BIOMETRICS 54, 1085–1096
September 1998

Estimation of Denominator Degrees of Freedom of F-Distributions for Assessing Wald Statistics for Fixed-Effect Factors in Unbalanced Mixed Models

D. A. Elston

Biomathematics and Statistics Scotland, MLURI,
Craigiebuckler, Aberdeen AB15 8QH, U.K.

SUMMARY

Tests for fixed-effect factors in unbalanced mixed models have previously used t-tests on a contrast-by-contrast basis or Wald statistics without a universally accepted method of calculating the denominator degrees of freedom. This situation has arisen because the variances of different contrasts are differently weighted sums of the variance components with associated degrees of freedom that are not necessarily equal. A simultaneous F-test for differences between all levels of a fixed-effect factor can be derived by forming new contrasts, by rotation of the original contrasts, with variances that are close to being the same weighted sum of variance components. The associated degrees of freedom for these new contrasts are nearly equal. A small simulation study shows the appropriateness of a X^2 approximation to the distribution of the weighted sums of variance components. Three simple examples are used to demonstrate the effects of rotation. The last of these examples is also used to compare the proposed simultaneous F-test with the distribution of the Wald statistic obtained by numerical simulation. The method of rotations is then applied to data on the range size of mountain hares (*Lepus timidus*) to assess the evidence for a two-way interaction between season and habitat.

1. Introduction

Analysis of data from balanced experiments is traditionally performed by an analysis of variance in which the total sum of squared deviations of observations about their mean is partitioned into sums of squares attributable to either the different treatment effects or the different random effects. Under standard distributional assumptions, each random effect sum of squares follows a multiple of a X^2-distribution, this multiple being the product of the degrees of freedom (the number of independent error contrasts whose square has the required expectation) and a linear combination of the different variance components, the coefficients in this linear combination being fixed by the design. Under any null hypothesis concerning the absence of treatment effects, the corresponding treatment sum of squares also follows such a multiple of a X^2-distribution. Furthermore, the linear combination of variance components in the multiple for the treatment sum of squares is identical to the linear combination for one of the random-effect sum of squares; hence, hypotheses about the treatment effects can be tested by dividing treatment mean squares by the appropriate error mean square to form a variance ratio with an F-distribution. The benefits of experimental design accrue, first, through the variance of all contrasts for any given treatment factor or interaction having the same expectation under the null hypothesis; second, through this variance being as small as possible; and, third, through there being as many error contrasts as possible whose variance is the same as that for the treatment factor or interaction.

Analysis of data that lacks these properties of balance is much less straightforward, yet there are many areas in which, due to the nature of experimental material, the treatment effects cannot be applied in a balanced fashion, and so such unbalanced data are the norm rather than the

Email address: D.Elston@bioss.sari.ac.uk

Key words: Degrees of freedom; REML; Rotation of contrasts; Wald statistic.

exception. Particular problems encountered are that the decomposition of the total sum of squares is no longer unique and that, in any given decomposition, corresponding to any particular treatment mean square there is not an error mean square with equal expectation under the null hypothesis. Hence, in the case of unbalanced data, it has become accepted practice to adopt a more explicit modelling approach, estimating fixed-effect parameters and variance components by maximum likelihood. Use of full maximum likelihood leads to biased estimates of the variance components, requires appeal to asymptotic arguments, and gives results that differ from the exact tests provided in the balanced case by an analysis of variance. These problems with full maximum likelihood have been overcome by the use of restricted maximum likelihood (REML), in which the variance components are estimated using only the likelihood in the residual space that is independent of the fixed effects (Patterson and Thompson, 1971). The differences between full and residual maximum likelihood are often quite small, but each method has slightly different properties (Searle, Casella, and McCulloch, 1992). However, the coincidence of REML and ANOVA for balanced data when the estimated variance components are positive and the presence of limited replication in the higher strata together provide strong reasons for preferring REML.

Testing even a single fixed-effect contrast in REML presents a problem since the estimated error variances are not sums of squares with X^2-distributions. For complete two-way classifications with one fixed-effect and one random factor, exact tests for fixed effects can be found (Khuri and Littell, 1987; Gallo and Khuri, 1990), but these results leave open the question of what to do in general. Giesbrecht and Burns (1985) have shown that, by accepting a X^2 approximation to the distribution of variances of contrast estimates, degrees of freedom for a t-test can be calculated using Sattherthwaite's approximation (Sattherthwaite, 1946) and that the resulting confidence intervals perform reasonably in a simulation study. In their paper, Giesbrecht and Burns do not address the issue of simultaneous testing for factors with more than two levels, preferring instead to perform t-tests on selected orthogonal contrasts that are not statistically independent. Similarly, McLean and Saunders (1988) show how to perform t-tests for contrasts involving levels of both fixed and random effects; again, they ignore the problem of simultaneous testing.

One statistic recently proposed for simultaneous testing of fixed-effect factors is the Wald statistic (Berk, 1987), this being a generalization of the Hotelling T^2 statistic without the requirement for the variance–covariance matrix to follow a Wishart distribution. The distribution of the Wald statistic can be evaluated empirically by simulation. For a theoretical evaluation of the distribution, it is natural to turn to the family of F-distributions with numerator degrees of freedom specified by the number of fixed-effect contrasts being considered. Determination of suitable denominator degrees of freedom is hampered by the different fixed-effect contrasts having unequal variances whose X^2 approximations have different numbers of degrees of freedom. The alternatives currently used (cf., Brown and Kempton, 1994) are to take the smallest stratum degrees of freedom over those strata that contribute to the contrast variances, to take the degrees of freedom for the lowest level stratum, or to use a X^2 approximation. The latter two alternatives are overly optimistic, while the first alternative is unnecessarily cautious if only a small amount of information is taken from a stratum with only a few degrees of freedom.

A completely different approach to the testing of fixed effects, which makes use of the residual likelihood after constraining certain fixed-effect parameters to be zero, has recently been developed by Welham and Thompson (1997). The primary disadvantage of the constrained likelihood approach is that it requires separate model fits for every comparison, even within a sequence of nested models. In addition to demonstrating the efficacy of their own methodology, Welham and Thompson (1997) demonstrate clearly the inadequacy of the X^2 approximation to the Wald statistic.

