GAMMA AND LOGNORMAL DISTRIBUTIONS
AS MODELS IN STUDYING
FOOD-CHAIN KINETICS

Battelle

Pacific Northwest Laboratories
Richland, Washington 99352

MAY 1973

Prepared for the U.S. Atomic Energy
Commission under Contract AT(45-1):1830

NOTICE

This report was prepared as an account of work sponsored by the United States Government. Neither the United States nor the United States Atomic Energy Commission, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately-owned rights.

PACIFIC NORTHWEST LABORATORY
operated by
BATTELLE
for the
U.S. ATOMIC ENERGY COMMISSION
Under Contract AT(45-1)-1830

AEC-RL RICHLAND, WASH.

3 3679 00062 3043

GAMMA AND LOGNORMAL DISTRIBUTIONS AS MODELS
IN STUDYING FOOD-CHAIN KINETICS

by

L. L. Eberhardt¹ and R. O. Gilbert²

May 1973

BATTELLE
PACIFIC NORTHWEST LABORATORIES
RICHLAND, WASHINGTON 99352

¹Research Associate
Ecosystems Department

²Senior Research Scientist
Statistics Section
Systems Department

TABLE OF CONTENTS

<u>Section</u>	<u>Page</u>
1. INTRODUCTION AND SUMMARY	1
2. BASIC DEFINITIONS AND METHODS	4
2.1 Density Functions	4
2.2 Generation of Lognormal and Gamma Variables	6
3. DISTINGUISHING BETWEEN LOGNORMAL AND GAMMA DISTRIBUTIONS	9
3.1 Skewness and Kurtosis	9
3.2 "Separate Families" Tests	10
4. DISTRIBUTION THEORY	32
4.1 The s^2 and z Statistics	32
4.2 Moments of the Lognormal Distribution	34
4.3 Expected Value of z for Gamma Variables	36
4.4 Variance of Log-transformed Random Variables from Gamma Distributions	40
4.5 Summary of Results on Expected Values	41
4.6 Tests of Significance for Equality of Coefficients of Variation	42
4.7 Bartlett's Test of Equality of Variances	44
4.8 Scheffé's Test of Equality of Variances	48
4.9 Confidence Limits on the Coefficient of Variation	53
5. SIMULATION RESULTS	55
5.1 Tests for Equality of Coefficients of Variation	55
5.2 Robustness of Bartlett's Test	60
5.3 Robustness of Scheffé's Test	62
5.4 Power of Scheffé's Test	64
6. CONCLUSIONS	70
7. REFERENCES	73
8. APPENDIX: LISTING OF COMPUTER PROGRAMS USED IN THE STUDY	75

TABLES

	<u>Page</u>
3.1 Measures of Skewness (γ_1) and Kurtosis (γ_2) for Simulated Gamma Distributions	11
3.2 Measures of Skewness (γ_1) and Kurtosis (γ_2) for Simulated Lognormal Distributions	12
3.3 "Actual" Type I Errors of T_{fN} as Compared with Nominal Levels	21
3.4 "Actual" Type I Errors of T_{gN} as Compared with Nominal Levels	22
3.5 The Mean, Variance, Skewness and Kurtosis of T_{fN} for Lognormal Observations	25
3.6 The Mean, Variance, Skewness and Kurtosis of T_{gN} for Gamma Observations	26
3.7 Power of T_{fN} to Detect Gamma Distributed Variates	27
3.8 Power of T_{gN} to Detect Lognormal Distributed Variates	28
3.9 The Mean, Variance, Skewness and Kurtosis of T_{fN} for Gamma Observations	29
3.10 The Mean, Variance, Skewness and Kurtosis of T_{gN} for Lognormal Observations	30
4.1 Comparison of τ and τ' Computed Using Equations (4.23) and (4.25) When $c = 0.20$	54
5.1 Simulated Power for Tests $F_{\ell n}$ and F_g Using s^2 and z as Criteria on Gamma and Lognormal Variates	58
5.2 Simulated and Expected Power of $F_{\ell n}$ on Log-transformed Variables from the Lognormal Distribution	61
5.3 Tests of Equality of Variances for Log-transformed Random Variables from a Simulated Gamma Distribution	63
5.4 Comparison of Simulated and Tabular Power of Scheffé's Test for Homogeneity of Variances	68

FIGURES

	<u>Page</u>
3.1 Density Functions of the Gamma and Lognormal Distributions for $c = 1.00$ and 0.707 When $E(x) = 1.0$	19
3.2 Density Functions of the Gamma and Lognormal Distributions for $c = 0.500$ and 0.301 When $E(x) = 1.0$	20
3.3 Power of T_{fN} to Detect Gamma Observations	24
3.4 Power of T_{gN} to Detect Lognormal Observations	24

1.0 INTRODUCTION AND SUMMARY

The passage of trace substances through food-chains is an important and a dynamic process. It is of concern both in understanding those basic processes lumped under the heading of "biological productivity" and in studying the dispersal and effects of contaminants like DDT. The analysis of field data is largely dependent on much the same kinds of kinetic models as are used in assessing movement of tracers within a single organism. The complications of "compartment models" are widely recognized, from both physiological and biometric points of view. Their ecological analogs admit additional difficulties by way of sampling problems and questions of system-identification.

This report deals with some aspects of a basic question -- the choice of a suitable frequency distribution to represent observations of the concentration of substances such as radionuclides and DDT in samples of biota. Quite a variety of data on radionuclides indicates that the frequency distributions are markedly skewed, even in the face of efforts to limit analysis to data taken in relatively small areas and short time spans. Such attempts to obtain uniform conditions as well as the relatively high costs of sample analyses usually results in small samples, and often only the coefficient of variation is available for comparisons. Data on radionuclides (principally from fallout due to nuclear explosions) indicate the coefficient of variation to be frequently in the range of 10 to 40 percent (Eberhardt, 1964; Remmenga and Whicker, 1967) and apparently constant over time for a given species and environment. Some data on concentrations of DDT in biota indicate that somewhat the same circumstances may obtain for pesticides.

The log-normal and gamma distributions have been proposed as appropriate models for frequency distributions of radionuclide concentrations in biota and substrates. Ellett and Brownell (1964) have examined one of the largest sets

of data available thus far, and conclude that the gamma distribution is to be preferred to the log-normal. However, it is convenient to assume lognormality since a logarithmic transformation then permits use of normal theory on the data. We are especially interested here in the consequences of analysis based on log-transformed data from a gamma distribution, i.e. incorrectly assuming lognormality. Accordingly, the present study is concerned in part with examining the behavior of two F tests, F_{ln} and F_g that utilize the criteria s^2 and z , respectively, for testing the hypothesis that two coefficients of variation are equal, where s^2 is the sample variance of log-transformed data, and

$$z = \ln \bar{x} - \overline{\ln x} \quad .$$

F_{ln} and F_g are appropriate tests if the data are from lognormal and gamma distributions, respectively. We attempt here to appraise the consequences of using an F test appropriate for one distribution when the other distribution in fact holds. Our overall conclusion from our limited simulation results is that either criterion give about the same performance in terms of power and robustness, no matter whether the data are lognormal or gamma. One may thus feel reasonably confident about making comparisons with either test. This in turn suggests we can feel somewhat better about using log-transformed data for "normal-theory" tests.

In addition to the above results, we examine here, using computer simulations, the robustness of Bartlett's (1937) test for homogeneity of variances and the power and robustness of Scheffé's (1959:83) test of equality of variances when the data are log-transforms of gamma variates. As might be expected, Bartlett's test proved not to be robust in this situation. Scheffé's test, however, appeared to be relatively robust and to have reasonably good power for the cases studied.

Also, we have approximated (using computer simulations) the power and Type I error of two "separate families" tests derived by Jackson (1969) for distinguishing between lognormal and gamma distributions. Our results indicate that the power of both tests can be rather low unless the number of observations is 200 or more, depending of course on the particular lognormal and gamma density curves being compared.

Some simulations were also conducted that illustrated the substantial sampling variability associated with the sample coefficients of skewness and kurtosis, suggesting that for the small sample sizes encountered in practice, these measures will not be very useful in distinguishing between lognormal and gamma distributions.

Many results in distribution theory that would be useful in this study are either not available or require approximations for practical use. It has thus been necessary to resort to simulation as a means of checking approximations and providing some hints as to what may happen in those cases where approximate results are not available. An alternative to simulation -- "moment" approximations to sampling distributions -- seems worth exploring, and it may well be that some exact results could be obtained for particular cases.

It may be worth emphasizing here that the present study is not, for the most part, directly concerned with the comparisons of means. The emphasis is instead on the comparison of coefficients of variation or of variances. This is, of course, a consequence of the fact that mean and variance are not independent in the lognormal and gamma distributions. The general observation of a tendency for a constant coefficient of variation thus leads one to want to test that hypothesis first, before going on to compare means.

