main effects (page 237). If two continuous variables are strongly correlated with each other, they may both show as non-significant even when each has a clear linear relationship to the response. This is because the effect of each is measured after adjusting for the other. SS should not be adjusted for higher-order interactions between their fixed effects.

The statistics package returns an F*-test denominator of zero, or zero error d.f.* Erase from the input model the last-entered (and highest-order) term, which then becomes the residual error term (see pages 57 and 258).

The statistics package returns an inexact F*-test.* The design permits only a quasi *F*-ratio, although *post hoc* pooling may present a viable alternative (see page 40 and footnotes to tables of the model structures).

The statistics package calculates the F*-ratio for some random factors using a different denominator to that prescribed in this book.* It may be using an unrestricted mixed model (see page 242 and footnotes to tables of the model structures).

In a design with many crossed factors, is there a problem with getting multiple P*-values in the ANOVA table of results?* Multiple *P*-values are not a problem when they are generated by an ANOVA that has partitioned sources of variation in the response, because each tests a different null hypothesis. This is true also of a priori contrasts, but unplanned *post hoc* tests must account for an inflated Type I error that results from multiple tests of the same null hypothesis (page 245).

My ANOVA on three samples is not significant, but when I do a t *test on each pair of samples, one of them does give a significant result.* All that the multiple tests have given you is an excessive Type I error rate. See the section on evaluating alternative designs (page 250, and particularly point 5 on page 252). Consider use of *post hoc* tests designed to account for the inflated error (page 245).

There are few error degrees of freedom for one or more F*-ratios.* Investigate options for *post hoc* pooling (see page 38 and footnotes to tables of the model structures). If pooling is not possible then reflect on the need to plan the analysis before collecting the data (see the example on page 51).

C.P. Doncaster, Analysis of Variance and covariance, Cambridge Univertity Press, 2007.

Glossary

Adjusted sums of squares Adjustment to the sum of squares used in general linear models (GLM) to account for designs without orthogonality. A Type II adjustment to the SS of a term involves adjusting for all other terms in the model that do not contain the term in question. A Type III adjustment involves adjusting for all other terms in the model, including those containing the term in question. Only Type II SS are suitable for models with fixed cross factors, and only Type III SS are suitable for models with random cross factors.

Analysis of covariance (ANCOVA) Analysis of variance on a model containing one or more covariates, usually in addition to one or more categorical factors. Each covariate X is tested for a linear trend with the continuous response Y.

Analysis of variance (ANOVA) An analysis of the relative contributions of explained and unexplained sources of variance in a continuous response variable. In this book, we use the term 'ANOVA' in its broad sense to include explanatory factors that vary on continuous as well as categorical scales, with a focus on balanced designs. Parametric ANOVA and GLM partition the total variance in the response by measuring sums of squared deviations from modelled values. Significant effects are tested with the *F* statistic, which assumes a normal distribution of the residual error, homogeneous variances and random sampling of independent replicates.

A priori tests Tests that are integral to the original hypothesis.

Assumptions These are the necessary preconditions for fitting a given type of model to data. No form of generalisation from particular data is possible without assumptions. They provide the context for, and the means of evaluating, scientific statements purporting to truly explain reality. As with any statistical test, ANOVA assumes unbiased sampling from the population of interest. Its other assumptions concern the error variation against which effects are tested by the ANOVA model. Underlying assumptions should be tested where possible, and otherwise acknowledged as not testable for a given reason of design or data deficiency.

Balance A balanced design has the same number of replicate observations in each sample. Balance is a desirable attribute particularly of cross-factored models, where loss of balance generally (though not inevitably) leads to loss of orthogonality. The consequent complications to the partitioning of sources of variance in the response are accommodated by general linear models.

Block A level of a random factor designated to sample unmeasured variation in the environment.

Blocking factor A random factor designated to sample unmeasured variation in the environment.

