5-1-1998

Analysis of heterogeneity of variance using anome on ln S^2

Casey Volino

ANALYSIS OF HETEROGENEITY OF VARIANCE USING ANOME ON ln S^2

by

Casey A. Volino

A Thesis Submitted
in
Partial Fulfillment
of the
Requirements for the Degree of
MASTER OF SCIENCE
in
Applied and Mathematical Statistics

Approved by:

Prof. Edward G. Schilling
(Thesis Advisor)

Prof. David R. Lawrence

Prof. Donald D. Baker
(Department Head)

GRADUATE STATISTICS
COLLEGE OF ENGINEERING
ROCHESTER INSTITUTE OF TECHNOLOGY
ROCHESTER, NEW YORK
MAY, 1998

Title of Thesis: ANALYSIS OF HETEROGENEITY OF VARIANCE USING ANOME ON ln S_i^2

I ________________________ hereby grant permission to the Wallace Memorial Library of RIT to reproduce my thesis in whole or in part. Any reproduction will not be for commercial use or profit.

Date: 6-1-98

Forward

While working at Corning, Incorporated, I had the good fortune of being able to work with several highly skilled senior statistical engineers, from whom I learned a great deal about various statistical techniques. One such statistician, Charles Comer, introduced me to an article by L. S. Nelson[1] which succinctly detailed the technique of using an analysis-of-variance-type approach to detect variance heterogeneity. Sometime later, when I learned about the analysis of means (ANOME) technique, I thought it would be interesting to use ANOME to analyze variance heterogeneity and then compare the results to the ANOVA approach outlined by L. S. Nelson.

Abstract

When n replicates are available from a factorial experiment, several methods exist for testing the validity of the assumption of equal variances within the "cells" or treatment combinations of the experiment. A new test is proposed for variances of random samples believed to be from normal populations. This new test combines both the familiar graphical analysis of means for treatment effects (ANOME) and the analysis of the logarithms of the within-group variances to produce a graphical display of the test for variance homogeneity. To determine robustness of the proposed test against departures from the underlying normality assumption, this new test is also evaluated for non-normal populations.

Another analysis-of-means-type test was developed by Wludyka and Nelson[2] which utilizes Dirichlet distributions and specially constructed tables. The new test, proposed herein, has an advantage in that it relies solely on critical values developed for the analysis-of-means procedure. As an added simplification, only those critical values corresponding to infinite degrees of freedom are required.

A ln ANOME analysis of Nelson's data (used to demonstrate the ln ANOVA procedure) yielded the same conclusion. Also, simulation results indicate that when the underlying assumption of normality is not feasible, the ln ANOME procedure demonstrated equivalent or superior Type-I error-rate stability and power among tests which rely on that assumption. However, when the underlying assumption of normality is tenable, Bartlett's test performs the best of all homogeneity-of-variance tests studied in maintaining stable Type-I errors and power.

Table of Contents

List of Tables

List of Figures

Introduction

In many statistical analyses, the statistician is interested in testing the hypothesis of homogeneity of variances of two or more populations. Often these populations represent the "groups" in an analysis of variance, and it is desirable to test the validity of the assumption of equal "within-group" variances. However, at times, we are interested in only a direct comparison of the variability, or spread, among several candidate populations (i.e. suppliers, measurement devices, etc.). In the latter situation, the variances can be thought of as those of random samples from populations represented by the "groups" or cells in a one-way or higher analysis of variance. Hence, our discussion will focus on the general case of variances of random samples from populations constituting the "groups" or "cells" in the analysis of variance.

Suppose we have a sample of n observations from each of k populations. We shall denote the random sample from population i by $\{Y_{i1}, Y_{i2}, \ldots, Y_{in}\}$ with sample mean

$$\overline{Y}_i = \frac{\sum_{j=1}^{n} Y_{ij}}{n} \qquad (1)$$

and sample variance

$$S_i^2 = \frac{\sum_{j=1}^{n} (Y_{ij} - \overline{Y}_i)^2}{n-1}. \qquad (2)$$

Then the hypothesis we would wish to test would be of the form

$$H_0\text{:}\ \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2, \qquad (3)$$

where σ_i^2 represents the variance of the i^{th} population. We would then test this against the alternate hypothesis

$$H_A\text{: Not } H_0 \text{ for at least one } \sigma_i^2. \tag{4}$$

When these populations are thought to be normally distributed, there exist several well-established techniques commonly accepted as appropriate tests of this hypothesis. Among these are the common F-test (for two variances), and Bartlett's test.

An approximate test, presented by Bartlett and Kendall[3], involves performing the usual analysis of variance on the logarithms of the "within-group" sample variances. Bartlett and Kendall point out that the variance of ln S^2 is independent of σ^2, and that a known theoretical error term is available which depends on only the number of replicates, n, for each treatment combination. An approximation to this theoretical error term given by L. S. Nelson[1] is of the following form:

$$\text{Var}(\ln S^2) \cong \frac{2}{n-2} \tag{5}$$

A more precise approximation of this error term is given by Wludyka and Nelson[2] as

$$\text{Var}(\ln S^2) \cong \frac{2}{n-1} + \frac{2}{(n-1)^2} + \frac{4}{3(n-1)^3} - \frac{16}{15(n-1)^5} \tag{6}$$

For $n \geq 5$ replicates per subgroup, the logarithms of the subgroup variances are approximately normally distributed. Thus the test of the null hypothesis in (3) above is essentially transformed into a test on means. This test relies on the fact , as Scheffé[4] indicates, that "the analysis of variance [procedure] is fairly insensitive to the shape of the distributions of the estimated means." Lorenzen and Anderson[5] add that it does not matter if we use $\log_e$ (natural logarithm) or $\log_{10}$ for this test, since both log functions are scale related and ANOVA is scale invariant.

For dealing with variances from non-normal but continuous distributions, Levene's[6] test was modified by Brown and Forsythe[7] from it's original form to serve as a non-parametric test of variances. The proposed procedure (that does analysis of means on the natural logarithms of the subgroup variances) will be compared to Bartlett and Kendall's approximate ANOVA test, Bartlett's[8] test, and a modified version of Levene's test due to Brown and Forsythe. The comparison will be made under varying conditions.

ANOME of ln S^2

Decision Limits for ANOME

In the usual analysis of means for treatment effects (ANOME) from a factorial experiment, E. G. Schilling[9] gives the following formula for the upper and lower decision limits,

$$0 \pm \hat{\sigma}_e h_\alpha \sqrt{\frac{q}{N}}, \qquad (7)$$

where N = total number of observations in the experiment,

q = degrees of freedom for the effect tested (same as ANOVA), and

k = number of points plotted (number of means to compare).

The values for h_α differ depending on whether the effect of interest is a main effect or an interaction effect. For main effects, the decision limits can be exactly specified as

$$h_\alpha = H_\alpha \sqrt{\frac{k}{k-1}}, \qquad (8)$$

since in the case of testing main effects, H_α is exact. For interaction effects, $h_\alpha = h_\alpha^*$, where the h_α^* are Sidak factors tabulated in Table A.9 in E. R. Ott and E. G. Schilling[10], as suggested by L. S. Nelson[11] H_α critical values for $\alpha = 0.10$, 0.05, 0.01, and 0.001 and infinite degrees of freedom are given in Appendix A-1. E. G. Schilling[9] justifies the use of Sidak factors with the statement,

> "For interactions and nested factors, h_α^* is used because of the nature of the correlation among points plotted."

For one-way layouts, it may by desirable for ANOME on log-variances to be constructed so that "effects" are centered about their average log-variance instead of about zero. One would then be able to easily retrieve the original subgroup variances from the "main effects" as $S_i^2 = e^{\ln S_i^2}$.

Procedure for ANOME on ln S^2

To adapt the ANOME procedure to the analysis of the logarithms of variances, we take the following steps:

Step 1. Compute the subgroup variances, and take their natural logarithms.

Step 2. Calculate the treatment effects for the main effects as the

Main Effects for Factor A:

$$A_i = \overline{\ln S_i^2} \quad \overline{\overline{\ln S^2}} \tag{9}$$

Main Effects for Factor B:

$$B_j = \overline{\ln S_j^2} \quad \overline{\overline{\ln S^2}} \tag{10}$$

Repeat this for each factor in the experiment.

Step 3. Calculate the treatment effects for the interactions as the difference between the average log-variance for each treatment combination of the factors and the grand average of all the log-variances, less any previously estimated lower-order effects. (See below for two-factor example.)

Interaction Effects for the AB Interaction:

$$AB_{ij} = \overline{\ln S_{ij}^2} \quad \overline{\overline{\ln S^2}} \quad A_i \quad B_j \tag{11}$$

Repeat this for each interaction in the experiment.

Step 4. Calculate the theoretical error variance, and take it's square root for use in computing the decision limits. Regard this estimate as having infinite degrees of freedom.

Approximation to the Theoretical Error Variance:

$$\hat{\sigma}_e^2 = \frac{2}{n-1} + \frac{2}{(n-1)^2} + \frac{4}{3(n-1)^3} - \frac{16}{15(n-1)^5} \tag{12}$$

$$\hat{\sigma}_e = \sqrt{\hat{\sigma}_e^2} \text{ with } \infty \text{ degrees of freedom}$$

Step 5. Compute the decision limits for main effects:

Decision Limits for Main Effects:

Note that by adding the average overall mean back into the treatment effect, we could essentially center the effects about the overall average. However, the decision limits for the main effects centered about zero would be computed as follows:

$$0 \pm \hat{\sigma}_e H_\alpha \sqrt{\frac{k}{k-1}} \sqrt{\frac{k-1}{N}}, \tag{13}$$

which reduces to

$$0 \pm \hat{\sigma}_e H_\alpha \sqrt{\frac{k}{N}}. \tag{14}$$

For a one-way analysis, N = k, so the limits are given by

$$0 \pm \hat{\sigma}_e H_\alpha. \tag{15}$$

(N = k for a one-way analysis, since after taking the natural logarithm of the subgroup variances, there is effectively only one replicate per cell.)

