

Kansas State University Libraries

New Prairie Press

Conference on Applied Statistics in Agriculture

1997 - 9th Annual Conference Proceedings

TESTING VARIANCE COMPONENTS USING THE GLM AND MIXED PROCEDURES OF SAS®

Ron McNew

Andy Mauromoustakos

Follow this and additional works at: <https://newprairiepress.org/agstatconference>



Part of the [Agriculture Commons](#), and the [Applied Statistics Commons](#)



This work is licensed under a [Creative Commons Attribution-Noncommercial-No Derivative Works 4.0 License](#).

Recommended Citation

McNew, Ron and Mauromoustakos, Andy (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>

This is brought to you for free and open access by the Conferences at New Prairie Press. It has been accepted for inclusion in Conference on Applied Statistics in Agriculture by an authorized administrator of New Prairie Press. For more information, please contact cads@k-state.edu.

TESTING VARIANCE COMPONENTS USING THE GLM AND MIXED PROCEDURES OF SAS®

Ron McNew and Andy Mauromoustakos
Agricultural Statistics Laboratory
University of Arkansas

Abstract

The test of a variance component in random and mixed normal linear models can be done using the F statistic from the analysis of variance or the Wald statistic which is the ratio of the variance component estimate to its estimated standard error. These are the methods used in the GLM and MIXED procedures of SAS®. We show that these two tests can give different results on the same data. For the one-way random model, the one-sided Wald test on the among group variance component can never be significant at the 0.05 probability level when the number of levels of the random factor is six or less. This is in contrast to the F test which, under the null model, will achieve the nominal level, even when using Satterthwaite's approximation for the distribution of the test statistic. The Wald test is conservative for even relatively large numbers of levels of the among group factor. Increasing the number of observations per level increases, rather than decreases, the difference between the actual and nominal significance levels. The reason that the Wald test is so conservative is that it uses the estimated standard error, which is a function of the variance component estimate. These results help explain why the F test and the Wald test can be so different.

1. Introduction

We were using a mixed model to analyze data for the purpose of making inferences about the fixed effects and variance components of the model. We were particularly interested in being able to demonstrate statistical significance for a variance component. From our familiarity with the situation, we believed the variance component was positive and expected a test to confirm our belief. We used the MIXED procedure of SAS® (SAS Institute Inc., 1996a) for the analysis, which provides a "Z Value" (the Wald statistic; Serfling, 1980) for the test on the variance component. The p-value for the test was large, indicating non-significance. Because of this contrary result, we also used the GLM procedure (SAS Institute Inc., 1989) in which we included a RANDOM statement and its TEST option to give a different test on

the variance component. For unbalanced data such as ours, this gives an approximate F test (using Satterthwaite's approximation; Satterthwaite, 1946) for the variance component. The result of the test was a small p-value, matching our expectation. So we were faced with the dilemma: Which test should we trust? From our past experiences and from work of others (e.g., Ames and Webster, 1991; Zhou and Mathew, 1994), we were confident that the approximate F test is reliable and therefore we could trust the p-value from this test. The properties of the Wald test are "large sample results" and small sample approximations may be poor. Documentation for PROC MIXED includes the caution that "Wald tests can be unreliable in small samples." Thus it seemed that we should not trust the Wald test. However, it left the unanswered questions:

Why are the results from the two tests so different?

When can we rely on the Wald test?

2. Example

The following example is simpler than our original problem in that there are only two factors, data are balanced, and the F tests are exact. However, it has the same contradictory results from the Wald test and the F test for a variance component. The situation is this: On each of 4 farms, seven genetic lines were randomly allocated to seven large plots. The response variable to be analyzed was grain yield; the data along with the marginal means for farms and lines are given in Table 1. It is clear from the marginal means that farm differences are large compared to differences among lines. Both factors "Farms" and "Lines" were considered random and variance components of these factors were of interest. The analysis of variance (Table 2) also reflects the large difference due to farms in that the Farm Mean Square is very large compared to either the Residual or the Line Mean Squares. The F test for Farms has a very small p-value but the Wald test does not.

