in statements about confidence limits in 'repeated measurement' experiments are then almost inevitable.

My pharmacologist friend is not be blamed for trusting a package that came from an apparently reputable source. STATVIEW may be arithmetically sound, but it lacks any warning against misuse, indeed its manual contains instructions advising the erroneous assumption that 'Occasions' should be regarded as one more factor in an analysis exactly as condemned above. A biologist employed in a research institute may have had no training fitting him to judge the appropriateness of the available software, nor access to anyone who can advise him or who can develop a program to meet a perceived special need. Faults lie not only in software manuals but also in some elementary textbooks that may be read as implying : "If you are able to construct a data file that software package XX will accept, go ahead". This uncritical message takes no account of whether the assumptions inherent in XX correspond with a model and error structure reasonable for the data. Gomez and Gomez [8] is a textbook commendable for its special attention to agricultural research in tropical regions; its explicit recommendation of the split-plot analysis for repeated measurements, without comment on its validity, is a deplorable flaw.

5. *Multiple Analyses ('B')*

The structure and randomization of design for the pr plots determines the form of the main analysis of variance. Repetition of this for each of τ variates is easy. Those who worship tests of statistical significance will now encounter difficulties, for one aspect of correlation between occasions is that the τ analyses are not logically independent. There can be no simple rule for correct inference, but an underlying tendency for the analysis at $\tau = 2$ to tell much the same story as that for $\tau = 1$ must not be ignored. Nevertheless, a broad inspection of the τ analyses may help the planning of a better analysis of type C.

6. *Analysis by Chosen Functions ('C')*

Analysis of type C conforms closely to what Rowell and Walters advised and illustrated. It can be well explained by consideration of the pharmacological experiment mentioned in Section 2 above. Although in respect of total number of plots, this was small relative to most field experiments, for pharmacokinetic research it was a large and costly experiment which contained many important features. The main design, which has $p = 6 (= 2 \times 3)$ and $r = 3$, was an unrestricted randomization for 18 animals, 9 of each sex, and 3 doses of a drug under study assigned to each set of 9. The doses, with amounts unchanged, were administered daily throughout the experiment. The repeated measure aspect of the experiment had $\tau = 4$, the standard system of blood sampling and measurement being used immediately before the first dosing (so as to give x_1) and at 1, 6, 12 months thereafter. Hence, the fundamental analysis of variance has the form :

	Sex (S)	Dose (D)	(S.D.)	Error	Total
d.f.	1	2	2	12	17

Of course, if desired, D could be divided into two orthogonal contrasts, D1(linear) and D2 (quadratic).

An obvious compound variate to study is the simple total for each animal :

$$L_0 = y_1 + y_2 + y_3 + y_4$$

analysis of variance of L_0, or of $L_0/4$, can be interpreted in terms of the long-term consequences of administration of a constant dose of the drug.

Adoption of a plan of repeated measurements presumably indicated interest in time trends, for which purpose one might look particularly at a second compound variate :

$$L_1 = -3y_1 + y_2 + y_3 + y_4$$

for each animal, this is 3 times the difference between the initial state and the average of all measurements after dosing begins. This can be examined with the same scheme of analysis of variance but with an unfamiliar modification:

	L_1	$L_1 \cdot S$	$L_1 \cdot D$	$L_1 \cdot S \cdot D$	Error	Total
d.f.	1	1	2	2	12	18

In this analysis, the quantity traditionally called 'correction for the mean' is entered as the square for L_1, and all subsequent items are computed by the standard rules as were those for the L_0 analysis. The error mean square is appropriate to any significant test or precision statement on the mean of the compound variate. The remaining items of the analysis aid examination of evidence on how the change measured by L_1 depends upon sex and size of dose.

A set of τ compound variates might be completed by defining :

$$L_2 = y_2 - 2y_3 + y_4$$

and

$$L_3 = -y_2 + y_4$$

each being analysed as was L_1. Here L_2 relates to curvature in the time-trend of the measured blood property, L_3 relates to the total change over a long interval.

