Paper 5088 -2020

A Warning about Wald Tests

David A. Dickey, NC State University

ABSTRACT

In mixed models, tests for variance components can be of interest. In some instances, several tests are available. For example, in standard balanced experiments like blocked designs, split plots and other nested designs, and random effect factorials, an F test for variance components is available along with the Wald test, Wald being a test based on large sample theory. In some cases, only the Wald test is available, so it is the default output when the COVTEST option is invoked in analyzing mixed models. However, one must be very careful in using this test. Because the Wald test is a large-sample test, it is important to know exactly what is meant by large sample. Does a field trial with 4 blocks and 80 entries (genetic crosses) in each block satisfy the "large sample" criterion? The answer is no, because, for testing the random block effects, it is the number of blocks (4) that needs to be large, not the overall sample size (320). Surprisingly it is not even possible to find significant block effects with the Wald test in this example, no matter how large the true block variance is. This problem is not shared by the F test when it is available as it is in this example. A careful look at the relationship between the F test and the Wald test is shown in this paper, through which the details of the above phenomenon is made clear. The exact nature of this problem, while important to practitioners is apparently not well known.

INTRODUCTION

This paper compares the F test for random effects in mixed models to the Wald test. In the MIXED procedure and its relatives (GLIMMIX, NLMIXED), SAS[1] software delivers the Wald test as an option. It gives no default tests for random effects. This is a reasonable strategy given the approximate nature of the Wald test and the fact that F is not available for some models. F is available, however, for many designed experiments. Both approaches, Wald and F, test the same hypothesis (0 variance component) but the tests can differ in their power to detect $\sigma^2>0$ and in their fidelity to the claimed false rejection rate. The F test has exactly the claimed null hypothesis rejection rate for normally distributed data in balanced experiments of any size and the Satterthwaite approximation, used when the data are not balanced, is quite accurate. The F tests have reasonably good power. In contrast, the Wald test is approximate in nature, requiring large samples to validate its use. Exactly what does "large sample" mean in the context of Wald? It is not the overall sample size that must be large as will be shown. An article by McNew and Mauromoustakos (1997) notes that the Wald test fails to work well in small samples but does not rigorously show why.

Consider a randomized complete block design with n=rt observations where r is the number of levels of the random block effect and t is the number of treatments. For example, in field trials of entries (genetic families) for a crop, it is not unusual to have a few fields, say r=4 with t=100 or more different entries in each field. The overall sample size, n=400, would usually be considered large, but in fact it is the number of random levels r=4, not the overall sample size n=400 that needs to be large. It is shown herein that the Wald test cannot exceed the square root of $(r-1)/2$, which is 1.22 in this example. With r=4, the Wald test CANNOT exceed the usual 5% critical value 1.645. This is true regardless of the true

[1] SAS is the registered trademark of SAS Institute, Cary, NC

block variance. On the other hand, as the ratio of block to error variance increases, the F test's power continues to increase. F is unbounded even with just 4 blocks.

The organization of this paper is as follows. In section 1 a trivial simple random sample is used to point out the advantage, in terms of bias, of REML estimation (the default in the MIXED procedures) over ordinary maximum likelihood. The example breaks down the ANOVA table sources into n individual one degree of freedom contrasts. This paper uses that methodology throughout. Section 2 reviews the Wald test. The test's construction simplifies nicely based on these one degree of freedom contrasts. The randomized complete block (RCB) case is illustrated in section 3 followed by a generalization of the Wald approach. Section 4 develops the key relationship between F and the Wald test when both are available. Section 5 shows a Monte Carlo illustration of the results. Readers not interested in the underlying math might go directly to Section 5. Section 6 explores more rigorously the relationship between F and the Wald test. It also indicates why the REML variance component estimate is sometimes the null hypothesis value 0. It gives insight on how to interpret the situation in which the Z test is missing. The remaining sections extend the results to several commonly encountered experimental designs including Latin squares, nested designs, split plots, and mixed factorials. Simulations in section 12 suggest that, for at least some cases, missing values may not change the results in any major way.

1. REML

The mixed model is $\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{Z}\boldsymbol{\gamma}+\mathbf{e}$ where $\boldsymbol{\gamma}$ and $\mathbf{e}$ are vectors of normal, mean 0 random variables, $\mathbf{X}$ and $\mathbf{Z}$ are known matrices of constants and $\boldsymbol{\beta}$ is a fixed effect parameter vector to be estimated. REML estimation reparameterizes the data into linear combinations, some of which involve the mean $\mathbf{X}\boldsymbol{\beta}$ and others that have mean 0 and involve only the random effects.

A trivial but informative example is a simple random sample {8, 12, 7, 13} from a $N(\mu,\sigma^2)$ distribution. The model is just $\mathbf{Y}=\mathbf{X}\boldsymbol{\beta}+\mathbf{e}$ where $\mathbf{X}$ is simply a column of 1s and $\boldsymbol{\beta}$ is the population mean μ. The ordinary likelihood is maximized by setting σ^2 equal to $((8-\mu)^2+(12-\mu)^2+(7-\mu)^2+(13-\mu)^2)/4$ which, being an average of $(Y-\mu)^2$ terms, is an unbiased estimate of $\sigma^2=E\{(Y-\mu)^2\}$. This result is impractical because the population mean μ is usually unknown. With μ unknown, the likelihood function is maximized by substituting the sample average for μ. The sample average fits the sample better, in terms of squared deviations, than any other number including the true mean μ. As a result, this substitution of the sample mean produces a downward bias in the variance estimate. The bias is removed if the sum of squared deviations from the sample mean is divided by $n-1$ instead of n where n is the sample size. Beginning courses in statistics typically present only the unbiased estimate. In our example the sample average is 10, the sum of squared deviations is SSq=26, the biased estimate is SSq/n=26/4=6.50 whereas the more common unbiased estimate is SSq/(n-1)=26/3=8.67, a 33% increase.

The REML method uses sets of orthogonal linear combinations to estimate variance components. There are several equivalent sets. We present two of these evaluated for the sample above. The first is

$$L=0.5\begin{pmatrix}1&1&1&1\\-1&-1&1&1\\-1&1&-1&1\\1&-1&-1&1\end{pmatrix}\begin{pmatrix}Y_1\\Y_2\\Y_3\\Y_4\end{pmatrix},\ L\sim N\left(\begin{pmatrix}2\mu\\0\\0\\0\end{pmatrix},\begin{pmatrix}\sigma^2&0&0&0\\0&\sigma^2&0&0\\0&0&\sigma^2&0\\0&0&0&\sigma^2\end{pmatrix}\right).$$

In this vector **L** of linear combinations, the last 3 have known mean 0. For this sample,

$$\mathbf{L}=0.5\begin{pmatrix}1&1&1&1\\-1&-1&1&1\\-1&1&-1&1\\1&-1&-1&1\end{pmatrix}\begin{pmatrix}8\\12\\7\\13\end{pmatrix}=\begin{pmatrix}20\\0\\5\\1\end{pmatrix}.$$

An equally valid representation is given by

$$L=0.5\begin{pmatrix}1&1&1&1\\-\sqrt{2}&\sqrt{2}&0&0\\0&0&-\sqrt{2}&\sqrt{2}\\-1&-1&1&1\end{pmatrix}\begin{pmatrix}Y_1\\Y_2\\Y_3\\Y_4\end{pmatrix},\ L\sim N\left(\begin{pmatrix}2\mu\\0\\0\\0\end{pmatrix},\begin{pmatrix}\sigma^2&0&0&0\\0&\sigma^2&0&0\\0&0&\sigma^2&0\\0&0&0&\sigma^2\end{pmatrix}\right).$$ In our sample

$$\mathbf{L}=0.5\begin{pmatrix}1&1&1&1\\-\sqrt{2}&\sqrt{2}&0&0\\0&0&-\sqrt{2}&\sqrt{2}\\-1&-1&1&1\end{pmatrix}\begin{pmatrix}8\\12\\7\\13\end{pmatrix}=\begin{pmatrix}20\\2\sqrt{2}\\3\sqrt{2}\\0\end{pmatrix}$$

In both cases, the squares of the last 3 elements of **L** sum to 26. These three elements have known population mean $\mu=0$ so their sample mean is irrelevant. Their average square, 26/3, is the previously seen unbiased estimate of σ^2. Orthogonality of the contrasts ensures uncorrelated variables and hence independence for normal data.

