

Review article

The application of analysis of variance (ANOVA) to different experimental designs in optometry

R. A. Armstrong, F. Eperjesi and B. Gilmartin

Vision Sciences, Aston University, Birmingham B4 7ET, UK

Abstract

Analysis of variance (ANOVA) is the most efficient method available for the analysis of experimental data. Analysis of variance is a method of considerable complexity and subtlety, with many different variations, each of which applies in a particular experimental context. Hence, it is possible to apply the wrong type of ANOVA to data and, therefore, to draw an erroneous conclusion from an experiment. This article reviews the types of ANOVA most likely to arise in clinical experiments in optometry including the one-way ANOVA ('fixed' and 'random effect' models), two-way ANOVA in randomised blocks, three-way ANOVA, and factorial experimental designs (including the varieties known as 'split-plot' and 'repeated measures'). For each ANOVA, the appropriate experimental design is described, a statistical model is formulated, and the advantages and limitations of each type of design discussed. In addition, the problems of non-conformity to the statistical model and determination of the number of replications are considered.

Keywords: Analysis of variance (ANOVA), experimental design, factorial experimental design, random effect factor, randomised blocks, repeated measures design, split-plot design.

Introduction

Analysis of variance (ANOVA) is the most efficient parametric method available for the analysis of data from experiments. It was devised originally to test the differences between several different groups of treatments thus circumventing the problem of making multiple comparisons between the group means using *t*-tests (Snedecor and Cochran, 1980). ANOVA is a method of great complexity and subtlety with many different variations, each of which applies in a particular experimental context. Hence, it is possible to apply the wrong type of ANOVA in a particular experimental situation and, as a consequence, draw the wrong conclusions from the data.

A previous article (Armstrong *et al.*, 2000) described the origin of ANOVA, the logic which underlies the method and the assumptions necessary to apply it to

data. In addition, the simplest form of the analysis, the one-way ANOVA in a randomised design, was described and illustrated using a clinical experiment drawn from optometric research. The various methods available for making specific comparisons between pairs of group means (also known as '*posthoc*' tests) were also considered.

The objective of the present article is to extend these techniques to different experimental designs. For the purpose of this article, these types of ANOVA will be considered under the following headings likely to cover many situations encountered in optometric research: (1) one-way ANOVA, 'random effects' model (2) two-way ANOVA in randomised blocks (3) three-way ANOVA (4) factorial ANOVA (5) factorial ANOVA, split-plot design, and (6) factorial ANOVA, repeated measures design. In each case, the type of experimental design is described, a statistical model is given and the advantages and limitations of the appropriate ANOVA discussed. In addition, the problems of non-conformity to the statistical model and determination of the number of replications are considered. The number of types of ANOVA described are illustrated with simple data sets drawn from clinical experiments in optometry. These data sets are used only to illustrate the methodology

Received: 11 June 2001

Revised Form: 28 November 2001

Accepted: 17 December 2001

Correspondence and reprint requests to: Dr. R. A. Armstrong, Vision Sciences, Aston University, Birmingham B4 7ET, UK.

and the results quoted may not be typical of those that would be observed in more extensive experiments.

Statistical models

In a previous article (Armstrong *et al.*, 2000), we described a commonly used notation to describe the basic model of an ANOVA. The subscript '*i*' is used to denote the group or class (i.e. the treatment group), '*i*' taking the values '1 to *a*', whereas the subscript '*j*' designates the members of the class, '*j*' taking the values '1 to *n*' (hence, '*a*' groups and '*n*' replicates per group). Within class '*i*', the observations x_{ij} are assumed to be normally distributed about a mean μ with variance s^2 . This linear model can be written:

$$x_{ij} = \mu + a_i + e_{ij} \quad (1)$$

Hence, an observed value x_{ij} is the sum of three parts: (1) the overall mean of the observations (μ) (2) a treatment or class deviation '*a*', and (3) a random element '*e*' drawn from a normally distributed population. The random element reflects the combined effects of natural variation between replications and errors of measurement. All the more complex types of ANOVA can be derived from this simple model by the addition of one or more further terms to equation 1 (Table 1).