In what follows, it is demonstrated that the denominator degrees of freedom problem for the Wald statistic can be largely overcome by considering suitable linear combinations of the fixed-effect contrasts used to define a given factor or interaction of factors. The purpose of redefining the fixed-effect contrasts is to find a construction for the Wald statistic that matches as closely as possible the nature of an F-statistic as a ratio of X^2 random variables. This requires the formation of fixed-effect contrasts whose variances come as close as possible to being the same linear combination of variance components. However, even if the variances of these new contrasts can be expressed as identically weighted sums of the different variance components, this weighted sum will not necessarily follow an exact X^2-distribution. Hence, the F-distribution will in general never prove exact for the Wald statistic, and the appropriateness of the estimated distributions is examined by a small simulation study. The effect of forming linear combinations of contrasts is demonstrated by way of three simple, constructed examples and by application to a real data set.

2. Methods

Consider the standard linear model, namely

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\alpha} + \sum \boldsymbol{U}_i\boldsymbol{\beta}_i,$$

with p linearly independent fixed effects $\boldsymbol{\alpha}$ and independent random effects $\boldsymbol{\beta}_i$ with corresponding variance components σ_i^2, $i = 1 \cdots b$. Then

$$E[\boldsymbol{y}] = \boldsymbol{X}\boldsymbol{\alpha},$$

and

$$\text{var}[\boldsymbol{y}] = \boldsymbol{V} = \sum \boldsymbol{U}_i\boldsymbol{U}_i^{\mathrm{T}}\boldsymbol{\sigma}_i^2.$$

If the variance components σ_i^2 are known, $\boldsymbol{\alpha}$ can be estimated by weighted regression with

$$\hat{\boldsymbol{\alpha}} = \boldsymbol{W}\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{y}, \tag{1}$$

where $\boldsymbol{W}$, the covariance matrix of $\boldsymbol{\alpha}$, is given by

$$\boldsymbol{W} = \left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}.$$

In practice, the variance components are not known and (1) has to be used after substitution of their estimates $\hat{\sigma}_i^2$. These estimates can be calculated iteratively either using the full likelihood or using the residual likelihood (normally calculated as the likelihood of $(\boldsymbol{I} - \boldsymbol{X}(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{X})^{-1}\boldsymbol{X}^{\mathrm{T}})\boldsymbol{y}$).

Let any nonsingular set of contrasts of the elements of $\boldsymbol{\alpha}$ be given by the r rows $\boldsymbol{c}_1^{\mathrm{T}} \cdots \boldsymbol{c}_r^{\mathrm{T}}$ of a contrast matrix $\boldsymbol{C}$. Then $\text{var}(\boldsymbol{C}\boldsymbol{\alpha}) = \boldsymbol{C}\boldsymbol{W}\boldsymbol{C}^{\mathrm{T}}$, and the Wald statistic T_C for the set of contrasts in $\boldsymbol{C}$ is given by

$$T_C = \hat{\boldsymbol{\alpha}}^{\mathrm{T}}\boldsymbol{C}^{\mathrm{T}}\left[\boldsymbol{C}\boldsymbol{W}\boldsymbol{C}^{\mathrm{T}}\right]^{-1}\boldsymbol{C}\hat{\boldsymbol{\alpha}}. \tag{2}$$

When $r = 1$ and writing $C = c_1^{\mathrm{T}}$, we can follow Giesbrecht and Burns (1985) and approximate the distribution of T_C by the $F_{1,g}$ distribution, where the denominator degrees of freedom are estimated using Sattherthwaite's approximation

$$g = 2\left\{\text{var}\left(\boldsymbol{c}_1^{\mathrm{T}}\hat{\boldsymbol{\alpha}}\right)\right\}^2 \Big/ \text{var}\left(\text{var}\left(\boldsymbol{c}_1^{\mathrm{T}}\hat{\boldsymbol{\alpha}}\right)\right), \tag{3}$$

with $\text{var}(\boldsymbol{c}_1^{\mathrm{T}}\hat{\boldsymbol{\alpha}}) = \boldsymbol{c}_1^{\mathrm{T}}\boldsymbol{W}\boldsymbol{c}_1$.

The denominator in (3) is evaluated by taking the linear approximation

$$\text{var}\left(\boldsymbol{c}_1^{\mathrm{T}}\hat{\boldsymbol{\alpha}}\right) \approx \sum l_{i1}\sigma_i^2 = \boldsymbol{l}_1^{\mathrm{T}}\boldsymbol{\sigma}^2, \tag{4}$$

where $\boldsymbol{\sigma}^2 = (\sigma_1^2 \cdots \sigma_b^2)^{\mathrm{T}}$,

$$l_{i1} = \boldsymbol{c}_1^{\mathrm{T}}\partial\boldsymbol{W}/\partial\sigma_i^2\boldsymbol{c}_1 = \boldsymbol{c}_1^{\mathrm{T}}\boldsymbol{W}_i\boldsymbol{c}_1, \tag{5}$$

say, with

$$\begin{aligned}
\boldsymbol{W}_i &= \partial\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}/\partial\sigma_i^2 \\
&= \left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\partial\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}/\partial\sigma_i^2\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1} \\
&= \left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}^{-1}\partial\boldsymbol{V}/\partial\sigma_i^2\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1} \\
&= \left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{U}_i\boldsymbol{U}_i^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1}\left(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{V}^{-1}\boldsymbol{X}\right)^{-1},
\end{aligned}$$

and using this linear approximation and an estimate of $\text{cov}(\hat{\boldsymbol{\sigma}}^2)$ to estimate

$$\text{var}\left(\widehat{\text{var}}\left(\boldsymbol{c}_1^{\mathrm{T}}\hat{\boldsymbol{\alpha}}\right)\right) \approx \boldsymbol{l}_1^{\mathrm{T}}\text{cov}\left(\hat{\boldsymbol{\sigma}}^2\right)\boldsymbol{l}_1. \tag{6}$$

When $r > 1$, there is currently no guidance about how to construct a reference distribution for the Wald statistic. Although the above formulas can be used to perform independent tests for each contrast, construction of a joint test is clearly preferable. The following sections redress this imbalance, first, by considering what we would like of the contrasts defining a factor with more than two levels, second, by describing the room for maneuver in the defining contrasts, and, finally, by considering how this room for maneuver might best be utilized.