The displayed contrasts are orthonormal but orthogonality is sufficient. Any contrast $L=\sum_{i=1}^{n} c_i Y_i$ with $\sum_{i=1}^{n} c_i = 0$ can be normalized by replacing c_i by $c_i/\sqrt{\sum_{j=1}^{n} c_j^2}$. The one degree of freedom sum of squares is $L^2/\sum_{i=1}^{n} c_i^2$, as is well known. REML ignores the first linear combination, which is used to estimate the mean, then takes advantage of the fact that the last n=3 linear combinations have known mean $\mu=0$. The unbiased variance estimate is (0+25+1)/3 or (8+18+0)/3 which is 26/3 in either case.

The example here simply shows the advantage of REML estimation in terms of bias. The REML estimate is unbiased in a simple random sample. REML variance parameter estimates are less biased than maximum likelihood estimates. REML estimates may not be exactly unbiased in general. The current example is too simple to be practical. The reason for looking at the Wald test in a simple random sample is to set the stage for the later, more practical, sections of this paper. The next section continues using the simple example.

2. WALD TESTS

The Wald test divides a parameter estimate by an estimated standard error. This estimated standard error is a function of the second derivative of the likelihood. The objective function -2 (REML log likelihood) is based on linear contrasts L_i with known means 0. There are n-k of these where k is the rank of the design matrix $\mathbf{X}$. In a simple random sample of size n, $\mathbf{X}$ is just a column of 1s, $\mathbf{X}=\mathbf{1}$, and n–k=n–1. The variance matrix of the L contrasts is, as shown in section 1, the diagonal matrix $\sigma^2\mathbf{I}_{(n-1)x(n-1)}$. Using n–1 orthonormal contrasts with known means 0, the objective function and its first two derivatives in a normal simple random sample with n observations are:

$$-2\text{ (REML log likelihood)} = (n-1)\ln(2\pi)+(n-1)\ln(\sigma^2)+SSq/\sigma^2 \text{ where } SSq=\sum_{i=1}^{n-1} L_i^2 .$$

$$\text{First derivative with respect to } \sigma^2\text{: } \frac{(n-1)}{\sigma^2}-\frac{SSq}{\left(\sigma^2\right)^2}$$

$$\text{Second derivative: } \frac{-(n-1)}{\left(\sigma^2\right)^2}+2\frac{SSq}{\left(\sigma^2\right)^3}$$

The first derivative set to 0 gives the unbiased estimate $\hat{\sigma}^2=SSq/(n-1)$ for a simple random sample. Inserting $\hat{\sigma}^2$ for σ^2 and $(n-1)\hat{\sigma}^2$ for SSq in the second derivative gives the Hessian matrix H. In this case, H has just one element, $H=\frac{-(n-1)}{\left(\hat{\sigma}^2\right)^2}+2\frac{(n-1)\hat{\sigma}^2}{\left(\hat{\sigma}^2\right)^3}=\frac{(n-1)}{\left(\hat{\sigma}^2\right)^2}$, a Wald variance $2H^{-1}=\frac{2\left(\hat{\sigma}^2\right)^2}{n-1}$, and thus a Wald test $Z=\sqrt{\frac{n-1}{2}}$ which, surprisingly and unfortunately, does not even involve the data!! This means that the "test" will give the same decision regardless of how large is the sample variance. Here we have simply followed the Wald formula, without regard to assumptions. This whole article is an illustration of the importance of understanding and checking assumptions. The reader can run the following SAS program to confirm the Z result above. The COVTEST option produces the Wald test:

```
Data A;
   do n=4, 9, 19; do rep = 1 to 3;
     do i=1 to n;
            Y= 1000 + 100*normal(12375);
     output; end;  end; end;
ods output CovParms=Wald;
PROC MIXED data=A COVTEST;
    by n rep;
    model Y= ;
PROC PRINT data=Wald;
  var n Zvalue probZ;
run;
```

The output below matches our theory in that $Z^2=(n-1)/2$ in every case regardless of the observations. Because $2(1.645)^2+1=6.41$ the level 0.05 Wald Z test will never reject the 0 variance hypothesis if n<7 even though we know the variance is positive when the observations differ. The 5% Wald test has no power if n<7.

Obs	n	ZValue	ProbZ
1	4	1.22	0.1103
2	4	1.22	0.1103
3	4	1.22	0.1103
4	9	2.00	0.0228
5	9	2.00	0.0228
6	9	2.00	0.0228
7	19	3.00	0.0013
8	19	3.00	0.0013
9	19	3.00	0.0013

3. THE RANDOMIZED COMPLETE BLOCK CASE

Our first example of a design of practical interest is the randomized complete block (RCB) design with r replicates in blocks. Each block contains responses for t treatments for a total of n=rt observations. Block effects are usually random. The assumption of no interaction between the blocks and treatments allows for an estimated error variance and hypothesis tests.

Consider r=4 blocks and t=3 levels of a fixed effect treatment. Assuming random block effects, our interest centers on the block and error variance components σ^2_{Block} and σ^2. The model is $Y_{ij}=\mu+\tau_i+B_j+e_{ij}$, with treatment levels i=1,2,3, block levels j=1,2,3,4, $B_j \sim N(0, \sigma^2_{\text{Block}})$ and $e_{ij} \sim N(0,\sigma^2)$. All random terms are independent of each other by assumption. We take a similar approach to that of the simple random sample case with Display 1 below showing one choice of linear combinations or "contrasts." With response vector $\mathbf{Y}$, the first three elements $\mathbf{C}_i'\mathbf{Y}$ of the vector $\mathbf{L}=\mathbf{C}\mathbf{Y}$ in Display 1 estimate fixed effects. Here $\mathbf{C}_i'$ is the i^{th} row of the contrast matrix $\mathbf{C}$. The next three contrasts $\mathbf{C}_i'\mathbf{Y}$, each with known mean 0, are linear combinations of block and error effects. The last six have known mean 0 and are linear combinations of the errors. The display breaks the data down completely into individual one degree of freedom contrasts. The error contrasts are free of fixed effects and block effects because their coefficients sum to 0 within each block and treatment. The last 9 contrasts come from distributions with known mean 0 and are independent. By definition of REML, they constitute the inputs to the REML likelihood, just as before. For each contrast, the quantity $(C_i'Y)^2/(C_i'C_i)$ is called the contrast "sum of squares," despite it being a sum of only one item. The contrasts are chosen to split into a block group and an independent error group. The sum of these 1 df contrast sums of squares is the ANOVA sum of squares.

$$\mathbf{CY} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & 0 & 1 & -1 & 0 & 1 & -1 & 0 & 1 & -1 & 0 & 1 \\ -1 & 2 & -1 & -1 & 2 & -1 & -1 & 2 & -1 & -1 & 2 & -1 \\ -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 \\ 1 & 0 & -1 & 1 & 0 & -1 & -1 & 0 & 1 & -1 & 0 & 1 \\ 1 & 0 & -1 & -1 & 0 & 1 & 1 & 0 & -1 & -1 & 0 & 1 \\ 1 & 0 & -1 & -1 & 0 & 1 & -1 & 0 & 1 & 1 & 0 & -1 \\ 1 & -2 & 1 & 1 & -2 & 1 & -1 & 2 & -1 & -1 & 2 & -1 \\ 1 & -2 & 1 & -1 & 2 & -1 & 1 & -2 & 1 & -1 & 2 & -1 \\ -1 & 2 & -1 & 1 & -2 & 1 & 1 & -2 & 1 & -2 & 2 & -1 \end{pmatrix} \begin{pmatrix} Y_{11} \\ Y_{21} \\ Y_{31} \\ Y_{12} \\ Y_{22} \\ Y_{32} \\ Y_{13} \\ Y_{23} \\ Y_{33} \\ Y_{14} \\ Y_{24} \\ Y_{34} \end{pmatrix} = \begin{pmatrix} L_{11} \\ L_{21} \\ L_{31} \\ L_{12} \\ L_{22} \\ L_{32} \\ L_{13} \\ L_{23} \\ L_{33} \\ L_{43} \\ L_{53} \\ L_{63} \end{pmatrix}$$

Display 1. RCB with fixed effect contrasts L_{k1}, random block effect contrasts L_{k2} and error effect contrasts L_{k3}.