3. A solution

We chose to use a one-way random-effects model in which all random effects are mutually independent and normally distributed as a starting point for obtaining answers to our questions. For this model, the ratio of the "Among" mean square (M_A) to the "Within" mean square (M_W) provides an exact F test for the "Among" variance component (σ_A^2). For "a" levels of the random factor, "n" observations per level, and $F = M_A/M_W$, the Wald statistic is

$$\frac{M_A - M_W}{\sqrt{\frac{2M_A^2}{a-1} + \frac{2M_W^2}{a(n-1)}}} = \frac{F - 1}{\sqrt{F^2 + \frac{a-1}{a(n-1)}}} \sqrt{\frac{a-1}{2}}$$

The first of the two factors in the right-hand expression is always less than 1; thus the Wald statistic never exceeds the second factor. For a one-sided test on σ_A^2 at $\alpha = 0.05$, for which the critical value is 1.645, a significant result cannot be obtained for 6 or fewer levels of the random factor. We calculated the probability of a significant result for the null model for several values of “a” and “n” for the one-sided case above. The results of these calculations given in Figure 1 make it clear that the number of levels of the random factor cannot be 10 or 20 but must be much larger for the actual significance level to be near the nominal level. A characteristic that may not be intuitive is that increasing “n” increases, rather than decreases, the distance between the nominal and actual levels.

4. The cause

One might conjecture that the very large numbers required for an adequate approximation for the Wald statistic is due to the fact that it involves estimating variance components rather than means. That this is not the whole story was suggested by the following result. If we consider the ratio of the variance component estimate to its actual standard error, rather than to its estimated standard error, the sampling distribution of this variable does not suffer the same problems as the Wald statistic. We calculated the probability of a significant result for the one-sided hypothesis described previously. The results displayed in Figure 2 show that the actual significance level is slightly larger than the nominal level; again the difference is aggravated by increasing “n.” Thus the problem of the Wald statistic results from having to use the estimated standard error of the variance component estimate, which is not independent of the variance component estimate, and not solely from the use of second-order statistics.

5. Recommendation

The above analyses give results for the one-way model. However, the results are not limited to this simple model; rather, they apply to any other linear model in which the variance component is estimated by a linear combination of two independent mean squares. Although we have not attempted a formal analysis of more complex models, we believe that the highly

conservative nature of the Wald test for a variance component would apply to most models and estimation methods. Therefore, we suggest that data analysts use the F test for variance components from the GLM procedure. If the model is such that only the MIXED procedure is appropriate for the analysis, then the Wald test for a variance component should be ignored unless the associated number of random effects is large. Our analysis of the one-way random model indicates that “large” may be much greater than 100.

6. Summary

The results from our study show that the Wald statistic for the one-way random model cannot be trusted, with respect to the p-value or test size, unless the number of levels of the random factor is very large. Version 6.12 of the MIXED procedure (SAS Institute Inc., 1996b) is the first version in which the Wald statistic is not an automatic output, a change presumably resulting from the recognition of its unreliability. Our results indicate that the Wald statistic for testing a variance component is an extremely conservative procedure. This is a consequence of using a non-independent estimate of the error in the variance component estimate. As other studies have shown, Satterthwaite's approximate F test from an ANOVA is an adequate testing procedure and, therefore, is preferable to using Wald's statistic. Although our "solution" was considered only in the one-way random model, it can be extended to other situations where the variance component estimate has a similar form, viz., it's proportional to the difference between two independent mean squares. It can even apply in the unbalanced one-way random model. We conjecture that the result would extend to most linear models in which random effects are normally distributed, regardless of the data structure or estimation method.

References

- Ames, Michael H. and John T. Webster. 1991. On estimating approximate degrees of freedom. *American Statistician*. 45: 45-50.
- SAS Institute Inc. 1989. *SAS/STAT® User's Guide, Version 6, Fourth Edition, Volume 2*. SAS Institute Inc.: Cary, NC.
- SAS Institute Inc. 1996a. *SAS/STAT® Software: Changes and Enhancements through Release 6.11*. SAS Institute Inc.: Cary, NC.
- SAS Institute Inc. 1996b. *SAS/STAT® Software: Changes and Enhancements for Release 6.12*. SAS Institute Inc.: Cary, NC.
- Satterthwaite, F. E. 1946. An approximate distribution of estimates of variance components. *Biometrics Bulletin*. 2:110-114.
- Serfling, Robert J. 1980. *Approximation Theorems of Mathematical Statistics*. John Wiley & Sons: New York, NY.
- Zhou, Leping and Thomas Mathew. 1994. Some tests for variance components using generalized p values. *Technometrics*. 36:394-402.