Table 1. Statistical models of the types of ANOVA described. For definition of terms see relevant text

Type of ANOVA	Y	Constant elements	Additional components	Interaction terms
One-way, fixed	x_{ij}	$\mu + e_{ij}$	a_i	
One-way, random	x_{ij}	$\mu + e_{ij}$	A_i	
Two-way, randomised blocks	x_{ij}	$\mu + e_{ij}$	$a_i + b_j$	
Three-way	x_{ijk}	$\mu + e_{ijk}$	$a_i + b_j + c_k + d_{ij}$	
Two-factor, factorial	x_{ijk}	$\mu + e_{ijk}$	$a_i + b_j$	$(ab)_{ij}$
Two-factor, factorial split-plot	x_{ijk}	$\mu + e_{ijk}$	$M_i + B_j + T_k + d_{ijk}$	$(MT)_{ik}$

In each case, the dependent variable (Y) can be considered to be the sum of the constant elements, the additional components and, where appropriate, the interaction terms in each row.

Table 2. The one-way ANOVA 'random effects model' (hierarchical or nested design) on five measurements of IOP made on a subject 1 min apart on 3 sample days (A)

Design					
Sample day	A1		A2		A3
Repeat	18		17		19
Measurements	19		18		18
	20		16		20
	19		17		20
	21		17		19
ANOVA table:					
Variation	SS	d.f.	MS	F	ExpMS
Between days	17.73	2	8.866	10.64	$\sigma_m^2 + 5\sigma_D^2$
Between repeat measurements within days	10.0	12	0.833		σ_m^2

ExpMS = expected mean square. Components of variance: between days (σ_D^2) = 1.607, between repeat measurements (σ_m^2) = 0.833.

Types of ANOVA

One-way ANOVA, 'random effects' model

In our previous article (Armstrong *et al.*, 2000), we described a one-way ANOVA in a randomised design which compared the reading rates of three groups of subjects, viz., young normal subjects, elderly normal subjects and subjects with age-related macular degeneration. This ANOVA is described as a 'fixed effects' model in which the objective is to estimate the differences between the subject groups and these are regarded as 'fixed' or discrete effects. There is, however, an alternative model called the 'random effects' model in which the objective is to estimate the degree of variation of a particular measurement and in many circumstances to compare different sources of variation in space and time.

Example

An example of an experiment of this type is shown in Table 2. Five measurements of intraocular pressure (IOP) were made on a subject, 1 min apart on 3 days chosen randomly. The objective was to determine, for individual subjects, the degree of variation in IOP from minute-to-minute compared with the variation between measurements made on different days. In this case,

estimates of variability are the objective rather than the determination of fixed effects. Based on these estimates of variability, a suitable protocol for measuring IOP in a clinical context might be devised. For example, if IOP varied considerably from minute to minute but, on average, little between days, several measurements of the IOP on a single occasion might be an appropriate strategy. By contrast, if minute to minute variation was negligible but there was significant day to day variation, it might be better to measure IOP only once, but on a sample of days. This type of experiment may also be described as a 'hierarchical' or 'nested type' design, especially if each sample is composed of subsamples and these in turn are subsampled. In addition, variation may be spatial rather than temporal, e.g. visual function may be quantified at different locations on the retina. In this case, there may be variation between eyes, between different locations within the same retina, and between sequential measurements at the same location.

Model

Equation 1 describes a 'fixed effect' model in which the a_i are fixed quantities to be estimated. The corresponding 'random effects' model is similar (*Table 1*), but the symbol A_i replaces a_i and represents the difference between the IOP of the i th patient and the mean. Hence, A_i is a random variable and the term e_{ij} refers to errors of measurement and to the fact that IOP is measured on a sample of occasions.

ANOVA

The ANOVA appropriate to this design provides an F -test of whether there is significant variation between days. When the null hypothesis is false, however, the mean square between classes (in this case between days) estimates the quantity, $\sigma_m^2 + n\sigma_D^2$ and it is possible to calculate the 'components of variance' σ_m and σ_D from the ANOVA table (*Table 1*). These are estimates of the variance of the measurements made between days and between determinations on a single day. In the example quoted (*Table 2*), the F -test suggests significant variation

in IOP between days. In this context, however, the components of variance are more useful and indicate that the component between days is approximately twice as great as that on a single day.