The linear contrast L_{43} has coefficients computed as in Table 1 below. The row labels show the L_{31} treatment contrast coefficients, -1, 2, and -1. The column labels show the L_{12} block coefficients -1, -1, 1, and 1. The table entries, the coefficients in the error contrast L_{43}, are the products of the row and column labels. Note that each block coefficient is repeated for all treatments within the block and each treatment coefficient is repeated r times, once per block, in **CY** above. By construction, the block coefficients and error coefficients (the Table 1 entries) sum to 0 within each row (each treatment level). The treatment coefficients and error coefficients sum to 0 within each column (each block level). Notice that L_{43} has known mean $\mu=0$. Compute it as though estimating a block by treatment contrast. The assumption of no true block by treatment interaction ensures that the associated contrast is a function only of e_{ij}. It has mean 0 and its square is an unbiased estimate of the error variance σ^2. There are $(r-1)(t-1)=(3)(2)=6$ such contrasts L_{k3} , $k=1,2,\ldots,6$, available from such tables, where 6 is also the ANOVA table error degrees of freedom.

Block Treatment	**–1**	**–1**	**1**	**1**
–1	1	1	–1	–1
2	–2	–2	2	2
–1	1	1	–1	–1

Table 1. Contrast coefficients for L_{43} in Display 1 (read down the columns)

Some details follow. Because the row (column) labels sum to 0 and are repeated within each block (treatment), the block and treatment contrasts are uncorrelated and thus are independent for normal data. Because the table entries sum to 0 across the rows and down the columns, the block and treatment effects are multiplied by 0 in each error contrast. This shows that error contrast L_{43} is independent of all row and column contrasts. Because the

row, column, and error contrasts are also independent within groups, all the contrasts are independent of each other. These last 3+6=9 contrasts are free of fixed effects and thus have means 0. By choosing contrasts in this manner, any RCB analysis breaks down into independent one degree of freedom contrasts. A more mathematically formal development of these ideas follows.

The model for a RCB is $Y_{ij}=\mu+\tau_t+B_j+e_{ij}$, i=1,2,…,t ; j=1,2,3,…,b with $B_j \sim N(0,\sigma^2_{Block})$ and $e_{ij} \sim N(0,\sigma^2)$. Let c_{ijk} represent the coefficients in row k of the contrast matrix **C** in Display 1. For rows k=1, 2, and 3 the contrasts take the form $L_{k1}=\sum_{i=1}^{t}\sum_{j=1}^{b} c_{ijk}(\mu+\tau_i+B_j+e_{ij}) = \sum_{i=1}^{t}\sum_{j=1}^{b} c_{ik}(\mu+\tau_i+e_{ij})$. Notation c_{ik} indicates that the coefficients c_{ijk} in L_{k1} change only with i. Because $\sum_{i=1}^{t} c_{ijk}=0$ for all j and k=2 or 3, we have $\sum_{j=1}^{b}(\sum_{i=1}^{t} c_{ijk})(\mu+B_j)=0$ so, for k=2 or 3, treatment contrast L_{k1} does not involve blocks or μ. The same argument shows that, for the block contrasts, we have $L_{k2} = \sum_{i=1}^{t}\sum_{j=1}^{b} c_{ijk}(\mu+\tau_i+B_j+e_{ij})=\sum_{i=1}^{t}\sum_{j=1}^{b} c_{jk}(B_j+e_{ij})$, an expression not involving any τ_i or μ. The expected square of block contrast L_{k2} is (by independence of B and e), $\sum_{j=1}^{b} E\left\{\sum_{i=1}^{t} c_{jk}\left(B_j+e_{ij}\right)\right\}^2 = t^2\sum_{j=1}^{b} c_{jk}^2\sigma^2_{Block}+\sum_{i=1}^{t}\sum_{j=1}^{b} c_{jk}^2\sigma^2 = (t\sigma^2_{Block}+\sigma^2)\sum_{i=1}^{t}\sum_{j=1}^{b} c_{jk}^2$. Dividing by $\sum_{i=1}^{t}\sum_{j=1}^{b} c_{jk}^2 = t\sum_{j=1}^{b} c_{jk}^2$ to get the contrast sum of squares, we see that each squared block contrast has expected value $t\sigma^2_{Block}+\sigma^2$. Recall that $t\sigma^2_{Block}+\sigma^2$ is the expected mean square for blocks from a standard ANOVA table.

The error contrasts have the form $L_{k3} = \sum_{i=1}^{t}\sum_{j=1}^{b} c_{ijk}(\mu+\tau_i+B_j+e_{ij})=\sum_{i=1}^{t}\sum_{j=1}^{b} c_{ijk}e_{ij}$, a function of the error terms only. This equality follows from the fact that the sum on i of the coefficients on B_j and the sum on j of the coefficients on τ_i are both 0. The block and error contrasts are independent because $E\left\{\left(\sum_{i=1}^{t}\sum_{j=1}^{b} c_{jk}(B_j+e_{ij})\right)\left(\sum_{i'=1}^{t}\sum_{j'=1}^{b} c_{i'j'k}e_{i'j'}\right)\right\} = \sigma^2\sum_{i=1}^{t}\sum_{j=1}^{b} c_{jk}c_{ijk'}$ which (from $\sum_{i=1}^{t} c_{ijk'}=0$) is 0 for every j by construction. Note that $E\{e_{ij}e_{i'j'}\}=\sigma^2$ only when $i=i'$ and $j=j'$ and is 0 otherwise.

4. A GENERAL PLAN FOR REML ANALYSIS

In Display 1, the contrast vector is $\mathbf{L}=\mathbf{CY}$. The rows C_i' of **C** were carefully chosen with certain subsets of coefficients summing to 0 to give independence. Rows 4-6 of **CY** show independent linear block contrasts L_i, each with a 1 df sum of squares whose expected square is $V=t\sigma^2_{Block}+\sigma^2$. Rows 7-12 show (r–1)(t–1) error contrasts, the square of each estimating $V=\sigma^2$. Contrasts were carefully chosen so that within a treatment or a block level, the c_{ijk} coefficients are either constant or they sum to 0. Notice that (r–1) and (t–1)(r–1) are ANOVA degrees of freedom, symbolized df in the formulas below. By independence of these contrasts, –2(REML log likelihood) becomes the sum of two separately maximizable parts, a block part and an error part. Each part has the form and derivatives given in formulas (1) below:

-2 (REML log likelihood) = $df(\ln(2\pi))+df(\ln(V))+SSq/V$ where $SSq=\sum_{i=1}^{df}(C_i'Y)^2/(C_i'C_i)$ (1)

First derivative: $\frac{df}{V}-\frac{SSq}{V^2}$

Second derivative: $\frac{-df}{V^2}+2\frac{SSq}{V^3}$

Note that df is both the number of contrasts and the ANOVA degrees of freedom for the source in question. The first derivative set to 0 gives the estimate $V=SSq/df$. This is the usual ANOVA table mean square for the source. Inserting this for V in the second derivative gives a Hessian $H=\frac{-df}{V^2}+2\frac{(df)(V)}{V^3}=\frac{df}{V^2}$, and a Wald variance $2H^{-1}=\frac{2V^2}{df}$.

Here V might estimate a single variance component or a linear combination of variance components given by the ANOVA expected mean square for the source. This approach will be applicable to many experimental designs. Whenever V estimates the error variance component σ^2, the Wald test will be $\sqrt{df_{error}/2}$ regardless of the data.