Table 1. Grain yields and marginal means from seven genetic lines grown on each of four farms.

	Line 1	Line 2	Line 3	Line 4	Line 5	Line 6	Line 7	Mean
Farm 1	42	50	55	47	47	51	66	51
Farm 2	44	32	40	41	36	31	54	40
Farm 3	18	24	18	16	15	14	30	19
Farm 4	68	65	65	69	64	66	69	67
Mean	43	43	45	43	41	41	55	44

Figure 1. Relation of the actual significance level of the Wald test of the among group variance component to the number of levels of the random factor and the number of observations (n) per level for a nominal significance level of 0.05.

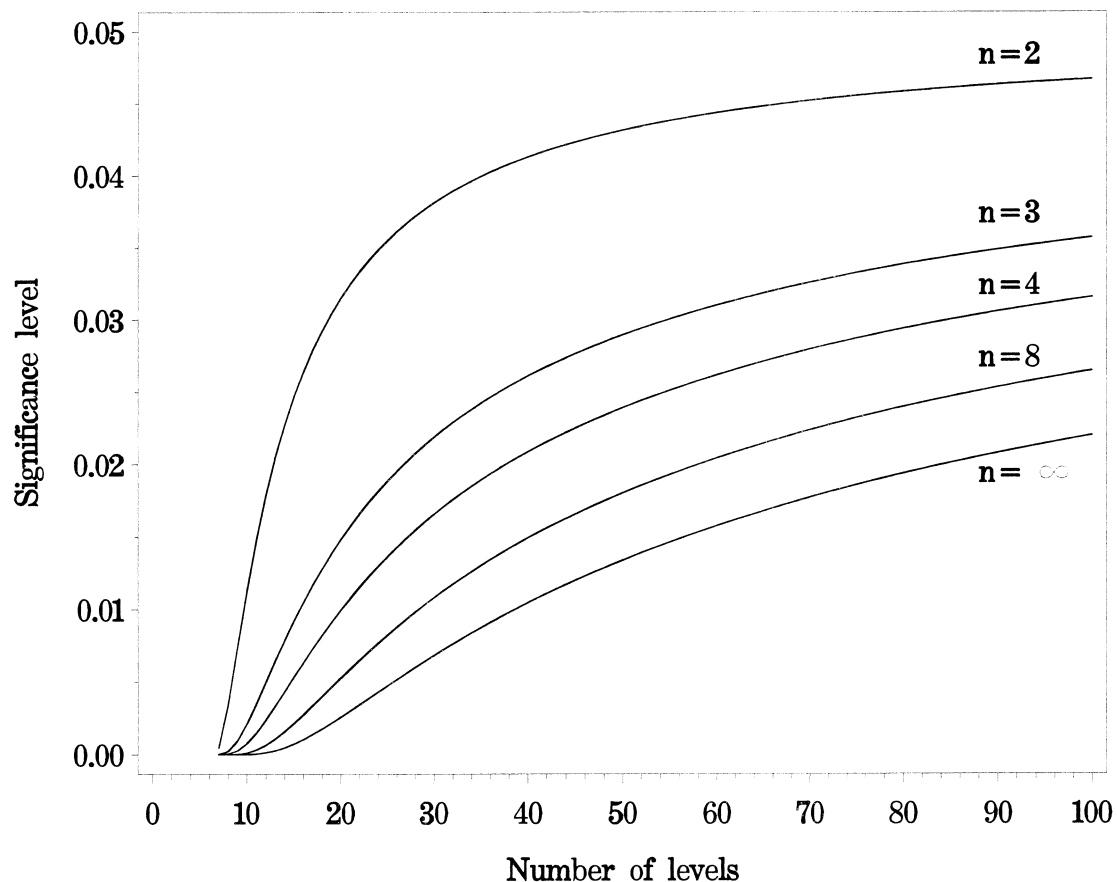


Table 2. Analysis of variance, F tests (F), and Wald tests (Z), with associated p-values for the grain yield data from seven genetic lines grown on each of four farms.

Source	df	Mean Square	F	p-value	Z	p-value
Farm	3	2776	132.0	<0.001	1.22	0.112
Line	6	95	4.5	0.006	1.34	0.090
Residual	18	21				

Figure 2. Relation of the actual significance level to the numbers of levels and observations (n) per level when the actual standard error replaces the estimated standard error in the Wald statistic at a nominal significance level of 0.05.