2.1 *Desirable Properties for Factors Defined by* $r > 1$ *Contrasts*

In the balanced analysis of variance situation, the null distribution for the sum of squares for each contrast is proportional to a X^2-distribution, where the constant of proportionality is a weighted sum of the variance components. For contrasts $\boldsymbol{c}_1 \cdots \boldsymbol{c}_r$ defining a factor with r degrees of freedom, the corresponding weights $\boldsymbol{l}_1 \cdots \boldsymbol{l}_r$ are then all equal regardless of the particular choice of contrasts. In the unbalanced situation, we wish to mimic this property by making the weights as similar as possible. One obvious goal then is to prefer sets of defining contrasts for which

$$\max_{i,j}(\boldsymbol{l}_i - \boldsymbol{l}_j)^{\mathrm{T}}(\boldsymbol{l}_i - \boldsymbol{l}_j) \tag{7}$$

is small. This criterion has been selected because, as will be shown below, it leads to a series of minimization problems, each of which has a closed-form solution. Note that equality of the $\boldsymbol{l}_i$, $i = 1 \cdots r$, is sufficient for equality of the contrast degrees of freedom. However, as will be shown in Example 2, this is not a necessary condition.

2.2 *Scope for Changing the Defining Contrasts*

Suppose as in Section 2 that a factor is defined by the contrasts $\boldsymbol{c}_1 \cdots \boldsymbol{c}_r$, which form the rows of the matrix $\boldsymbol{C}$. Then for any invertible $r \times r$ matrix $\boldsymbol{M}$, the $r \times p$ matrix $\boldsymbol{MC}$ is a new contrast matrix with identical Wald statistic to that for the contrast matrix $\boldsymbol{C}$. (Substitute $\boldsymbol{MC}$ for $\boldsymbol{C}$ in equation (1) and use the matrix identity $(\boldsymbol{FGH})^{-1} = \boldsymbol{H}^{-1}\boldsymbol{G}^{-1}\boldsymbol{F}^{-1}$ for invertible matrices $\boldsymbol{F}$, $\boldsymbol{G}$, and $\boldsymbol{H}$.) Furthermore, if $\boldsymbol{M}$ is chosen such that the matrix $\boldsymbol{D}^2 = \mathrm{var}(\boldsymbol{MC}\hat{\boldsymbol{\alpha}}) = \boldsymbol{MCWC}^{\mathrm{T}}\boldsymbol{M}^{\mathrm{T}}$ is diagonal with elements d_i^2, then the corresponding calculation of the Wald statistic T_{MC} is

$$T_{MC} = \hat{\gamma}^{\mathrm{T}}\boldsymbol{D}^{-2}\hat{\gamma} = \sum \hat{\gamma}_i^2/d_i^2, \tag{8}$$

where $\hat{\gamma} = \boldsymbol{MC}\hat{\boldsymbol{\alpha}}$. Thus, T_{MC} is a sum of r ratios, the numerators being squares of contrast values and the denominators being their corresponding variances. Each such ratio is distributed approximately as $F_{1,gi}$, where g_i can be estimated as outlined above. By construction, the estimated correlation between any pair of numerator contrast values is zero.

There is considerable flexibility in the choice of $\boldsymbol{M}$, some of which can be removed by suitable scaling constraints. One way to make this reduction without any loss of generality involves taking the Choleski decomposition $\boldsymbol{CWC}^{\mathrm{T}} = \boldsymbol{LL}^{\mathrm{T}}$ and then writing $\boldsymbol{M} = \boldsymbol{RL}^{-1}$ for any rotation matrix $\boldsymbol{R}$ with $\boldsymbol{RR}^{\mathrm{T}} = \boldsymbol{I}$. This removes all superfluous variation while leaving us to determine $\boldsymbol{R}$. In what follows, the rows of the matrix $\boldsymbol{L}^{-1}\boldsymbol{C}$ will be treated as being the original contrasts.

2.3 *Algorithms for Determining* $\boldsymbol{R}$

When $r = 2$, we can choose $\boldsymbol{R}$ to rotate the contrasts $\boldsymbol{c}_1$ and $\boldsymbol{c}_2$ by an angle θ in their plane until reaching the angle θ_0 at which (7) reaches a minimum. To find θ_0, we consider the rotated contrasts $\boldsymbol{c}_1(\theta) = \boldsymbol{c}_1\cos(\theta) + \boldsymbol{c}_2\sin(\theta)$ and $\boldsymbol{c}_2(\theta) = \boldsymbol{c}_2\cos(\theta) - \boldsymbol{c}_1\sin(\theta)$. Suppose they have variances that are linear combinations of the variance components with weights $\boldsymbol{l}_1(\theta)$ and $\boldsymbol{l}_2(\theta)$, respectively. Then, for $j = 1 \cdots b$,

$$\begin{aligned} l_{1j}(\theta) &= \cos^2(\theta)\boldsymbol{c}_1^{\mathrm{T}}\boldsymbol{W}_j\boldsymbol{c}_1 + \sin^2(\theta)\boldsymbol{c}_2^{\mathrm{T}}\boldsymbol{W}_j\boldsymbol{c}_2 + 2\cos(\theta)\sin(\theta)\boldsymbol{c}_1^{\mathrm{T}}\boldsymbol{W}_j\boldsymbol{c}_2 \\ l_{2j}(\theta) &= \sin^2(\theta)\boldsymbol{c}_1^{\mathrm{T}}\boldsymbol{W}_j\boldsymbol{c}_1 + \cos^2(\theta)\boldsymbol{c}_2^{\mathrm{T}}\boldsymbol{W}_j\boldsymbol{c}_2 - 2\cos(\theta)\sin(\theta)\boldsymbol{c}_1^{\mathrm{T}}\boldsymbol{W}_j\boldsymbol{c}_2, \end{aligned}$$

and so

$$l_{1j}(\theta) - l_{2j}(\theta) = \cos(2\theta)\left(\boldsymbol{c}_1^{\mathrm{T}}\boldsymbol{W}_j\boldsymbol{c}_1 - \boldsymbol{c}_2^{\mathrm{T}}\boldsymbol{W}_j\boldsymbol{c}_2\right) + 2\sin(2\theta)\boldsymbol{c}_1^{\mathrm{T}}\boldsymbol{W}_j\boldsymbol{c}_2. \tag{9}$$