No doubt it should also be noted that neither the lognormal nor the gamma distribution may be "the" appropriate distribution in studies of food-chain

kinetics. However, it seems abundantly clear that the normal distribution is ruled out of consideration. A compelling argument along those lines is that the coefficient of variation seems roughly constant for sets of data for which the means may range over many orders of magnitude of concentration of some trace substance. For example, concentrations of radionuclides may readily be studied at picocurie levels (10^{-12} curies) and appear to have similar behavior up to microcurie (10^{-6} curies) and higher concentrations. Much the same situation holds with DDT which can now be reliably measured down to parts per trillion. Normal additive errors are unlikely to be appropriate in such circumstances.

2.0 BASIC DEFINITIONS AND METHODS

2.1 Density Functions

The lognormal distribution is described in some detail in the book by Aitchison and Brown (1969). The density function is given by

$$f(x) = \frac{1}{\sigma x \sqrt{2\pi}} \exp \left[-\frac{1}{2\sigma^2} (\ln x - \mu)^2 \right] \quad (2.1)$$

for $x > 0$, $-\infty < \mu < \infty$, $\sigma > 0$. We have that

$$E(x) = \exp\left[\mu + \frac{1}{2}\sigma^2\right]$$

and

$$\text{Var}(x) = [\exp(2\mu + \sigma^2)] [\exp(\sigma^2) - 1] .$$

Hence the squared coefficient of variation is given by

$$c^2 = \frac{\text{Var}(x)}{[E(x)]^2} = \exp(\sigma^2) - 1 . \quad (2.2)$$

The coefficients of skewness and kurtosis for any distribution are defined by

$$\gamma_1 \equiv \mu_3 \mu_2^{-3/2}$$

and

(2.3)

$$\gamma_2 \equiv (\mu_4/\mu_2^2) - 3$$

respectively, where μ_i is the i^{th} moment about the mean. In the case of the lognormal distribution we find that

$$\gamma_1 = 3c + c^3$$

and

$$\gamma_2 = 16c^2 + 15c^4 + 6c^6 + c^8 .$$

The gamma distribution may be defined as (Hahn and Shapiro (1967))

$$f(x) = \frac{\lambda}{\Gamma(\eta)} (\lambda x)^{\eta-1} e^{-\lambda x} \quad (2.4)$$

for $x \geq 0$, $\lambda > 0$, $\eta > 0$, where $\Gamma(\eta)$ is the well known gamma function. It follows that

$$E(x) = \eta/\lambda \quad \text{and} \quad \text{Var}(x) = \eta/\lambda^2$$

which implies

$$c^2 = 1/\eta .$$

Furthermore, for the gamma distribution we have that

$$\gamma_1 = 2c \quad \text{and} \quad \gamma_2 = 6c^2 .$$

The gamma distribution may also be expressed in the form

$$f(x) = \frac{x^{\beta/2-1} e^{-x/2\alpha}}{(2\alpha)^{\beta/2} \Gamma(\beta/2)} , \quad (2.5)$$

where

$$E(x) = \alpha\beta , \quad \text{Var}(x) = 2\alpha^2\beta \quad \text{and} \quad c^2 = 2/\beta .$$

The relationship between the parameters in equations (2.4) and (2.5) is

$$\beta = 2\eta \quad \text{and} \quad \alpha = (2\lambda)^{-1}.$$

For later use we note that if we let $\alpha = 1$ and $\beta = \nu$, then the density function given by (2.5) becomes

$$f(x^2) = \frac{x^{\nu/2-1} e^{-x^2/2}}{2^{\nu/2} \Gamma(\nu/2)}, \quad (2.6)$$

which is the well known chi-square distribution with ν degrees of freedom (d.f.).

2.2 Generation of Lognormal and Gamma Variates

The various simulations utilized in this report required the generation of lognormal and gamma variates on the computer. Pseudo-lognormal variates were obtained for the simulation studies in Sections 4 and 5 by first generating a pseudo-random $N(0,1)$ variate y using the formula

$$y = (-2 \ln u_1)^{1/2} \cos 2\pi u_2$$

suggested by Box and Muller (1958), where u_1 and u_2 are two uniformly distributed random variables generated using the uniform number generator RN4. The lognormal variates utilized in Section 3 were generated using the subroutine RN2 to generate two uniform numbers u_1 and u_2 which were then transformed to two $N(0,1)$ variates y_1 and y_2 using

$$y_1 = (-2 \ln u_1)^{1/2} \cos 2\pi u_2$$

and

$$y_2 = (-2 \ln u_1)^{1/2} \sin 2\pi u_2$$

(Box and Muller (1958)). Both RN2 and RN4 are listed in the Appendix. These two generators have been evaluated by Gilbert (1973) for randomness and periodicity. Both appear to have desirable statistical properties.

Variates x following the lognormal distribution (equation (2.1)) with parameters μ and σ were obtained using the transformation

$$x = \exp(\mu + \sigma y) .$$

The parameters μ and σ were determined by first specifying the coefficient of variation c , then solving for σ using the equation

$$\sigma = [\ln(1 + c^2)]^{1/2} ,$$

which follows directly from equation (2.2).

Virtually all of the analyses considered here depend on σ (or its equivalent in the gamma distribution), so the value of μ is not of much concern. Hence, μ was arbitrarily set at

$$\mu = \frac{\sigma}{c} .$$

We note, however, this implies $c = \sigma/\mu$ which is the coefficient of variation of normal variates obtained by taking logarithms of lognormal observations. We note that when c is small then

$$\ln(1 + c^2) \doteq c^2$$

and σ is thus approximately equal to c which implies μ is roughly unity.

Pseudo-gamma random variables x were obtained using the expression

$$x = -\lambda^{-1} \sum_{i=1}^n \ln(1 - u_i) , \quad (2.7)$$

where u_i is a random variable from a uniform distribution on the interval (0,1) generated using the RN2 generator listed in the Appendix. For the simulations in Sections 4 and 5 we have arbitrarily set $\lambda = 1$ which implies $E(x) = \eta$. In Section 3 we have let $\lambda = \eta$ so that $E(x) = 1$.

We recall from above that $c^2 = 1/\eta$ for the gamma distribution as defined in equation (2.4). Since it is clear from equation (2.7) that η must be an integer, our choice of values of

$$c = \eta^{-1/2}$$

is not completely arbitrary. In virtually all of the work here, the main interest is in comparing lognormal and gamma distributions having approximately the same coefficients of variation. A restriction on comparisons is thus imposed by the necessity that η be an integer.

We note in passing that $x = -\lambda^{-1} \ln(1 - u)$ has the negative exponential distribution given by

$$f(x) = \lambda e^{-\lambda x}$$

which implies that a gamma variate with parameters λ and η generated using equation (2.7) is simply the sum of η independent random variables from the exponential distribution with parameter η . This is easily shown by recalling that the random variable u defined as the value of the cumulative distribution function of any continuous distribution has the uniform distribution on the interval (0,1). Thus if x follows the exponential distribution we have that

$$u = F(x) = \int_0^x \lambda e^{-\lambda \tau} d\tau = 1 - e^{-\lambda x}.$$

Solving for x we find that $x = -\frac{1}{\lambda} \ln(1 - u)$.

3.0 DISTINGUISHING BETWEEN LOGNORMAL AND GAMMA DISTRIBUTIONS

3.1 Skewness and Kurtosis

The generators described above were incorporated into the simulation programs LOGN (lognormal) and SKEW (gamma) listed in the Appendix. These programs include provision for calculating the coefficients of skewness and kurtosis (equations (2.3)), as estimated from sample moments calculated from simulated data and for computing expected values of γ_1 and γ_2 . They also are arranged so that a series of "replications" can be run and the results used to calculate a standardized measure of deviations from expectation:

$$T = \frac{\hat{\gamma}_i - E(\gamma_i)}{\widehat{\text{Var}}(\gamma_i)} \quad \text{for } i = 1, 2,$$

where $\widehat{\text{Var}}(\gamma_i)$ is calculated "between replicates". If substantial numbers of replicates are used, one can thus roughly interpret the above measure in terms of a unit normal distribution.

Initial runs with small to moderate sample sizes demonstrated the substantial sampling variability associated with γ_1 and γ_2 , and the impracticality of using these measures with the sample sizes encountered in practice. Even quite large samples show substantial variation. The results are thus mostly useful as a check of sorts on the behavior of the simulation generators. Tables 3.1 and 3.2 show results for 30 replicates each based on 1000 random variables. Even with samples of this size, variances for the higher values of c in the lognormal results are still substantial. Since each value of a gamma random variable for $c = 0.10$ requires 100 random variables from the uniform distribution ($c^2 = 1/n$), the computations for this one value of c alone require 3×10^6 calls to RN2 (the total is about 5×10^6 calls). Gilbert (1973) showed that the period of RN2 is at least 6×10^6 .

The overall evidence seems to be that the simulated distributions are in accord with expectation in terms of skewness and kurtosis.