Step 6. Compute the decision limits for interaction effects

***Decision Limits for Interaction Effects*:**

The decision limits for interaction effects centered about zero are computed as follows:

$$0 \pm \hat{\sigma}_e h^*_\alpha \sqrt{\frac{q}{N}}, \tag{16}$$

where N = total number of treatment combinations in the experiment,

q = degrees of freedom for the interaction effect tested, and

h^*_α = Sidak factor for k means and infinite degrees of freedom.

The next section presents a worked example to further illustrate the technique.

Example of ANOME on ln S^2

An example of the ANOME on ln S_i^2 (henceforth called ln ANOME) procedure is now presented using the same data given by L. S. Nelson[1] to illustrate ANOVA on ln S_i^2. The experimental design involved three factors (A, B, and C), with k = 3, 2, and 4 levels, respectively. The entire experiment was replicated six times. The original data is shown below.

	A1		A2		A3	
	B1	B2	B1	B2	B1	B2
C1	53.0	49.8	53.0	51.7	51.3	53.2
	52.1	53.2	52.5	50.1	48.9	51.9
	55.9	51.3	50.4	52.9	51.7	53.1
	53.0	52.6	51.5	49.8	53.4	49.6
	55.0	51.7	52.6	52.3	52.6	54.1
	52.0	53.7	53.5	51.9	50.1	53.1
C2	59.3	52.3	55.0	54.1	51.5	57.5
	51.4	56.4	57.0	54.7	56.4	55.0
	58.0	53.5	55.6	54.7	51.3	52.3
	54.1	54.2	50.3	56.7	49.4	54.1
	58.3	53.7	51.4	54.4	55.4	50.8
	56.1	51.8	57.5	53.2	53.1	53.1
C3	55.7	55.3	58.9	48.2	57.7	54.3
	53.5	55.9	57.0	56.5	61.9	54.6
	55.4	54.7	58.0	59.0	49.6	54.7
	54.5	54.1	57.7	52.9	57.0	56.7
	53.5	55.2	57.3	54.5	56.6	57.6
	59.4	59.3	52.2	56.8	56.8	58.1
C4	54.0	59.3	62.0	57.5	57.2	56.8
	54.7	59.5	58.5	58.1	62.4	60.3
	57.9	55.7	57.9	56.3	63.1	60.9
	58.8	56.9	59.7	63.4	56.5	52.3
	59.1	58.4	60.8	54.6	60.7	61.0
	64.8	56.1	57.3	56.9	55.1	61.1

Step 1. Compute the subgroup variances, and take their natural logarithms.

A	B	C	Var(Y)	ln(Var(Y))
A1	B1	C1	2.544	0.934
A1	B1	C2	8.944	2.191
A1	B1	C3	4.819	1.572
A1	B1	C4	14.942	2.704
A1	B2	C1	2.019	0.703
A1	B2	C2	2.627	0.966
A1	B2	C3	3.391	1.221
A1	B2	C4	2.695	0.991
A2	B1	C1	1.259	0.230
A2	B1	C2	8.791	2.174
A2	B1	C3	5.619	1.726
A2	B1	C4	3.255	1.180
A2	B2	C1	1.527	0.423
A2	B2	C2	1.335	0.289
A2	B2	C3	14.331	2.662
A2	B2	C4	8.968	2.194
A3	B1	C1	2.691	0.990
A3	B1	C2	7.059	1.954
A3	B1	C3	15.700	2.754
A3	B1	C4	11.159	2.412
A3	B2	C1	2.508	0.919
A3	B2	C2	5.392	1.685
A3	B2	C3	2.800	1.030
A3	B2	C4	12.603	2.534
			Average:	1.518

Step 2. Calculate the treatment effects for the main effects as the difference between the average log-variance for each level of the factor and the grand average of all the log-variances.

Main Effects for Factor A:

$$A_i = \overline{\ln S_i^2} \quad \overline{\overline{\ln S^2}}$$

$A_1 = (0.934 + 2.191 + 1.572 + 2.704 + 0.703 + 0.966 + 1.221 + 0.991)/8 - 1.518 = -0.108$

$A_2 = (0.230 + 2.174 + 1.726 + 1.180 + 0.423 + 0.289 + 2.662 + 2.194)/8 - 1.518 = -0.158$

$A_3 = (0.990 + 1.954 + 2.754 + 2.412 + 0.919 + 1.685 + 1.030 + 2.534)/8 - 1.518 = 0.266$

Main Effects for Factor B:

$$B_j = \overline{\ln S_j^2} \quad \overline{\overline{\ln S^2}}$$

B_1 = (0.934 + 2.191 + 1.572 + 2.704 + 0.230 + 2.174 + 1.726 + 1.180 + 0.990 + 1.954 + 2.754 + 2.412)/12 – 1.518 = 0.217

B_2 = (0.703 + 0.966 + 1.221 + 0.991 + 0.423 + 0.289 + 2.662 + 2.194 + 0.919 + 1.685 + 1.030 + 2.534)/12 – 1.518 = -0.217

Main Effects for Factor C:

$$C_m = \overline{\ln S_m^2} \quad \overline{\overline{\ln S^2}}$$

C_1 = (0.934 + 0.703 + 0.230 + 0.423 + 0.990 + 0.919)/6 – 1.518 = -0.818

C_2 = (2.191 + 0.966 + 2.174 + 0.289 + 1.954 + 1.685)/6 – 1.518 = 0.025

C_3 = (1.572 + 1.221 + 1.726 + 2.662 + 2.754 + 1.030)/6 – 1.518 = 0.309

C_4 = (2.704 + 0.991 + 1.180 + 2.194 + 2.412 + 2.534)/6 – 1.518 = 0.484

Step 3. Calculate the treatment effects for the interactions as the difference between the average log-variance for each combination of the factors and the grand average of all the log-variances, less any previously estimated lower-order effects.

Interaction Effects for the AB Interaction:

$$AB_{ij} = \overline{\ln S_{ij}^2} \quad \overline{\overline{\ln S^2}} \quad A_i \quad B_j$$

AB_{11} = (0.934 + 2.191 + 1.572 + 2.704)/4 – 1.518 – (-0.108) – (0.217) = 0.223

AB_{12} = (0.703 + 0.966 + 1.221 + 0.991)/4 – 1.518 – (-0.108) – (-0.217) = -0.223

AB_{21} = (0.230 + 2.174 + 1.726 + 1.180)/4 – 1.518 – (-0.158) – (0.217) = -0.249

AB_{22} = (0.423 + 0.289 + 2.662 + 2.194)/4 – 1.518 – (-0.158) – (-0.217) = 0.249

AB_{31} = (0.990 + 1.954 + 2.754 + 2.412)/4 – 1.518 – (0.266) – (0.217) = 0.026

AB_{32} = (0.919 + 1.685 + 1.030 + 2.534)/4 – 1.518 – (0.266) – (-0.217) = -0.026

Interaction Effects for the AC Interaction:

$$AC_{im} = \overline{\ln S^2_{im}} \quad \overline{\overline{\ln S^2}} \quad A_i \ - C_m$$

AC_{11} = (0.934 + 0.703)/2 – 1.518 – (-0.108) – (-0.818) = 0.226

AC_{12} = (2.191 + 0.966)/2 – 1.518 – (-0.108) – (0.025) = 0.143

AC_{13} = (1.572 + 1.221)/2 – 1.518 – (-0.108) – (0.309) = -0.323

AC_{14} = (2.704 + 0.991)/2 – 1.518 – (-0.108) – (0.484) = -0.047

AC_{21} = (0.230 + 0.423)/2 – 1.518 – (-0.158) – (-0.818) = -0.215

AC_{22} = (2.174 + 0.289)/2 – 1.518 – (-0.158) – (0.025) = -0.153

AC_{23} = (1.726 + 2.662)/2 – 1.518 – (-0.158) – (0.309) = 0.525

AC_{24} = (1.180 + 2.194)/2 – 1.518 – (-0.158) – (0.484) = -0.157

AC_{31} = (0.990 + 0.919)/2 – 1.518 – (0.266) – (-0.818) = -0.012

AC_{32} = (1.954 + 1.685)/2 – 1.518 – (0.266) – (0.025) = 0.010

AC_{33} = (2.754 + 1.030)/2 – 1.518 – (0.266) – (0.309) = -0.202

AC_{34} = (2.412 + 2.534)/2 – 1.518 – (0.266) – (0.484) = 0.204

Interaction Effects for the BC Interaction:

$$BC_{jm} = \overline{\ln S^2_{jm}} \quad \overline{\overline{\ln S^2}} \quad B_j \quad C_m$$

BC_{11} = (0.934 + 0.230 + 0.990)/3 – 1.518 – (0.217) – (-0.818) = -0.199

BC_{12} = (2.191 + 2.174 + 1.954)/3 – 1.518 – (0.217) – (0.025) = 0.346

BC_{13} = (1.572 + 1.726 + 2.754)/3 – 1.518 – (0.217) – (0.309) = -0.027

BC_{14} = (2.704 + 1.180 + 2.412)/3 – 1.518 – (0.217) – (0.484) = -0.121

BC_{21} = (0.703 + 0.423 + 0.919)/3 – 1.518 – (-0.217) – (-0.818) = 0.199

BC_{22} = (0.966 + 0.289 + 1.685)/3 – 1.518 – (-0.217) – (0.025) = -0.346

BC_{23} = (1.221 + 2.662 + 1.030)/3 – 1.518 – (-0.217) – (0.309) = 0.027

BC_{24} = (0.991 + 2.194 + 2.534)/3 – 1.518 – (-0.217) – (0.484) = 0.121

Interaction Effects for the ABC Interaction:

$$ABC_{ijm} = \ln S^2_{ijm} - \overline{\overline{\ln S^2}} - A_i - B_j - C_m - AB_{ij} - AC_{im} - BC_{jm}$$

