Ecology, 75(3), 1994, pp. 717–722

USE AND MISUSE OF MIXED MODEL ANALYSIS OF VARIANCE IN ECOLOGICAL STUDIES[1]

CYNTHIA C. BENNINGTON
Department of Biology, West Virginia University, Morgantown, West Virginia 26506-6057 USA

WILLIAM V. THAYNE
Department of Statistics, West Virgnia University, Morgantown, West Virginia 26506-6330 USA

Abstract. Analysis of variance is one of the most commonly used statistical techniques among ecologists and evolutionary biologists. Because many ecological experiments involve random as well as fixed effects, the most appropriate analysis of variance model to use is often the mixed model. Consideration of effects in an analysis of variance as fixed or random is critical if correct tests are to be made and if correct inferences are to be drawn from these tests. A literature review was conducted to determine whether authors are generally aware of the differences between fixed and random effects and whether they are performing analyses consistent with their consideration. All articles (excluding Notes and Comments) in *Ecology* and *Evolution* for the years 1990 and 1991 were reviewed.

In general, authors that stated that their model contained both fixed and random effects correctly analyzed it as a mixed model. There were two cases, however, where authors attempted to define fixed effects as random in order to justify broader generalizations about the effects. Most commonly (63% of articles using two-way or greater ANOVA), authors neglected to mention whether they were dealing with a completely fixed, random, or mixed model. In such instances, it was not clear if the author was aware of the distinction between fixed and random effects, and it was often difficult to ascertain from the article whether their analysis was consistent with their experimental methods. These findings suggest several statistical guidelines that should be followed. In particular, the inclusion of explicit consideration of effects as fixed or random and clear descriptions of F tests of interest would provide the reader with confidence that the author has performed the analysis correctly. In addition, such an explicit statement would clarify the limits of the inferences about significant effects.

Key words: Ecology; Evolution; *fixed effects; mixed model analysis of variance; random effects; statistical inference.*

INTRODUCTION

Analysis of variance is one of the most commonly used statistical techniques in ecological and evolutionary studies. In many cases, more than one explanatory variable is of interest as are the interactions among those variables. These analyses can quickly become complex, particularly when a model contains both "fixed" and "random" effects. Such "mixed model" analyses are widely used in biological research as a result of the types of questions that are addressed. Because the analysis of mixed models is different than that for models which include only fixed effects or only random effects, proper recognition of effects as fixed or random is critical at all stages of the experimental design. Only when fixed and random effects are assigned correctly may the appropriate expected mean squares for hypothesis tests be determined. Prior to conducting an experiment, determining these expected mean squares increases the ability of the experimenter to maximize power to test hypotheses of interest. At the analysis stage, correct tests of hypotheses are dependent upon the use of the appropriate denominator mean square in the F test. When interpreting the results of an analysis of variance, the inferences drawn about a significant F value will differ depending upon whether the effect was fixed or random. Thus, ecologists planning to use analysis of variance must consider whether effects are fixed or random prior to conducting an experiment to ensure that the analysis is powerful, performed correctly, and can be legitimately interpreted in the manner originally intended.

The differences between analysis of variance models employing fixed and random effects were first defined by Eisenhart (1947). Eisenhart's paper and others that have followed (e.g., Henderson 1953, Wilk and Kempthorne 1955, Searle 1971*a*) have described the assumptions made and tests used for fixed and random effects. Most statistical textbooks (e.g., Searle 1971*b*, Steel and Torrie 1980, Sokal and Rohlf 1981, Zar 1984) provide a list of rules for determining whether an effect is fixed or random and describe the derivation of expected mean squares for mixed models. Unfortunately, ambiguity regarding the correct application of these

[1] Manuscript received 18 January 1993; revised 6 July 1993; accepted 27 July 1993.

rules to biological experiments has remained. Incorrect assumptions about fixed and random effects can lead to an improper analysis and ultimately to erroneous results and conclusions. Specifically, an F test that is performed incorrectly because of a lack of regard for fixed and random effects may lead an experimenter to conclude that there are differences among levels of an effect, when in fact, there are none. Alternatively, differences that actually exist may be obscured by an inappropriate F test.

Fixed effects are those explanatory variables for which the levels of the effect in the experiment were specifically chosen by the investigator. Every level of interest has been included in the experiment. No other levels are of interest, and multiple range tests are often employed to determine which pairs of means are different from each other. The null hypothesis for a fixed effect is that the dependent variable of interest does not differ in its response to the different levels of that effect. There are multiple populations for which all possible comparisons are made. Fixed effects may include factors, such as species, temperature, diet, or water availability, for which the experimenter is interested in testing the null hypothesis that the effects of specific species, temperatures, etc., are equal. These effects are called "fixed" because the same levels of the effect would be used again if the experiment was repeated.