Squaring (9) and summing over j gives

$$(\boldsymbol{l}_1 - \boldsymbol{l}_2)^{\mathrm{T}}(\boldsymbol{l}_1 - \boldsymbol{l}_2) = \cos^2(2\theta)S_1 + \cos(2\theta)\sin(2\theta)S_2 + \sin^2(2\theta)S_3, \tag{10}$$

where

$$S_1 = \sum\left(\boldsymbol{c}_1^{\mathrm{T}}\boldsymbol{W}_j\boldsymbol{c}_1 - \boldsymbol{c}_2^{\mathrm{T}}\boldsymbol{W}_j\boldsymbol{c}_2\right)^2,$$

$$S_2 = 4\sum \left(c_1^{\mathrm{T}}W_jc_1 - c_2^{\mathrm{T}}W_jc_2\right) c_1^{\mathrm{T}}W_jc_2,$$
$$S_3 = 4\sum \left(c_1^{\mathrm{T}}W_jc_2\right)^2.$$

Differentiating with respect to θ and equating to zero gives turning points for (10) at

$$4\theta = \tan^{-1}(S_2/(S_1 - S_3)),$$

from which the root for θ corresponding to the minimum can by determined by evaluation of (10).

Note that, if the number of variance components is three, the minimized value of equation (7) may not be zero, as equality of $\boldsymbol{l}_1$ and $\boldsymbol{l}_2$ requires the solution of a six-dimensional problem with two constraints, two simultaneous equations, and just one parameter.

When $r > 2$, a simple inductive argument shows that the number of angles involved in constructing $\boldsymbol{R}$ is $r(r-1)/2$. This poses problems for minimizing (7), as the the dimension increases markedly with r. However, the construction of (7) is appropriate for an iterative algorithm that addresses in turn the most discordant pairs of contrasts. Each iteration is simple and quick to perform, as it has a closed-form solution, although many iterations may be required. Such an algorithm consists of the following steps.

Step 1*:* Find from among the r contrasts $\boldsymbol{c}_1 \cdots \boldsymbol{c}_r$ the ones for which $(\boldsymbol{l}_i - \boldsymbol{l}_j)^{\mathrm{T}}(\boldsymbol{l}_i - \boldsymbol{l}_j)$ is a maximum, say $\boldsymbol{c}_h$ and $\boldsymbol{c}_l$.

Step 2*:* Fix all contrasts except $\boldsymbol{c}_h$ and $\boldsymbol{c}_l$, then rotate by an angle θ in their plane until reaching the angle θ_0 that minimizes $(\boldsymbol{l}_h - \boldsymbol{l}_l)^{\mathrm{T}}(\boldsymbol{l}_h - \boldsymbol{l}_l)$. Replace the original contrasts $\boldsymbol{c}_h$ and $\boldsymbol{c}_l$ with their rotated counterparts $\boldsymbol{c}_h\cos(\theta_0) + \boldsymbol{c}_l\sin(\theta_0)$ and $\boldsymbol{c}_l\cos(\theta_0) - \boldsymbol{c}_h\sin(\theta_0)$.

Step 3*:* If the new value of (7) is sufficiently small or insufficient reduction to (7) has been made, stop; otherwise, return to step 1.

This algorithm forms the required contrasts directly. If the rotation matrix $\boldsymbol{R}$ is required, it can easily be calculated as the product of matrices defining the two-dimensional rotations used by the algorithm.

Again, equality of variance component weights cannot always be achieved. Indeed, the optimization problem has $r \times b$ variables with r constraints and $(b-1)$ unknowns. Hence, there are $(r-1)\times(b-1)$ equations with $r(r-1)/2$ parameters, and so there will in general never be an exact solution when $b > r/2 + 1$.

Where the fixed effects have a factorial treatment structure, the Wald statistics can be evaluated by taking the appropriate terms from (8) and the denominator degrees of freedom determined by minimizing (7), restricting to particular groups of contrasts. Note that, to ensure the correct Wald statistic is calculated for each combination of treatment factors, it is desirable to begin by taking $\boldsymbol{C}$ to be the contrast matrix for all fixed degrees of freedom and to take the Choleski matrix $\boldsymbol{L}$ to be upper triangular. This ensures that the Wald statistics are calculated for adding terms according to their position from left to right in the $\boldsymbol{X}$ matrix.

3. A Comparison of Single Degree of Freedom Wald Statistics and Their Reference Distributions

It is instructive at this point to see the results of a simulation study of the effectiveness of the approximate distributions of Wald statistics proposed by Giesbrecht and Burns (1985) for single degree of freedom contrasts. This simulation study was carried out for an experiment with 20 plots in 10 blocks of 2 plots. There was one treatment factor with 2 levels, each of which was always allocated to 10 plots. Varying degrees of between- and within-block information was achieved by having either 0, 2, 4, 6, 8, or 10 blocks containing both treatment levels. The between-plot variance component was always 1, and the between-block variance component took each of the values 1, 2, 4, and 8. For each combination of design and variance component, 1000 simulations were conducted with a null treatment effect. Simulations were repeated if the estimated between-block variance component was negative. For each simulation, the Wald statistic and estimated denominator degrees of freedom were stored. An upper p-value was obtained by rounding the degrees of freedom to the nearest integer and referencing each Wald statistic against the appropriate F-distribution.