Categorical factor A factor with levels that are classified by categories (e.g., factor Sex with levels male and female). A factor may vary on a continuous scale (e.g., distance in km, or time in hours) but still be treated as categorical if it is measured at fixed intervals (e.g., before and after a place or event).

Control A treatment level used to factor out extraneous variation by mimicking the test procedure in all respects other than the manipulation of interest. For example, a liquid fertiliser applied to a crop needs to be tested against a control of an equal quantity of liquid without the fertiliser ingredients. Failure to do so can result in a false positive induced by the carrier medium alone.

Correlation Any co-variation of continuous factors with each other or with a continuous response. Correlation between explanatory factors is a form of non-orthogonality.

Covariate A factor X that varies at least on an ordinal scale, and usually on a continuous scale (such as time, distance, etc.) and is therefore a covariate of the response Y. Analysis of covariance assumes that the response has a linear relation to the covariate, and transformations of response or covariate may be necessary to achieve this prior to analysis.

Crossed factor One factor is crossed with another when each of its levels is tested in each level of the other factor. For example, watering regime is crossed with sowing density if the response to the wet regime is tested at both high and low sowing density, and so is the response to the dry regime (assuming both factors have just two levels).

Data The measurements of the response at given levels of factors of interest.

Degrees of freedom (d.f.) The number of independent pieces of information required to measure the component of variation, subtracted from the total number of pieces contributing to that variation. Analysis of variance always has two sets of d.f.: the first informs on the number of test samples and the second informs on the amount of replication available for testing the effect. For example, a result $F_{2,12} = 3.98$, $P < 0.05$ indicates a significant effect with three levels allocated between 15 sampling units.

Effect A term in the statistical model accounting for one of several independent sources of variance in the response. For example the cross-factored model $Y = B|A + \varepsilon$ has two main effects (A and B) and one interaction effect (B*A).

Effect size The magnitude of an effect, measured in terms of deviations of sample means from the grand mean (fixed factor), or the steepness of the regression slope from horizontal (covariate), or the magnitude of between-sample variance (random factor). The significance of an effect depends upon a combination of its size, the amount of background variation and the sample size. An effect of small magnitude can thus be strongly significant if it is sampled with little residual variation from many replicates. Conversely, an apparently large effect may have no significance if it is sampled with large residual variation or from few replicates.

Error variance The random variation in the response against which an effect is tested, containing all of the same components of variation estimated in the population except for the test effect. The validity of ANOVA depends on three assumptions about the error variance: (i) that the random variation around fitted values is the same for all sample means of a factor, or across the range of a covariate; (ii) that the residuals contributing to this variation are free to vary independently of each other; (iii) that the residual variation approximates to a normal distribution.

Experiment A manipulative study involving the application of one or more treatments under controlled conditions. Where possible, treatment levels are randomly assigned to sampling units, and effects compared against a control.

Factor A source of variance in the response. A categorical factor is measured in categorical levels, whereas a covariate factor is measured on a scale of continuous (or sometimes ordinal) variation. A model might be constructed to test the influence of a factor as the sole explanation ($Y = A + \varepsilon$) or as one of many factors variously crossed with each other or nested within each other.

Factorial model A model containing crossed factors in which every level of each factor is tested in combination with every level of the other factors. Fully replicated factorial designs test whether the effect of one factor is moderated by interaction with another.

False negative The result of making a Type II error by accepting a false null hypothesis. This type of error can incur severe consequences for sampling units, such as patients being screened for a disease or rivers being screened for a pollutant. The Type II error rate β can be minimised by using a design with sufficient replication to ensure high power for distinguishing true effects.

False positive The result of making a Type I error by rejecting a true null hypothesis. Tests that are deemed significant if $P < 0.05$ must sanction a false positive arising once in every 20 runs on average. This causes problems particularly in studies that apply the same test to a large number of datasets to screen for a phenomenon with low incidence in the population. A positive identification is more likely false than true if incidence $< \alpha$, the Type I error rate.