An effect is considered "random" if the experimenter has not specifically chosen levels of the effect to be in the experiment, but has drawn a random sample from a larger population of possible levels. Thus, he wishes to draw inferences about the entire population from which he has sampled. In most cases, the experimenter is interested in obtaining an estimate of a variance component, or the magnitude of variability due to a particular effect in the model. Unlike fixed effects, there are no comparisons among populations, rather there is a single population for which an estimate of variance is of interest. Examples of random effects common in biological research are family, genotype, and individual. If the experiment were repeated, the experimenter would choose a sample of different (or new) levels for family, genotype, or individual.

These rules condense to three main criteria for determining whether an effect is fixed or random: (1) Were individual levels of the effect selected because they are of particular interest, or were they chosen completely at random? (2) Will conclusions be confined to those levels of the effect actually studied, or will they be applied to a larger population? (3) If the experiment were repeated, would the same levels of the effect be studied again, or would new samples be drawn from the larger population of possible samples (Eisenhart 1947)?

There are fairly straightforward rules that can be applied to any effect, and there are certain effects that are virtually always either fixed or random. In some cases, however, the decision as to fixed or random for any given effect is equivocal (Li 1964). Problems most often arise for explanatory variables that do not fit the idea of a "treatment" because the effect is inherent in the experimental system. Examples of such effects would be species, variety, population, and environment. When dealing with effects of time and place the decision as to fixed or random is particularly difficult (Searle 1971*a*, *b*). For example, it is not always clear whether the years over which a study was conducted are of specific interest or whether they can be considered a random sample of many possible years. Similarly, a number of sites over which an experiment is conducted may be a random sample from a larger population of sites about which inferences can be drawn, or sites may have been chosen in such a way that it is necessary to confine conclusions to those particular sites in the study.

For experiments involving both fixed and random effects, the appropriate model for the analysis of variance is the mixed model. The differences in the correct interpretation for fixed effects and random effects can perhaps be best expressed in terms of the null hypotheses for each effect in a simple case of a mixed model. Consider two effects, A and B, where A is fixed, B is random, and there is an interaction (A × B) possible between them. For a given dependent variable, the null hypothesis concerning A is that there is no difference in means among the levels of A in the experiment. For B, the null hypothesis is that there is no variability among all possible levels of B (including those not sampled), not that there are no differences among levels of that effect included in the experiment. For the interaction term (A × B), the null hypothesis is that variability among levels of B is the same for all levels of A. This differs from the case for fixed effects in that the null hypothesis for an interaction between two fixed effects (A and C) is that the response of the dependent variable is not different among specific levels of A depending upon the particular level of C.

This paper addresses some of the problems commonly encountered in the analysis of biological data with respect to fixed and random effects. We reviewed some recent ecological and evolutionary literature to address two main objectives: (1) determine whether authors are generally aware of random effects in their analysis of variance models, and (2) determine whether the analysis of mixed models is being performed correctly. We present examples from this literature review to illustrate proper and improper consideration of effects and to examine the consequences of improper consideration.

METHODS

We reviewed all articles (excluding Notes and Comments) in the 1990 and 1991 issues of *Ecology* and *Evolution* and placed the statistics employed in each paper into one of several categories. If analysis of variance was not used or if only one-way analysis of variance was used, we gave no further consideration to that paper. No attempt was made to determine whether analysis of variance would have been a more appropriate form of analysis when it was not used. Likewise, if analysis of variance was employed, we did not attempt to determine whether other statistical tests (e.g., nonparametric tests) would have been more appropriate. Where more than one kind of analysis of variance was used, we categorized the paper by the most complex model if all were correct and by the incorrect model if one or more of the others was correct. If a two-way or greater analysis of variance was used, we investigated several points: (1) Was there any specific consideration of whether effects in the model were fixed or random in the descriptions of the methods and/or the results? (2) If there was a statement concerning each of the effects, was the treatment of random and/or fixed effects consistent with sampling and inferences? (3) If there was no such statement, were all effects clearly fixed, such that explicit consideration was not crucial? (4) Was it possible to tell from results whether the correct error terms were used where random effects were employed in the model (whether explicitly stated or not)? (5) If enough information was provided, were the appropriate F tests used in the mixed model analysis?