A common situation in experimental designs is that in which the usual unbiased variance component estimate in the ANOVA table is of the form $(MS_N-MS_D)/c$ where $F= MS_N/MS_D$ and MS_N and MS_D are independent ANOVA mean squares (by Cochran's Theorem). Here $E\{(MS_N-MS_D)/c\}$ is the variance component of interest. The first derivative in formula (1) shows that ANOVA table mean squares MS_N and MS_D are also REML estimates. The mean squares have degrees of freedom df_N and df_D. If the variance component in question is 0, the ratio $F=MS_N/MS_D$ is a ratio of independent Chi-square variables so, with normal data, this F has exactly an F distribution under the 0 variance component hypothesis. F is the standard ANOVA F test for a 0 variance component. Recall that in PROC GLM, a RANDOM/TEST statement is required to get F tests with proper denominators in models with multiple variance components.

The Wald variance of $(MS_N-MS_D)/c$, by formula (1), is $\frac{2MS_N^2}{c^2df_N}+\frac{2MS_D^2}{c^2df_D}=\left(\frac{2MS_D^2}{c^2df_N}\right)\left(F^2+\frac{df_N}{df_D}\right)$

and the variance component estimate is $\left(\frac{MS_N-MS_D}{c}\right)=\frac{MS_D}{c}(F-1)$ where $F=MS_N/MS_D$.

The Wald test Z thus satisfies formula (2) below:

$$Z=\sqrt{\frac{df_N}{2}}\frac{(F-1)}{\sqrt{F^2+df_N/df_D}}\ ,\ F\geq 1 \qquad (2)$$

Z = missing otherwise

The REML likelihood function is defined on the domain $\sigma^2_{Source}\geq 0$ for all random effect sources. The reason for the $F\geq 1$ caveat is that, when $MS_N-MS_D<0$ (i.e. $F<1$) the REML likelihood is maximized at the boundary of its domain, resulting in a 0 variance component

estimate, no derivative, and thus no standard error. SAS prints out a dot for Z to indicate a missing value whenever a variance component estimate hits the domain boundary 0. Almost certainly the estimation was stopped at the boundary when its iterative algorithm's next step would have taken the estimate into the negative numbers. This suggests that, for cases with both Wald and F available, a missing Z is associated with F<1 and indicates insignificant evidence of a positive variance component. Importantly, this gives an interpretation to an otherwise uninterpretable situation.

For the RCB design, the estimate $(MS_N - MS_D)/c$ is unbiased but is negative whenever F<1. Because REML replaces negative values by 0, the REML estimate cannot also be unbiased, illustrating the previous comment that REML estimates are not always unbiased.

We return to analyzing the block design before applying REML contrasts and formula (2) to other designs. For the error variance, recall that the Wald test is $Z=\sqrt{df_{error}/2}$.

The block variance is $\sigma^2_{Block}=((t\sigma^2_{Block}+\sigma^2)-\sigma^2)/t$. Its estimate is $(MS_{Block}-MS_{error})/t$ $= MS_{error}(F_{Block}-1)/t$ so formula (2) applies giving $Z=\sqrt{\frac{r-1}{2}}\frac{(F_{Block}-1)}{\sqrt{F^2_{Block}+\frac{(r-1)}{(r-1)(t-1)}}}=$ $\sqrt{\frac{r-1}{2}}\frac{(F_{Block}-1)}{\sqrt{F^2_{Block}+\frac{1}{(t-1)}}}$ for $F_{Block} \geq 1$. If $F_{Block}< 1$ the block variance estimate is 0 and Z is missing because the REML likelihood function is restricted to the domain $\sigma^2 \geq 0$. Recall that the exception for F<1 applies to all examples herein.

5. EMPIRICAL CHECK AND DEMONSTRATION

As an illustration and numeric check on the relationship between the Wald Z and F tests, we generate 2000 data sets in SAS. Without loss of generality we use error variance 1. Four different block to error standard deviation ratios sd are used. The graphs show the superior power of F compared to the Wald test.

Each data set has data from a randomized complete block design with r=19 blocks and t=15 treatments. There are n=rt=285 observations in each data set. Our theory shows that the Wald Z for blocks is bounded by $\sqrt{(19-1)/2}=3$. The (approximate) Wald Z and the (exact) ANOVA F test for blocks are computed for each sample. In Figure 1 each panel uses one ratio of standard deviations, block to error. Reading left to right within each row the ratios are 0.25, 0.5, 1 and 2. Variance ratios are thus 1/16, 1/4, 1, and 4.

On each vertical axis is the calculated block F statistic, truncated at F=75 to retain reasonable resolution in the interesting parts of the graphs. A horizontal reference line appears at the 5% critical F. Data sets producing (red or blue) points above this line have F tests that reject the zero block variance hypothesis. These are correct rejections of the (false) null hypothesis. The horizontal (X) coordinate of each point is the corresponding Wald Z for that data set. There are two vertical reference lines. The leftmost thin line marks the 95th percentile, 1.645, of the N(0,1) asymptotic distribution of Z. Points (red) lying to the right of this critical value represent rejections by Z of the 0 variance component

hypothesis at the 5% level. As has been shown analytically, all points lie to the left of the thicker red reference line at $\sqrt{(r-1)/2}=3$.

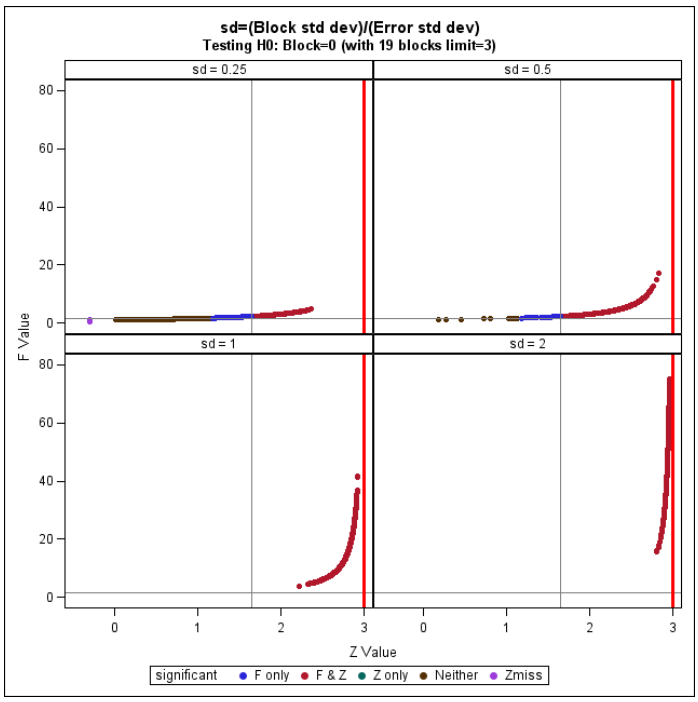

Figure 1: F test statistic versus Wald Z. Ratios of block to error standard deviations are sd=0.25, 0.5, 1, and 2. There are r=19 blocks and t=15 treatment levels.

Points in the upper left quadrant formed by the F and Z critical value reference lines are blue. They represent cases with F significant but not Z. Dark brown points, to the left and below these are cases in which neither Z nor F is significant. Some Wald Zs are missing (Zmiss) when the ratio of standard deviations is 0.25. The following table gives the counts. When missing Zs occur, as in the upper left quadrant, a point is displayed in purple at Z= −0.3 to so indicate. There is no point in the bottom right quadrant. There is no case with only Z significant.

	std deviation ratio				
p<0.05	0.25	0.5	1	2	Total
F & Z	571	1926	2000	2000	6497
F only	690	62	0	0	752
Zmiss	109	0	0	0	109
neither	630	12	0	0	642
Total	2000	2000	2000	2000	8000

Table 2. Significance counts for F and Z (RCB, 19 blocks).