Fitted values The values of the response predicted by the model for each data point. Fitted values are the sample means for categorical factors, or points on a regression line for a covariate.

Fixed factor A factor with levels that are fixed by the design and could be repeated without error in another investigation. The factor has a significant effect if sample means differ by considerably more than the background variation, or for a covariate, if the variation of the regression line from horizontal greatly exceeds the variation of data points from the line.

F statistic The test statistic used in ANOVA and GLM, named in honour of R. A. Fisher, who first described the distribution and developed the method of analysis of variance in the 1920s. The continuous *F* distribution for a given set of test and error d.f. is used to determine the probability of obtaining at least as large a value of the observed ratio of explained to unexplained variation, given a true null hypothesis. The associated *P*-value reports the significance of the test effect on the response.

Fully replicated design A design with replicate sampling units at each factor level, or for designs with more than one factor, each combination of factor levels. Such designs have residual variation given by these nested random sampling units, which permits the ANOVA to test all sources of variance in the response. A design without full replication has the random sampling unit cross-factored with other terms, contributing to the variance in the response having one or more untestable sources.

General linear model (GLM) Generic term for parametric analyses of variance that can accommodate combinations of factors and covariates, and unbalanced and non-orthogonal designs. GLMs generally use an unrestricted model for analysing combinations of fixed and random factors.

Generalised linear model (GLIM) Generic term for analyses of variance that can accommodate combinations of factors and covariates, and can permit the residuals to follow any distribution from the exponential family, which includes Gaussian, Poisson, binomial and gamma distributions. Components of variation are partitioned using maximum likelihood rather than sums of squares.

Hypothesis test An analysis of data to test for pattern against the null hypothesis H_0: no pattern. Analysis of variance subjects a dataset to one or more test hypotheses, described by a model. For example, a test of the model $Y = B|A + \varepsilon$ may reject or accept the null hypothesis of no effect of A on the response. Likewise, it rejects or accepts the null hypotheses of no B effect and of no interaction effect. A decision to reject H_0 is taken with some predefined probability α of making a Type I error by rejecting a true null hypothesis. A decision to accept H_0 is taken with a probability β of making a Type II error by accepting a false null hypothesis.

Independent replicates The power of any statistical test to detect an effect depends on the accumulation of independent pieces of information. ANOVA assumes that replicate data points are independent of each other in the sense that the value of one data point at a given factor level has no influence on the value of another sampled at the same level. The assumption is often violated by the presence of confounding factors. For instance, a sample of ten subjects will not provide ten independent pieces of information about a response if it comprises five pairs of siblings. Independence can be restored by declaring a factor Sibling, or by measuring just one individual at random of each pair. Likewise, a response of leaf area to soil type is tested with replicates given by the number of independent plants not, by the number of leaves.

Interaction An interaction tests whether the effect of one factor on a response depends on the level of another factor. For example, students may respond to different tutorial systems according to their age, indicated by a significant interaction effect Age*System on the response. If one factor is a covariate, the interaction is illustrated by different regression slopes at each level of the categorical factor. Two covariates show a significant interaction in a curved plane for their combined effect on the response. An interaction effect must always be interpreted before its constituent main effects, because its impact influences interpretation of the main effect.

Linear model A model with linear (additive) combinations of parameter constants describing effect sizes. Linear models can describe non-linear trends in covariates, for example by transformation of the data or fitting a polynomial model. All the models in this book are linear.

Main effect A main effect tests whether the effect of one factor on a response occurs irrespective of the level of another factor. For example, students may respond to different tutorial systems regardless of their age. Main effects must always be interpreted after interpreting any interactions.

Marginality The fundamental principle of ANOVA that terms be tested in hierarchical order. This becomes an issue in non-orthogonal designs, where the variance due to an interaction must be estimated after factoring out the variance due to the terms contained within it.