It is important to identify whether a 'fixed' or 'random' effect model is appropriate in each case. This is particularly important in more complex factorial-type designs in which there may be a mixture of 'fixed' and 'random' effects (Snedecor and Cochran, 1980). One way of deciding whether a factor is 'fixed' or 'random' is to imagine the effect of changing one of the levels of the factor (Ridgman, 1975). If this makes it a different experiment, for example, by substituting a different group of subjects, then it is a 'fixed' effect factor. By contrast, if we considered it the same experiment, for example, by substituting a different sample day, as in the example described above, it would be a 'random' effect factor.

Two-way ANOVA in randomised blocks

In the one-way, 'fixed' effects ANOVA described previously (Armstrong *et al.*, 2000), each observation was classified in only one way, i.e. in which treatment or subject group the observation fell. Replicates were either allocated to treatment groups at random or subjects within a group were a random sample of a particular population. Such an experiment is often described as a 'randomised design'. More complex experimental designs are possible, however, in which an observation may be classified in two or more ways.

Example

An example of an experiment in which the observations are classified in two ways is shown in *Table 3*. This experiment studied the effect of four coloured filters on the reading rate of 12 patients grouped by age. Such an experiment is often called a 'randomised blocks design' because there is a restriction in how subjects are randomised to treatments: (1) subjects are first grouped into 'blocks' of similar age and (2) treatments are applied at random to the subjects within each block

Design	Red	Yellow	Green	Blue
Age 1	84.5	72.0	70.5	39.5
Age 2	79.3	68.2	62.6	47.2
Age 3	36.0	46.1	48.9	38.0
ANOVA table:				
Variation	SS	d.f.	MS	F
Treatments	1102.95	3	367.65	3.40
Blocks	1448.98	2	724.49	6.71*
Error	647.91	6	107.98	

* $p < 0.05$

Table 3. Two-way ANOVA in randomised blocks with four treatment groups (coloured filters) and with replicates also classified into three age groups (blocks). Data are the reading rate of the patient (number of correct words per minute)

separately. Another possible application in optometric research is if a number of treatments are to be given to, or tests made on, each of a sample of subjects. In this case, the individual subject is the 'block' and the treatments are given in a random order to each subject.

Model

In the example given in *Table 3*, the fact that subjects are also grouped into age classes gives a more complex model (*Table 1*). In addition to the treatment effect 'a', this design also includes a term for the age effect 'b' and the ANOVA table includes an extra term for subject age. In addition to the assumptions made in the randomised design, i.e. homogeneity of variance, additive class effects and normal distribution of errors (Armstrong *et al.*, 2000), this type of design makes the additional assumption that the difference between two treatments is consistent across all replications.

ANOVA

The ANOVA appropriate to the two-way design in randomised blocks is shown in *Table 3*. This design is often used to remove the effect of a particular source of variation from the analysis. For example, if there was significant variation because of the age of the subjects and, if subjects had been allocated to treatments at random, then all the between subject age variation would have been included in the pooled error variance. The effect of this would be to increase the error variance, reduce the 'power' of the experiment and therefore, make it more difficult to demonstrate a possible treatment effect. In a two-way randomised blocks design, however, variation between subjects, attributable to their age, is calculated as a separate effect and therefore, does not appear in the error variance. This may increase the 'power' of the experiment and make it more probable that a treatment effect would be demonstrated.

In the example quoted (*Table 3*), despite the blocking by age, there is no evidence for an effect of coloured filter on reading rates but significant effects were present between 'blocks' presumably reflecting the effect of age.

A comparison of the ANOVA table in *Table 3* with that for a one-way ANOVA in a randomised design (Armstrong *et al.*, 2000) demonstrates that reducing the error variance by blocking has a cost, viz., a reduction in the degrees of freedom (d.f.) of the error variance which makes the estimate of the error variation less reliable. Hence, an experiment in randomised blocks would only be effective if the blocking by age or some other factor reduced the pooled error variance sufficiently to counter the reduction in d.f. It has been estimated that a block-type experiment needs to produce a 14% reduction in the error variance in order to offset the additional unreliability in estimating the pooled error (Cochran and Cox, 1957).