Results

We reviewed 675 papers in *Ecology* and *Evolution* for the years 1990 and 1991. Of these, almost half (303) used some form of analysis of variance, and 226 (33.5%) used a model that was two-way or greater. Of these 226 papers, only 84 (37.2%) provided an explicit consideration of whether the effects in their model were fixed or random. In two of these cases, mixed models were described that incorrectly assumed an effect to be random which was, in fact, fixed. In all other cases where there was an explicit description of effects provided by the author(s), there was a clear recognition of the distinction between fixed and random effects, and the analyses were performed correctly. Overall, the majority of authors did not provide an explicit consideration of their effects, but the majority of those that used a mixed model analysis of variance did describe their effects either in the Methods or Results sections. Fourteen authors (6.2% of those using models that were two-way or greater) whose studies required a mixed model analysis of variance either did not recognize it as such, and performed the analysis as if all effects were fixed, or did not provide enough information in their results for us to ascertain that a mixed model was used in the analysis.

In order to present real examples of the misuse and misunderstandings surrounding mixed model analysis of variance, several of the analyses performed in the papers reviewed from *Ecology* and *Evolution* will be described. To avoid casting statistical stones at particular individuals, titles, authors, and specific details have been removed from descriptions of papers found to have faults. Sufficient details are supplied to illustrate where the problems lie without placing blame.

Examples from the literature

Perhaps the most common mistake encountered in this literature review was that of nested random effects being incorrectly treated as fixed. Nested analyses of variance are common in biological data analysis as they arise whenever major groupings of a factor are divided into smaller subgroups. According to Sokal and Rohlf (1981), all nested effects must be randomly chosen. In reality, exceptions to this may occur. However, when nested effects are treated as fixed, it is critical that inferences made from the analysis are limited to those specific subgroups included in the experiment. Such situations are relatively rare, and, in general, a nested analysis of variance is either a completely random model (if all levels of classification are random) or a mixed model (if the highest level of classification is a fixed effect).

One example from the literature where nested effects were treated as fixed involved clones sampled from three source populations (A, B, and C). A total of 35 clones from the three populations were collected with 17, 8, and 10 collected from A, B, and C respectively. Clearly, the author was interested specifically in those three source populations from which clones were collected, and this effect was correctly considered to be fixed. However, differences among particular clones were not of interest. It is doubtful, for example, that comparisons among each of the 17 clones from source population A would be meaningful. In fact, there is little mention of the effect of the nested term clone-within-source population in the discussion of results except in terms of variation among clones (for development times). This suggests that clones were being used as representative random samples of each of the source populations and that the quantity of interest was the variance among clones, not the absolute differences in their means. The description of clone collection was not described precisely in the paper, making it impossible to determine whether it was necessary to consider clone to be a fixed effect given the constraints of the sampling procedure. If clone were considered random, the correct error mean square for the F test to detect significant differences among source populations would then have been the nested "clone (source population)" term. In neither of the analyses presented would the correct test have changed the conclusion of the significance of source population, although the magnitude of the F value would have been decreased (i.e., $F_{2,73} = 24.1$ would change to $F_{2,32} = 10.9$). It is also apparent that this change in the analysis greatly reduces error degrees of freedom, resulting in a substantial loss of power to test the null hypothesis of no differences among the three source populations.

In general, when authors stated specifically whether each of their effects was fixed or random, the consideration was correct. However, in both of the cases for which this was not true, effects were considered to be random when they did not legitimately fit the criteria for random effects. For example, in one study, seeds of an annual plant were collected from 40 maternal sibships from two populations (A and B) known to differ in annual temperature and rainfall. Two watering treatments (weekly and biweekly) were then imposed, where three plants from each sibship were exposed to each of the watering treatments. In the analysis of this experiment, watering treatment was correctly considered fixed and both population and family nested within population were considered to be random effects. Clearly, family was correctly considered a random effect as specific differences among family means in response to watering were not of interest. An estimation of variance among families within each of the populations was obtained. It is not appropriate, however, to consider population to be random when the two populations were presumably selected specifically because they were from different physical environments. The author was interested in obtaining an estimate of among-population variance in flowering time, but the experimental design was not consistent with this type of analysis. Populations were not chosen at random, but were expected a priori to differ. Thus, it is not legitimate to draw inferences about a larger set of all possible populations based on these data. In the Results section, the author discussed specific differences in the response of plants from the two populations to the watering treatments. For example, plants from Population A "began budding at significantly smaller size than [plants from Population B] in all cases." Such a direct comparison between populations does not seem consistent with the treatment of population as a random effect.