3.2 "Separate Families" Tests

Expanding on earlier work by Cox (1961, 1962), Jackson (1968) derived two tests for distinguishing between the lognormal and gamma distributions. He applied his tests (Jackson (1969)) to fibre-diameter measurements on wool tops, and found they had greater power of discrimination than the χ^2 test or a graphical method (Monfort, 1964). The "separate families" tests are asymptotic in the sense that the null distributions of the test statistics are known only for large sample sizes. We use computer simulations here to examine how large the samples must be for the asymptotic distribution to hold, and to approximate the power of the tests.

3.2.1 Definition of Test Statistics

Let us suppose that $x_1, x_2, \dots, x_n$ are independent and identically distributed. Denoting by H_0 and H_A the null and alternative hypothesis, respectively, the two tests considered by Jackson (1968: 361) are (i) $H_0: x_i$ are lognormal vs $H_A: x_i$ are gamma, and (ii) $H_0: x_i$ are gamma vs $H_A: x_i$ are lognormal. The lognormal density function in the notation used by Jackson is given by

$$f(x) = \frac{1}{x\sqrt{2\pi\alpha_2}} \exp \left[-\frac{1}{2\alpha_2} (\ln x - \alpha_1)^2 \right], \quad x > 0 \quad (3.1)$$

TABLE 3.1
Measures of Skewness (γ_1) and Kurtosis (γ_2) for
Simulated Gamma Distributions *

True Coef. of Variation c	Sample Coef. of Variation	Skewness		Kurtosis		T	Simulated Variances	
		Sample $\hat{\gamma}_1$	True γ_1	Sample $\hat{\gamma}_2$	True γ_2		$\widehat{\text{Var}}(\gamma_2)$	$\widehat{\text{Var}}(\gamma_2)$
0.100	0.099	0.264	0.200	0.611	0.060	0.218	0.011	0.087
0.200	0.221	0.434	0.400	0.182	0.240	0.027	0.036	1.575
0.301	0.302	0.627	0.603	0.175	0.545	0.031	0.019	0.618
0.408	0.401	0.846	0.816	0.210	1.000	0.027	0.019	1.156
0.500	0.492	1.028	1.000	0.160	1.500	0.022	0.031	1.604
0.577	0.591	1.138	1.155	-0.117	2.000	-0.019	0.020	0.756

*
$$T = \frac{\hat{\gamma}_i - E(\gamma_i)}{\widehat{\text{Var}}(\gamma_i)} \quad \text{for } i = 1, 2.$$

TABLE 3.2
Measures of Skewness (γ_1) and Kurtosis (γ_2) for
Simulated Lognormal Distribution *

True Coef. of Variation c	Sample Coef. of Variation	Skewness			Kurtosis			Simulated Variances	
		Sample $\hat{\gamma}_1$	True γ_1	T	Sample $\hat{\gamma}_2$	True γ_2	T	$\widehat{\text{Var}}(\gamma_1)$	$\widehat{\text{Var}}(\gamma_2)$
0.100	0.100	0.293	0.301	-0.090	0.143	0.162	-0.085	0.007	0.048
0.200	0.212	0.601	0.608	-0.079	0.696	0.664	0.078	0.007	0.167
0.300	0.304	0.952	0.927	0.204	1.688	1.566	0.151	0.015	0.654
0.400	0.376	1.264	1.264	-0.002	2.956	2.969	-0.007	0.039	3.888
0.500	0.483	1.566	1.625	-0.232	4.566	5.035	-0.186	0.064	6.346
0.600	0.624	2.119	2.016	0.143	9.749	8.001	0.141	0.526	152.618

*

$$T = \frac{\hat{\gamma}_i - E(\gamma_i)}{\widehat{\text{Var}}(\gamma_i)} \quad \text{for } i = 1, 2.$$

and that of the gamma distribution by

$$f(x) = \frac{\beta_2}{\beta_1 \Gamma(\beta_2)} \left(\frac{\beta_2 x}{\beta_1} \right)^{\beta_2 - 1} \exp \left(- \frac{\beta_2 x}{\beta_1} \right), \quad x > 0. \quad (3.2)$$

Equations (3.1) and (2.1) are identical except that $\alpha_1 = \mu$ and $\alpha_2 = \sigma^2$.

Similarly, equations (3.2) can be obtained from equation (2.4) by letting $\eta = \beta_2$ and $\lambda = \beta_2/\beta_1$.

The test statistics we investigate in this section are denoted by T_{fN} and T_{gN} given by equations (3.9) and (3.13) below. Their computation involves the use of several series approximations as indicated below.

For testing the null hypothesis H_0 :lognormal against the alternative H_A :gamma, Jackson (1968), derives the log likelihood ratio test statistic T_f , where

$$T_f = n \left[\beta_{2,\hat{\alpha}} \left(\ln \frac{\beta_{2,\hat{\alpha}}}{\beta_{1,\hat{\alpha}}} + \hat{\alpha}_1 - 1 \right) - \hat{\beta}_2 \left(\ln \frac{\hat{\beta}_2}{\hat{\beta}_1} + \hat{\alpha}_1 - 1 \right) + \ln \Gamma(\hat{\beta}_2) - \ln \Gamma(\beta_{2,\hat{\alpha}}) \right]. \quad (3.3)$$

Estimates of α_1 , β_1 and $\beta_{1,\alpha}$ to be used in equation (3.3) are given by

$$\hat{\alpha}_1 = n^{-1} \sum_{i=1}^n \ln x_i = \overline{\ln x}, \quad (3.4)$$

$$\hat{\beta}_1 = n^{-1} \sum_{i=1}^n x_i = \bar{x},$$

$$\hat{\beta}_{1,\hat{\alpha}} = \exp (\hat{\alpha}_1 + \hat{\alpha}_2/2),$$

where

$$\hat{\alpha}_2 = n^{-1} \sum_{i=1}^n (\ln x_i - \hat{\alpha}_1)^2 \quad . \quad (3.5)$$

The estimate $\hat{\beta}_2$ is the solution to the equation

$$\ln \hat{\beta}_2 - \psi(\hat{\beta}_2) = \ln \hat{\beta}_1 - \hat{\alpha}_1 \quad , \quad (3.6)$$

where ψ is the Psi or Digamma function defined by (Albramowitz and Stegun (1964:258))

$$\begin{aligned} \psi(x) &= \frac{\partial \ln \Gamma(x)}{\partial x} \\ &\doteq \ln x - \frac{1}{2x} - \frac{1}{12x^2} + \frac{1}{120x^4} - \frac{1}{252x^6} + \dots \end{aligned} \quad (3.7)$$

Using the first three terms in equation (3.7), we may solve equation (3.6) using the familiar quadratic formula yielding

$$\hat{\beta}_2 \doteq \frac{1 + [1 + (3/4) (\ln \hat{\beta}_1 - \hat{\alpha}_1)]^{1/2}}{4 (\ln \hat{\beta}_1 - \hat{\alpha}_1)} \quad . \quad (3.8)$$

Similarly, $\hat{\beta}_{2,\hat{\alpha}}$ is the solution to

$$\ln \hat{\beta}_{2,\hat{\alpha}} - \psi(\hat{\beta}_{2,\hat{\alpha}}) = \frac{1}{2} \hat{\alpha}_2 \quad ,$$

which, using the first three terms in equation (3.7) and the quadratic formula can be shown to be

$$\hat{\beta}_{2,\hat{\alpha}} \doteq \frac{1 + [1 + (2/3) \hat{\alpha}_2]^{1/2}}{2\hat{\alpha}_2} \quad .$$

The only remaining quantities needed to compute T_f are $\ln \Gamma(\hat{\beta}_2)$ and $\ln \Gamma(\hat{\beta}_{2,\hat{\alpha}})$ which are conveniently approximated by the series expansion (Albramowitz and Stegun (1964:257))

$$\ln \Gamma(x) \doteq \frac{1}{2} \ln(2\pi) - x + (x - 1/2) \ln x + \frac{1}{12x} - \frac{1}{360x^3} \quad .$$

The variance of T_f when n is large was found by Jackson (1968) to be given by

$$\text{Var}(T_f) = n \hat{\beta}_{2,\hat{\alpha}}^2 [\exp(\hat{\alpha}_2) - 1 - \hat{\alpha}_2 - (1/2) \hat{\alpha}_2^2] \quad .$$

Finally, it follows from the work of Cox (1961, 1962) that the asymptotic null distribution of

$$T_{fN} = T_f / [\text{Var}(T_f)]^{1/2} \quad (3.9)$$

(the statistic used in our simulations) is normal with mean zero and variance 1, where negative values are expected under H_A .

Turning now to the test of the null hypothesis H_0 :gamma, against the alternative H_A :lognormal, Jackson (1968) derives the log-likelihood ratio test statistic

$$T_g = \left(\frac{n}{2}\right) \ln \frac{\hat{\alpha}_2}{\alpha_{2,\hat{\beta}}} + n \hat{\beta}_2 (\hat{\alpha}_1 - \alpha_{1,\hat{\beta}}) \quad , \quad (3.10)$$

where $\hat{\alpha}_1$, $\hat{\alpha}_2$ and $\hat{\beta}_2$ are given by equations (3.4), (3.5) and (3.8).