The four L-variates may not be the best possible choice among an infinity of possibilities. Their symbolic orthogonality is scarcely important, but to anyone familiar with the intricacies of analysis of variance they can provide various confirmations of arithmetic. Indeed, an automatic following of standard scaling of sums of squares will give four separate analyses that can be summed and rearranged in obvious manner so as to agree with the long and potentially misleading analysis of variance that results from type A, or split-plot, analysis as output from uncritical use of STATVIEW or similar software. Some users might prefer to replace L_1 by an alternative estimator of linear trend, for example the common :

$$L = -3y_1 - y_2 + y_3 + 3y_4$$

The error estimates may be better understood by reference to an additive linear model, of the kind that usually underlies analysis of variance. Suppose that for the blood property under study there is a value for each animal an added (or subtracted) contribution for its sex, dose and occasion status in addition to a component of random error. The random error components on different occasions may be correlated. As a simple first attempt to be general, write σ^2 for the variance among animals in respect of their initial values and ρ_i for the correlation coefficient between random errors on the same animal i time units apart; the inequality of lapse of time between successive measurements may render this assumption incorrect, but it should be adequate for qualitative indications here. By simple elementary algebra for variances of

linear functions of correlated variables, the expectations of the error mean squares in the correctly scaled L_0 to L_3 analyses are :

$$E_0 = \sigma^2 [\, 1 + (3\rho_1 + 2\rho_2 + \rho_3)/2 \,] \qquad E_1 = \sigma^2 [\, 1 - (\rho_1 + 2\rho_2 + 3\rho_3)/6 \,]$$

$$E_2 = \sigma^2 [\, 1 - (4\rho_1 - \rho_2)/3 \,] \qquad E_3 = \sigma^2 [\, 1 - \rho_2 \,]$$

In most experimental situations, one would expect the ρ_i to be positive; the above formulae then suggest that it will be usual for E_0 to exceed σ^2 but for E_1, E_2, E_3 to be less than σ^2, some indication of the manner in which a split-plot analysis may mislead its user.

Various special cases deserve attention as possibly reasonable in some circumstances. One corresponds with a continuous flow of correlative influences that could produce :

$$\rho_1 = \rho; \quad \rho_2 = \rho^2; \quad \rho_3 = \rho^3$$

This does not make the formulae appreciably simpler. Alternatively, one might speculate that the correlation coefficient is independent of the time interval, so that all ρ_i are equal, whence :

$$E_0 = \sigma^2 [\, 1 + 3\rho \,] \qquad E_1 = \sigma^2 [\, 1 - \rho \,]$$

$$E_2 = \sigma^2 [\, 1 - \rho \,] \qquad E_3 = \sigma^2 [\, 1 - \rho \,]$$

If all $\rho_i = 0$, the four error mean squares have expectation σ^2, necessary and sufficient conditions for the split-plot analysis to be valid. At the other extreme, if all $\rho_i = 1$, the definitions make clear that every value of L_1, L_2, L_3 must be zero, and therefore the corresponding error mean squares must be zero. With actual data, one might try to use the empirical values of the least squares in order to find values of σ^2 and the ρ_i that are in reasonable numerical agreement; this is unlikely to be very satisfactory when each mean square is based on as few as 12 d.f.

In the use of a type C analysis, the usefulness of interpretation of data that is achieved may greatly depend upon successful choice of one or more compound variates, in difficult cases possibly aided by a preliminary type B analysis. Simplicity, allied to meaningfulness, should be the rule. When, as in the experiment discussed above the treatments remain fixed and are applied continuously throughout, choice may be easy. Rotation and change-over designs can introduce much greater complexity of treatment sequence on a plot, possibly with a cyclic repetition.