The three-way ANOVA

In the two-way ANOVA in randomised blocks, when treatments are given sequentially to a subject, there is a possible 'carry-over' effect of one treatment on to the next or the subject may become fatigued as the tests proceed. An example of the former might include the sequential application of two drugs without a sufficient recovery period between them and of the latter, reading tests with different filters or magnifiers applied sequentially to the same subject. The solution is to have each combination of treatments given to the same number of subjects such that systematic effects due to treatment order will not create bias in the comparison of the treatment means.

Example

Examples of this type of design are shown in *Table 4*. With two treatments (A and B) and 'n' subjects, each of the treatment orders AB and BA would be given to $n/2$ subjects. In this case, it has been suggested that the data

Table 4. A three-way ANOVA with different treatments applied in sequence to the same subject

Example 1. Two treatments with 'n' subjects						
Combinations	AB	BA				
Subjects	$n/2$	$n/2$				
Example 2. Three treatments with 'n' subjects						
Combinations	ABC	ACB	BAC	BCA	CAB	CBA
Subjects	$n/6$	$n/6$	$n/6$	$n/6$	$n/6$	$n/6$
Structure of ANOVA table for three treatments and 36 subjects						
Variation	SS	d.f.	MS	F		
Total		107				
Treatments (a)		2		F_{trts}		
Order (c)		5		F_{order}		
Subjects (b)		35		F_{subj}		
Error		65				

n = number of subjects, trts = treatments, subj = subjects.

could be analysed using 't' tests as it may allow for incomplete and non-orthogonal designs (Armitage and Berry, 1987). With three treatments (A, B and C), if all treatment combinations were used, the order of treatments would be ABC, ACB, BAC, BCA, CAB and CBA and each would be given to $n/6$ subjects.

Model

The model for this type of design is given in *Table 1*. In this case, a third term c_k is added which represents the order of the treatment. The term d_{ij} represents the fact that the effect of a treatment may vary from subject to subject while the error term represents measurement errors and non-systematic variation between subjects (Snedecor and Cochran, 1980). Note that this model, as it stands, does not allow for a test of the possible interaction between treatment and order of treatment because this would require a larger factorial-type experiment and a different ANOVA.

ANOVA

In the ANOVA table (*Table 4*), variation attributable to the order of the treatments now appears as a specific experimental effect. This variation will not appear in the pooled error variance or affect the comparison of the treatment means. A limitation of this design, however, is that the number of replications 'n' must be a multiple of the number of treatments. Hence, with many treatments there will be a large number of possible orders of these treatments and the level of replication will increase accordingly. One method of solving this problem would be to use an incomplete design in which only some of the combinations of treatments would be given. For example, it would be possible to ensure that each treatment was given first, second and third to an equal number of

subjects, e.g. only the combinations ABC, BCA, CAB would be used (Snedecor and Cochran, 1980).

Factorial ANOVA

In a factorial experiment, the effects of a number of different factors can be studied at the same time. Combining factors usually requires fewer experimental subjects or replications than studying each factor individually in a separate experiment. In addition, by contrast with the three-way design, the between treatments or groups sums of squares is partitioned into specific comparisons or 'contrasts' (Ridgman, 1975; Armstrong *et al.*, 2000) which reveal the possible interactive effects between the factors. The interactions between factors often provide the most interesting information from a factorial experiment.

Example

Consider an example of the simplest type of a factorial experiment involving the application of two drugs (A and B) each given at two 'levels' (given or not given) to 24 subjects, randomly allocated, six to each treatment combination (*Table 5*). There are four treatment combinations, no drugs given, either 'A' or 'B' are given separately, or both drugs are given. This type of design is called a 2^2 factorial, i.e. two factors with two levels of each factor.

Model

The model for this design is given in *Table 1*. In this model, x_{ijk} is the value of the 'kth' replicate of the 'ith' level of A and the 'jth' level of B, a_i and b_j are the main effects and $(ab)_{ij}$ represents the two factor interaction between A and B.

Design: treatment combinations and orthogonal coefficients				
	None (1)	+A	+B	+AB
Treatment totals	832	853	881	966
Factorial effects				
A	-1	+1	-1	+1
B	-1	-1	+1	+1
AB	+1	-1	-1	+1
Structure of ANOVA table				
Variation	SS	d.f.	MS	F
Total (Drugs)		11 (3)		
A	468	1	468	2.38 ns
B	1094	1	1094	5.57*
AB	171	1	171	0.87 ns
Error	3926	20	196.3	

A, B main effects, AB interaction effect, * $p < 0.05$, ns = not significant ($p > 0.05$).