Jackson shows that

$$\alpha_{1,\hat{\beta}} = \psi(\hat{\beta}_2) - \ln\left(\frac{\hat{\beta}_2}{\hat{\beta}_1}\right)$$

and

$$\alpha_{2,\hat{\beta}} = \psi'(\hat{\beta}_2) ,$$

where the prime denotes differentiation. We used the first four terms of equation (3.7) to approximate $\psi(\hat{\beta}_2)$. $\psi'(\hat{\beta}_2)$ was calculated using the series approximation (Abramowitz and Stegun (1964:260))

$$\psi'(x) \doteq \frac{1}{x} + \frac{1}{2x^2} + \frac{1}{6x^3} - \frac{1}{30x^5} + \dots , \quad (3.11)$$

retaining the first four terms. The asymptotic variance of T_g is

$$\begin{aligned} \text{Var}(T_g) &= n \left[\frac{\psi'''(\hat{\beta}_2)}{4 \{\psi'(\hat{\beta}_2)\}^2} - \frac{\hat{\beta}_2 \{\psi''(\hat{\beta}_2)\}^2}{4 \{\psi'(\hat{\beta}_2)\}^2 \{\hat{\beta}_2 \psi'(\hat{\beta}_2) - 1\}} + 1/2 \right] \\ &= n \phi(\hat{\beta}_2), \text{ say,} \end{aligned}$$

where primes denote differentiation. Jackson gives an approximation to

$$\phi(x) \text{ as } \phi(x) = \frac{1}{2} \left\{ 1 - \frac{1 + \frac{2}{3x} - \frac{1}{18x^2} - \frac{4}{9x^3} + \frac{5}{36x^4} + \dots}{\left(1 + \frac{1}{2x} + \frac{1}{6x^2} - \frac{1}{30x^4} + \dots\right)^2} \right\} . \quad (3.12)$$

Similarly to T_{fN} , the results of Cox (1961, 1962) indicate

$$T_{gN} = T_g / [\text{Var}(T_g)]^{1/2} \quad (3.13)$$

is asymptotically $N(0,1)$ with negative values expected under H_A . In the next section we describe our simulation results on T_{fN} and T_{gN} .

3.2.2 Simulations

The approximate Type I error of the test statistic T_{fN} used to test H_0 :lognormal versus H_A :gamma was obtained by generating 1000 or 2000 samples $x_1, x_2, \dots, x_n$ of size n from a lognormal distribution and computing T_{fN} on each sample. The "actual" Type I error rate corresponding to a nominal α level test was taken to be the proportion of the empirical null distributions of T_{fN} falling in the upper and lower nominal 1%, 2.5% and 5% levels of the $N(0,1)$ distribution. The upper and lower percentages were also combined to give 2%, 5% and 10% actual significance levels for two-tailed tests. We also computed the sample mean and variance, and the coefficients of skewness and kurtosis of the 1000 or 2000 generated values of T_{fN} to check against their expected values of 0, 1, 0 and 0, respectively.

Results were obtained for sample sizes of $n = 30, 50, 70, 100$ and 200 . We anticipated that the actual and nominal α levels would become more nearly equal as n becomes large. 2000 iterations were used for all sample sizes except $n = 200$ when 1000 iterations were used. The above procedure was repeated for c equal to 0.301, 0.500, 0.707 and 1.000. The lognormal variates were generated using RN2 as outlined in section 2.2. Our results apply to the case where $E(x) = 1$ which implies $\mu = -\frac{1}{2}\sigma^2$.

The power of T_{fN} to detect gamma distributed observations was approximated by generating samples of gamma variates and computing T_{fN} on each sample, the whole procedure being iterated 1000 or 2000 times depending on

the size of n as indicated above. The approximate power at the nominal α level was then simply the proportion of the generated distribution of T_{fN} falling in the nominal rejection region. Estimates of power were obtained for the same sample sizes and values of c used in determining the actual Type I errors described above. The mean, variance and coefficients of skewness and kurtosis were computed on the empirical distribution of T_{fN} generated for each combination of n and c to compare with their expected values under H_0 . Pseudo-gamma variates were generated as described in section (2.2). As with the lognormal distribution in this section, we set $E(x) = 1$ for convenience sake. This in turn implies $\lambda = \eta$.

Approximations to the actual Type I error and power of T_{gN} were obtained in the same manner as for T_{fN} , i.e., the Type I error was obtained by generating gamma distributed samples; and the power likewise from empirical distributions of T_{gN} on lognormal data. The same values of n and c were used for both T_{fN} and T_{gN} .

In Figures 3.1 and 3.2 we have plotted the density functions of the lognormal and gamma distributions for $c = 0.301, 0.500, 0.707$ and 1.000 when $E(x) = 1$. We see that the lognormal and gamma become more nearly alike as c becomes small. Hence, we might expect the power of T_{fN} and T_{gN} to increase as c approaches 1.00 .

The "actual" and nominal Type I errors of T_{fN} and T_{gN} are compared in Tables 3.3 and 3.4. We are mostly concerned with the lower tail since we expect negative values of T_{fN} and T_{gN} if H_A is true. We see that the actual levels are often times significantly less than the nominal levels

FIGURE 3.1. DENSITY FUNCTIONS OF THE GAMMA AND LOGNORMAL DISTRIBUTIONS FOR $c = 1.00$ AND 0.707 WHEN $E(x) = 1$

FIGURE 3.2. DENSITY FUNCTIONS OF THE GAMMA AND LOGNORMAL DISTRIBUTIONS FOR $c = 0.500$ AND 0.301 WHEN $E(x) = 1$

TABLE 3.3

"Actual Type I Errors of T_{fN} as Compared
with Nominal Levels

c	n	Lower Tail			Upper Tail			Combined		
					Nominal α Level (%)					
		1	2.5	5	5	2.5	1	2	5	10
1.000	30	0.25*	0.45*	1.55*	2.30*	1.50	0.80	1.05*	1.95*	3.85*
	50	0.10*	0.40*	1.20*	2.80*	1.60	1.05	1.15*	2.00*	4.00*
	70	0.15*	0.70*	1.85*	3.60*	2.05	1.35	1.50	2.75*	5.45*
	100	0.10*	0.85*	1.90*	4.35	3.05	1.90*	2.00	3.90*	6.25*
	200	0.20	0.70*	1.80*	5.90	3.20	2.00*	2.20	3.90	7.70*
0.707	30	0.55	1.25*	2.90*	4.40	2.60	1.45	2.00	3.85*	7.30*
	50	0.20*	0.95*	2.50*	3.95*	2.55	1.45	1.65	3.50*	6.45*
	70	0.20*	0.70*	2.25*	4.15	2.90	1.70*	1.90	3.60*	6.40*
	100	0.40*	1.20*	2.65*	5.05	3.40	2.15*	2.55	4.60	7.70*
	200	0.40	0.30*	3.70	6.20	4.00	2.50*	2.90	5.30	9.90
0.500	30	0.25*	1.15*	2.40*	2.90*	1.55	0.60	0.85*	2.70*	5.30*
	50	0.65	1.45*	3.15*	3.80*	2.40	1.25	1.90	3.85*	6.95*
	70	0.40*	1.00*	2.80*	5.30	2.95	1.70*	2.10	3.95*	8.10*
	100	0.25*	1.30*	3.35*	5.15	2.85	1.75*	2.00	4.15	8.50*
	200	0.60	1.60*	3.90	6.30	3.70	2.00*	2.60	5.30	10.20
0.301	30	0.35*	0.95*	2.60*	3.50*	2.60	1.35	1.70	3.55*	6.10*
	50	0.75	2.05	3.95*	3.90*	2.15	1.20	1.95	4.20	7.85*
	70	0.60	1.55	2.70*	4.90	2.60	1.45	2.05	4.15	7.60*
	100	0.50	1.30*	4.00*	5.65	3.40	1.55	2.05	4.70	9.65
	200	0.80	2.00	3.60*	5.00	2.30	1.50	2.30	4.30	8.60

*Significantly different ($\alpha = .05$) than the nominal significant level.