The choice of compound variate should always take account of its intrinsic scientific interest. In the example discussed above, there is no absolute reason against using such a compound as

$$L = \log(y_1 + y_2) / (15.0 - y_3/y_4)$$

if that were thought to have biological meaning, although analysis of variance and any tests of statistical significance might demand consideration of additivity of plot errors or a transformation to deal with non-Normality of error distribution. There is no requirement that choice be restricted to functions linear in the recorded measurements, although often linear functions can capture the main features of the time trend. For the extreme case of $\tau = 2$, the only linear function must be essentially the difference between y_1 and y_2; in this situation, and in no other, a split-plot analysis may be all that is required since it is equivalent to analysis in the manner of Rowell and Walters.

In agricultural rotation experiments, it may be important to adopt a parametrization that takes account of residual effects from one season of treatment and cropping upon the future performance of a plot; such effects may be consequences of physico-chemical changes in the soil or of cumulative fertility benefits after a leguminous crop. Much has been published about particular examples, but some of the ideas presented here may be helpful in the face of new difficulties or in planning for a major new experiment.

7. *Multivariate ('D')*

Undoubtedly the data from an experiment for which repeated measurements have been recorded are intrinsically multivariate, pr observations structured for a chosen and imposed design but each observation relating to a set of τ variates. Any standard method for multivariate analysis might legitimately be tried. Such an analysis may lead to principal components, or to other functions of the τ variates, determined so as to have certain optimal properties. There can be no assurance that such functions will have any self-evident interpretation in relation to the objectives of the experiment.

In a field of research where many closely related experiments with repeated measurements are being conducted, a formal multivariate analysis might produce useful ideas for functions worth future study but there may rarely be an experiment large enough to provide a definitive interpretation on internal evidence alone. The many procedures for multivariate analysis that can be found in the literature of statistics have the merit of reducing the dimensionality of the data below τ, but because they handle all the variates as of equal status they can take no account of order in time which is likely to be a primary interest of the agricultural scientist or other investigator.

8. *Software*

There does not appear to be any general software written to suit the needs of experiments with repeated measurements. The plausibility but potential faults of STATVIEW have been emphasized; a user prepared to do some *ad hoc* programming to overcome its weaknesses should be able fairly rapidly to develop an analysis of type C, or similarly to built upon what can be done with other standard analysis of variance software. GENSTAT also ought to be able to do much of what is required for a comprehensive type C analysis, based upon the user's specification of the variates to be analyzed. A research worker who undertakes many such experiments might be well advised to stimulate statistical colleagues to write a good general package; the task should not prove an excessive labour.

REFERENCES

[1] Anon, 1992. Abacus Concepts STATVIEW. Abacus Concepts Inc., Berkeley, California.

[2] Crowder, M. J. and Hand, D. J., 1990. Analysis of repeated measures. *Monographs on Statistics and Applied Probability*, **41**, Chapman & Hall, London.

[3] Finney, D. J., 1982. Intervention, and correlated sequences of observations. *Biometrics*, **38**, 255-267.

[4] Finney, D. J., 1990. Repeated measurements : What is measured and what repeats? *Statistics in Medicine*, **9**, 639-644.

[5] Finney, D. J., 1995. Statistical science and effective scientific communication. *J. Appl. Statist.*, **22**, 293-308.

[6] Fisher, R. A., 1921. Studies in crop variation. I. An examination of the yield of dressed grain from roadbalk. *J. Agril. Sci.*, **11**, 107-135.

[7] Fisher, R.A., 1924. The influence of rainfall on the yield of wheat at Rothamsted. *Phil. Transac. Roy. Soc.*, **B213**, 89-142.

[8] Gomez, K. A. and Gomez, A. A., 1984. *Statistical Procedures for Agricultural Research*, 2nd ed. John Wiley and Sons, New York.

[9] Rowell, J. G. and Walters, D.E., 1976. Analysing data with repeated observations on each experimental unit. *J. Agril. Sci.*, **87**, 423-432.

[10] Turner, H. N., 1939. The residual effect of organic fertilizers: A preliminary report on the Rothamsted four-course rotation experiment. *Empire Journal of Experimental Agriculture*, **7**, 343-349.

[11] Yates, F., 1982. Regression models for repeated measurements. *Biometrics*, **38**, 850-853.