Mean The arithmetic average value of the responses in a sample. The sample means provide the fitted values from which effect size is measured in analyses of categorical factors. In covariate analysis, the linear regression pivots on the coordinate for the sample means of response and covariate: $(\bar{x}, \bar{y})$.

Mean square (MS) The variance measured as variation per degree of freedom. The *F*-ratio is the ratio of explained to unexplained MS, where the numerator is the MS explained by the model and the denominator is the error MS left unexplained by the model.

Mensurative study A study that tests the effect of one or more factors on a response without controlled manipulation.

Mixed model A model with random and fixed factors.

Model The hypothesised effect(s) on a response, which can be tested with ANOVA for evidence of pattern in the data. An ANOVA model contains one or more terms, each having an effect on the response that is tested against unmeasured error or residual variation. A model with a single factor (whether categorical or covariate) is written: $Y = A + \varepsilon$, and the ANOVA tests the term A against the residual ε. Models with multiple factors require care with declaring all terms in a statistical package. For example, the cross-factored with nesting model: $Y = C|B'(A) + \varepsilon$ is analysed by declaring the terms: C|A + C|B(A). The two-factor randomised-block model: $Y + S'|B|A$ is analysed by declaring the terms: S|B|A – S*B*A for a Model 1 analysis, or the terms: S + B|A for a Model 2 analysis.

Model 1 In designs without full replication, an ANOVA model that assumes the presence of block-by-treatment interactions, even though the design has not allowed for their estimation. Randomised-block designs may be analysed by Model 1 or Model 2. Repeated-measures designs are generally analysed by Model 1.

Model 2 In designs without full replication, an ANOVA model that assumes the absence of block-by-treatment interactions, even though the design has not allowed for any direct test of this assumption. Randomised-block designs may be analysed by Model 1 or 2. Split-plot designs are generally analysed by Model 2.

Multiple tests Multiple tests of the same hypothesis cause inflation of the Type I error rate. The problem arises in data probing, involving an unplanned search for any significant differences amongst a set of samples. For example, if replicate measures are taken from two levels of a factor A and from three levels of a factor B, then a search for any differences between the five samples might involve a total of ten independent *t* tests (A_1 versus A_2, A_1 versus B_1, ... etc.). If each has a Type I error rate of 0.05, then the ensemble of ten tests has a probability of $1 - 0.95^{10} = 0.40$ of mistakenly rejecting the null hypothesis of no single difference between any sample means. This unacceptably high rate is avoided only by using a statistical model that respects a planned design of data collection. A cross-factored ANOVA would partition the total variance in the response into three testable sources: A, B, and B*A, each with their own *P*-value testing a specific null hypothesis.

Nested factor One factor is nested within another when each of its levels are tested in (or belong to) only one level of the other. For example a response measured per leaf for a treatment factor applied across replicate trees must declare the trees as a random factor nested in the treatment levels. The sampling unit of Leaf is then correctly nested in Tree nested in Treatment.

Nuisance variable Factors or covariates holding no interest in their own right, but requiring inclusion in the model in order to factor out their contributions to variation in the data.

Null hypothesis (H_0) The statistically testable hypothesis of no pattern in the data. The null hypothesis is the proposal that nothing interesting is happening, against which to test a model hypothesis of trend in a sample or differences between samples. If the test upholds the null hypothesis, then we conclude that the ANOVA model takes the form $Y = \varepsilon$; otherwise we infer a significant effect of a factor of interest on the response. A null hypothesis must be open to falsification. For example, a null hypothesis of zero difference between samples is capable of falsification. A suitable ANOVA will evaluate the evidence for a difference and accept or reject the null hypothesis accordingly. In contrast, a null hypothesis of a difference between samples is not capable of falsification, because absence of evidence (for a difference) is not evidence of absence.