TABLE 3.4

"Actual" Type I Errors of T_{gN} as Compared
With Nominal Levels

c	n	Lower Tail			Upper Tail			Combined		
		Nominal α Level (%)								
		1	2.5	5	5	2.5	1	2	5	10
1.000	30	0.05*	0.55*	1.7*	2.4*	1.3*	0.85	0.9*	1.85*	4.1*
	50	0.45	1.20*	3.2*	3.95*	2.4	0.85	1.3	3.60*	7.15*
	70	0.45	1.15*	3.05*	4.55	2.35	1.00	1.45	3.50*	7.60*
	100	0.45	1.80	4.10	5.00	3.15	1.45	1.90	4.95	9.10
	200	0.40	1.00*	2.80*	5.00	2.90	1.50	1.90	3.90	7.80*
0.707	30	0.20*	1.05*	3.45*	2.85*	1.55*	0.65	0.85*	2.60*	6.30*
	50	0.40*	1.20*	2.70*	2.90*	1.55*	0.90	1.20	2.75*	5.60*
	70	0.40*	1.60*	4.35	4.60	2.75	1.60*	2.00	4.35	8.95
	100	0.70	1.85	3.90*	5.20	2.95	1.30	2.00	4.80	9.10
	200	0.30	1.80	5.20	5.20	2.70	1.90*	2.20	4.50	10.40
0.500	30	0.85	1.80	4.10	3.95*	2.45	1.30	2.15	4.25	8.05*
	50	0.60	2.35	4.65	4.60	2.90	1.35	1.95	5.25	9.25
	70	0.60	1.95	3.65*	6.05*	3.40*	1.75*	2.35	5.35	9.70
	100	0.65	1.60	4.05	5.90	4.00*	2.30*	2.95*	5.60	9.95
	200	0.90	2.50	4.50	7.20*	4.60*	2.50*	3.40*	7.10*	11.00
0.301	30	0.85	1.75	4.05	4.50	2.85	1.60*	2.45	4.60	8.55*
	50	1.05	2.80	4.95	7.00*	3.95*	2.30*	3.35*	6.75*	11.95*
	70	1.20	2.70	4.80	7.05*	4.45*	3.15*	4.35*	7.15*	11.85*
	100	1.45	3.05	5.40	7.95*	5.35*	3.35*	4.80*	8.40*	13.35*
	200	1.40	2.90	6.20	10.60*	8.00*	5.20*	6.60*	10.90*	16.00*

* Significantly different ($\alpha = .05$) than the nominal significance level.

with some improvement for larger sample sizes. The differences in the lower tail appear to be somewhat larger for T_{fN} than for T_{gN} . For both T_{fN} and T_{gN} , the differences appear to be greater when $c = 1.00$ than for smaller values of c . The actual and nominal levels for T_{fN} appear to be quite close for an $\alpha = 0.05$ two-tailed test when $n = 200$ for all c used except $c = 1.00$. For T_{gN} the actual 2-tailed level is significantly greater than the nominal $\alpha = 0.05$ level for $c = 0.301$ and the difference appears to increase with sample size. In general, we conclude that the actual and nominal α levels of T_{gN} appear to be nearly equal for $n \geq 100$, particularly for small c . For T_{fN} it appears the sample size may need to be closer to 200 or more before the differences are sufficiently small. The estimated standard deviations of the estimated proportions for $\alpha = 0.05$ are 0.0069 and 0.0049 for empirical distributions based on 1000 and 2000 iterations, respectively.

In Tables 3.5 and 3.6 we show the mean, variance and coefficients of skewness and kurtosis for the simulated distributions of T_{fN} and T_{gN} when H_0 is true. These indicate the actual null distributions to be skewed to the right with a mean slightly less than zero. The variance of both statistics tends towards 1 as n gets large. There is some evidence to suggest the variance of T_{gN} becomes greater than 1 as n becomes large, particularly for small c .

Data relating to the simulated power of T_{fN} and T_{gN} is given in Tables 3.7 through 3.10 as well as illustrated in Figures 3.3 and 3.4. It is clear that one-tailed tests (lower tail) are appropriate. The power of T_{fN} to detect gamma distributions is apparently greater than that of T_{gN} to detect

FIGURE 3.3 POWER* OF T_{fN} TO DETECT GAMMA ALTERNATIVES FIGURE 3.4. POWER* OF T_{gN} TO DETECT LOGNORMAL ALTERNATIVES

*POWER IS FOR ONE-SIDED TESTS IN LOWER TAIL. THE POWER IS NEARLY IDENTICAL FOR THE CORRESPONDING 2α TWO-TAILED TEST. POWER WAS APPROXIMATED USING THE NOMINAL RATHER THAN ACTUAL α LEVELS.

TABLE 3.5

The Mean, Variance, Skewness and Kurtosis of
 T_{fN} for Lognormal Observations[†]

		$E(T_{fN})$	$Var(T_{fN})$	$\gamma_1(T_{fN})$	$\gamma_2(T_{fN})$
Expected Under H_0 :		0.00	1.00	0.00	0.00
<u>c</u>	<u>n</u>				
1.000	30	-0.128	0.637	0.541	0.940
	50	-0.102	0.710	0.904	2.382
	70	-0.084	0.793	0.768	2.020
	100	-0.034	0.943	1.199	4.858
	200	0.004	0.975	0.721	1.368
0.707	30	-0.054	0.824	0.487	1.005
	50	-0.066	0.850	0.899	3.577
	70	-0.050	0.829	0.652	1.156
	100	-0.024	1.027	1.110	4.547
	200	0.037	1.127	0.902	2.803
0.500	30	-0.056	0.725	0.232	0.348
	50	-0.067	0.853	0.473	1.621
	70	-0.007	0.924	0.648	1.724
	100	-0.007	0.959	0.429	0.585
	200	0.017	1.039	0.456	0.973
0.301	30	-0.018	0.801	0.562	1.938
	50	-0.049	0.885	0.196	0.585
	70	-0.001	0.930	0.399	0.902
	100	0.021	0.980	0.313	0.520
	200	0.010	0.957	0.280	0.776

[†]Each datum is based on 2000 values of T_{fN} when $n = 30, 50, 70$ and 100 ; and on 1000 values when $n = 200$.

TABLE 3.6

The Mean, Variance, Skewness and Kurtosis of

 T_{gN} for Gamma Observations[†]

		$E(T_{gN})$	$Var(T_{gN})$	$\gamma_1(T_{gN})$	$\gamma_2(T_{gN})$
Expected Under H_0 :		0.00	1.00	0.00	0.00
<u>c</u>	<u>n</u>				
1.000	30	-0.130	0.680	0.530	0.770
	50	-0.113	0.815	0.431	0.528
	70	-0.073	0.839	0.432	0.627
	100	-0.043	0.937	0.323	0.313
	200	0.010	0.913	0.378	0.341
0.707	30	-0.103	0.756	0.359	0.649
	50	-0.123	0.763	0.530	1.529
	70	-0.054	0.921	0.496	1.333
	100	-0.090	0.967	0.546	1.654
	200	-0.085	1.049	0.472	0.643
0.500	30	-0.089	0.905	0.434	1.018
	50	-0.060	1.002	0.473	0.998
	70	-0.033	1.059	0.488	0.969
	100	0.021	1.064	0.462	0.747
	200	0.048	1.173	0.255	0.716
0.301	30	-0.054	0.953	0.368	0.980
	50	0.027	1.165	0.391	1.403
	70	0.080	1.214	0.408	1.449
	100	0.090	1.347	0.452	1.568
	200	0.157	1.731	0.742	3.522

[†] Each datum is based on 2000 values of T_{gN} when $n = 30, 50, 70$ and 100 ; and on 1000 values when $n = 200$.

TABLE 3.7

Power of T_{fN} to Detect Gamma Distributed Variates

Nominal α Levels (%)										
c	n	Lower Tail			Upper Tail			Combined		
		1	2.5	5	5	2.5	1	2	5	10
1.000	30	0.90	5.30	17.70	0.00	0.00	0.00	0.90	5.30	17.70
	50	7.45	25.40	51.50	0.00	0.00	0.00	7.45	25.45	51.50
	70	23.80	55.40	79.70	0.00	0.00	0.00	23.80	55.40	79.70
	100	61.55	84.35	93.85	0.00	0.00	0.00	61.50	84.35	93.85
	200	99.20	100.00	100.00	0.00	0.00	0.00	99.20	100.00	100.00
0.707	30	5.20	13.65	25.95	0.05	0.00	0.00	5.20	13.65	26.00
	50	14.75	28.30	45.45	0.00	0.00	0.00	14.75	28.30	45.45
	70	29.70	49.20	66.00	0.00	0.00	0.00	29.70	49.20	66.00
	100	48.60	66.70	79.55	0.00	0.00	0.00	48.60	66.70	79.55
	200	88.20	95.50	98.50	0.00	0.00	0.00	88.20	95.50	98.50
0.500	30	5.80	11.90	19.20	0.50	0.20	0.05	5.85	12.10	19.70
	50	12.05	21.65	31.90	0.05	0.00	0.00	12.05	21.65	31.95
	70	18.40	30.95	44.30	0.10	0.10	0.05	18.45	31.05	44.40
	100	31.70	46.20	59.15	0.00	0.00	0.00	31.70	46.20	59.15
	200	63.80	77.30	85.80	0.10	0.00	0.00	63.80	77.30	85.00
0.301	30	3.55	6.80	12.05	1.15	0.70	0.30	3.85	7.50	13.20
	50	7.95	13.60	20.85	0.90	0.60	0.35	8.30	14.20	21.75
	70	10.55	18.35	26.70	0.60	0.40	0.25	10.80	18.75	27.30
	100	15.60	24.30	34.40	0.45	0.25	0.15	15.75	24.55	34.85
	200	32.70	43.50	53.00	0.50	0.40	0.30	33.00	43.90	53.00