ABC_{111} = (0.934) – 1.518 – (-0.108) – (0.217) – (-0.818) – (0.223) – (0.226) – (-0.199) = -0.126

ABC_{112} = (2.191) – 1.518 – (-0.108) – (0.217) – (0.025) – (0.223) – (0.143) – (0.346) = -0.174

ABC_{113} = (1.572) – 1.518 – (-0.108) – (0.217) – (0.309) – (0.223) – (-0.323) – (-0.027) = -0.237

ABC_{114} = (2.704) – 1.518 – (-0.108) – (0.217) – (0.484) – (0.223) – (-0.047) – (-0.121) = 0.537

ABC_{121} = (0.703) – 1.518 – (-0.108) – (-0.217) – (-0.818) – (-0.223) – (0.226) – (0.199) = 0.126

ABC_{122} = (0.966) – 1.518 – (-0.108) – (-0.217) – (0.025) – (-0.223) – (0.143) – (-0.346) = 0.174

ABC_{123} = (1.221) – 1.518 – (-0.108) – (-0.217) – (0.309) – (-0.223) – (-0.323) – (0.027) = 0.237

ABC_{124} = (0.991) – 1.518 – (-0.108) – (-0.217) – (0.484) – (-0.223) – (-0.047) – (0.121) = -0.537

ABC_{211} = (0.230) – 1.518 – (-0.158) – (0.217) – (-0.818) – (-0.249) – (-0.215) – (-0.199) = 0.135

ABC_{212} = (2.174) – 1.518 – (-0.158) – (0.217) – (0.025) – (-0.249) – (-0.153) – (0.346) = -0.890

ABC_{213} = (1.726) – 1.518 – (-0.158) – (0.217) – (0.309) – (-0.249) – (0.525) – (-0.027) = -0.409

ABC_{214} = (1.180) – 1.518 – (-0.158) – (0.217) – (0.484) – (-0.249) – (-0.157) – (-0.121) = -0.354

ABC_{221} = (0.423) – 1.518 – (-0.158) – (-0.217) – (-0.818) – (0.249) – (-0.215) – (0.199) = -0.135

ABC_{222} = (0.289) – 1.518 – (-0.158) – (-0.217) – (0.025) – (0.249) – (-0.153) – (-0.346) = -0.628

ABC_{223} = (2.662) – 1.518 – (-0.158) – (-0.217) – (0.309) – (0.249) – (0.525) – (0.027) = 0.409

ABC_{224} = (2.194) – 1.518 – (-0.158) – (-0.217) – (0.484) – (0.249) – (-0.157) – (0.121) = 0.354

ABC_{311} = (0.990) – 1.518 – (0.266) – (0.217) – (-0.818) – (0.026) – (-0.012) – (-0.199) = -0.009

ABC_{312} = (1.954) – 1.518 – (0.266) – (0.217) – (0.025) – (0.026) – (0.010) – (0.346) = -0.454

ABC_{313} = (2.754) – 1.518 – (0.266) – (0.217) – (0.309) – (0.026) – (-0.202) – (-0. 027) = 0.646

ABC_{314} = (2.412) – 1.518 – (0.266) – (0.217) – (0.484) – (0.026) – (0.204) – (-0.121) = -0.183

ABC_{321} = (0.919) – 1.518 – (0.266) – (-0.217) – (-0.818) – (-0.026) – (-0.012) – (0.199) = 0.009

ABC_{322} = (1.685) – 1.518 – (0.266) – (-0.217) – (0.025) – (-0.026) – (0.010) – (-0.346) = 0.454

ABC_{323} = (1.030) – 1.518 – (0.266) – (-0.217) – (0.309) – (-0.026) – (-0.202) – (0.027) = -0.646

ABC_{324} = (2.534) – 1.518 – (0.266) – (-0.217) – (0.484) – (-0.026) – (0.204) – (0.121) = 0.183

Step 4. Calculate the theoretical error variance, and take it's square root for use in computing the decision limits. Regard this estimate as having infinite degrees of freedom.

$$\hat{\sigma}_e^2 = \frac{2}{6-1} + \frac{2}{(6-1)^2} + \frac{4}{3(6-1)^3} - \frac{16}{15(6-1)^5} = 0.4903$$

$$\hat{\sigma}_e = \sqrt{\hat{\sigma}_e^2} = \sqrt{0.4903} = 0.70021$$

Step 5. Compute the decision limits for main effects (shown for $\alpha = 0.05$).

Decision Limits for Main Effects: $0 \pm \hat{\sigma}_e H_\alpha \sqrt{\frac{k}{N}}$

A: k = 3 levels, $H_\alpha = H_{0.05,3,\infty} = 1.91$, N = 24. Hence,

$$0 \pm \hat{\sigma}_e H_\alpha \sqrt{\frac{k}{N}} \Rightarrow 0 \pm 0.70021 * 1.91 * \sqrt{\frac{3}{24}} \Rightarrow 0 \pm 0.4728$$

B: k = 2 levels, $H_\alpha = H_{0.05,2,\infty} = 1.386$, N = 24. Hence,

$$0 \pm \hat{\sigma}_e H_\alpha \sqrt{\frac{k}{N}} \Rightarrow 0 \pm 0.70021 * 1.386 * \sqrt{\frac{2}{24}} \Rightarrow 0 \pm 0.2801$$

C: k = 4 levels, $H_\alpha = H_{0.05,4,\infty} = 2.14$, N = 24. Hence,

$$0 \pm \hat{\sigma}_e H_\alpha \sqrt{\frac{k}{N}} \Rightarrow 0 \pm 0.70021 * 2.14 * \sqrt{\frac{4}{24}} \Rightarrow 0 \pm 0.6116$$

Step 6. Compute the decision limits for interaction effects (shown for $\alpha = 0.05$).

AB: k = 3 * 2 = 6 levels, q = ab – a – b + 1 = 6 – 3 – 2 + 1 = 2,

$h^*_{0.05,6,\infty} = 2.631$, N = 24. Hence, the decision limits are

$$0 \pm \hat{\sigma}_e h^*_\alpha \sqrt{\frac{q}{N}} \Rightarrow 0 \pm 0.70021 * 2.631 * \sqrt{\frac{2}{24}} \Rightarrow 0 \pm 0.5317$$

AC: k = 3 * 4 = 12 levels, q = ac – a – c + 1 = 12 – 3 – 4 + 1 = 6,

$h^*_{0.05,12,\infty} = 2.858$, N = 24. Hence, the decision limits are

$$0 \pm \hat{\sigma}_e h^*_\alpha \sqrt{\frac{q}{N}} \Rightarrow 0 \pm 0.70021 * 2.858 * \sqrt{\frac{6}{24}} \Rightarrow 0 \pm 1.0003$$

BC: k = 2 * 4 = 8 levels, q = bc – b – c + 1 = 8 – 2 – 4 + 1 = 3,

$h^*_{0.05,8,\infty} = 2.727$, N = 24. Hence, the decision limits are

$$0 \pm \hat{\sigma}_e h^*_\alpha \sqrt{\frac{q}{N}} \Rightarrow 0 \pm 0.70021 * 2.727 * \sqrt{\frac{3}{24}} \Rightarrow 0 \pm 0.6749$$

ABC: k = 3 * 2 * 4 = 24 levels, q = abc – ab – ac – bc + a + b + c - 1 = 6,

$h^*_{0.05,24,\infty} = 3.071$, N = 24. Hence, the decision limits are

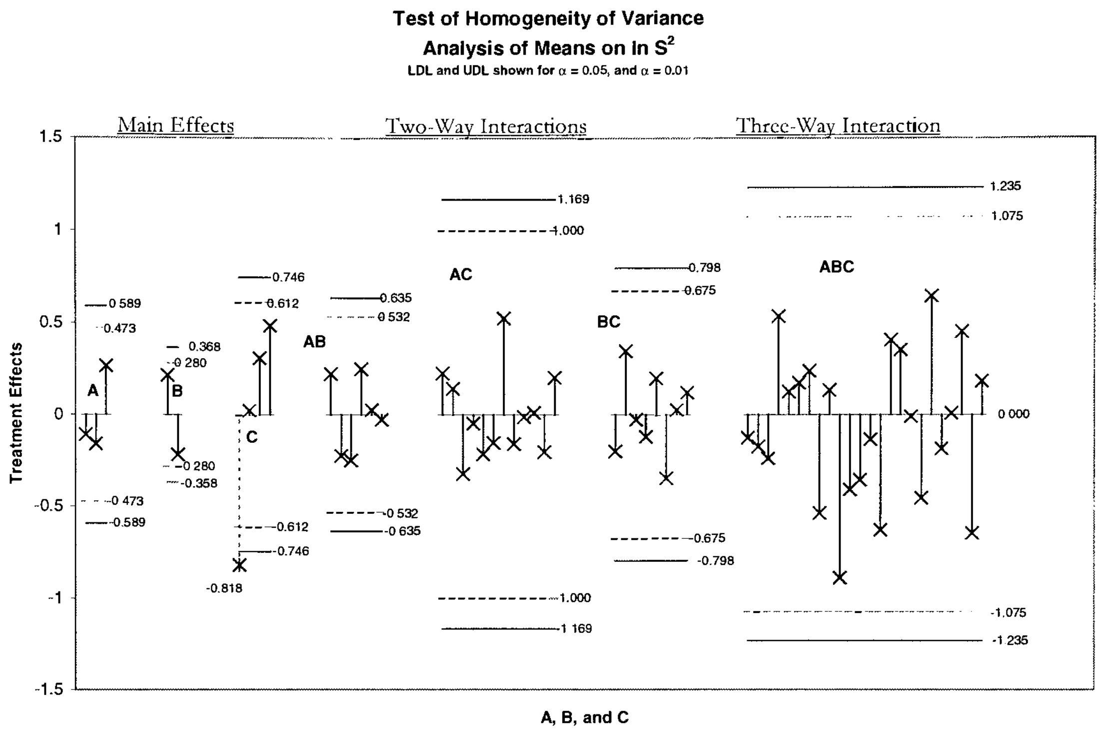
$$0 \pm \hat{\sigma}_e h^*_\alpha \sqrt{\frac{q}{N}} \Rightarrow 0 \pm 0.70021 * 3.071 * \sqrt{\frac{6}{24}} \Rightarrow 0 \pm 1.0749$$

Step 7. Construct the chart.

Note that the chart (next page) displays decision limits at the $\alpha = 0.01$ level, as well as at the $\alpha = 0.05$ level as calculated above.