Table 5. Factorial ANOVA with two drugs (A and B) given at two levels, given (+) or not given (-) (a 2^2 factorial) with six replications

ANOVA

As in previous examples, the total sums of squares can be broken down into that associated with differences between the effects of the drugs and error (Table 5). In this case, the between treatments sums of squares can be broken down further into 'contrasts' which describe the main effects of 'A' and 'B' and the interaction effect 'A × B'. These effects are linear combinations of the means, each being multiplied by a number or 'coefficient' to calculate a particular effect. In fact, the meaning of an effect can often be appreciated by studying these coefficients (Table 5). The main effect of drug 'A' is calculated from those groups of subjects which receive drug 'A' (+) compared with those which do not (-). Note that in a factorial design, every observation is used in the estimate of the effect of every factor. Hence, factorial designs have 'internal replication' and this may be an important consideration in deciding the number of subjects to use. The main effect of 'B' is calculated similarly to that of 'A'. By contrast, the two-factor interaction ('A × B') can be interpreted as a comparison of the effect of the combined action of 'A' and 'B' with the individual effects of 'A' and 'B'. A significant interaction term would imply that the effects of 'A' and 'B' were not additive, i.e. the effect of the combination 'AB' would not be predictable from knowing the individual effects of 'A' and 'B'. In the quoted example, there is no significant effect of drug A, the effect of drug B is significant and the non-significant interaction indicates that the effect of B was consistent regardless of whether A was given or not.

Note that in a 2² factorial, partitioning the treatments sums of squares into factorial effects provides all the information necessary for interpreting the results of the experiment and further 'posthoc' tests would not be needed. However, with more complex factorial designs, e.g. those with more than two levels of each factor,

further tests may be required to interpret the main effect or an interaction. With factorial designs, it is better to define specific comparisons before the experiment is carried out rather than to rely on 'posthoc' tests. Factorial experiments can be carried out in a completely randomised design, in randomised blocks or in a more complex design. The relative advantages of these designs are the same as for the one-way design.

Factorial ANOVA, split-plot design

In the 2² factorial described above, the experimental subjects were assigned at random to all possible combinations of the two factors. However, in some designs, the two factors are not equivalent to each other. A common case, called a split-plot design, arises when one factor can be considered to be a major factor and the other a minor factor.

Example

In optometry research, this situation could arise if measurements were made on both the right and left eyes of subjects employing two different subject groups (Table 6). As an example, an investigator wished to study whether IOP in right and left eyes was elevated in patients with high blood pressure. Alternatively, a measurement may be made on subjects within two treatment groups employing different methods or equipment. The problem that arises in these types of design is the dependence or correlation between the measurements made on a subject (Rosner, 1982). In these experiments, the subject group would be the major factor while 'which eye' or 'which method' would be regarded as a minor factor. The difference between this and an ordinary factorial design is that previously, all treatment combinations were assigned at random to replicates whereas in a split-plot design, the subplot

Table 6. Factorial ANOVA, split-plot design with two subject groups (normal and elevated blood pressure) with three subjects (P1, P2, P3) in each group and the left (L) and right (R) eye studied from each patient. Data are intraocular pressures

Design	P1	P2	P3	
Control				
R eye	17.3	16.9	14.7	
L eye	17.1	16.5	14.3	
ARMD				
R eye	21.4	24.3	21.4	
L eye	20.7	22.1	24.2	
ANOVA table:				
Variation	SS	d.f.	MS	F
Subject group	115.94	1	115.94	34.4**
Main plot error	13.48	4	3.37	
Right/Left eye	0.10	1	0.10	0.06 ns
Interaction	0.067	1	0.067	0.04 ns
Sub plot error	6.59	4	1.65	

** $p < 0.01$; ns = not significant.

treatments are randomised over the whole large block but only within each main plot. In some applications, experimenters may subdivide the subplots further to give a split-split-plot design (Snedecor and Cochran, 1980).