TABLE 3.8

Power of T_{gN} to Detect Lognormal Distributed Variates

c	n	Lower Tail			Upper Tail			Combined		
					Nominal Level (%)					
		1	2.5	5	5	2.5	1	2	5	10
1.000	30	11.15	21.80	33.80	0.15	0.00	0.00	11.15	21.80	33.95
	50	26.40	41.65	57.05	0.00	0.00	0.00	26.40	41.65	57.05
	70	42.85	59.75	73.40	0.00	0.00	0.00	42.85	59.75	73.40
	100	65.75	79.15	88.15	0.00	0.00	0.00	65.75	79.15	88.15
	200	96.40	99.10	99.60	0.00	0.00	0.00	96.40	99.10	99.60
0.707	30	8.85	15.95	26.75	0.30	0.15	0.05	8.90	16.10	27.05
	50	16.25	28.10	41.15	0.05	0.05	0.00	16.25	28.15	41.20
	70	25.80	41.35	54.90	0.00	0.00	0.00	25.80	41.35	54.90
	100	43.35	58.60	72.00	0.00	0.00	0.00	43.35	58.60	72.00
	200	82.60	90.20	95.00	0.00	0.00	0.00	82.60	90.20	95.00
0.500	30	4.40	9.70	17.45	0.20	0.05	0.05	4.45	9.75	17.65
	50	8.55	16.55	27.90	0.30	0.10	0.00	8.55	16.65	28.20
	70	16.10	26.95	39.30	0.05	0.05	0.00	16.10	27.00	39.35
	100	26.25	40.65	53.10	0.00	0.00	0.00	26.25	40.65	53.10
	200	55.50	70.90	81.40	0.00	0.00	0.00	55.50	70.90	81.40
0.301	30	3.15	5.65	10.80	0.65	0.20	0.05	3.20	5.85	11.45
	50	4.55	9.15	15.60	0.65	0.25	0.15	4.70	9.40	16.25
	70	8.30	13.80	22.35	0.35	0.20	0.00	8.30	14.00	22.70
	100	11.15	19.20	30.00	0.15	0.00	0.00	11.15	19.20	30.15
	200	23.60	35.30	48.60	0.10	0.00	0.00	23.60	35.30	48.70

TABLE 3.9

The Mean, Variance, Skewness and Kurtosis
of T_{fN} for Gamma Observations[†]

		$E(T_{fN})$	$Var(T_{fN})$	$\gamma_1(T_{fN})$	$\gamma_2(T_{fN})$
Expected Under H_0 :		0.00	1.00	0.00	0.00
<u>c</u>	<u>n</u>				
1.000	30	-1.227	0.221	0.238	0.611
	50	-1.649	0.240	0.198	0.663
	70	-2.006	0.236	0.225	0.947
	100	-2.454	0.265	0.147	0.444
	200	-3.549	0.236	0.008	0.055
0.707	30	-1.166	0.549	0.208	0.248
	50	-1.559	0.544	-0.017	0.519
	70	-1.935	0.648	0.003	0.588
	100	-2.315	0.677	-0.105	0.424
	200	-3.349	0.729	-0.224	0.228
0.50	30	-0.899	0.827	-0.061	0.537
	50	-1.239	0.909	-0.167	0.453
	70	-1.522	0.953	-0.222	0.608
	100	-1.894	0.955	-0.269	0.476
	200	-2.722	1.066	-0.119	0.590
0.301	30	-0.571	0.934	-0.183	0.721
	50	-0.849	1.134	-0.227	1.130
	70	-1.062	1.181	-0.275	1.242
	100	-1.273	1.310	-0.364	1.358
	200	-1.846	1.692	-0.654	2.831

[†]Each datum is based on 2000 values of T_{fN} when $n = 30, 50, 70$ and 100; and on 1000 values when $n = 200$.

TABLE 3.10

The Mean, Variance, Skewness and Kurtosis
of T_{gN} for Lognormal Observations[†]

		$E(T_{gN})$	$Var(T_{gN})$	$\gamma_1(T_{gN})$	$\gamma_2(T_{gN})$
Expected Under H_0 :		<u>0.00</u>	<u>1.00</u>	<u>0.00</u>	<u>0.00</u>
<u>c</u>	<u>n</u>				
1.000	30	-1.337	0.710	-0.162	0.472
	50	-1.839	0.771	-0.532	1.196
	70	-2.230	0.854	-0.411	0.809
	100	-2.748	0.993	-0.737	2.556
	200	-3.976	1.051	-0.540	0.872
0.707	30	-1.124	0.807	-0.061	0.728
	50	-1.499	0.827	-0.450	1.601
	70	-1.816	0.821	-0.413	0.692
	100	-2.222	0.989	-0.685	2.362
	200	-3.242	1.089	-0.679	1.865
0.500	30	-0.871	0.715	-0.020	0.255
	50	-1.151	0.834	-0.187	1.089
	70	-1.449	0.892	-0.407	1.001
	100	-1.744	0.932	-0.268	0.371
	200	-2.516	1.020	-0.325	0.774
0.301	30	-0.584	0.778	-0.360	1.256
	50	-0.746	0.878	-0.070	0.545
	70	-0.950	0.919	-0.278	0.723
	100	-1.164	0.967	-0.216	0.414
	200	-1.642	0.943	-0.205	0.714

[†]Each datum is based on 2000 values of T_{gN} when $n = 30, 50, 70$ and 100; and on 1000 values when $n = 200$.

lognormal variates for given sample sizes and coefficient of variation. Unfortunately the power of either test does not appear to be very large unless $n \geq 100$ and c is near 1. As expected from Figures 3.1 and 3.2, the power of T_{fN} and T_{gN} increases as c increases when $E(x) = 1$. The unusually low power of T_{fN} for $n = 30$ when $c = 1.0$ is probably related to the fact that the gamma distribution degenerates into the exponential when $c = 1.0$. The density functions of the exponential and the lognormal when $c = 1.0$ are illustrated in Figure 3.1. The reader is referred to Cox (1961, 1962) who developed separate families tests specifically for differentiating between lognormal and exponential distributions, and to Jackson (1968) who give power functions of the tests.

Finally, we emphasize that the power results presented here were obtained by comparison with the nominal α -levels of the $N(0,1)$ distribution, rather than against their actual Type I errors as approximated here using simulations. Since the actual levels were in general less than the nominal levels, this implies the power of the tests if measured at their true α critical-levels would probably be somewhat larger than reported here.

4. DISTRIBUTION THEORY

In section 4 we discuss and/or derive some useful and interesting results necessary for an understanding of our simulations in section 5. Some of the derivations have been obtained previously by others, but we include them here for completeness.

4.1 The s^2 and z Statistics

We are interested in constructing tests of significance for equal coefficients of variation based on s^2 and z for both lognormal and gamma data, and then determining the effect on these tests of switching the underlying distribution from that assumed in deriving the test. These tests are discussed explicitly in section (4.6) below. They are based on s^2 and z for which some background discussion seems appropriate at this stage.

Interest here is in the sampling distribution of two quantities:
(1) the sample variance s^2 of log-transformed random variables from the lognormal and gamma distributions, where

$$s^2 = (n-1)^{-1} \sum_{i=1}^n (\ln x_i - \overline{\ln x})^2 \quad , \quad (4.1)$$

and (2) the sampling distribution of z computed from samples drawn from the (untransformed) lognormal and gamma distributions, where

$$z = \ln \bar{x} - \overline{\ln x} \quad . \quad (4.2)$$

One reason for interest in z is the prospect that it may shed some light on the effects of "pooling" or "compositing" samples to reduce analytical costs. The result of a given pooled analysis might be considered to be the mean ($\bar{x}$) of several individual organisms (or tissues, soil samples, etc.) and z thus expresses the difference between the logarithm of such a pooled sample and the result that would have been obtained had individual samples been analyzed and a logarithmic transformation applied to the resulting data (as is quite commonly done).

The z statistic was used by Linhart (1965) in obtaining approximate confidence limits for the coefficient of variation of a gamma distribution, and traces back to Bartlett's (1937) study of tests of homogeneity of variances and to Bartlett and Kendall (1946).

One of the four distributions mentioned above (2 statistics for each of 2 distributions) is immediately available; that is, log-transformed random variables from a lognormal distribution are normally distributed so that s^2 computed from n observations has the distribution

$$s^2 \sim [\sigma^2/(n-1)] \chi_{n-1}^2, \quad (4.3)$$

where σ^2 is the variance of the normal distribution. It is also known (Linhart (1965)) that the distribution of z when computed on n observations from a gamma distribution is given approximately by

$$(n\beta)^{-1} \left[1 + \frac{n+1}{3n\beta} \right] \chi_{n-1}^2.$$

However, the distribution of s^2 computed on log-transforms of gamma variates, and the distribution of z on lognormal observations are apparently unknown.