As can be seen from the ln ANOME chart, the variability differed for the different levels of factor C. In particular, we see that the variability for the first level of factor C is significantly less than the variability for the other three levels of factor C.

Figure 1: Example of ANOME on ln S^2.



Bartlett & Kendall's

ln ANOVA Test of Homogeneity of Variance

Bartlett and Kendall's[3] ANOVA-type test of homogeneity of variance involves first computing the subgroup variances and then computing their natural logarithms, as was shown previously for the ANOME on ln S^2. Next, the theoretical error variance is computed based on the number of replications as shown in equation (12). Then the usual sum of squares, degrees of freedom, and mean squares are computed for every term in the original model. Usually, we would not have an error term for this situation, since there is now effectively only one replicate of the entire experiment and the model would therefore be completely specified. However, we make use of the theoretical error variance with infinite degrees of freedom to compute values for the F-statistics and perform tests of hypothesis. The analysis of the example data from L. S. Nelson is shown below.

ln ANOVA on L. Nelson's Data

Source	DF	Adj SS	Adj MS	F	p	
A	2	0.86224	0.43112	0.87925	0.4151	
B	1	1.12867	1.12867	2.30188	0.1292	
C	3	6.00366	2.00122	4.08141	**0.0066**	**
A*B	2	0.90023	0.45012	0.91800	0.3993	
A*C	6	1.26216	0.21036	0.42902	0.8601	
B*C	3	1.04869	0.34956	0.71291	0.5441	
A*B*C	6	3.50764	0.58461	1.19229	0.3069	
Error	∞		0.49033			

Again we arrive at the conclusion that the variability differed for the different levels of factor C. This time, however, further investigation is required to determine the nature of this difference.

Bartlett's Test of Homogeneity of Variance

Bartlett's[8] test of homogeneity of variance is a modification of Neyman and Pearson's[12] generalized likelihood-ratio test (L1 test, 1931). This modification involved replacing the biased maximum-likelihood estimators of the variances with unbiased estimators and substituting n_i-1 for n_i in the weights. Bartlett's test is known to rely heavily on the assumption of normality of the underlying distributions.

The value of the test statistic is determined from the data by first computing the sample variances of each of the k subgroups and then computing the subsequent pooled sample variance as

$$s_{pooled}^2 = \frac{\sum_{i=1}^{k} (n_i - 1)s_i^2}{N - k} \tag{17}$$

Then base-10 logarithms are taken of each of the k sample variances and of the pooled variance. Ultimately, the test statistic is given by

$$\chi_0^2 = 2.3026 \cdot \frac{(N-k)\log_{10} s_{pooled}^2 - \sum_{i=1}^{k} (n_i - 1)\log_{10} s_i^2}{1 + \frac{1}{3(k-1)}\left(\sum_{i=1}^{k} \frac{1}{n_i - 1} - \frac{1}{N-k}\right)} \tag{18}$$

The value of χ_0^2 is then compared to the critical chi-square value with k-1 degrees of freedom.

Levene's Test with Brown and Forsythe's Modification

Levene's[6] test with Brown and Forsythe's[7] modification is essentially a non-parametric test of homogeneity of variance. The test statistic is constructed as follows: Let

$$V_{ij} = \left|Y_{ij} - \tilde{Y}_i\right| , \tag{19}$$

where

Y_{ij} = the jth observation in the ith group and

$\tilde{Y}_i$ = the median of the ith group.

Then form the one-way ANOVA statistic

$$F(calc) = \frac{\dfrac{\sum_{i=1}^{k} n_i (\overline{V}_{i.} - \overline{V}_{..})^2}{k-1}}{\dfrac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} (V_{ij} - \overline{V}_{i.})^2}{N-k}}, \tag{20}$$

where

$$\overline{V}_{i.} = \frac{\sum_{j=1}^{n_i} V_{ij}}{n_i}, \tag{21}$$

$$\overline{V}_{..} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n_i} V_{ij}}{N}, \tag{22}$$

and

$$N = \sum_{i=1}^{k} n_i \tag{23}$$

is the total number of observations in the experiment. The value of the test statistic, F(calc), is compared to a critical value of the F-distribution, with k-1 numerator degrees of freedom and N-k denominator degrees of freedom.

Monte Carlo Simulations

Simulations provided the means for assessing, under controlled conditions, the ability of these individual tests to reject correctly or incorrectly the null hypothesis. All simulations involved a balanced, one-way layout. Seven cases were considered, as indicated in Table 1.

Table 1 Cases Used For Monte Carlo Simulation

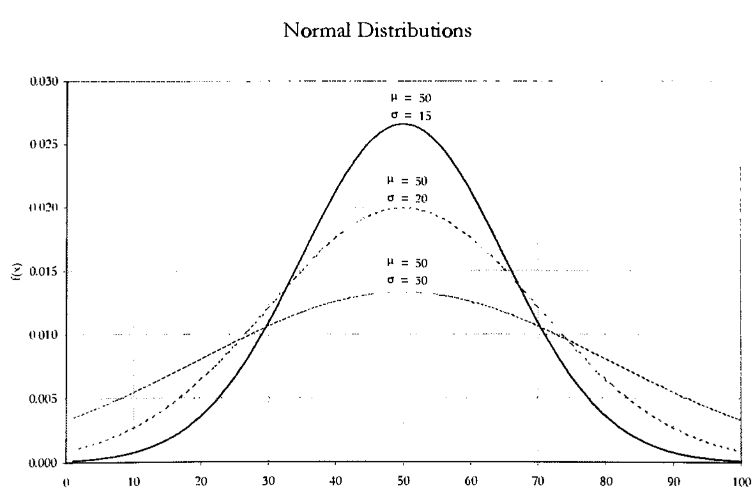
Case	Conditions	Null Hypothesis
Case 1	$\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$, X~Normal($\mu$=50, σ=15)	True (Assessment of Type-I Error Rate)
Case 2	$\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$, X~Weibull($\omega$=1, ϕ=1.5)	True (Assessment of Type-I Error Rate in Non-Normal Situations)
Case 3	$\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$, X~Gamma($\psi$=2.5, λ=2)	True (Assessment of Type-I Error Rate in Non-Normal Situations)
Case 4	$\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_{k-1}^2$, X~Normal($\mu$=50, σ=15) σ_k^2 not equal since, X~Normal(μ=50, σ=20)	False (Assessment of Type-II Error Rate and Power)
Case 5	$\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_{k-1}^2$, X~Normal($\mu$=50, σ=15) σ_k^2 not equal since, X~Normal(μ=50, σ=30)	False (Assessment of Type-II Error Rate and Power)
Case 6	$\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_{k-1}^2$, X~Weibull($\omega$=1, ϕ=1.5) σ_k^2 not equal since, X~Weibull(ω=2, ϕ=1.5)	False (Assessment of Type-II Error Rate and Power in Non-Normal Situations)
Case 7	$\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_{k-1}^2$, X~Gamma($\psi$=2.5, λ=2) σ_k^2 not equal since, X~Gamma(ψ=5.0, λ=2)	False (Assessment of Type-II Error Rate and Power in Non-Normal Situations)

To assess the Type-I error rate in Cases 1, 2, and 3, k = 2, 4, 6, 8, and 10 subgroups were generated and compared, with n = 2, 3, …, 10 replicates each. The other four cases assessed the power of the homogeneity of variance tests using comparisons of k = 2, 4, 6, 8, and 10 subgroups with n = 2, 3, …, 10 replicates.

One thousand simulations per condition (case, subgroup, and replicate combination) were conducted in accordance with random-number-generating procedures outlined by Dodson and Nolan[13] These procedures make use of the fact that all BASIC-type programming languages are capable of producing uniformly-distributed pseudo-random numbers on the [0, 1] interval. These uniformly-distributed random numbers can be used to generate other random numbers for almost any distribution. In the simplest form of random-number generation, this is done by setting the cumulative distribution function of the desired density function equal to the uniformly-distributed random number and then solving for the random variable of the new distribution. An example of this procedure for the two-parameter Weibull distribution is shown in Appendix A-3. When a closed form does not exist for the cumulative distribution function, special algorithms must be employed. Two such algorithms were used to generate Normal-distributed and Gamma-distributed random numbers. The code for all random numbers generated is available in Appendix A-4.

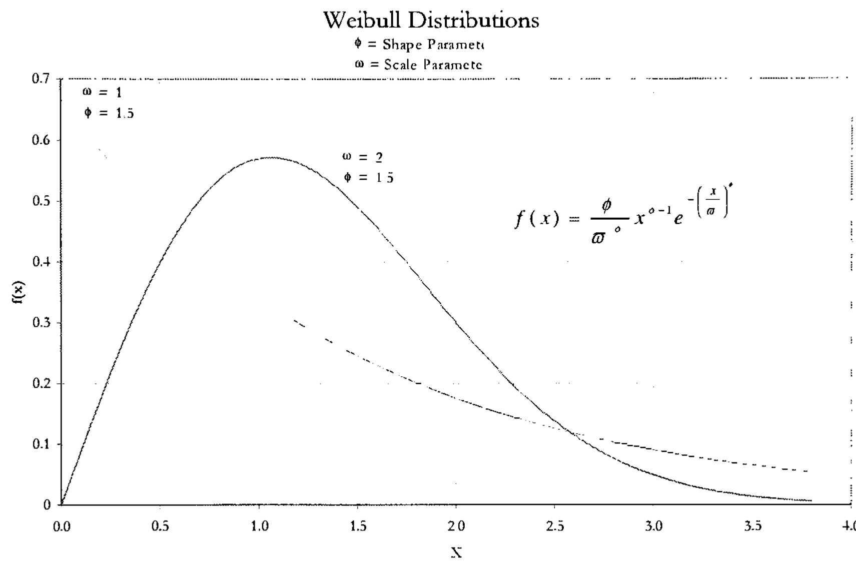
The three Normal distributions chosen for this study are shown in Figure 2.