Ordinary least squares (OLS) A method of estimating the values of parameters in linear models by minimising the sum of squared deviations of each observation of the response from the model estimate. In ANOVA, this sum is known as the residual sum of squares, $SS_{residual}$, and it partitions out the variation left unexplained by the model.

Orthogonality A cross-factored design is orthogonal if each of its factors are independent of each other. Two categorical cross factors are orthogonal by design if each level of one is measured at every level of the other. Orthogonal designs partition total variation in the response straightforwardly into testable components using sequential sums of squares for each effect in turn.

Although a balanced design generally (but not inevitably) ensures orthogonality, this can be difficult to achieve in practice, especially with covariates. Two covariates are only orthogonal if they have a correlation coefficient of zero. Loss of orthogonality can reduce or enhance the power of a design to detect effects, and usually requires analysis with the aid of adjusted sums of squares calculated in a GLM.

Parsimony The principle of sampling the minimum number of factors necessary to answer the question of interest with a single model. Each additional cross factor adds an extra dimension to the design and multiplies up the number of potential sources of variation in the response. For example the one-way model $Y = A + \varepsilon$ has one testable source (A); the two-factor model: $Y = B|A + \varepsilon$ has three testable sources ($A + B + B*A$); the three factor model $Y = C|B|A + \varepsilon$ has seven testable sources, and so on. Parsimony is not improved by ignoring any nuisance factors that contribute to variation in the data. These must be included in the analysis.

Placebo A treatment used in medicinal trials to control for extraneous variation by mimicking the test procedure in all respects other than the therapeutic benefit of interest. For example, a drug trial for the effectiveness of a medicinal pill requires a treatment with two levels: drug and placebo, where the placebo is a dummy pill of the same shape and colour as the drug pill except that it does not contain the drug. The need for a control is well illustrated by the 'placebo effect' – the psychological boost to health that can be stimulated by an environment of medical care. For this reason, the treatment levels usually need to be allocated in a 'double blind' trial, whereby neither doctor nor patient can distinguish drug from placebo.

Polynomial predictor A polynomial equation describes a curvilinear relationship with one or more exponents. Polynomials can be tested with linear models by declaring the covariate more than once. For example, the relationship: $y = a + bx + cx^2$ is tested in GLM by requesting the polynomial predictor in the form: X|X and taking sequential SS.

Pooling The construction of an error term from more than one source of variance in the response. A priori pooling occurs in designs without full replication, where untestable interactions with random factors are pooled into the residual variation. The analysis then proceeds on the assumption that the interactions are either present (Model 1) or absent (Model 2). Planned *post hoc* pooling is applied to mixed models by pooling a non-significant error term with its own error term. The design is thereby influenced by the outcome of the analysis (in terms of whether or not an error term is itself significant). More generally, pooling can describe the process of joining together samples, for example in calculating a main effect MS by pooling across levels of a cross factor.

Population In a statistical model for analysis of variance, the population is the theoretical complete set of units from which we sample replicate independent and randomly selected measures for the purposes of testing treatment effects. Any random factor requires a clear definition of the population it describes, so that a given sampling regime can be seen to fairly represent it. For

example, do the subjects for a treatment come from a particular age, gender or ethnic group?

Post hoc tests Tests that are supplementary to the original hypothesis.

Power The capacity of a statistical test to detect an effect if it truly occurs. A test with high power has a low probability of mistakenly accepting a false null hypothesis (i.e., a low Type II error rate). Power increases with more replication, provided it is applied at an appropriate scale. For example a response measured per leaf for a treatment applied across replicate trees includes trees as a random factor nested in the treatment levels. The power of the design depends on the number of replicate trees per treatment level, and not directly on the number of replicate leaves per tree.

Pseudoreplication The result of replicates in a sample not being truly independent of each other, which inflates the Type I error rate. ANOVA models are particularly prone to pseudoreplication if they omit to declare sources of nuisance variation in addition to the effects of interest.