When the ratio of standard deviations is 0.25 we see that there are 109 out of 2000 cases in which the Z statistic is missing ($F_{Block}<1$). For these, the search for a likelihood maximizing

value has been stopped at the domain boundary. With 19 blocks we see 690 and 62 cases, respectively (for the two lowest sd values) in which F is significant but Z is not.

Rerunning the simulation with r=6 blocks, the Z limit is $\sqrt{5/2}=1.58$ so Z cannot exceed 1.645 and thus cannot be significant at the 5% level. Figure 2 illustrates this case. The red boundary line is now to the left of the critical Z value showing that Z cannot be significant even if the block variance is much larger than the error variance. The F statistic, however, has reasonable power. When the block variance is 4 times the error variance (bottom right panel) the empirical power is 100% for F, 0 for the Wald Z statistic.

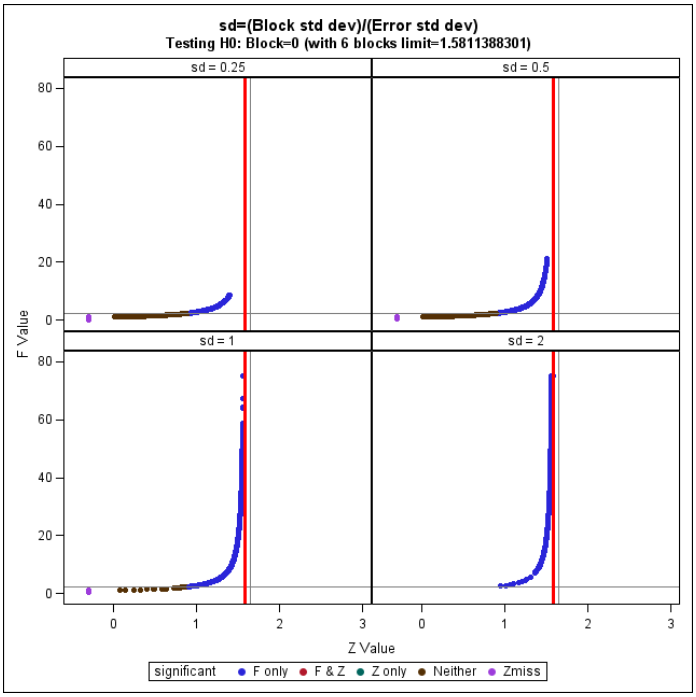

Figure 2: F test statistic versus Wald Z. Ratios of block to error standard deviations are sd=0.25, 0.5, 1, and 2 and there are r=6 blocks.

The corresponding table of counts appears below.

	std deviation ratio				
P<0.05	0.25	0.5	1	2	Total
F only	621	1539	1963	2000	6123
Neither	885	381	32	0	1298
Zmiss	494	80	5	0	579
Total	2000	2000	2000	2000	8000

Table 2. Significance counts for F and Z (RCB, 6 blocks).

6. MORE ON THE RELATIONSHIP OF F AND Z

The relationship

$$Z_{Block} = \sqrt{\frac{(r-1)}{2}} \frac{(F_{Block}-1)}{\sqrt{F_{Block}^2 + \frac{1}{(t-1)}}} \quad \text{for } F_{Block} \geq 1$$

is a monotone continuous relationship between F_{Block} and Z when $F_{Block} \geq 1$. An $F_{Block}<1$ is not significant at any reasonable level. In Figure 2, when sd = 0.25, there are 1379 cases with F_{Block} insignificant including 494 associated with missing Zs. Every missing Z is associated with $F_{Block}<1$ and vice versa.

Further light is cast on the missing Z problem by noting that, in a randomized complete block design, the variance structure is an nxn block diagonal matrix (n=tr). It has t by t compound symmetric submatrices on the diagonal. Such a submatrix for t=4 treatments is

$$\begin{pmatrix} \sigma^2_{Block}+\sigma^2 & \sigma^2_{Block} & \sigma^2_{Block} & \sigma^2_{Block} \\ \sigma^2_{Block} & \sigma^2_{Block}+\sigma^2 & \sigma^2_{Block} & \sigma^2_{Block} \\ \sigma^2_{Block} & \sigma^2_{Block} & \sigma^2_{Block}+\sigma^2 & \sigma^2_{Block} \\ \sigma^2_{Block} & \sigma^2_{Block} & \sigma^2_{Block} & \sigma^2_{Block}+\sigma^2 \end{pmatrix} \text{ and is } \begin{pmatrix} a+b & b & b & b \\ b & a+b & b & b \\ b & b & a+b & b \\ b & b & b & a+b \end{pmatrix} \text{ in general.}$$

In addition to thinking of a randomized complete block (RCB) design as a two way (block by treatment) structure this gives an alternate representation as a one way analysis with a compound symmetric block diagonal variance matrix. A notable restriction in the RCB design is that all elements are nonnegative. In general a compound symmetric matrix is $\mathbf{V}= a\mathbf{I}+b\mathbf{1}\mathbf{1}'$ as shown in the rightmost matrix above, where $\mathbf{I}$ is the identity matrix and $\mathbf{1}$ is a column of 1s. Any t by t compound symmetric matrix $\mathbf{V}$ satisfies $\mathbf{V1} = (a+tb)\mathbf{1}$ showing that $\mathbf{1}$ is an eigenvector with eigenvalue (a+tb). We can find t−1 mutually orthonormal eigenvectors (orthogonal contrast columns) $\boldsymbol{\lambda}$ with $\mathbf{1}'\boldsymbol{\lambda}=0$. For each of these, $\mathbf{V}\boldsymbol{\lambda} = a\boldsymbol{\lambda}+b\mathbf{1}\mathbf{1}'\boldsymbol{\lambda}=a\boldsymbol{\lambda}+(b\mathbf{1})(\mathbf{1}'\boldsymbol{\lambda})=a\boldsymbol{\lambda}+(b\mathbf{1})(0)=a\boldsymbol{\lambda}$. $\mathbf{V}$ is a valid covariance matrix as long as all eigenvalues, (a+tb) and a, are nonnegative. We only need $b \geq -a/t$. In contrast we must have $b=\sigma^2_{Block} \geq 0$ in a RCB design. Every randomized complete block design has a block diagonal compound symmetric variance structure, but not every block diagonal compound symmetric variance structure can be associated with a randomized complete block design.

In SAS PROC MIXED for example, the RANDOM BLOCK statement assumes the design is a randomized complete block design with $b=\sigma^2_{Block} \geq 0$. Its REML likelihood exists only on the domain of possible variances, sometimes leading to estimates of σ^2_{Block} that are 0, and missing Wald Z tests as seen above. This is most likely to happen when the number of blocks and the block variance are small. Had REPEATED / SUBJECT=BLOCK been used rather than the RANDOM statement, there would be no such restriction on the domain of the REML likelihood. Doing this, the search always converges successfully to parameter estimates that occasionally imply negative b values off the diagonal.

Negative covariances do occur in practice. Consider inoculating birds within a cage, where cage is supposedly a blocking factor. If birds within each cage are inoculated with different

growth stimulants (treatments), the natural within cage competition for food can produce a negative covariance in weight gains. This means the correct analysis is not that of a RCB. The experimenter intended this to be a RCB design but the appropriate model is a one way ANOVA with compound symmetric covariance matrix.

7. NESTED VARIANCE COMPONENT MODEL

There are several designs related to the RCB for which the REML contrast approach yields an informative relationship between the ANOVA F test and the Wald test. A RCB with no treatments becomes a nested design with two variance components. In genetics, the blocks might be family effects, the errors might be sibling effects, and the response might be intelligence in humans or weight gain in animals. Such a trait would be highly heritable if the intraclass correlation coefficient $\frac{\sigma^2_{\text{family}}}{\sigma^2_{\text{family}}+\sigma^2_{\text{sibling}}}$ (which is used as a simple heritability coefficient in genetics) were large. That is, a trait that varies quite a bit from family to family but varies little within families appears to be inherited and gives a large heritability coefficient. The design is called a nested design with siblings nested within families. Estimation and testing of variance components is the main goal in such an experiment.