Many of the papers reviewed had effects in their models involving time and place. In general, these were treated by the authors as fixed effects without an explanatory statement as to why this was so. In many cases, the description of methods was not sufficiently complete to decide whether such effects could be considered random. An example of the difficulty which arises when dealing with effects of time comes from a study of tail length in birds measured over 7 yr. A three-way analysis of variance was employed to determine the effect of sex, age, and year on tail length. From the analysis of variance table provided in the Results section, it was obvious that all three effects were considered fixed, although this was never explicitly stated. Age and sex are clearly fixed effects, but the classification of year is not so straightforward. The criteria for random effects cannot be completely met as years of an experiment are virtually never chosen completely at random. However, certain years may or may not be of particular interest, and certainly the same years would not be repeated in another experiment. Because time proceeds in an orderly progression over which conditions are likely to change, differences among specific times are often of interest. Environmental conditions may be measured from year to year such that the cause of differences among years in some response variable may be explicitly tested. For these reasons, year is often considered a fixed effect and it is assumed that this was the rationale in the previously described analysis. In this particular example, the author's interpretation of year as a fixed effect is consistent with his analysis of it as such. For example, he found a correlation between precipitation (which varied among years) and tail length. Thus, he measured a specific environmental variable known to differ among years and found a relationship that suggests that differences in tail length may be due to differences in precipitation among years. Therefore, it seems as though the author is not attempting to draw conclusions about a larger sample of possible years over which the study could have been conducted. However, because seven different years were involved, it may have been possible to consider these as representative of a larger population of years and to use the year term to obtain an estimate of the magnitude of variation in tail length associated with year. Such reasoning would have led to year being considered a random effect. Thus, perhaps the most important issue for determining whether or not the effect of year can be considered fixed or random is whether enough years have been sampled to reflect actual amounts of annual variation. If this is not the case, the experimenter cannot legitimately generalize his results to all possible years. Although the analysis appears to have been performed correctly, the decision to consider year fixed may not have been easy. In such situations, an explicit consideration of the effects in an analysis of variance model by the author would allow the reader to understand at the outset what assumptions are being made and to interpret the results accordingly.

Place, as well as time, is often difficult to assign as a fixed or random effect. Difficulty in assigning fixed or random status to a "place" effect can occur when dealing with blocks. In a truly randomized complete block (RCB) design where there are replicates of each treatment in each block such that a treatment × block interaction is possible, blocks are generally considered to be random. This is usually the most desirable case, as experimenters are rarely interested in the effect of block, but use block as a way of removing extraneous variability from main effects. In practice, however, it may be difficult to choose blocks at random, leading block to be treated as a fixed effect.

Two related papers describing separate experiments provide examples of block being considered fixed in one instance and random in another. In both cases, an explicit consideration of fixed and random effects was given in the description of the analysis. Artificial ponds were set up in both studies to investigate amphibian population interactions, and these ponds were arranged in blocks in a large field. In the first of these papers, ponds were arranged into blocks in the field to "account for unknown physical gradients at the site." In this case, block was considered random. Presumably block positions were chosen at random, and different positions would be chosen if the experiment were repeated. The experimental design was consistent with the interpretation that blocks are representative of a larger population of possible blocks. In the second paper, there were a total of 10 blocks. Each block incorporated the effect of both time and local environment, as the same experiment was performed twice in one summer on the same five artificial ponds. Block was considered to be a fixed effect in the analysis of variance employed. Since dates were not randomly chosen and because there may be predictable differences among blocks due to date, it was deemed necessary to consider block as a fixed effect in this case. There are problems with this consideration, however. When block is treated as a fixed effect, the experimenter has decided that inferences will be confined to those blocks in the analysis, and the effect of treatment is tested over the residual error term, not the block × treatment term. The test of the treatment effect is then a test of whether there are treatment differences given those specific blocks included in the experiment. This is generally not of interest, particularly when the dependent variable may be responding differently to the treatment depending upon the block. The authors of these papers recognized that effects such as block may be either random or fixed depending upon the method used to select levels of that factor. However, when block was considered fixed, the authors should have justified their test of the treatment effect. The danger of considering block to be a fixed effect should be considered when an experiment is being designed, as doing so may make it impossible to test for the effect of the treatment of interest.

Discussion

The total number of mistakes in the statistical analysis of published papers reviewed was relatively small. Many of the authors employing mixed models recognized the distinction between fixed and random effects and correctly analyzed their data. However, for those authors who did not provide an explicit consideration of effects, mistakes were common. For example, of the nine *Ecology* papers that did not consider whether their effects were fixed or random, all (five) of those papers for which it was possible to tell how the analysis was performed incorrectly considered a random effect to be fixed. Very few papers provided a clear rationale for the consideration of their effects, and many provided no way for the reader to determine how the analysis was conducted. Our review of the literature indicates a distinct need for authors to become more aware of the effects in their models and to incorporate a consideration of these effects into the description of their statistical tests. The results of this review suggest a list of several rules for authors to follow when they are using analysis of variance.