Random factor A factor with levels that sample at random from a defined population. A random factor will be assumed to have a normal distribution of sample means, and homogeneous variance of means, if its MS is the error variance for estimating other effects (e.g., in nested designs). The random factor has a significant effect if the variance among its levels is considerably greater than zero.

Random sampling Replicate measures of a response to a given factor level must be taken at random if they are to represent the population that is being sampled. As with any statistical test, ANOVA assumes random sampling. This assumption is violated for instance if a test for a gender effect of body weight samples heavier males and lighter females.

Randomised-blocks A design containing a random blocking factor, crossed with other factor(s) that have a randomised order of levels within each block.

Regression Analysis of a covariate, or multiple covariates in the case of multiple regression. In this book we refer to such analyses as analyses of covariance, regardless of whether or not the model also includes categorical factors.

Repeated-measures A design containing a random factor (usually Subject) crossed with one or more treatments having levels that are applied in a fixed sequence (usually temporal). For example, the performance of subjects may be tested before and after imbibing a treatment drink with two levels: tonic and control. The repeated-measures factor is Time with two levels: before and after. The design has no degrees of freedom for testing residual variation.

Repeated-measures factor A factor (usually temporal) with a fixed sequence of levels that are crossed with a random factor (usually Subject).

Replicates Randomly selected and independent measurements of the response that together make up a sample of the population of interest.

Replication A model is fully replicated if it has true residual variation, given by a nesting of sampling units in samples. Full replication requires taking more than one independent, randomly selected measurement of the response at each level of each categorical factor, or at each combination of levels of crossed factors. The true residual variation allows estimation of all the explained components of variation.

Residual variation All ANOVA models have residual variation defined by the variation amongst sampling units within each sample. This is always given by the last mean square in ANOVA tables, and denoted 'ε' (epsilon) in the descriptions of fully replicated models. Models without full replication may have no degrees of freedom for measuring residual variation (e.g., randomised-block, split-plot and repeated-measures models).

Response The continuous variable on which data are collected to test for sources of variance. The response is the variable Y on the left of the equals sign in the model equation: $Y = A + \varepsilon$, etc.

Restricted model A mixed model (i.e., with random and fixed factors) is termed restricted if a random factor is not allowed to have fixed cross factors amongst its components of variation estimated in the population. This restriction influences the choice of error MS for random effects. The ANOVA tables in this book are all constructed with the restricted model.

Sample A group of replicate measures of the response taken at the same level of a categorical factor (or combination of factor levels if several categorical factors are present), or across a range of values of a covariate.

Sampling unit The basic unit from which is recorded a single measure or observation of the response.

Significance The strength of evidence for an effect, measured by a *P*-value associated with the *F*-ratio from analysis of variance. A significant effect has a small *P*-value indicating a small chance of making a Type I error. For example, $P < 0.05$ means a less than 5% chance of mistakenly rejecting a true null hypothesis. For many tests this would be considered a reasonable level of safety for rejecting the null hypothesis of no effect, in favour of the model hypothesis of a significant effect on the response. The significance of an effect is not directly informative about the size of the effect. Thus an effect may be statistically highly significant as a result of low residual variation, yet have little biological significance as a result of a small effect size in terms of the amount of variation between sample means or the slope of a regression. A non-significant effect should be interpreted with reference to the Type II error rate, which depends on the power of the test to detect significant effects.

Split-plot A design with two or more treatment factors, and the levels of one factor applied at a different scale to those of another. For example, whole blocks might be allocated to wet or dry watering regime, and plots within blocks allocated to dense or sparse sowing density.

Sum of squares (SS) The sum of squared deviations of each independent piece of information from its modelled value. Analysis of variance partitions the total variation in the response into explained and unexplained sums of squares.

Test hypothesis H_1 The hypothesis describing the statistical model to be tested by analysis of variance. The hypothesis H_1 may have several partitions (e.g., A + B + B*A), which describe putative pattern in the data. The evidence for pattern is tested against the null hypothesis H_0 of no pattern.