The results of these simulations are summarized in Table 1. The balance of information within and between blocks has a much greater influence on estimated denominator degrees of freedom than the value of the variance component. The degrees of freedom are estimated quite stably but by no means without variation. They are greatest with the smallest nonzero amount of within-block information and, for any mixture of information, are greater than available from either source in isolation. The p-values overall are not too far from their nominal values, but there is clear evidence

Table 1

Summaries of sets of 1000 simulations for each combination of design and between-block variance component: (a) mean estimated degrees of freedom; (b) standard deviation of estimated degrees of freedom; (c), (d), and (e) proportion of Wald statistics exceeding the 90th, 95th and 99th percentiles of their estimated reference distributions, respectively

Number of blocks containing both treatment levels	Between block component of variance			
	1	2	4	8
(a)				
0	8.0	8.0	8.0	8.0
2	16.3	16.4	15.4	13.6
4	15.7	14.3	12.4	11.0
6	13.1	12.0	10.7	9.9
8	10.7	10.1	9.6	9.2
10	9.0	9.0	9.0	9.0
(b)				
0	0.00	0.00	0.00	0.00
2	1.69	1.70	2.21	2.26
4	2.14	2.38	2.07	1.52
6	2.07	1.84	1.24	0.84
8	1.00	0.87	0.65	0.41
10	0.00	0.00	0.00	0.00
(c)				
0	0.11	0.10	0.10	0.12
2	0.16	0.13	0.12	0.11
4	0.13	0.12	0.14	0.11
6	0.12	0.11	0.11	0.10
8	0.12	0.12	0.09	0.10
10	0.09	0.10	0.10	0.11
(d)				
0	0.041	0.057	0.049	0.057
2	0.084	0.076	0.063	0.061
4	0.075	0.080	0.068	0.057
6	0.068	0.055	0.057	0.051
8	0.065	0.062	0.046	0.055
10	0.050	0.050	0.055	0.050
(e)				
0	0.005	0.010	0.007	0.013
2	0.020	0.019	0.012	0.017
4	0.023	0.023	0.017	0.009
6	0.022	0.012	0.013	0.012
8	0.013	0.016	0.006	0.013
10	0.012	0.011	0.006	0.009

that the tests are inclined to be too optimistic when the smaller variance component values are combined with a mixture of between- and within-block information. Examination of *p-p* plots constructed from the lower tail probabilities showed small but significant departures from the 1:1 line that were present in the center as well as the extremes of the simulated distributions.

4. Illustrative Examples

4.1 *Example* 1

Consider the example given by the first three columns of Table 2. This is a partially balanced, incomplete block design, with eight blocks of four plots and a single treatment factor having four levels. The response variable has been simulated from a normal distribution using variance components of 100 for between blocks (σ_1^2) and 1 for between plots within blocks (σ_2^2). The treatment factor has one contrast that is estimated in the block stratum, whose error term is estimated with 6 d.f., and two contrasts that are estimated in the plot stratum, whose error term has 22 d.f. For this example, REML estimates of the variance components are 202.0 and 0.74 for σ_1^2 and σ_2^2, respectively.

Table 2

Fixed-effect factors ($f1, f2, f3$) and random factors ($b1, b2, sb3, b3$) and test data ($y1, y2, y3$) for the three examples of calculating degrees of freedom for the Wald statistic

$f1$	$b1$	$y1$	$f2$	$b2$	$y2$	$f3$	$sb3$	$b3$	$y3$
1	1	−7.11	1	1	−1.84	1	1	1	−19.09
2	1	−7.12	2	1	−0.59	2	2	1	−25.29
1	1	−8.07	1	2	−2.08	1	3	2	−5.88
2	1	−5.63	2	2	−3.37	2	4	2	−11.90
1	2	13.26	1	3	−10.56	1	5	3	−1.82
2	2	12.85	2	3	−13.15	2	6	3	−6.88
1	2	14.50	1	4	12.56	1	7	4	−14.92
2	2	13.96	2	4	11.66	2	8	4	−10.80
1	3	9.98	1	5	3.65	1	9	5	5.47
2	3	8.57	2	5	3.69	2	1	5	7.84
1	3	9.00	1	6	−0.22	1	2	6	−0.03
2	3	10.02	2	6	−1.44	2	3	6	10.88
1	4	27.56	1	7	16.03	3	4	7	17.01
2	4	26.63	2	7	14.72	4	5	7	14.79
1	4	26.58	1	8	11.99	3	6	8	−3.09
2	4	28.07	2	8	9.73	4	7	8	−4.30
3	5	−14.24	3	9	6.79	1	8	9	8.43
4	5	−12.99	4	9	7.41	2	9	9	6.64
3	5	−14.46	3	10	−7.12	1	1	10	2.02
4	5	−15.75	4	10	−7.58	2	2	10	−8.06
3	6	8.41	3	11	−1.91	3	3	11	−2.62
4	6	5.82	4	11	−3.11	4	4	11	−5.18
3	6	7.15	3	12	7.75	3	5	12	−13.90
4	6	6.24	4	12	6.53	4	6	12	−16.70
3	7	20.52	3	13	2.41	3	7	13	−16.42
4	7	20.57	4	13	1.75	4	8	13	−15.37
3	7	20.70	3	14	−8.79	3	9	14	−7.84
4	7	19.47	4	14	−7.72	4	1	14	−5.55
3	8	7.39	3	15	−9.47	3	2	15	1.89
4	8	6.08	4	15	−9.95	4	3	15	13.12
3	8	8.72	3	16	5.34	3	4	16	−5.33
4	8	6.46	4	16	6.87	4	5	16	−6.37

The variances of the within-block contrasts have no contribution from the between-block component of variance, while the variance of the between-block contrast initially has a very small contribution from the within-block component of variance. However, when we form new contrasts by rotation, starting with one contrast in the block stratum and one contrast in the plot stratum, the new contrasts have variance component weights and degrees of freedom that are a function of the angle of rotation. The variance component weights and degrees of freedom, derived from $\boldsymbol{W}, \boldsymbol{W}_1, \boldsymbol{W}_2$, and $\text{cov}(\hat{\boldsymbol{\sigma}}^2)$ in Table 3, are uniquely equal at a rotation of 45° (Figure 1). After rotation, there are estimated to be 18.9 degrees of freedom.