A small modification of Display 1 serves to illustrate this situation. In Display 1, rows 2 and 3 of $\mathbf{CY}$ were used to compute treatment effects. Without treatment effects these contrasts, which are still independent of the r–1 block (now family) contrasts, would be additional contrasts in the among sibling effects (errors) that are independent of family effects (blocks) by construction. Now there are r(t–1) error contrasts, coming from the (t–1) siblings in each of r families, rather than (r–1)(t–1) RCB error contrasts. The expected mean squares from ANOVA still provide the expectations for each squared contrast and thus $(MS_{Family}–MS_{Sibling})/t$ is still the estimate of the family (block) variance. Applying formula (2) with $df_N = (r–1)$ and $df_D = r(t–1)$ we have, when $F_{family} \geq 1$, $Z = \sqrt{\frac{r-1}{2}}\,\frac{(F_{\text{family}}-1)}{\sqrt{F^2_{\text{family}}+\frac{r-1}{r(t-1)}}}$. The (r–1)/r factor arises when t siblings are nested within each of r families.

8. PSEUDOREPLICATION, HIERARCHICAL MODELS, & GENERAL FORM

Suppose, in the RCB example above, the r=4 blocks represent r pots each with t plants. With no treatments this is just another nested design like the heritability one. If, however, fertilizers are applied to the pots, the pots are the (random) experimental units and the plants are termed "pseudoreplicates." If f=2 levels of fertilizer are assigned at random to the 4 experimental units (the pots) then there are r–f=2 degrees of freedom for experimental error and f–1=1 df for fertilizer. The t plants within each pot are pseudoreplicates. If, in addition, different levels of a treatment, e.g. t different chemicals rubbed on leaves of the t individual plants within each pot, this would become a standard split plot. Split plots will be discussed later.

The fertilizer example has r–2=2 df for estimating pot to pot variation among pots treated alike, this being the experimental error term. The fertilizer F test is the fertilizer mean square divided by the pot(fertilizer) mean square. Using the REML approach, the F denominator is the average of 2 pot contrast sums of squares where the contrasts are

orthogonal to the fertilizer contrast. For example with 3 plants per plot, where pots 1 and 2 get one fertilizer and pots 3 and 4 get the other, the $\mathbf{CY}=\mathbf{L}$ shown in Display 2 below could be used. The orthogonal $\mathbf{C}$ matrix has first row associated with the intercept and second row contrasting fixed fertilizer effects. The third and fourth rows of $\mathbf{C}$ are pot contrasts orthogonal to the mean and fertilizer contrasts. The two pot rows are the source of the experimental error variance estimate. In the first four rows the coefficients are constant within each experimental unit (pot). The remaining rows are within pot contrasts of plants. The coefficients sum to 0 within each pot. In Display 2, the plant contrasts in rows 5-12 consist of pairs of within pot contrasts, one pair per pot, to emphasize the nested nature but any orthogonal set of r(t–1) linear combinations that sum to 0 within pots would suffice. Even the $\mathbf{CY}$ in Display 1 would work. The t–1 former treatment columns would supply (t–1) additional pseudoreplication contrasts, giving (r–1)(t–1)+(t–1) = r(t–1) plant(pot) contrasts.

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 \\ -1 & -1 & -1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & -1 \\ -1 & -1 & -1 & 1 & 1 & 1 & -1 & -1 & -1 & 1 & 1 & 1 \\ -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -1 & 2 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -1 & 2 & -1 \end{pmatrix} \begin{pmatrix} Y_{11} \\ Y_{21} \\ Y_{31} \\ Y_{12} \\ Y_{22} \\ Y_{32} \\ Y_{13} \\ Y_{23} \\ Y_{33} \\ Y_{14} \\ Y_{24} \\ Y_{34} \end{pmatrix} = \begin{pmatrix} L_{11} \\ L_{21} \\ L_{31} \\ L_{12} \\ L_{22} \\ L_{32} \\ L_{13} \\ L_{23} \\ L_{33} \\ L_{14} \\ L_{24} \\ L_{34} \end{pmatrix}$$

Display 2. Orthogonal contrasts for pseudoreplication with r=4 pots and t=3 plants per plot. Y_{ij} is the plant i, pot j response.

The proper fertilizer F test, as delivered by a RANDOM POT/TEST statement in the GLM procedure, will have 1 numerator and r–2=2 denominator df and the F test for pots within fertilizer will have r–2 numerator and r(t–1) denominator df. The Wald Z for the pot variance component will be, for $F_{pot} \geq 1$,

$$Z = \sqrt{\frac{r-2}{2}} \frac{(F_{pot}-1)}{\sqrt{F_{pot}^2 + \frac{r-2}{r(t-1)}}}$$

Based on formula (2), a 5 step strategy for studying further designs begins to appear:

(step 1) Write out orthogonal linear contrasts for the fixed treatment effects.

(step 2) Write out carefully chosen orthogonal contrasts in the random effects, orthogonal to the fixed effect contrasts. Create them in sets associated with the analysis of variance sources so that each 1 df contrast sum of squares, $(\mathbf{C}_i'\mathbf{Y})^2\,(\mathbf{C}_i'\mathbf{C}_i)$, has expected value equal to the associated ANOVA expected mean square and all of the contrasts are independent.

The RCB example shows how this can be done. Set the contrast coefficients to be constant within each level of the source and make sure the coefficients for any source lower in the hierarchy (like an interaction of the source with another factor or like the levels of a second factor nested within the source factor) sum to 0 within each source level to ensure independence.

In the RCB example, the error contrast coefficients summed to 0 within each block level and within each treatment level because, in Table 1, elementwise products were used. The block coefficients summed to 0 within each treatment because each treatment appeared in every block.

(step 3) Using the random effect contrasts as data, construct the REML likelihood. If contrasts as in (step 2) can be found, the REML likelihood will factor into separately maximizable factors, each corresponding to a source in the ANOVA table and having a variance V equal to the expected mean square for that source.

(step 4) Notice that each ANOVA source is associated with a term in –2(REML log likelihood) of the form

$$-2\ (\text{REML log likelihood}) = \ df(\ln(2\pi))+df(\ln(V))+SSq/V \text{ where } SSq=\sum_{i=1}^{df}(\mathbf{C}_i'\mathbf{Y})^2(\mathbf{C}_i'\mathbf{C}_i)$$

where df is both the number of contrasts and the ANOVA degrees of freedom for the source. Note that the resulting REML estimate, V=SSq/df, is the mean square for that source in the ANOVA table. It estimates the linear combination of variance components given in the expected mean square column of the ANOVA table.

(step 5) If the ANOVA F test is the ratio of two mean squares, apply formula (2) from section 4. If not, see section 11 below.

The challenging part is step 2. To illustrate this approach, consider $\mathbf{CY}$ in Display 2 for the fertilizer example. Plants within pots are pseudoreplicates. The fact that the coefficients in each of the last 8 rows of $\mathbf{C}$ sum to 0 within each pot ensures that the REML likelihood factors as desired. Looking at row 3 we see that $\mathbf{C}_3'\mathbf{Y}$ is a linear combination of block totals where each block total has variance $9\sigma^2_{\text{Block}}+3\sigma^2$. Because $\mathbf{C}_3'$ is a row of pot contrast coefficients, each coefficient is repeated 3 times. It follows that the expected squared contrast is $E\{(\mathbf{C}_3'\mathbf{Y})^2/(\mathbf{C}_3'\mathbf{C}_3)\}= 3\sigma^2_{\text{Pot}}+\sigma^2$. The same is true for row 4. For rows 5-12 the same logic shows that the average of $(\mathbf{C}_i'\mathbf{Y})^2/(\mathbf{C}_i'\mathbf{C}_i)$ values for these 8 rows maximizes the REML likelihood and gives the 8 df mean square for plants, MS(plant(pot)), from the ANOVA table.