Model

The model for a two factor split-plot design is given in *Table 1*. In this case, M represents main-plot treatments, B blocks and, T subplot treatments. The symbols '*i*' and '*j*' indicate the main-plots while '*k*' identifies the subplot within the main-plot. The two components of error e_{ij} and d_{ijk} represent the fact that the error variation between main-plots is likely to be different from that between subplots. For example, one might expect there to be less natural variation when comparing eyes within a subject (subplot) than the eyes of different subjects (main-plot). An alternative model suggested by Rosner (1982) would be to consider the x_{ij} as distributed about the same mean and variance but for any two members of the same class to be correlated, the degree of correlation being measured by the intra-class correlation coefficient (Snedecor and Cochran, 1980).

ANOVA

The resulting ANOVA (*Table 6*) is more complex than that of a simple factorial design because of the different error terms. Hence, in a two-factor, split-plot ANOVA, two errors are calculated, the main-plot error is used to test the main effect of subject group while the subplot error is used to test the main effect of eyes and the possible interaction between the factors. In the quoted example, there is a significant increase in IOP in patients with elevated blood pressure but no difference between eyes; the non-significant interaction suggesting that the elevation in IOP was consistent in both eyes. The subplot error is usually smaller than the main-plot error and also has more d.f. Hence, such an experimental design will usually estimate the main effect of the subplot factor and its interaction more accurately than the main effect of the major factor. Some experimenters will deliberately design an experiment as a 'split-plot' to take advantage of this property.

A disadvantage of such a design, however, is that occasionally, the main effect of the major factor may be large but not significant while the main effect of the minor factor and its interaction may be significant but too small to be important. In addition, a common mistake is for researchers to analyse a split-plot design as if it were a fully randomised two-factor experiment. In this case, the single pooled error variance will either be too small or too large for testing the individual treatment effects and the wrong conclusions could be drawn from the experiment. To decide whether a particular experiment is a split-plot it is useful to

consider the following: (1) Are the factors equivalent or does one appear to be subordinate to the other? (2) Is there any restriction in how replicates were assigned to the treatment combinations? and (3) Is the error variation likely to be the same for each factor?

Caution should also be employed in the use of '*posthoc*' tests in the case of a split-plot design. '*Posthoc*' tests assume that the observations taken on a given subject are uncorrelated so that the subplot factor group means are not related. This is not likely to be the case in split-plot experiment because some correlation between measurements made on the right and left eye is inevitable. Standard errors appropriate to the split-plot design can be calculated (Cochran and Cox, 1957; Freese, 1984) and can be used, with caution, to make specific comparisons between the treatment means. However, a better method is to partition the sums of squares associated with main effects and interaction into specific contrasts and to test each against the appropriate error (Snedecor and Cochran, 1980).

Factorial ANOVA, repeated measures design

The repeated measures factorial design is a special case of the split-plot type experiment in which measurements on the experimental subjects are made sequentially over several intervals of time. The ANOVA is identical to the preceding example but with time constituting the subplot factor. Repeated measurements made on a single individual are likely to be highly correlated and therefore the usual '*posthoc*' tests cannot be used. Nevertheless, it is possible to partition the main effects and interaction sums of squares into contrasts. In a repeated measures design, the shape of the response curve, i.e. the regression of the measured variable on time, may be of particular interest. A significant interaction between the main-plot factor and time would indicate that the response curve with time varied at different levels of the main-plot factor.

Non-conformity to model

In the ANOVAs described above, the model specifies that the effects of different factors should be additive and that the errors are normally distributed with the same variance. It is unlikely that these conditions are ever realised in real experiments and it is reasonable to consider the consequences of failure of these assumptions.

A major error in the measurement of the data can distort the mean of a treatment considerably and, by increasing the error variance, can have a profound effect on the experiment as a whole. An investigation into 'suspicious' values should always be made. Such a process may reveal a mistake and enable the correct value to be substituted or if not, rejecting the aberrant

observation and analyzing the data without it. A number of rules for rejecting such observations are given by Snedecor and Cochran (1980).

Lack of independence of the errors can arise if a faulty experimental design is employed. For example, if all replications of a given treatment were processed at the same time by the same person or by the same machine; different technicians or machines being employed for the other treatments, then there will be positive correlations between replicates within treatments. The most effective precaution against this type of effect is to use appropriate randomization at every stage of the experiment. If different treatments have different error variances (heterogenous variances), the *F*-test may indicate a significant result when no true effects are present. One method of solving this problem is to omit those treatments which are substantially different from the rest. In addition, if within the different treatments, the standard deviation is proportional to the mean, such that effects are proportional rather than additive, a logarithmic transformation may stabilise the variance (Armstrong *et al.*, 2000).