Figure 2: Normal Distributions Used for Simulations



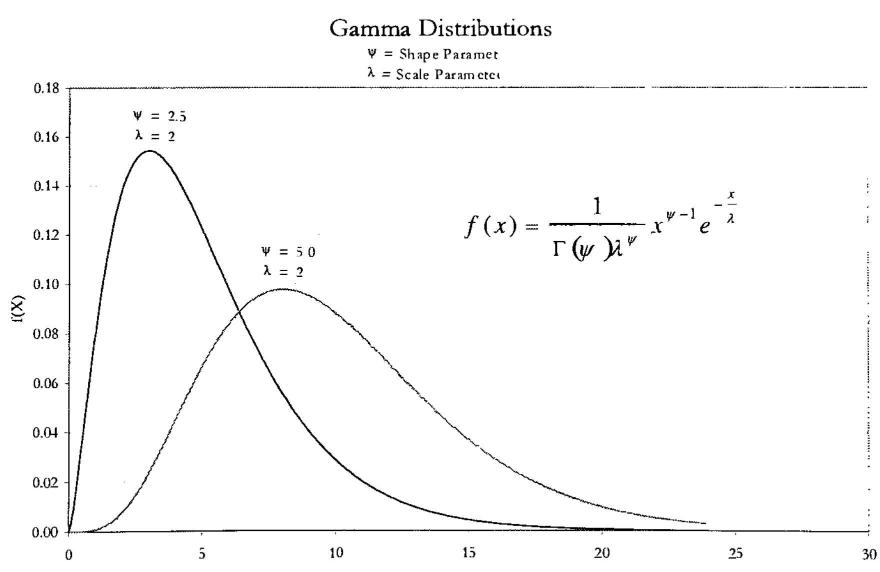
The two forms of the Weibull distribution used for simulations are shown in Figure 3.

Figure 3: Weibull Distributions Used for Simulations



The two forms of the Gamma distribution (both Chi-Square, in this case, since $\lambda = 2$) are shown in Figure 4.

Figure 4: Gamma Distributions Used for Simulations



Once random numbers of the appropriate type were generated, the number of rejections of the null hypothesis (out of one-thousand simulations) was recorded for each set of conditions. This was done concurrently on the data for each of the following statistical tests:

1. the standard ANOVA for testing means,
2. the ANOVA on ln S_i^2 (ln ANOVA) to test for homogeneity of variance,
3. the proposed ANOME on ln S_i^2 (ln ANOME) to test for homogeneity of variance,
4. Bartlett's Test of homogeneity of variance,
5. the standard F-test (for k = 2 variances only), and
6. Levene's test with Brown and Forsythe's modification.

In situations were the null hypothesis of equal variances was true, the tests were compared on the basis of their ability to maintain, over all conditions, the expected Type-I error rate. When the null hypothesis was false (that is, when the last of k variances was not equal to the other k-1 variances), the tests were compared on the basis of power. This is more easily understood considering the following familiar diagram:

Figure 5: Types of Error In Hypothesis Testing

<u>Types of Error:</u>

	H_0 True	H_0 False
Reject H_0:	Type-I Error α	Correct Decision Power $(1-\beta)$
Accept H_0:	Correct Decision	Type-II Error β

<u>Simulation Conditions</u>

H_0 True	H_0 False
Case1 Case 2 Case 3	Case 4 Case 5 Case 6 Case 7

Simulation Results

The behavior of the tests of homogeneity of variance in conforming to the Type-I error rate at the five percent level are presented in Tables 2, 3, and 4. Each of the tables that follow display the number of rejections (at the $\alpha = 0.05$ significance level) of the null hypothesis of equal variances out of one-thousand simulations.

Table 2: Results for Case 1, Normal Distribution with $\mu = 50$, and $\sigma = 15$

(Number of rejections of the null hypothesis out of 1000)

		n								
HOV Test	k	2	3	4	5	6	7	8	9	10
ln ANOVA	2	68	35	37	40	53	44	44	50	35
(Bartlett and Kendall)	4	108	63	47	48	45	46	50	69	71
	6	113	65	56	58	45	51	42	47	38
	8	120	69	56	46	44	53	57	55	41
	10	116	50	55	56	48	51	69	51	69
ln ANOME	2	68	35	37	40	53	44	44	50	35
(Volino)	4	119	70	50	49	46	52	53	70	66
	6	142	82	66	77	54	53	46	65	47
	8	172	98	85	66	75	77	67	62	50
	10	177	92	93	88	70	72	94	64	76
Bartlett's Test	2	34	28	35	36	51	40	43	48	35
	4	31	39	27	33	38	41	39	54	63
	6	36	32	31	41	40	34	33	38	38
	8	19	30	26	20	34	34	50	40	33
	10	16	21	25	32	28	39	50	39	47
F-Test (Two Variances)	2	45	30	35	37	51	41	43	48	35
Levene's Test	2		0	49	7	46	9	32	18	33
	4		0	59	1	30	9	31	15	40
	6		0	55	0	25	10	31	12	23
	8		0	50	0	27	6	53	15	17
	10		0	58	0	27	6	42	14	36

Even though the assumption of approximate normality of $\ln S^2$ is stated in many references as holding for $k \geq 5$, the number of replicates within each cell was examined for $k < 5$, as well as $k = 5, 6, \ldots, 10$. In fact, "real world" experiments frequently have fewer than five replicates. It is of interest, therefore, to study the behavior of these statistics under less-than-ideal conditions.

Both the ln ANOVA and ln ANOME techniques are more apt to reject the hypothesis of equal variances than is the F-test, Bartlett's test, or Levene's test. Between ln ANOME and ln ANOVA, as k increases, ln ANOME rejects slightly more often than ln ANOVA. This makes ln ANOME in its present form less attractive than other homogeneity-of-variance tests for $k > 6$ or 7. Bartlett's test, given the condition of normality, performed closest to the expected nominal rejection rate of 50 out of 1000, or 5%.

Curiously, in Levene's test, there is a kind of "odd-even" effect that could be due to the modification of Levene's[6] original test by Brown and Forsythe[7] (Brief mention of this "odd-even" effect was made in Conover, Johnson, and Johnson.[14]) For odd numbers of replicates, the median is the value of the "middle" observation in terms of magnitude. Considering deviations from the median, based on an odd number of replicates, one term in equation (19) is always zero. Hence, averages of the V_{ij} are smaller making the test overly conservative relative to the nominal Type-I rejection rate. In fact, for $n = 3$ replicates per cell, the null hypothesis was never rejected! (Calculations from the simulation programs were cross-checked frequently with the results obtained from both SAS and MINITAB, and all test-statistics and p-values agreed.) For an even number of replicates, the results from Levene's test were less conservative than results for odd numbers of replicates, but were still below the nominal rejection rate of five percent.

The results for the Weibull and Gamma distributions show the dependence of all of these tests, except Levene's, on the underlying assumption of normality. Type-I error rates for all but Levene's test, which again gives evidence of an "odd-even" effect, are at times more than triple or quadruple the nominal five percent rejection rate one would expect.

Table 3: Results for Case 2, Weibull Distribution with ω = 1, and ϕ = 1.5

(Number of rejections of the null hypothesis out of 1000)

		n								
HOV Test	k	2	3	4	5	6	7	8	9	10
ln ANOVA	2	75	89	77	98	93	104	97	87	108
(Bartlett and Kendall)	4	123	101	73	112	104	137	112	140	140
	6	152	130	91	137	138	153	151	173	176
	8	186	163	96	145	149	174	197	191	196
	10	201	163	104	170	182	207	205	227	238
ln ANOME	2	75	89	77	98	92	104	97	87	108
(Volino)	4	118	104	66	106	92	137	111	119	134
	6	154	136	78	133	148	149	162	152	165
	8	211	172	77	137	135	152	183	150	143
	10	225	186	95	172	169	168	177	184	182
Bartlett's Test	2	30	73	67	91	91	102	96	87	107
	4	47	89	72	105	98	137	105	138	142
	6	49	78	88	134	133	150	148	158	179
	8	55	85	107	131	151	161	194	189	199
	10	43	101	127	163	176	207	199	219	245
F-Test (Two Variances)	2	40	79	68	93	91	102	96	87	107
Levene's Test	2		0	81	7	42	21	50	26	42
	4		0	84	2	52	21	26	17	50
	6		0	98	4	38	14	37	19	40
	8		0	84	4	51	10	42	15	24
	10		0	86	2	48	7	28	10	30

For tests of equal variances in non-normal situations, only Levene's test, for even numbers of replicates, yielded a Type-I rejection rate comparable to the nominal 5% rate. All other tests resulted in greatly inflated Type-I error rates relative to the nominal. In the case of an underlying Weibull distribution, ln ANOME was comparable to or better than ln ANOVA and Bartlett's Test for $n \geq 4$ replicates per subgroup. Also, as the number of variances, k, increased, ln ANOME showed slightly better Type-I error rate stability than did either ln ANOVA or Bartlett's Test.

Table 4: Results for Case 3, Gamma Distribution with $\psi = 2.5$, and $\lambda = 2$

(Number of rejections of the null hypothesis out of 1000)

		n								
HOV Test	k	2	3	4	5	6	7	8	9	10
ln ANOVA	2	88	75	89	105	109	116	117	139	127
(Bartlett and Kendall)	4	114	98	122	134	139	154	165	181	185
	6	160	130	125	177	186	217	233	207	228
	8	170	135	163	188	210	223	253	262	278
	10	184	134	162	199	224	263	289	306	312
ln ANOME	2	88	75	88	105	109	116	117	139	127
(Volino)	4	129	95	124	122	129	132	157	165	178
	6	184	125	139	157	154	203	202	188	210
	8	186	153	158	182	178	205	224	222	254
	10	220	164	162	176	172	224	233	233	254
Bartlett's Test	2	49	61	83	99	103	111	112	134	122
	4	58	84	110	124	139	156	165	180	187
	6	63	103	115	176	197	231	232	225	234
	8	68	111	170	204	217	246	277	292	286
	10	67	124	195	232	242	283	315	329	329
F-Test (Two Variances)	2	59	67	84	99	104	111	112	134	122
Levene's Test	2		0	82	5	46	24	38	29	49
	4		0	109	5	44	15	43	19	43
	6		0	98	7	43	16	48	12	40
	8		0	83	2	47	9	41	20	22
	10		0	93	4	37	12	33	18	30

The performance of the tests of homogeneity of variance in maintaining power (1-β) at the five-percent level are presented in Tables 5, 6, 7, and 8. Tables 5 and 6 show the results under normality assumptions, while Tables 7 and 8 show the results under non-normality.