9. SPLIT PLOT DESIGN IN BLOCKS

The split plot design in blocks has a nested random structure with ra whole plot treatment units in r blocks of a units, b split plot units within each whole plot unit, and an ANOVA table of this form where Q(effect) represents a quadratic form in a fixed effect:

Source	df	expected mean square
Block	r–1	$\sigma^2+b\sigma_D^2+ab\sigma_{block}^2$
Factor A	a–1	$\sigma^2+b\sigma_D^2+Q(A)$
Error A	(a–1)(r–1)	$\sigma^2+b\sigma_D^2$
Factor B	b–1	$\sigma^2+Q(B)$
AB	(a–1)(b–1)	$\sigma^2+Q(AB)$
Error B	a(r–1)(b–1)	σ^2

The model is $Y_{ijk}=\mu+B_i+\alpha_j+D_{ij}+\beta_k+(\alpha\beta)_{jk}+e_{ijk}$, with random block effects B_i, random error A effects D_{ij}, and error B effects e_{ijk}. We assume $B_i \sim N(0,\sigma_B^2)$, $D_{ij} \sim N(0,\sigma_D^2)$, and $e_{ijk} \sim N(0,\sigma^2)$. As before, the Wald Z for error B is a constant. For block and error A, the F tests are ratios of two mean squares. From the expected mean squares, the error A mean square is the appropriate denominator for blocks and for Factor A while Error B is appropriate for testing error A, fixed effects of factor B, and interaction AB. Appropriate contrasts in random effects can be found and thus formula (2) applies. We have these Wald Z statistics for variance components in terms of the corresponding F statistics:

$$Z_{Block}=\sqrt{\frac{r-1}{2}}\,\frac{(F_{Block}-1)}{\sqrt{F_{Block}^2+\frac{(r-1)}{(a-1)}}},\quad Z_D=\sqrt{\frac{(a-1)(r-1)}{2}}\,\frac{(F_D-1)}{\sqrt{F_D^2+\frac{(a-1)}{a(b-1)}}}$$ for error A,

$$Z_{errorB}=\sqrt{\frac{a(r-1)(b-1)}{2}}$$ for error B.

To show how appropriate random effect contrasts can be found, suppose there are r=5 blocks. We can use 4 contrast rows based on –1 1 0 0 0, 0 0 0 –1 1, –1 –1 0 1 1, and –1 –1 4 –1 –1 with each entry repeated within each block. Alternatively use a set comparing each block to its predecessors: –1 1 0 0 0, –1 –1 2 0 0, –1 –1 –1 3 0, and –1 –1 –1 –1 4. This second set easily generalizes to any number of levels of any factor. Similarly, for whole plot levels, there are an infinite number of sets of a–1 orthogonal contrasts. For D contrasts, we simply take the elementwise products of the block and factor A coefficients just as in Table 1. Each block coefficient gets replicated ab times, each A contrast coefficient rb times and each D coefficient b times in the rows of contrast matrix **C**. Likewise a set of b–1 factor B contrasts and, from elementwise products, a set of (a–1)(b–1) contrast coefficients for AB can be computed. For the error B terms, the (r–1)(b–1) elementwise products of the block by factor B coefficients and the (r–1)(a–1)(b–1) elementwise products of the block by AB contrast coefficients combine to give a(r–1)(b–1) error B contrast columns.

By construction, appropriate subsets of these contrast coefficients sum to 0 in such a way that the block contrasts, the D contrasts, and the error B contrasts are independent, each with expected square equal to the expected mean square for the associated ANOVA source. For many balanced experiments, sets of elementwise crossproduct contrast coefficients

(e.g. error A) or combinations of them (e.g. error B) provide the appropriate groups of independent contrasts. Once we see that such contrasts exist, formula (2) gives the relationship between the F and Wald Z tests. We again see that Z does not exist when F<1 and that, in general balanced designs, Z^2 is bounded by df/2 where df is the degrees of freedom for the variance component of interest.

10. LATIN SQUARE

When the rows and columns of an rxr Latin Square represent blocking factors, each is considered random at r levels. There are no interaction terms in the Latin Square model. There are r–1 df for rows, columns, and treatments and $r^2–1–3(r–1)=(r–1)(r–2)$ df for error. All three F tests have r–1 numerator df. The error Wald Z will be the constant $\sqrt{(r-1)(r-2)/2}$ regardless of the data. The row and column Wald Z tests will have the same form, differing only in the numerator mean square for F. The common form, from formula (2), is $Z=\sqrt{\frac{r-1}{2}}\frac{(F-1)}{\sqrt{F^2+\frac{r-1}{(r-1)(r-2)}}}=\sqrt{\frac{r-1}{2}}\frac{(F-1)}{\sqrt{F^2+\frac{1}{(r-2)}}}$. Only Latin squares of size 3x3 or higher have enough data to supply an error term. If the Latin square treatment is random, its F and Z tests will also have this relationship.

11. MIXED FACTORIAL TREATMENTS

An example for which F is not a simple ratio of mean squares (formula (2) will be slightly modified) is a factorial with some factors random. These may be the treatments in a Latin Square, RCB, completely randomized design, split plot, etc. In some of these cases, a mean square, say that for random factor A, will have an expected mean square involving a multiple of the A variance component plus a linear combination of other variance components that is not matched by the expected mean square for any other single source. In that case, a linear combination of mean squares can usually be found whose expected value is the same as what remains in the expected mean square for A after the A variance component is set to 0. That linear combination forms the denominator for a "pseudo F test" whose distribution is approximately F and whose denominator degrees of freedom are approximated by a formula due to Satterthwaite. For example, in the SAS GLM procedure the RANDOM statement with its TEST option will produce the pseudo F test and degrees of freedom. Because the Satterthwaite df formula involves calculated mean squares, it is a random variable, varying from sample to sample whereas formula (2) involves constant degrees of freedom.

As an example, suppose drivers at a levels, factor A, along with b cars of the same model, factor B, and c locations, factor C, all picked at random from large populations, are used in an experiment. Assume each (driver, car, location) combination is used r times for a total of abcr observations in a completely randomized design. Because all factors are random, so are all of their interactions. The model is $Y_{ijkm} = \mu + A_i + B_j + C_k + (AB)_{ij} + (AC)_{ik}, + (BC)_{jk} + (ABC)_{ijk} + e_{ijkm}$ with all random terms independent, normal, and each having mean 0. Using contrasts for A, B, and C and the elementwise product method of constructing interaction coefficients, orthogonal contrasts for all random sources satisfying (step 2)

above can be found. Formula (2) holds only after df_D is replaced by Satterthwaite's df. Hence the previous comment that formula (2) does not exactly hold. Formula (2) is technically appropriate for random components whose F tests are ratios of 2 mean squares. The log likelihood form in Formula (1) still holds. First, consider a subset of the rows of the ANOVA table, including the rows used in constructing the usual pseudo F test for factor A.

Source	df	Expected mean square
A	a–1	$\sigma^2+r\sigma^2_{ABC}+rb\sigma^2_{AC}+rc\sigma^2_{AB}+rbc\sigma^2_{A}$
AB	(a–1)(b–1)	$\sigma^2+r\sigma^2_{ABC}+rc\sigma^2_{AB}$
AC	(a–1)(c–1)	$\sigma^2+r\sigma^2_{ABC}+rb\sigma^2_{AC}$
ABC	(a–1)(b–1)(c–1)	$\sigma^2+r\sigma^2_{ABC}$
Error	abc(r-1)	σ^2

Looking at the expected mean squares, if σ^2_A were 0 the expected mean square for A would become $\sigma^2+r\sigma^2_{ABC}+rb\sigma^2_{AC}+rc\sigma^2_{AB}$. We could estimate $rbc\sigma^2_A$ by subtracting, from the mean square for A, an estimate of $\sigma^2+r\sigma^2_{ABC}+rb\sigma^2_{AC}+rc\sigma^2_{AB}$. Unlike previous examples, no single mean square has this expected value. Using ANOVA notation, the expected value of $MS_{AB}+MS_{AC}–MS_{ABC}$ is seen to be $\sigma^2+r\sigma^2_{ABC}+rb\sigma^2_{AC}+rc\sigma^2_{AB}$. The resulting pseudo F statistic, $F_A = MS_A/(MS_{AB}+MS_{AC}–MS_{ABC})$, is produced with the GLM procedure's RANDOM statement. Using the first derivative in formula (2) the REML estimate of the A variance component is $MS_A–(MS_{AB}+MS_{AC}–MS_{ABC})$, matching the ANOVA table estimate. Differences between F and Z^2 are due to their denominators rather than to the estimated variance component common to the numerators.