Non-normal distribution of the errors will also tend to produce too many significant results in an ANOVA. The solution may be to transform the data, e.g. counts of rare events may be distributed according to the Poisson distribution and a transformation to square roots may be appropriate. In addition, if the data are proportions or percentages, an angular transformation may be necessary (Snedecor and Cochran, 1980).

Number of replications in an experiment

In our previous article (Armstrong *et al.*, 2000), we described a simple method of obtaining an approximate estimate of the number of replications required in an experiment to have a good chance of revealing a particular effect. Such an approach may provide a reasonable approximation in simple experiments, such as the one-way design, but are unlikely to be appropriate in more complex experiments with many treatments or factorial combinations. In more complex experiments, it is better to consider the number of d.f. of the error term rather than the number of replications per treatment group. The number of d.f. depends on both the number of treatments and the number of replicates. It is not possible to define the lower limit of the d.f. for all experiments, but a good standard would be to try and achieve at least 15 d.f. for the error term if possible (Ridgman, 1975). Hence, in an experiment with four treatments, five replicates per treatment would be required to provide approximately 15 d.f. for the error term. In more complex designs, such as the split-plot or repeated measures design, which usually have more precision in the estimation of the

subplot factor and interaction, then it would be appropriate to have at least 15 d.f. for the subplot error.

Conclusions

The key to the correct application of ANOVA in optometric research is careful experimental design and matching the correct analysis to that design. The following points should therefore, be considered before designing any experiment:

- (1) In a single factor design, ensure that the factor is identified as a 'fixed' or 'random effect' factor.
- (2) In more complex designs, with more than one factor, there may be a mixture of fixed and random effect factors present so ensure that each factor is clearly identified.
- (3) Where replicates can be grouped or blocked, the advantages of a randomised blocks design should be considered. There should be evidence, however, that blocking can sufficiently reduce the error variation (by at least 14%) to counter the loss of d.f. compared with a randomised design.
- (4) Where different treatments are applied sequentially to a patient, the advantages of a three-way design in which the different orders of the treatments are included as an 'effect' should be considered.
- (5) If treatments can be expressed as factors then combining different factors to make a more efficient experiment and to measure possible factor interactions should always be considered.
- (6) The effect of 'internal replication' should be taken into account in a factorial design in deciding the number of replications to be used. Where possible, each error term of the ANOVA should have at least 15 d.f.
- (7) Consider carefully whether a particular factorial design can be considered to be a split-plot or a repeated measures design. If such a design is appropriate, consider how to continue the analysis bearing in mind the problem of using *post hoc* tests in this situation.
- (8) Data should be checked for gross errors, lack of independence of the errors, non-additivity and non-normality. In some cases, transformation of the data may be necessary before the analysis is carried out.
- (9) If there is any doubt about the above issues, the researcher should seek advice from a statistician with experience of optometric research before carrying out the experiment. It is particularly important to check for model assumptions and that the d.f. and error terms are correct. An erroneous design will not test a null hypothesis adequately and once committed to such a design, there may be little a statistician can do to help.

References

- Armitage, P. and Berry, G. (1987) *Statistical Methods in Medical Research*, 2nd edn. Blackwell Scientific Publications, Oxford and London.
- Armstrong, R. A., Slade, S. V. and Eperjesi, F. (2000) An introduction to analysis of variance (ANOVA) with special reference to data from clinical experiments in optometry. *Ophthal Physiol. Opt* **20**, 235–241.
- Cochran, W. G. and Cox, G. M. (1957) *Experimental Designs*, 2nd edn. John Wiley, New York, London and Sydney.
- Freese, F. (1984) *Statistics for Land Managers*. Paoeny Press, Jedburgh, Scotland.
- Ridgman, W. J. (1975) *Experimentation in Biology*. Blackie, London.
- Rosner, B. (1982) Statistical methods in ophthalmology: an adjustment for the intraclass correlation between eyes. *Biometrics* **38**, 105–114.
- Snedecor, G. W. and Cochran, W. G. (1980) *Statistical Methods*, 7th edn. Iowa State University Press, Ames, Iowa.