In section 4.4 we do, however, obtain an expression for $E(s^2)$ for log-transformed gamma variates, and we conclude this section by obtaining an approximation of $E(z)$ for lognormal variates.

We have that

$$E(z) = E(\ln \bar{x}) - E(\ln x) \quad ,$$

where $E(\ln x) = \mu$ for $x \sim \text{lognormal}$. The sampling distribution of $\ln \bar{x}$ is not readily available and it appears necessary to substitute $\ln E(x)$ in place of $E(\ln \bar{x})$, where from equation (2.1) we have $E(x) = \exp[\mu + \sigma^2/2]$. Hence,

$$\begin{aligned} \ln E(x) &\doteq \ln \left\{ \exp(\mu + \sigma^2/2) \right\} \\ &= \mu + \sigma^2/2 \end{aligned}$$

so that

$$\begin{aligned} E(z) &\doteq \mu + \sigma^2/2 - \mu \\ &= \sigma^2/2 \end{aligned} \quad . \quad (4.4)$$

We note in passing that z may be written as

$$z = \ln \left[\frac{\sum_{i=1}^n x_i}{n \left(\prod_{i=1}^n x_i \right)^{1/n}} \right] \quad .$$

4.2 Moments of the Lognormal Distribution

For purposes of completeness and easy availability for future use, we derive in this section the moments of the lognormal distribution using

its moment generating function $M_t(x)$. The moments are given without proof in, e.g., Aitchison and Brown (1969). These moments may be useful in attempting to obtain the sampling distribution of z for lognormal variates. $M_t(x)$ is given by

$$M_t(x) = E(e^{tx}) = \frac{1}{\sigma\sqrt{2\pi}} \int_0^{\infty} \frac{1}{x} \exp \left[tx - \frac{1}{2\sigma^2} (\ln x - \mu)^2 \right] dx .$$

Let

$$u = \ln x - \mu$$

so that

$$du = (1/x)dx \quad \text{and} \quad x = e^{u+\mu} .$$

Then

$$\begin{aligned} M_t(x) &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp \left[te^{u+\mu} - \frac{u^2}{2\sigma^2} \right] du \\ &= \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp \left[-u^2/2\sigma^2 \right] \left[1 + t e^{u+\mu} + t^2 e^{2(u+\mu)} + \dots \right] du \\ &= \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=0}^{\infty} \frac{e^{k\mu} t^k}{k!} \int_{-\infty}^{\infty} e^{-u^2/2\sigma^2 + ku} du \\ &= \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=0}^{\infty} \frac{t^k}{k!} e^{k\mu + \frac{k^2\sigma^2}{2}} \int_{-\infty}^{\infty} \exp \left[- \left(\frac{u}{\sqrt{2\sigma^2}} - \frac{k}{2\sqrt{1/2\sigma^2}} \right)^2 \right] du . \end{aligned}$$

Now let

$$y = \frac{u}{\sqrt{2\sigma^2}} - \frac{k}{2\sqrt{1/2\sigma^2}} \quad \text{so that} \quad dy = \sqrt{1/2\sigma^2} du .$$

Then we have

$$\begin{aligned}
 M_t(x) &= \frac{1}{\sigma\sqrt{2\pi}} \frac{1}{\sqrt{1/2\sigma^2}} \sum_{k=0}^{\infty} \frac{t^k}{k!} e^{k\mu + \frac{k^2\sigma^2}{2}} \int_{-\infty}^{\infty} e^{y^2} dy \\
 &= \sqrt{2} \sum_{k=1}^{\infty} \frac{t^k}{k!} e^{k\mu + \frac{k^2\sigma^2}{2}} \\
 &= \sqrt{2} \left[1 + t e^{\mu + \sigma^2/2} + \frac{t^2}{2!} e^{2(\mu + \sigma^2)} + \dots \right]
 \end{aligned}$$

The moments $E(x^k)$ are given by the coefficients of $t^k/k!$

4.3 Expected Value of z For Gamma Variates

In this section we find an approximation to the expected value of z , i.e., an approximation of

$$E(z) = E(\ln \bar{x}) - E(\ln x) ,$$

when the x_i are gamma distributed. We turn our attention first to $E(\ln x)$.

The following deviation of $E(\ln x)$ follows closely that given by Bartlett and Kendall (1946). Let $y = \ln x$, where x is defined by equation (2.5).

Then the moment generating function of y is

$$\begin{aligned}
 M_t(y) &= E(e^{ty}) \\
 &= \frac{(2\alpha)^{-\beta/2}}{\Gamma(\beta/2)} \int_0^{\infty} e^{t \ln x} x^{\beta/2-1} e^{-x/2\alpha} dx \\
 &= \frac{(2\alpha)^{-\beta/2}}{\Gamma(\beta/2)} \int_0^{\infty} x^{t + \beta/2-1} e^{-x/2\alpha} dx \\
 &= (2\alpha)^t \frac{\Gamma(\beta/2 + t)}{\Gamma(\beta/2)} .
 \end{aligned} \tag{4.5}$$

The cumulant generating function is given by

$$\begin{aligned} L_t(y) &= \ln M_t(y) \\ &= k_1 t + k_2 t^2/2! + k_3 t^3/3! + \dots, \end{aligned}$$

where the k_i are the cumulants of the distribution. Of interest to us is that $k_1 = \mu_1 = E(y)$ and $k_2 = \mu_2 = \text{Var}(y)$. (The results below for μ_2 will be utilized in section 4.4). From equation (4.5) we find that

$$L_t(y) = t \ln 2\alpha + \ln \Gamma(\beta/2 + t) - \ln \Gamma(\beta/2) \quad . \quad (4.6)$$

We obtain k_1 and k_2 by setting $t = 0$ in the first and second differentiations of (4.6) with respect to t . We find that

$$\frac{\partial L_t(y)}{\partial t} = \ln 2\alpha + \psi(\beta/2 + t)$$

and

$$\frac{\partial^2 L_t(y)}{\partial t^2} = \psi'(\beta/2 + t) \quad ,$$

where $\psi(x)$, the Psi (Digamma) function, is defined by

$$\psi(x) = \frac{\partial \ln \Gamma(x)}{\partial x} = \frac{\psi'(x)}{\Gamma(x)} \quad ,$$

and

$$\psi'(x) = \frac{\partial \psi(x)}{\partial x}$$

($\psi(x)$ and $\psi'(x)$ were also used in section 3.21).

Thus

$$E(\ln x) = \mu_1 = k_1 = \ln 2\alpha + \psi(\beta/2)$$

and

$$\text{Var}(\ln x) = \mu_2 = k_2 = \psi'(\beta/2) \quad , \quad (4.7)$$

where $\psi(x)$ and $\psi'(x)$ may be approximated by equations (3.7) and (3.11), respectively.

Now we direct our attention to finding $E(\ln \bar{x})$, where as above, we assume the x_i are gamma distributed in the form of equation (2.5). Recall that the sum of random variables from a gamma distribution also has a gamma distribution; i.e., the gamma distribution is "reproductive" with respect to summation. Thus, if

$$v = \frac{1}{n} \sum_{i=1}^n x_i = \bar{x}$$

then

$$f(n\bar{x}) = \frac{(n\bar{x})^{(n\beta/2)-1} e^{-n\bar{x}/2\alpha}}{(2\alpha)^{n\beta/2} \Gamma(n\beta/2)}$$

and

$$f(v) = \frac{\left(\frac{n}{2\alpha}\right)^{n\beta/2} v^{(n\beta/2)-1} e^{-nv/2\alpha}}{\Gamma(n\beta/2)} \quad . \quad (4.8)$$

We then have that

$$E(\ln v) = \frac{\left(\frac{n}{2\alpha}\right)^{n\beta/2}}{\Gamma(n\beta/2)} \int_0^\infty v^{(n\beta/2)-1} e^{-nv/2\alpha} \ln v \, dv \quad .$$

Now, let

$$u = \frac{nv}{2\alpha} \quad \text{for which} \quad du = \frac{n}{2\alpha} dv \quad .$$

Then

$$\begin{aligned} E(\ln v) &= \frac{1}{\Gamma(n\beta/2)} \int_0^\infty u^{(n\beta/2)-1} e^{-u} \left[\ln \frac{2\alpha}{n} + \ln u \right] du \\ &= \frac{1}{\Gamma(n\beta/2)} \left[\ln \frac{2\alpha}{n} \Gamma(n\beta/2) + \int_0^\infty u^{(n\beta/2)-1} e^{-u} \ln u \, du \right] \quad . \end{aligned}$$

Since

$$\begin{aligned} \Gamma'(x) &= \frac{\partial \Gamma(x)}{\partial x} \\ &= \int_0^\infty u^{x-1} e^{-u} \ln u \, du \end{aligned}$$