It is no coincidence that the best angle of rotation turns out to be 45°; this is the angle that ensures that both variance expansions have equal coefficients for the between-block variance component. The contrast c_2 that is initially estimated between plots within blocks has $l_{21} = 0$ and, consequently, $c_2^{\mathrm{T}} \boldsymbol{W}_1 c_2 = 0$. Since $\boldsymbol{W}_1$ is symmetric and of full rank, $\boldsymbol{W}_1 c_2 = 0$, and so, from (9), $l_{11}(\theta) - l_{21}(\theta) = 0$ when $\cos(2\theta) = 0$, and hence $\theta = 45^\circ$. As there are just two variance components in this example, $l_{11}(\theta) - l_{21}(\theta) = 0$ immediately implies $l_{12}(\theta) - l_{22}(\theta) = 0$, and so this defines the rotation that minimizes (7). The occurrence of 45° as the optimal angle of rotation is due to the extremity of the example, with one initial contrast estimated entirely within blocks and the other initial contrast estimated entirely between blocks. Furthermore, $l_{11}(\theta)/\cos^2(\theta) = l_{21}(\theta)/\sin^2(\theta)$, explaining the symmetry that is evident in Figure 1.

The superficial conclusion from this example, namely that rotation leads to an averaging of the error degrees of freedom for the original contrasts, is erroneous. In fact, 18.9 is better seen as being effectively the greatest number of independent error contrasts that can be constructed with an equal contribution from the two strata. Realizing this, it is possible to construct the following example in which degrees of freedom are apparently additive.

Table 3

Intermediate outputs from Example 1 (on left) and Example 2 (on right). The fixed-effects matrix, $\boldsymbol{X}$, was set to have columns 1–4 corresponding to the overall mean and contrasts between factor levels (1, 2) vs. (3, 4), 3 vs. 4, and 1 vs. 2, respectively.

Estimated Covariance Matrix of Fixed Effects (W)

25.3	0	0	0	4.22	0	0	0
0	25.3	0	0	0	4.22	0	0
0	0	0.046	0	0	0	0.0404	0
0	0	0	0.046	0	0	0	0.0404

Derivatives of W with Respect to Variance Components for Blocks ($W1$) and Plots Within Blocks ($W2$)

$W1$

0.125	0	0	0	0.0624	0	0	0
0	0.125	0	0	0	0.0624	0	0
0	0.125	0	0	0	0	0	0
0	0	0	0	0	0	0	0

$W2$

0.0303	0	0	0	0.0312	0	0	0
0	0.0303	0	0	0	0.0312	0	0
0	0	0.0624	0	0	0	0.0628	0
0	0	0	0.0624	0	0	0	0.0628

Contrasts Scaled to Have Variance 1
Before ($c1$, $c2$) Rotation and After ($c1'$, $c2'$) Rotation

$c1$	$c2$	$c1'$	$c2'$	$c1$	$c2$	$c1'$	$c2'$
0	0	0	0	0	0	0	0
0.199	0	0.141	0.141	0.487	0	0.344	0.344
0	4.65	−3.29	3.29	0	4.97	−3.52	3.52
0	0	0	0	0	0	0	0

Covariance Matrix of Random Effects

	Blocks	Plots		Blocks	Plots
Blocks	13629.		Blocks	653.	
Plots	−0.0124	0.0496	Plots	−0.0299	0.0598

4.2 *Example* 2

Consider now the design given in columns 4–6 of Table 2, with 16 blocks of 2 plots, which is again of the form of a partially balanced, incomplete block design. The response y was generated using normal block and plot effects with variances 100 and 1, respectively. Again, one treatment contrast is estimated between blocks while the other two are estimated within blocks, and so the corresponding entries of $\boldsymbol{W}_1$ are zero (Table 3). This time, however, degrees of freedom for the between-block contrast and for any two within-block contrasts are equal. Indeed, a rotation in the plane of the block contrast and either of the within-block contrasts has the degrees of freedom for the rotated contrasts always equal, although the number of degrees of freedom varies as a function of the angle of rotation. In fact, the most satisfactory angle of rotation in terms of (7) is 45°, whereupon $l_1(\theta) = l_2(\theta)$ and the rotated contrasts have 28 degrees of freedom.

This example illustrates that equality of degrees of freedom is a necessary, but not sufficient, condition for equality of variance component weights. It also illustrates that the rotations are doing more than averaging the degrees of freedom. For this design, if we denote independent between-block error contrasts by $e_{b1} \cdots e_{b14}$ and independent within-block error contrasts as $e_{w1} \cdots e_{w14}$, then $e_{bi} \pm e_{wi}$, $i = 1 \cdots 14$, will give 28 independent error contrasts with equal variance, whose existence was implied by the denominator degrees of freedom associated with the best angle of rotation.

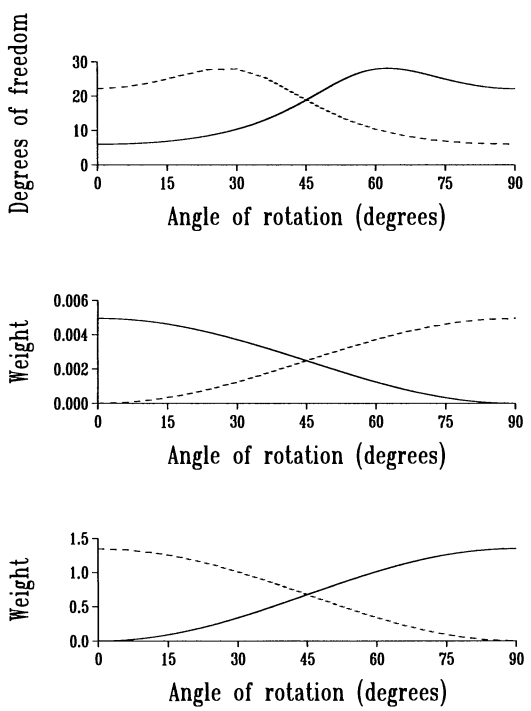

Figure 1. Effect of angle of rotation on contrasts estimated without rotation from between-block (solid line) and within-block (dashed line) information. Top panel, estimated degrees of freedom of contrast variance; central and bottom panels, coefficients of the between- and within-block variance components, respectively, in the expansion of contrast variances as a weighted sum of variance components.