Table 5: Results for Case 4, Normal Distribution with $\mu = 50$, and $\sigma_1 = \sigma_2 = \ldots = \sigma_{k-1} = 15$, $\sigma_k = 20$

(Number of rejections of the null hypothesis out of 1000)

		n								
HOV Test	k	2	3	4	5	6	7	8	9	10
ln ANOVA	2	77	45	41	55	86	95	135	114	117
(Bartlett and Kendall)	4	99	75	51	58	68	91	106	98	103
	6	115	65	56	76	73	71	107	82	111
	8	105	68	46	56	77	63	72	99	100
	10	133	57	51	50	72	63	79	67	108
ln ANOME	2	77	45	41	55	86	95	135	114	116
(Volino)	4	111	75	54	66	73	100	102	100	104
	6	132	75	79	83	77	88	104	89	113
	8	150	95	80	76	99	81	93	111	122
	10	188	88	84	79	79	83	97	79	115
Bartlett's Test	2	37	30	37	51	81	87	133	110	116
	4	31	47	37	58	76	93	98	101	109
	6	27	30	29	61	71	78	99	94	125
	8	27	36	32	44	73	71	67	102	106
	10	40	23	27	47	63	61	83	76	110
F-Test (Two Variances)	2	47	35	38	51	81	88	133	110	116
Levene's Test	2		0	70	10	67	35	62	61	84
	4		0	66	5	80	22	68	37	71
	6		0	54	4	76	26	56	39	83
	8		0	68	5	52	19	50	33	58
	10		0	60	1	39	12	63	23	76

From Tables 5 and 6, it is clear that all tests respond to the increase from 20 to 30 in the k^{th} σ. In Table 5, all tests that rely on the normality assumption performed about equally well in terms of power. However, the ln ANOME test seemed to detect the difference in variability of the k^{th} group more readily than either ln ANOVA or Bartlett's Test. The F-test results are generally so close to those of Bartlett's test that there is little value in mentioning both. Judging from Table 6, Bartlett's test seems to be the best at detecting true differences in variability, followed by ln ANOME. Hence, the faith of many authors in Bartlett's test when a normality assumption is tenable seems justified.

Table 6: Results for Case 5, Normal Distribution with $\mu = 50$, and $\sigma_1 = \sigma_2 = \ldots = \sigma_{k-1} = 15$, $\sigma_k = 30$

(Number of rejections of the null hypothesis out of 1000)

		n								
HOV Test	k	2	3	4	5	6	7	8	9	10
ln ANOVA	2	80	95	114	197	264	318	419	436	504
(Bartlett and Kendall)	4	108	88	121	183	257	314	392	436	504
	6	129	95	114	161	226	308	364	404	451
	8	145	95	94	131	210	226	366	365	417
	10	151	95	102	133	167	219	305	333	398
ln ANOME	2	80	95	114	197	264	318	419	436	504
(Volino)	4	122	72	132	179	286	343	402	457	512
	6	142	104	140	177	272	334	406	451	516
	8	184	110	115	170	254	295	439	471	509
	10	194	134	122	184	241	293	397	456	519
Bartlett's Test	2	45	84	101	185	257	310	412	427	499
	4	61	90	166	227	328	386	463	500	566
	6	55	78	155	216	299	387	474	479	568
	8	43	116	129	205	328	348	503	501	545
	10	34	93	135	215	276	335	442	479	550
F-Test (Two Variances)	2	57	84	104	185	258	311	412	427	499
Levene's Test	2		0	102	25	151	106	268	192	308
	4		0	158	55	237	165	326	282	389
	6		0	140	48	198	180	305	299	407
	8		0	134	43	191	142	351	310	407
	10		0	146	41	198	125	302	278	398

For Cases 6 and 7, it was not possible in simulations to alter the variance of a group without also changing the group's mean. This is due to the fact that for the Weibull and Gamma distributions, both the mean and the variance of the population are direct functions of the parameters that describe these distributions. Nonetheless, the results are shown in Tables 7 and 8 for k-1 groups from the same population and the k[th] group different.

Table 7: Results for Case 6, Weibull Distributions with ω = 1, and ϕ = 1.5 (First k-1 groups), ω = 2 (k[th] group)

(Number of rejections of the null hypothesis out of 1000)

		n								
HOV Test	k	2	3	4	5	6	7	8	9	10
ln ANOVA	2	118	135	202	271	326	358	421	460	501
(Bartlett and Kendall)	4	167	152	163	278	342	394	479	510	535
	6	155	175	202	276	346	426	456	508	577
	8	93	205	235	313	373	425	507	528	573
	10	145	221	240	298	350	422	456	542	574
ln ANOME	2	118	135	202	271	326	358	421	460	501
(Volino)	4	165	140	162	269	336	378	481	503	553
	6	161	177	195	270	336	424	446	501	580
	8	115	207	264	316	362	421	471	546	581
	10	182	236	279	278	323	386	426	533	546
Bartlett's Test	2	48	107	188	260	317	351	410	456	494
	4	92	168	195	305	379	435	509	549	583
	6	80	193	258	320	411	491	503	568	633
	8	38	201	272	393	435	494	573	598	646
	10	69	205	297	359	443	506	549	610	629
F-Test (Two Variances)	2	60	117	191	262	318	352	413	456	494
Levene's Test	2		0	139	28	161	117	191	199	268
	4		0	181	59	170	142	288	239	326
	6		0	163	50	187	141	233	237	346
	8		0	135	41	190	120	285	230	328
	10		0	157	42	155	112	251	217	325

For the simulations of Case 6, involving Weibull-distributed data with the k[th] group having mean and variance different from the other k-1 groups, the number of rejections of the null hypothesis was on average double the rejection rate in the homogenous case (see Table 2: all k groups the same). This was a fairly common phenomenon across all tests. Bartlett's test was the most sensitive to the difference in the k[th] group, followed by Bartlett and Kendall's ln ANOVA. The results of Levene's test once again revealed the "odd-even" effect mentioned earlier and, as usual, was the most conservative in declaring differences among the within-group variances.

Note that the actual variance for the first k-1 Weibull-distributed groups was 0.376, as opposed to a variance in the k^{th} group of 1.503, determined by the formula for Weibull variance, namely

$$\sigma^2 = \varpi^2 \left\{ \Gamma\left[1+\frac{2}{\phi}\right] - \left\{ \Gamma\left[1+\frac{1}{\phi}\right] \right\}^2 \right\}. \tag{24}$$

Table 8: Results for Case 7, Gamma Distributions with ψ = 2.5, and λ = 2 (1^{st} k-1 groups), ψ = 5.0 (k^{th} group)

(Number of rejections of the null hypothesis out of 1000)

		n								
HOV Test	k	2	3	4	5	6	7	8	9	10
ln ANOVA	2	82	84	103	130	167	185	216	204	248
(Bartlett and Kendall)	4	124	113	145	175	190	240	234	267	305
	6	136	105	183	181	249	265	282	302	335
	8	152	134	166	219	253	303	322	349	364
	10	189	174	199	245	274	301	389	397	409
ln ANOME	2	82	84	103	130	167	185	216	204	248
(Volino)	4	132	111	129	162	166	218	232	264	287
	6	152	116	176	179	211	245	261	280	300
	8	171	137	165	204	226	254	270	299	305
	10	217	185	197	206	232	238	309	311	317
Bartlett's Test	2	46	68	94	118	162	181	214	201	240
	4	49	104	142	177	187	236	251	265	314
	6	45	103	192	214	257	284	302	324	346
	8	46	134	187	246	274	324	357	377	378
	10	68	160	216	283	280	316	425	408	456
F-Test (Two Variances)	2	62	74	97	118	163	181	214	202	241
Levene's Test	2		0	86	13	70	44	93	81	110
	4		0	107	19	80	34	103	65	112
	6		0	99	10	72	33	74	55	99
	8		0	113	10	51	24	81	48	84
	10		0	95	9	69	33	73	50	73

In an examination of power (1-β) of the tests performed on variances from Gamma distributions, of tests that rely on normality (including ln ANOVA, ln ANOME, and Bartlett's test), Bartlett's test again provided the most power when k > 4 and n > 3. ln ANOVA and ln ANOME were nearly as powerful. Levene's test was conservative to the extent that it was almost worthless as a test of equality of variances. Not surprisingly, this same test was much less powerful for an odd number of replicates than for an even number of replicates.

Summary and Conclusions

An analysis-of-means-type test (ln ANOME) for determining differences in variability between subgroups from normal populations was presented. Its merits parallel those of the usual analysis of means in that the result of the test is a graphical representation of the differences due to the various combinations of the variables. ln ANOME has the advantage of providing its own estimate of the error variances and of relying on the same tables that are commonly available for the standard ANOME procedure.

In cases of normality, the ln ANOME test, in its present form, is less able to maintain stable Type-I error rates than is the commonly accepted Bartlett's test. In cases of non-normality, in terms of both Type-I error rate and power, ln ANOME is comparable to and sometimes better than other tests like Bartlett's test and ln ANOVA. Moreover, the results from this study suggest that the ln ANOME procedure would allow for fewer than the "n = 5 replicates" cutoff that Bartlett and Kendall recommended for approximate normality of ln S^2 The expected Type-I and Type-II error rates were maintained for n = 4 replicates, and sometimes even for n = 3 replicates.

Bartlett's test is usually preferred in the literature for comparing variances from normal distributions. Results obtained in this investigation confirm that Bartlett's test provides good power when the assumption of normality is tenable. When that assumption is in doubt, ln ANOME or ln ANOVA may provide slightly more stable error rates of both types.

Levene's test, with Brown and Forsythe's modification, is plagued by an "odd-even" effect in its ability to maintain both the Type-I error rate and power. For n=3 replicates, the test never led to a rejection of the null hypothesis for any simulation. This test is an option both in the MEANS statement of PROC GLM in SAS and as part of the standard output in MINITAB, so an understanding of its questionable performance is very important in its application.