As before, the mean squares can be computed as averages of independent individual one df sums of squares computed in (step 2) of our REML analysis strategy. Using the Hessian derived variance of the estimated factor A variance component given by $MS_A–(MS_{AB}+MS_{AC}–MS_{ABC})$, we have, similar to our development of formula (2), the Wald variance estimate

$$\frac{2MS^2_A}{(a-1)}+\frac{2MS^2_{AB}}{(a-1)(b-1)}+\frac{2MS_{AC}}{(a-1)(c-1)}+\frac{2MS^2_{ABC}}{(a-1)(b-1)(c-1)}=$$
$$\frac{2}{(a-1)}\left(F^2_A D^2_A+\frac{MS^2_{AB}}{(b-1)}+\frac{MS^2_{AC}}{(c-1)}+\frac{MS^2_{ABC}}{(b-1)(c-1)}\right)$$

where $D_A = MS_{AB}+MS_{AC}-MS_{ABC}$ is the denominator for the F_A pseudo F test for factor A. In order to produce a formula similar to formula (2), write the Wald variance as

$$\frac{2D^2_A}{(a-1)}\left(F^2_A+\frac{MS^2_{AB}}{(b-1)D^2_A}+\frac{MS^2_{Ac}}{(c-1)D^2_A}+\frac{MS^2_{ABC}}{(b-1)(c-1)D^2_A}\right)=\frac{2D^2_A}{(a-1)}\left(F^2_A+\frac{1}{W}\right)$$

where W is

$$W = df_{Satterthwaite} = \frac{D_A^2}{\frac{MS_{AB}^2}{(b-1)} + \frac{MS_{Ac}^2}{(c-1)} + \frac{MS_{ABC}^2}{(b-1)(c-1)}} = \frac{(MS_{AB} + MS_{AC} - MS_{ABC})^2}{\frac{MS_{AB}^2}{(b-1)} + \frac{MS_{AC}^2}{(c-1)} + \frac{MS_{ABC}^2}{(b-1)(c-1)}}.$$

Interestingly, W is Satterthwaite's approximate degrees of freedom formula, $df_{Satterthwaite}$ for the pseudo F_A test. The formula for the Wald statistic thus has the form

$$Z = \sqrt{\frac{a-1}{2}} \frac{(F_A - 1)}{\sqrt{F_A^2 + \frac{(a-1)}{df_{Satterthwaite}}}} \text{ for } F_A \geq 0$$

A few details should be mentioned before showing a graph. First, $df_{Satterthwaite}$ is a random variable. No constant F critical value is available for a horizontal reference line. Instead, color will be used to distinguish significant (red) from non significant (blue) F_A tests. Second if any estimated variance component hits the boundary 0 it can affect <u>all</u> the other estimates and tests. All estimates may differ from what would have been found had the boundary restriction been removed. In Figures 5A and 5B, 6290 data sets were generated in order to get 5000 in which <u>all</u> PROC MIXED variance component estimates exceed 0. In Figure 5A and 5B there were 5 levels of A, 4 of B, and 3 of C with 2 replicates of each of the 60 A,B,C factorial combinations. Standard deviations of the 7 random effects (A, AB, C, …, ABC) are shown in the titles. The left figure 5A graphs all pseudo F_A tests from the SAS GLM procedure versus the Wald Z from the SAS procedure PROC MIXED. Figure 5B shows the subset of these points having Z between 0.92 and 0.99.

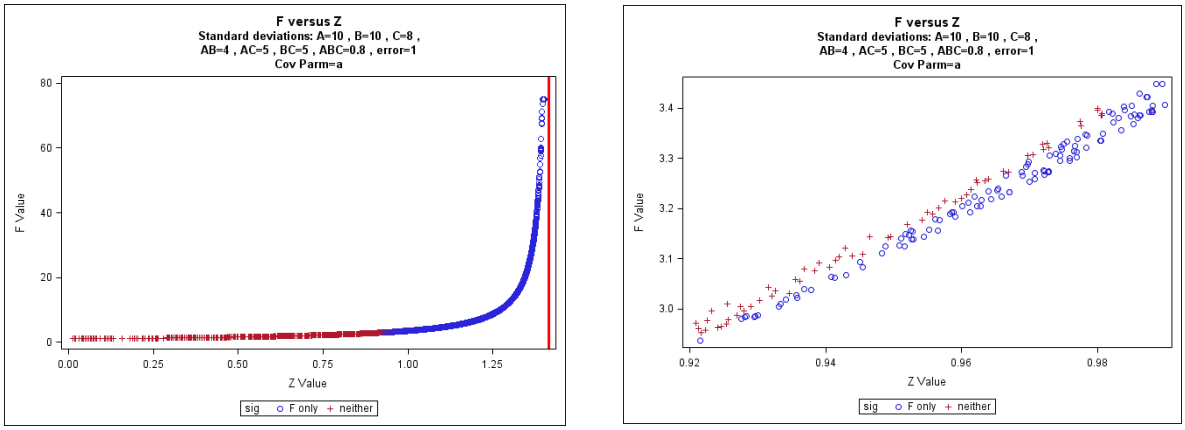

Figure 3. (**A**) F versus Z with red vertical limit line, and (**B**) the same graph restricted to a small interval of Z.

Figure 3A is hardly distinguishable in form from Figures 1 and 2 in which the denominator degrees of freedom for F were constant. Figure 5B shows the effect of the random $df_{Satterthwaite}$. The effect is minor but not zero. For a given Z and hence a given F numerator, a larger denominator is associated with a smaller F but larger Satterthwaite degrees of freedom. A smaller F could be significant while a slightly larger F might not, this being the predominant result in Figure 3B.

12. SOME ADDITIONAL EMPIRICAL RESULTS

We have not covered all mixed models, for example unbalanced data, repeated measures with autoregressive errors, and random coefficient models among others. Unbalanced data in general may differ with each pattern of missingness and extreme missingness could lead to non estimable parameters. Finding appropriate contrasts in our step (2) could be difficult or impossible.

As a quick empirical look, we reran the 2000 RCB analyses in each panel of Figure 1, this time with points set to missing whenever a uniform [0,1] random variable (independently generated for each point) was less than 0.10. There was very little effect of missing values for our particular RCB example. A plot and table similar to those of Figure 1 matched that figure quite closely. A visual assessment of one particular case falls far short of a general proof. Such a proof, however, may be difficult or impossible to obtain as it is not clear how to construct appropriate contrasts in the presence of missing values.

13. CONCLUSION

We have shown that, in many commonly encountered balanced experimental designs, there is an informative common functional relationship between the ANOVA F test for random effects and the Wald Z test. This relationship results in a severe lack of power for Z in cases in which there are only a few levels of the random factor. We have shown that the large sample reputation of the Wald test needs clarification. Large, in fact, refers to the number of random effect levels, not the total number of observations. The relationship between Z and F shows that missing Wald tests in these most common designed experiments are associated with $F<1$ which suggests that a missing Wald test can be taken as a failure to reject the 0 variance component null hypothesis. In general if there is an F test available for a variance component it is preferred to the Wald Z test.

REFERENCES

McNew, Ron and Mauromoustakos, A. (1997). "TESTING VARIANCE COMPONENTS USING THE GLM AND MIXED PROCEDURES OF SAS®," Conference on Applied Statistics in Agriculture. https://doi.org/10.4148/2475-7772.1311

CONTACT INFORMATION

Dr. David A. Dickey
Distinguished Professor (Emeritus)
NC State University
dickey@ncsu.edu
https://www4.stat.ncsu.edu/~dickey/