4.3 *Example* 3

Finally, we consider the design of columns 7–10 of Table 2. As before, this has one fixed factor with four levels, but now there are two random factors. The response variable y has been simulated using normal random effects with variances 25, 100, and 1. The fixed-effect contrasts are estimated across strata. Normalizing, then rotating the contrasts (1 and 2 vs. 3 and 4) and (3 vs. 4), we find the coefficients in the expansion of the contrast variances are not all equal at the same angle. Minimizing equation (7) gives an optimal angle of 45.2° with contrast degrees of freedom 18.0 and 18.2. Repeating this minimization iteratively over pairs of contrasts, the solution reached after eight iterations has $l_1 = (0.0004, 0.0033, 0.4055)^{\mathrm{T}}$, $l_2 = (0.0037, 0.0030, 0.4055)^{\mathrm{T}}$, and $l_3 = (0.0026, 0.0031, 0.4032)^{\mathrm{T}}$, with 13.3, 13.7, and 13.7 degrees of freedom, respectively. This example confirms the earlier assertion that equality of variance component weights cannot always be achieved.

To test the suitability of the $F_{3,13}$ distribution for this Wald statistic, 199 simulations were conducted under the null hypothesis by simulating independent random effects and calculating the Wald statistic in each case. The distribution of Wald statistics has been compared against the X_3^2, $F_{3,6}, F_{3,13}$, and $F_{3,28}$ distributions (Table 4). These denominator degrees of freedom were selected as being the smallest in any one stratum, the estimated number for the Wald statistic, and total number available. The test statistics used are the Kolmogorov–Smirnov, Anderson–Darling, and Neyman smoothing goodness-of-fit tests (see, e.g., Pearson and Hartley, 1976, pp. 117–119, Table 54; Kendall and Stuart, 1967, p. 444). The test statistic values (Table 4) indicate that the X_3^2 and $F_{3,28}$ distributions are significantly different from the empirical distribution but that either $F_{3,6}$ or $F_{3,13}$ are reasonable candidates. Thus, the $F_{3,13}$ distribution suggested by the method of rotation appears satisfactory, although it is certainly not the only reference distribution consistent with the simulations.

5. Application

To demonstrate the use of rotations in practice, consider their application to the analysis of the area of home ranges of mountain hares (*Lepus timidus*) in Glenbuchat, Scotland (Hulbert et al.,

Table 4
Goodness of fit statistics, with critical values, for a comparison of the simulated distribution of Wald statistics and four theoretical distributions

	Statistic		
	Kolmogorov–Smirnov	Anderson–Darling	Neyman smooth
$F_{3,6}$	0.75	0.53	1.4
$F_{3,13}$	0.58	0.80	4.0
$F_{3,28}$	0.95	2.44	13.5
χ^2_3	1.31	5.40	30.1
Upper tail quantiles (%)			
10	1.07	1.93	7.8
5	1.22	2.49	9.5

1996). Home range areas were estimated after radio-tracking and transformed by taking square roots to obtain an effective diameter with better distributional properties. It was expected that the area of home range might be a function of food availability and hence of habitat, as each range could broadly be classified as one of moorland, forest, and pasture; also, the behavior pattern of animals depends on climatic and reproductive factors, both of which depend on the time of year. Twenty animals were tracked over a period of 2 years, giving rise to a total of 40 estimates of home range areas in different season by animal combinations. A REML analysis of the effective diameters, treating season and habitat as fixed effects and individual animal as a random effect, led to the two-way table of means in Table 5 and to estimates of 0.64 and 0.17 for the between- and within-animal variance components, respectively.

To assess the evidence for an interaction between season and habitat, we need to find a reference distribution for the appropriate Wald statistic whose value is 20.8. The main contribution to this comes from the three ranges on moorland in the postbreeding season, all of which are large. After the Choleski transformation to ensure the interaction contrasts are independent of the main effects, the four interaction contrasts have similar degrees of freedom, ranging from 13.5 to 15.7, but have variances that differ by a factor of three in the weights accorded to the between-animal variance component (Table 6a). After four iterations of the pairwise rotation algorithm, the degrees of freedom and weights in the variance expansions for each contrast are within 1% of the corresponding values for the other contrasts (Table 6b); in particular, the appropriate test for the Wald statistic is to compare $20.8/4 = 5.2$ against the reference distribution $F_{4,15}$, which leads to a p-value of 0.008. Hence, in contrast to Example 2, the appropriate denominator degrees of freedom are similar to what we would have guessed by simply averaging the degrees of freedom for the normalized contrasts.

6. Discussion

The use of contrasts combined with a suitable rotation provides a single estimate of the denominator degrees of freedom for the F approximation to the null distribution of the Wald statistic. Although finding the best rotation is an iterative process, each step in the iteration has a closed-form solution and there is no requirement to create exact equality among the coefficients of the variance

Table 5
REML estimates of mean effective diameters (100 *m*) *of mountain hare home ranges, classified by season and habitat*

	Habitat		
	Moorland	Forest	Pasture
Season			
Winter	3.65	3.42	2.46
Breeding	4.36	3.63	3.46
Postbreeding	6.07	3.87	3.26
Standard error of differences: 0.26–0.80			

Table 6

Weights of variance components for contrast variances and associated degrees of freedom for the season by breed interaction of the mountain hare range diameters: (a) after transformation to being independent of the main effect contrasts and to have identity covariance matrix; (b) after rotation to remove discrepancies of the weights of the variance components in the contrast variances

	Contrast number			
	1	2	3	4
(a) Before Rotation				
Weights of variance components in contrast variances				
Between animal	0.118	0.138	0.041	0.116
Within animal	5.376	5.301	5.667	5.384
Approximate d.f.	15.25	15.74	13.47	15.20
(b) After Rotation				
Weights of variance components in contrast variances				
Between animal	0.104	0.103	0.103	0.103
Within animal	5.430	5.432	5.432	5.434
Approximate d.f.	14.90	14.89	14.89	14.88

components. However, Example 2 demonstrated that there is little room for complacency; the initial contrasts had equal degrees of freedom but sticking at this point would have been wrong. The procedure for rotating contrasts outlined above, namely building up the appropriate rotation out of a series of rotations in the two dimensions of greatest discrepancy, is effective and simple to program. Despite the additional computational load required for the rotations, this method is considerably less computer intensive than that required for performing a model-based simulation, which is a distributionally robust but computer-intensive method for determining a reference distribution on a case-by-case basis.