Future Work

G. E. P. Box[15] suggested a stability adjustment for ln ANOVA that calls for randomly assigning replicates within the "cells" of the experiment to two or more subgroups, and then computing their natural logarithms so as to have more than one ln S^2 estimate per cell. He then recommended using these as the replicates of the experiment and computing the error variance from these estimates as opposed to using the theoretical error variance. Although this investigation did not explore such an approach, it could easily be incorporated into the ln ANOME procedure.

Also, more simulation work could be done for those cases studied in order to give a broader, more complete coverage to differences between group variances. Then empirically-derived OC-curves could be constructed to better track the performance of ln ANOME and the other tests in this study.

Finally, Bartlett & Kendall and L. S. Nelson all made mention of the fact that these tests need not be only of the form ln S^2, but could also be based on ln S, ln Range, etc. It would be interesting to investigate these other forms in the graphical setting of the ANOME type.

References

1. Nelson, L. S. (1981), "Analysis of Heterogeneity of Variation," *Journal of Quality Technology* 13, pp. 143-144
2. Wludyka, P. S., and Nelson, P. S. (1997), "An Analysis-of-Means-Type Test for Variances From Normal Populations," *Technometrics* 39, pp. 274-285
3. Bartlett, M. S., and Kendall, D. G. (1946), "The Statistical Analysis of Variance-Heterogeneity and the Logarithmic Transformation," *Journal of the Royal Statistical Society* 8, Supplement, pp. 128-138
4. Scheffé, H. (1959), *The Analysis of Variance*, 1st ed.; John Wiley & Sons, New York, New York
5. Lorenzen, T. J., and Anderson, V. L. (1993), *Design of Experiments: A No-Name Approach*, 1st ed.; Marcel Dekker, Inc., New York, New York
6. Levene, H. (1960), "Robust Tests for Equality of Variances," *in Contributions to Probability and Statistics*, ed. I. Olkin, Palo Alto, Calif.: Standford University Press, pp. 278-292
7. Brown, M. B., and Forsythe, A. B. (1974), "Robust Tests for Equality of Variances," *Journal of the American Statistical Association* 69, pp. 364-367
8. Bartlett, M. S. (1937), "Properties of Sufficiency and Statistical Tests," *Proceedings of the Royal Statistical Society of London*, Ser. A 160, pp. 268-282
9. Schilling, E. G. (1973), "A Systematic Approach to the Analysis of Means: Part I. Analysis of Treatment Effects," *Journal of Quality Technology* 5, pp. 93-108

10. Ott, E. R., and Schilling, E. G. (1990), *Process Quality Control: Troubleshooting and Interpretation of Data*, 2nd ed.; McGraw-Hill, Inc., New York, New York

11. Nelson, L. S. (1983), "Exact Critical Values for Use with the Analysis of Means," *Journal of Quality Technology* 15, pp. 40-44

12. Neyman, J, and Pearson, E. S. (1931), "On the Problem of k Samples," *Bulletin de l'Academie Polonaise des Sciences et des Lettres*, A, pp. 460-481

13. Dodson, B, and Nolan, D (1995), *Reliability Engineering Bible*, 1st ed.; Quality Publishing, Tucson, Arizona

14. Conover, W. J., Johnson, M. E., Johnson, M. M. (1981), "A Comparative Study of Tests for Homogeneity of Variances, with Applications to the Outer Continental Shelf Bidding Data," *Technometrics* 23, pp. 351-361

15. Box, G. E. P. (1953), "Non-Normality and Tests on Variances," *Biometrika* 40, pp. 318-335

Appendix A-1

Exact Factors for One-Way Analysis of Means, H_α

Note: Two-Sided, and for Infinite Error Degrees of Freedom

Table A-1. Exact Factors for One-Way Analysis of Means, H_α

	Significance Level, α			
k	**0.10**	**0.05**	**0.01**	**0.001**
2	1.163	1.386	1.822	2.327
3	1.67	1.91	2.38	2.92
4	1.90	2.14	2.61	3.17
5	2.05	2.29	2.75	3.33
6	2.15	2.39	2.87	3.43
7	2.24	2.48	2.94	3.52
8	2.31	2.54	3.01	3.59
9	2.38	2.60	3.07	3.65
10	2.42	2.66	3.12	3.69
11	2.47	2.70	3.17	3.73
12	2.51	2.74	3.20	3.76
13	2.55	2.77	3.23	3.80
14	2.57	2.79	3.26	3.83
15	2.60	2.83	3.28	3.85
16	2.63	2.86	3.31	3.87
17	2.66	2.88	3.34	3.90
18	2.68	2.90	3.35	3.92
19	2.70	2.92	3.38	3.93
20	2.72	2.94	3.39	3.96

The values in this table were obtained from Table A.8, E. R. Ott and E. G. Schilling[12].

Appendix A-2

Sidak Factors for Analysis of Means for Treatment Effects, h_α^*

Note: Two-Sided, and for Infinite Error Degrees of Freedom

Table A-2. Sidak Factors for Analysis of Means for Treatment Effects, h_α^*

	Significance Level, α			
k	**0.10**	**0.05**	**0.01**	**0.001**
2	1.645	1.960	2.576	3.291
3	2.114	2.388	2.934	3.588
4	2.226	2.491	3.022	3.662
5	2.311	2.569	3.089	3.719
6	2.378	2.631	3.143	3.765
7	2.434	2.683	3.188	3.803
8	2.481	2.727	3.226	3.836
9	2.523	2.766	3.260	3.865
10	2.560	2.800	3.289	3.891
11	2.592	2.830	3.316	3.914
12	2.622	2.858	3.340	3.935
13	2.649	2.883	3.362	3.954
14	2.674	2.906	3.383	3.972
15	2.697	2.928	3.402	3.988
16	2.718	2.948	3.419	4.004
17	2.738	2.966	3.436	4.018
18	2.757	2.984	3.451	4.031
19	2.774	3.000	3.466	4.044
20	2.791	3.016	3.480	4.056
24	2.849	3.071	3.528	4.099
30	2.920	3.137	3.587	4.150
40	3.008	3.220	3.661	4.215
60	3.129	3.335	3.764	4.306

The values in this table were obtained from Table A.19, E. R. Ott and E. G. Schilling[12].

Appendix A-3

Example of Generating Random Numbers (Two-Parameter Weibull)

The cumulative distribution function of the two-parameter Weibull density is given by

$$F(y) = 1 - e^{-\left(\frac{y}{\varpi}\right)^{\phi}} \qquad \text{(A-3-1)}$$

where

ω = Weibull Scale Parameter (Characteristic-Life), and

ϕ = Weibull Shape Parameter.

If we let r be a uniformly-distributed computer-generated random number on the [0, 1] interval, we can set this equal to the c.d.f. of the two-parameter Weibull density to give us,

$$r = 1 - e^{-\left(\frac{y}{\varpi}\right)^{\phi}} \qquad \text{(A-3-2)}$$

Then, we can solve for y as follows:

$$1 - r = e^{-\left(\frac{y}{\varpi}\right)^{\phi}} \qquad \text{(A-3-3)}$$

$$\ln(1 - r) = -\left(\frac{y}{\varpi}\right)^{\phi} \qquad \text{(A-3-4)}$$

$$[-\ln(1 - r)]^{\frac{1}{\phi}} = \frac{y}{\varpi} \qquad \text{(A-3-5)}$$

$$\varpi[-\ln(1 - r)]^{\frac{1}{\phi}} = y \qquad \text{(A-3-6)}$$

But since 1-r is also uniformly distributed on the [0, 1] interval, we can further simplify this expression by replacing 1-r with r to yield

$$y = \varpi[-\ln(r)]^{\frac{1}{\phi}} \tag{A-3-7}$$

Therefore, the ability of the various versions of BASIC-type programming languages to produce uniformly-distributed random numbers (with the RND function) allows us to construct random numbers from the Weibull density (and many others).

Appendix A-4

Visual Basic Subroutines for Simulations in Excel

<u>Subroutines</u>:

RandNormal:	Generates a specified number of Normal-distributed random numbers.
RandWeibull:	Generates a specified number of Weibull-distributed random numbers.
RandGamma:	Generates a specified number of Gamma-distributed random numbers.
Bartlett:	Computes Bartlett's test statistic and p-value.
Levene:	Computes Levene's test statistic and p-value.
ANOVA:	Performs a one-way ANOVA, and computes the F-statistic and p-value.
lnANOVA:	Performs Bartlett and Kendall's log ANOVA, and computes the F-statistic and p-value.

```
Sub RandNormal(nVars As Integer, nRand As Integer, dblMean As Double, _
    dblStandardDev As Double, rngOutput As Range, Optional nSeed As Integer)
    Dim i As Integer, j As Integer, Pi As Double
    Pi = Application.Pi()

    On Error GoTo StartOver

    If nSeed <> 0 Then
        Randomize (nSeed)
    Else
        Randomize
    End If
    For i   1 To nVars
        For j   1 To nRand
StartOver:
            If Rnd <= 0.499999 Then
                rngOutput.Offset(j   1, i   1).Value  _
                    (Sqr(-2 * Log(Rnd)) * Cos(2 * Pi * Rnd)) * dblStandardDev + dblMean
                'Debug.Print "Cosine"
            Else
                rngOutput.Offset(j   1, i   1).Value  _
                    (Sqr(-2 * Log(Rnd)) * Sin(2 * Pi * Rnd)) * dblStandardDev + dblMean
                'Debug.Print "Sine"
            End If
        Next j
    Next i
End Sub
```

```
Sub RandWeibull(nVars As Integer, nRand As Integer, dblScale As Double, _
    dblShape As Double, rngOutput As Range, Optional nSeed As Integer)
    Dim i As Integer, j As Integer

    On Error GoTo ReDo

    If nSeed <> 0 Then
        Randomize (nSeed)
    Else
        Randomize
    End If
    For i   1 To nVars
        For j   1 To nRand
ReDo:
            rngOutput.Offset(j   1, i   1).Value   _
                 dblScale * (-Log(Rnd)) ^ (1 / dblShape)
        Next j
    Next i
End Sub

Sub RandGamma(nVars As Integer, nRand As Integer, dblAlpha As Double, _
    dblBeta As Double, rngOutput As Range, Optional nSeed As Integer)
    Dim i As Integer, j As Integer, nA As Integer, k As Integer
    Dim dblProd As Double, dblA As Double, dblB As Double
    Dim dblq As Double, dblRndy1 As Double, dblRndy2 As Double
    Dim dblz As Double, dblW As Double, dblXGamma As Double

    'Note dblAlpha must be non-integer

    On Error GoTo RepeatThis

    If nSeed <> 0 Then
        Randomize (nSeed)
    Else
        Randomize
    End If

    nA = Int(dblAlpha)
    For i = 1 To nVars
        For j   1 To nRand
RepeatThis:
            dblProd = 1
            For k   1 To nA
                dblProd   dblProd * Rnd
            Next k
            dblq   -(Log(dblProd))
            dblA   dblAlpha   nA
            dblB   1   dblA
TryItAgain:
            dblRndy1 = Rnd
            dblRndy1 = dblRndy1 ^ (1 / dblA)
            dblRndy2 = Rnd
            dblRndy2 = dblRndy2 ^ (1 / dblB)
            If dblRndy1 + dblRndy2 <= 1 Then
                dblz   dblRndy1 / (dblRndy1 + dblRndy2)
            Else
                GoTo TryItAgain
            End If
            dblW = Rnd
            dblW = -(Log(dblW))
            dblXGamma   (dblq + dblz * dblW) * (1 / dblBeta)
            rngOutput.Offset(j   1, i   1).Value   dblXGamma
        Next j
    Next i
End Sub
```

```
Function Bartlett(nGroups As Integer, nReplicates As Integer) As Double
    Dim i As Integer, j As Integer, rngData As Range
    Dim N As Integer, a As Integer
    Dim ni() As Integer, Sums() As Double, SumsSq() As Double
    Dim SubVar() As Double, Subgroups() As String, dblSPooled As Double
    Dim q As Double, c As Double

    Set rngData   Worksheets("Data Sheet").Range("B2:" & Chr(65 + nGroups) & _
                  1 + nReplicates)
    N   nGroups * nReplicates

    ReDim ni(nGroups) As Integer
    ReDim Sums(nGroups) As Double
    ReDim SumsSq(nGroups) As Double
    ReDim SubVar(nGroups) As Double

    For i   1 To nGroups
        ni(i) = nReplicates
        For j = 1 To nReplicates
            Sums(i)   Sums(i) + rngData(j, i).Value
            SumsSq(i)   SumsSq(i) + (rngData(j, i).Value) ^ 2
        Next j
    Next i

    For i   1 To nGroups
        SubVar(i)   (SumsSq(i)   ((Sums(i)) ^ 2 / ni(i))) / (ni(i)   1)
        dblSPooled   dblSPooled + (ni(i)   1) * SubVar(i)
    Next i

    a   nGroups

    dblSPooled = dblSPooled / (N   a)

    For i   1 To a
        q   q + (ni(i)   1) * Log10(SubVar(i))
        c   c + 1 / (ni(i)   1)
    Next i
    q   (N   a) * Log10(dblSPooled)   q
    c   1 + 1 / (3 * (a   1)) * (c   1 / (N   a))
    Bartlett = 2.3026 * q / c
    BartlettTS = Bartlett
    Bartlettp = Application.ChiDist(BartlettTS, a   1)
End Function
```

```
Function Levene(nGroups As Integer, nReplicates As Integer) As Double
    Dim i As Integer, j As Integer
    Dim N As Integer, a As Integer
    Dim ni() As Integer, nTemp As Integer
    Dim Subgroups() As String, dblTemp() As Double
    Dim Medians() As Double, GroupedData() As Double
    Dim Maxni As Integer, rngTemp As Range, dblTemp2 As Double
    Dim Vij() As Double, Vidot() As Double, Vdotdot As Double

    N   nGroups * nReplicates
    a   nGroups

    ReDim ni(nGroups) As Integer
    ReDim Medians(a) As Double
    For i   1 To nGroups
        ni(i) = nReplicates
        Set rngTemp   Worksheets("Data Sheet").Range("B2:B" & nReplicates + 1)
        Medians(i)   Application.Median(rngTemp.Offset(0, i   1))
    Next i

    ReDim Vij(nGroups, nReplicates) As Double
    For i   1 To nGroups
        For j = 1 To nReplicates
            Set rngTemp   Worksheets("Data Sheet").Range("B2:B" & nReplicates + 1)
            Set rngTemp   rngTemp.Offset(0, i   1)
            Vij(i, j)   Abs(rngTemp.Range("A" & j).Value   Medians(i))
        Next j
    Next i
    ReDim Vidot(a) As Double
    For i   1 To nGroups
        For j = 1 To nReplicates
            Vidot(i)   Vidot(i) + Vij(i, j)
        Next j
        Vdotdot = Vdotdot + Vidot(i)
        Vidot(i) = Vidot(i) / ni(i)
    Next i
    Vdotdot   Vdotdot / N
    For i   1 To nGroups
        Levene = Levene + nReplicates * (Vidot(i)   Vdotdot) ^ 2
    Next i
    Levene = Levene / (nGroups   1)
    dblTemp2 = 0
    For i = 1 To nGroups
        For j = 1 To nReplicates
            dblTemp2   dblTemp2 + (Vij(i, j)   Vidot(i)) ^ 2
        Next j
    Next i
    dblTemp2 = dblTemp2 / (N   a)
    Levene = Levene / dblTemp2
    LeveneTS = Levene
    Levenep   Application.FDist(LeveneTS, a   1, N   a)
End Function
```

```
Function ANOVA(nGroups As Integer, nReplicates As Integer) As Double
    Dim dblSSTr As Double, dblSSTO As Double, dblSSE As Double
    Dim dblFStar As Double, Sums() As Double, dblOSum As Double
    Dim i As Integer, j As Integer, k As Integer, N As Integer
    Dim rngTemp As Range, shtDataSheet As Worksheet

    k = nGroups
    N = k * nReplicates

    ReDim Sums(k) As Double
    Set shtDataSheet = Worksheets("Data Sheet")
    Set rngTemp   shtDataSheet.Range("B2:" & Chr(65 + nGroups) & nReplicates + 1)

    dblOSum   Application.Sum(rngTemp)

    For i = 1 To k
        Sums(i)   Application.Sum(rngTemp.Columns(i))
        dblSSTr   dblSSTr + (Sums(i) ^ 2)
        For j   1 To nReplicates
            dblSSTO = dblSSTO + rngTemp(j, i) ^ 2
        Next j
    Next i
    dblSSTr   dblSSTr / nReplicates   (dblOSum ^ 2) / N
    dblSSTO   dblSSTO   (dblOSum ^ 2) / N
    dblSSE   dblSSTO   dblSSTr
    dblFStar   (dblSSTr / (k   1)) / (dblSSE / (N - k))
    ANOVAp   Application.FDist(dblFStar, k   1, N   k)
    ANOVA = dblFStar
End Function

Function lnANOVA(nGroups As Integer, nReplicates As Integer) As Double
    Dim i As Integer, j As Integer, k As Integer, N As Integer
    Dim rngTemp As Range, shtDataSheet As Worksheet
    Dim dblErrorTerm As Double, SampVar() As Double
    Dim dblSSTr As Double
    Dim dblFStar As Double, dblOSum As Double

    k = nGroups
    N = k * 1

    ReDim SampVar(nGroups) As Double
    Set shtDataSheet = Worksheets("Data Sheet")
    Set rngTemp = shtDataSheet.Range("B2:" & Chr(65 + nGroups) & nReplicates + 1)

    For i   1 To k
        SampVar(i) = Application.Var(rngTemp.Columns(i))
        SampVar(i) = Log(SampVar(i))            'ln si^2
        dblSSTr = dblSSTr + (SampVar(i) ^ 2)
        dblOSum = dblOSum + SampVar(i)
    Next i

    dblSSTr = dblSSTr   (dblOSum ^ 2) / N

    dblErrorTerm = 2 / (nReplicates   1) + 2 / ((nReplicates   1) ^ 2) + _
        4 / (3 * (nReplicates   1) ^ 3)   16 / (15 * (nReplicates   1) ^ 5)

    dblFStar   (dblSSTr / (k   1)) / dblErrorTerm

    lnANOVAp   Application.FDist(dblFStar, k   1, 100000000)
    lnANOVA   dblFStar
End Function
```

Appendix A-5

Type-I Error Rates for ANOVA

While it was not the intent of this work to judge the robustness of the usual ANOVA for testing means, these analyses were performed nonetheless to examine the effects of non-normality on the Type-I error rates. In each cell in the table below, the number of times out of one thousand that the null hypothesis of equal treatment means was rejected at the $\alpha = 0.05$ level was tallied and recorded. That ANOVA is robust in the face of departures from the normality assumption is common knowledge. Not surprisingly, then, we see that ANOVA has little trouble maintaining the Type-I Error rate (here versus the 50 out of 1000, or the 5% level) for non-normal distributions.

Type-I Error Rate for ANOVA

		n								
Case	k	2	3	4	5	6	7	8	9	10
1	2	52	56	47	46	54	52	43	58	58
Normal	3	48	40	42	50	52	49	53	61	52
	4	35	55	57	40	53	54	42	49	54
	5	58	44	46	47	55	56	51	54	62
	6	63	47	50	39	47	43	46	48	60
	7	44	54	55	41	44	45	47	52	49
	8	50	50	64	53	53	43	61	54	47
	9	46	54	53	49	51	51	54	59	47
	10	64	55	46	47	54	54	41	40	54
2	2	51	57	46	44	41	43	55	47	47
Weibull	4	49	40	51	56	57	47	51	37	52
	6	52	44	35	62	52	39	59	42	52
	8	62	51	41	43	49	48	43	45	47
	10	69	68	63	56	42	40	48	52	59
3	2	53	43	46	51	57	45	45	51	56
Gamma	4	54	51	58	30	39	38	54	52	41
	6	52	39	38	35	51	46	41	40	44
	8	51	44	39	41	52	59	50	63	35
	10	48	48	51	60	47	46	49	44	45