Transformation A re-scaling procedure applied systematically to the response and/or covariates with the purpose of meeting the assumptions of the analysis. For example, measurements of volume and length might be log-transformed to linearise the relationship between them.

Treatment A factor with levels that are applied in a manipulative experiment. More loosely, a factor of interest (as opposed to a nuisance variable).

Type I error The mistake of rejecting a true null hypothesis. A maximum acceptable Type I error rate should be set a priori; in the biological sciences it is often taken to be $\alpha = 0.05$. An effect is then considered significant if it returns a $P < 0.05$. The Type I error is particularly susceptible to inflation in multiple tests of the same hypothesis. It is an unavoidable cause of false positives in screening programmes for rare phenomena.

Type II error The mistake of accepting a false null hypothesis. An acceptable Type II error should be set a priori; in the biological sciences it is often taken to be $\beta = 0.20$. The power of a test is greater the smaller is β. Models without full replication are particularly susceptible to Type II error, as a result of not testing higher-order interactions.

Unrestricted model A mixed model (i.e., with random and fixed factors) is termed unrestricted if a random factor is allowed to have fixed cross factors amongst its components of variation estimated in the population. This freedom influences the choice of error MS for random effects. The unrestricted model is not used in this book to construct ANOVA tables, though differences are noted in footnotes to the tables. It is generally used for unbalanced designs analysed with GLM.

Variance The variation in the data, measured as the average squared deviation of the data from the mean. Analysis of variance partitions the total variance into explained and unexplained components and estimates these variances as mean squares (MS).

References

Allison, D. B., Allison, R. L., Faith, M. S., Paultre, F. and Pi-Sunyer, F. X. (1997) Power and money: designing statistically powerful studies while minimizing financial costs. *Psychological Methods*, **2**, 20–33.

Beck, M. W. (1997) Inference and generality in ecology: Current problems and an experimental solution. *Oikos*, **78**, 265–73.

Carey, J. M. and Keough, M. J. (2002) The variability of estimates of variance, and its effect on power analysis in monitoring design. *Environmental Monitoring and Assessment*, **74**, 225–41.

Cohen, J. (1988) *Statistical Power Analysis for the Behavioral Sciences, Second Edition*. New Jersey: Lawrence Erlbaum.

Cohen, J. (1992) Quantitative methods in psychology: a power primer. *Psychological Bulletin*, **112**, 155–9.

Colegrave, N. and Ruxton, G. D. (2003) Confidence intervals are a more useful complement to nonsignificant tests than are power calculations. *Behavioral Ecology*, **14**, 446–450.

Crawley, M. J. (2002) *Statistical Computing. An Introduction to Data Analysis using S-Plus*. Chichester: Wiley.

Darlington, R. B. and Smulders, T. V. (2001) Problems with residual analysis. *Animal Behaviour*, **62**, 599–602.

Davey, A. J. H. (2003) *Competitive interactions in stream fish communities*. Ph.D. thesis, University of Southampton, Southampton.

Day, R. W. and Quinn, G. P. (1989) Comparisons of treatments after an analysis of variance in ecology. *Ecological Monographs*, **59**, 433–63.

Dayton, P. K. (1998) Reversal of the burden of proof in fisheries management. *Science*, **279**, 821–2.

Di Stefano, J. (2003) How much power is enough? Against the development of an arbitrary convention for statistical power calculations. *Functional Ecology*, **17**, 707–9.

Dytham, C. (1999, 2003) *Choosing and Using Statistics: A Biologist's Guide*. Oxford: Blackwell.

Field, S. A., Tyre, A. J., Jonzén, N., Rhodes, J. R. and Possingham, H. P. (2004) Minimizing the cost of environmental management decisions by optimizing statistical thresholds. *Ecology Letters*, **7**, 669–75.