The simulation study demonstrates that the F-distribution provides a reasonable, though not perfect, reference distribution for the Wald statistics, at least over the range of circumstances considered. As mentioned previously, the F-distribution can in general only be an approximation to the distribution of the Wald statistic, first, because the estimates of contrast variances do not have exact X^2-distributions and, second, because, even after rotation, the contrast variances may not be formed from exactly the same weighted sum of the variance components. An additional problem is caused by the estimated variance components being used to form weights for estimating the fixed effects without a corresponding increase in their variances. Furthermore, the coefficients in the expansion of the contrast variances are themselves estimated, as are the corresponding degrees of freedom. The combined effect of all of these shortcomings requires further research.

The three examples demonstrate the method for the case of one fixed-effect factor and either one or two random-effect factors. In the case of one random effect, we know that an exact solution exists and the rotations find it. Indeed, the results of Example 2 are counterintuitive in that they demonstrate a case where, while the degrees of freedom are split evenly between strata, after rotation, the contrast variances are estimated with degrees of freedom equal to the combined total across strata. In the case of two random effects, no exact solution exists but the rotations find a good approximation. Simulations in this case exclude some of the less conservative candidate distributions but are in reasonable accord with the F-distribution indicated by the best rotation. The application to the home range areas of hares also provides some comfort in that, after rotation, the contrast degrees of freedom are little changed despite a substantial initial discrepancy in the composition of the contrast variances.

Although other statistics may be introduced for testing fixed-effect factors, the Wald statistic is attractive because it exactly reproduces the analysis of variance tests for balanced designs. Because performance in small samples is the reason why REML is preferred over ML, the Wald statistic should only be ousted by other statistics that also coincide exactly with analysis of variance for balanced designs.

Acknowledgements

I thank Ian Hulbert for providing access to his data on mountain hares. Fabian Nabugoomu and two anonymous referees made many helpful comments on draft manuscripts. This work was supported by the Scottish Office of Agriculture, Environment and Fisheries Department.

Résumé

Des tests pour les effets fixes de modèles mixtes dans les cas déséquilibrés ont utilisé auparavant des tests de Student sur une base de contraste par contraste ou des statistiques de Wald sans une méthode universellement acceptée de calcul des nombres de degrés de liberté du dénominateur. Cette situation venait du fait que les variances des différents contrastes sont des sommes pondérées différemment des composantes de la variance avec des nombres de degrés de liberté qui ne sont pas nécessairement égaux. Un test F simultané pour des différences entre tous les niveaux d'un facteur fixe peut être obtenu en formant de nouveaux contrastes, par rotation des contrastes d'origine, avec des variances qui sont presque les mêmes sommes pondérées des composantes de la variance. Les degrés de liberté associés de ces nouveaux contrastes sont à peu près égaux. Une petite étude de simulation montre l'adéquation de l'approximation par un χ^2 pour la somme pondérée des composantes de la variance. Trois exemples simples sont utilisés pour démontrer les effets de la rotation. Le dernier de ces exemples est aussi utilisé pour comparer le test F simultané avec la distribution de la statistique de Wald obtenue par une simulation numérique. La méthode des rotations est alors utilisée à des données de tailles de domaines de lièvres des montagnes (*Lepus timidus*) pour mettre en évidence une interaction entre saison et habitat.

References

Berk, K. (1987). Computing for incomplete repeated measures. *Biometrics* **43,** 385–398.

Brown, H. K. and Kempton, R. A. (1994). The application of REML in clinical trials. *Statistics in Medicine* **13,** 1601–1617.

Gallo, J. and Khuri, A. I. (1990). Exact tests for the random and fixed effects in an unbalanced mixed two-way cross-classification model. *Biometrics* **46,** 1087–1095.

Giesbrecht, F. G. and Burns, J. C. (1985). Two-stage analysis based on a mixed model: Large-sample asymptotic theory and small-sample simulation results. *Biometrics* **41,** 477–486.

Hulbert, I. A. R., Iason, G. R., Elston, D. A., and Racey, P. A. (1996). Home-range sizes in a stratified upland landscape of two lagomorphs with different feeding strategies. *Journal of Applied Ecology* **33,** 1479–1488.

Kendall, M. G. and Stuart, A. (1961). *The Advanced Theory of Statistics,* Volume II. London: Griffin.

Khrui, A. I. and Littel, R. C. (1987). Exact tests for the main effects variance components in an unbalanced random two-way model. *Biometrics* **43,** 545–560.

McLean, R. A. and Saunders, W. L. (1988). Approximating degrees of freedom for standard errors in mixed linear models. *Proceedings of the Statistical Computing Section,* 50–59. New Orleans: American Statistical Association.

Patterson, H. D. and Thompson, R. (1971). Recovery of inter-block information when block sizes are unequal. *Biometrika* **58,** 93–108.

Pearson, E. S. and Hartlet, H. O. (eds). (1976). *Biometrika Tables for Statisticians,* Volume II. Cambridge, U.K.: Cambridge University Press.

Press, W. H., Flannery, B. P., Teukolsky, S. A., and Vetterling, W. T. (1986). *Numerical Recipies.* Cambridge, U.K.: Cambridge University Press.

Sattherthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin* **2,** 110–114.

Searle, S. R., Casella, G., and McCulloch, C. E. (1992). *Variance Components.* New York: Wiley.

Welham, S. J. and Thompson, R. (1997). Likelihood ratio tests for fixed model terms using residual maximum likelihood. *Journal of the Royal Statistical Society, Series B* **59,** 701–714.

Received April 1994; *revised April* 1997; *accepted November* 1997.