1. Consult a statistician prior to designing an experiment. First, this will ensure that the experimental design enables one to meet the objectives of the study. Second, a statistician can assist in determining the correct F test for each of the effects of interest when a mixed model analysis of variance is called for. Even for relatively simple models with only one random effect, determination of the correct denominator MS can become complex when there are interactions between fixed and random effects. While statistical programs like SAS (SAS 1985) make it relatively simple to analyze data, assignment of effects as random must be specified as must the appropriate F test for hypotheses that include the random effect in the denominator. In addition, there are two models that may be employed in the analysis of mixed models, and the interpretation of the F tests from these will differ. Comparisons between these models have been discussed in detail in several papers (Hocking 1973, Ayres and Thomas 1990, Fry 1992).

2. Provide an explicit consideration of each of the effects in an analysis of variance so that it is clear to the reader what assumptions are being made about each effect. Fowler (1990) provided a list of 10 suggestions for authors to follow to avoid statistical errors. The first of these was to explain clearly what was done. She provided a set of questions that a reader should be able to answer about the experimental design and analysis of data. A thorough description of the statistical analysis that answers questions about whether and why effects are considered fixed or random is crucial for understanding the analysis that follows.

3. For particularly complicated tests, it is best to include a means by which the reader may determine which MS is being used in the denominator of the F test for each of the effects in the model. An extra column (headed "denominator MS") in the analysis of variance table is one way of presenting this clearly. In addition, when the design is unbalanced, as is often the case in field experiments, it would be useful to present the expected mean square (with the appropriate coefficient) for each effect in the model. Lack of balance requires caution in analysis and interpretation (Searle 1987; see also Shaw and Mitchell-Olds 1993 for a detailed consideration of this topic in fixed effects models), and the consequences of this should be recognized by authors whose experiments are affected by unbalanced designs.

4. Consider whether effects are fixed or random *before* performing an experiment. The power with which hypotheses can be tested depends upon the assignment of effects as fixed or random. In addition, it may be impossible to test certain effects or interactions of interest in mixed models if the design is inappropriate. Effects that are random should not be treated as fixed to simplify the analysis. Similarly, an estimate of vari-

ance should not be obtained and discussed for an effect if the criteria for random effects are not met.

Acknowledgments

We would like to thank J. McGraw, T. Shaffer, and an anonymous reviewer for their valuable comments on the manuscript and K. Garbutt for providing the impetus for this work. This work was supported by NSF Dissertation Improvement Grant BSR-9107304 to C. C. Bennington.

Literature Cited

Ayres, M. P., and D. L. Thomas. 1990. Alternative formulations of the mixed-model anova applied to quantitative genetics. Evolution **44**:221–226.

Eisenhart, C. 1947. The assumptions underlying the analysis of variance. Biometrics **3**:1–21.

Fowler, N. 1990. The 10 most common statistical errors. Bulletin of the Ecological Society of America **71**:161–163.

Fry, J. D. 1992. The mixed model analysis of variance applied to quantitative genetics: biological meaning of the parameters. Evolution **46**:540–550.

Henderson, C. R. 1953. Estimation of variance and covariance components. Biometrics **9**:226–252.

Hocking, R. R. 1973. A discussion of the two-way mixed model. American Statistician **27**:148–152.

Li, C. C. 1964. Introduction to experimental statistics. McGraw-Hill, New York, New York, USA.

SAS. 1985. SAS user's guide: statistics. Version 5 edition. SAS Institute, Cary, North Carolina, USA.

Searle, S. R. 1971*a*. Topics in variance component estimation. Biometrics **27**:1–76.

———. 1971*b*. Linear models. John Wiley & Sons, New York, New York, USA.

———. 1987. Linear models for unbalanced data. John Wiley & Sons, New York, New York, USA.

Shaw, R., and T. Mitchell-Olds. 1993. ANOVA for unbalanced data: an overview. Ecology **74**:1638–1645.

Sokal, R. R., and F. J. Rohlf. 1981. Biometry. Second edition. W. H. Freeman, New York, New York, USA.

Steel, R. G. D., and J. H. Torrie. 1980. Principles and procedures of statistics: a biometrical approach. McGraw-Hill, New York, New York, USA

Wilk, M. B., and O. Kempthorne. 1955. Fixed, mixed, and random models. Journal of the American Statistical Association **50**:1144–1167.

Zar, J. H. 1984. Biostatistical analysis. Second edition. Prentice-Hall, Englewood Cliffs, New Jersey, USA.