(Olmstead, 1959, p. 518), we have the result that

$$\begin{aligned} E(\ln \bar{x}) &= E(\ln v) \\ &= \ln \frac{2\alpha}{n} + \frac{\Gamma'(n\beta/2)}{\Gamma(n\beta/2)} \\ &= \ln \frac{2\alpha}{n} + \psi(n\beta/2) \quad . \end{aligned}$$

Hence,

$$\begin{aligned} E(z) &= E(\ln \bar{x}) - E(\ln x) \\ &= \ln \frac{2\alpha}{n} + \psi(n\beta/2) - \ln 2\alpha - \psi(\beta/2) \\ &= \psi(n\beta/2) - \psi(\beta/2) - \ln n \quad . \end{aligned} \tag{4.9}$$

If we approximate $\psi(x)$ by the first 3 terms of the series expansion given in equation (3.7), then an approximation to $E(z)$ is given by

$$E(z) \doteq \frac{n-1}{n\beta} \left(1 + \frac{n+1}{3n\beta} \right) \quad . \quad (4.10)$$

If β is reasonably large, a further approximation is just

$$E(z) \doteq \frac{n-1}{n\beta} \quad . \quad (4.11)$$

We recall from equation (2.5) that $c^2 = 2/\beta$, so the error in the approximation in going from equation (4.10) to equation (4.11) can be judged from knowledge of the likely value of the coefficient of variation c .

4.4 Variance of Log Transformed Random Variates From Gamma Distributions

The remaining result needed for an appraisal of the performance of z and s^2 computed on data from both the lognormal and gamma distributions is the expected value of the variance of log-transformed random variables drawn from a gamma distribution. This is readily obtained from the cumulant generating function for $\ln x$ where x is drawn from a gamma distribution. In the previous section we showed that using equation (4.6) (the above mentioned cumulant generating function) that

$$\begin{aligned} E(s^2) &= \mu_2 = \psi'(\beta/2) \\ &\doteq \frac{2}{\beta} \left(1 + \frac{1}{\beta} + \frac{2}{3\beta^2} \right) \end{aligned} \quad (4.12)$$

where the approximation is obtained by retaining the first 3 terms on the right in equation (3.11). If β is large, a simple approximation is just

$$E(s^2) \doteq 2/\beta \quad . \quad (4.13)$$

4.5 Summary of Results on Expected Values

The results of sections 4.1 through 4.4 are summarized in the following table. (Recall that s^2 is the variance of log-transformed random variables).

<u>Distribution</u>	<u>$E(z)$</u>	<u>$E(s^2)$</u>
Lognormal	$\sigma^2/2$	σ^2
Gamma	$(n-1)/n\beta$	$2/\beta$

Note that varying degrees of approximation are involved for all but $E(s^2)$ for the lognormal distribution.

Interpretation of the simulation results described in section 5 below is aided by noting that the squared coefficient of variation, c^2 , for the lognormal distribution (equation (2.1)) is approximated by σ^2 ; i.e.,

$$c^2 = e^{\sigma^2} - 1$$

$$\approx \sigma^2$$

For the gamma distribution (equation (2.5)), we recall that

$$c^2 = 2/\beta$$

Hence, the entries in the above table are approximated on the left by $c^2/2$ and on the right by c^2 .

4.6 Tests of Significance for Equality of Coefficients of Variation

If we draw sample values from two lognormal distributions, we may test whether the coefficients of variation of these two distributions are equal by computing a one-tailed F-test of homogeneity of variance. It follows from equation (2.2) that homogeneity of variance implies equal coefficients of variation since if two lognormal distributions have equal coefficients of variation, they have equal values of the parameter σ^2 , where σ^2 is the variance of the normal distribution resulting from taking logarithms of observations from a log-normal distribution with parameters σ and μ .

The test for homogeneity of variances is accomplished by computing the statistic

$$F_{\ell n} = s_1^2 / s_2^2$$

where s_1^2 and s_2^2 are sample variances with $n_1 - 1$ and $n_2 - 1$ d.f., respectively, computed on log-transformed lognormal variates. It follows that

$$\frac{(n_i - 1)s_i^2}{\sigma_i^2} \sim \chi_{n_i-1}^2, \quad i = 1, 2,$$

and under the null hypothesis that $\sigma_1^2 = \sigma_2^2 = \sigma^2$, the statistic $F_{\ell n}$ has the F-distribution with $n_1 - 1$ and $n_2 - 1$ d.f.. A one-tailed test is appropriate since in our simulations to determine the power of $F_{\ell n}$, the variance s_1^2 was always computed from observations drawn from the population with larger coefficient of variation, i.e., with larger variance. Hence, large values of $F_{\ell n}$ indicate $\sigma_1^2 > \sigma_2^2$.

When sampling from a gamma distribution, we note that a test of homogeneity of variances on log-transformed gamma variates is equivalent to a test of equal coefficients of variation in the original (untransformed) scale. This is a consequence of the fact that if x is a gamma variate, then $c^2 = 2/\beta \doteq \text{Var}(\ln x)$ (see equation (4.13) and the table in section (4.5)). Hence, there is some rationale for testing equality of coefficients of variation using $F_{\ell n}$ if $F_{\ell n}$ is robust when sampling from gamma distributions.

An F test for the equality of two coefficients of variation on gamma data may be constructed utilizing the statistic z by using a result given by Linhart (1965) and attributed to Bartlett (1937); namely, that if a sample of size n is drawn from a gamma distribution (equation (2.5)), then z is approximately distributed as

$$(n\beta)^{-1} \left[1 + \frac{n+1}{3n\beta} \right] \chi_{n-1}^2 \quad . \quad (4.14)$$

Now, if we replace the term in brackets in equation (4.14) by unity (note the remark following equation (4.11)), we have the result that $zn\beta$ is approximately distributed as χ_{n-1}^2 . Hence, if samples of size n_1 and n_2 are drawn from two gamma distributions and the statistic $n_i\beta_i z_i$ ($i = 1, 2$) computed, then under the null hypothesis that $\beta_1 = \beta_2 = \beta$ we have that

$$F_g = \frac{n_1\beta_1 z_1 / (n_1 - 1)}{n_2\beta_2 z_2 / (n_2 - 1)}$$

has an approximate F distribution with $n_1 - 1$ and $n_2 - 1$ d.f.

In section 5.1 below, we use computer simulations to check on the behavior of F_g when sampling from gamma distributions. We also examine the behavior of $F_{\ell n}$ and F_g when their underlying distributional assumptions are not met; i.e., the behavior of $F_{\ell n}$ when the underlying distribution is

gamma, and that of F_g when sampling from lognormal distributions.

4.7 Bartlett's Test of Equality of Variance

It may be useful to show the connection between the result given above in equation (4.14) and Bartlett's (1937) test for heterogeneity of variances (which results from a modification of the Neyman-Pearson likelihood ratio statistic). Bartlett's test has been widely used, and is sometimes prescribed as a preliminary to conducting an analysis of variance. Scheffé (1959:83) notes that the method is extremely sensitive to non-normality and recommends against its use for "screening" purposes, or the use of any preliminary test of equality of variances, for that matter. However, the context of Scheffé's discussion has to do with testing equality of means by analysis of variance techniques, whereas the present study is largely concerned with variances rather than with means.

Following Pearson and Hartley (1958:57) Bartlett's test is based on a statistic M given by

$$M = N \ln \left[N^{-1} \sum_{i=1}^k v_i s_i^2 \right] - \sum_{i=1}^k v_i \ln s_i^2, \quad (4.15)$$

where s_i^2 , $i = 1, 2, \dots, k$ are k variance estimates based on $v_1, v_2, \dots, v_k$ degrees of freedom, respectively, and where

$$N = \sum_{i=1}^k v_i.$$

Under the assumption of normality the s_i^2 are distributed as $\chi^2_{\sigma_i^2/v_i}$ with v_i degrees of freedom, respectively, and under the null hypothesis that $\sigma_i^2 = \sigma^2$ for all i , where none of the v_i are small, the statistic M was shown by Bartlett to have an approximate χ^2 distribution with $k-1$ degrees of freedom.

It is apparent from equation (4.15) that M is closely related to $z = \ln \bar{x} - \overline{\ln x}$. Indeed, if $v_1 = v_2 = \dots = v_k = v$ then we may write M as

$$\begin{aligned} M &= N \ln \left[k^{-1} \sum_{i=1}^k s_i^2 \right] - N k^{-1} \sum_{i=1}^k \ln s_i^2 \\ &= N \left[\ln \bar{x} - \overline{\ln x} \right] \\ &= Nz \end{aligned}$$

where

$$\bar{x} = k^{-1} \sum_{i=1}^k s_i^2$$

and

$$\overline{\ln x} = k^{-1} \sum_{i=1}^k \ln s_i^2.$$

Furthermore, if the k variances s_i^2 , $i = 1, 2, \dots, k$ are drawn from the same gamma distribution then $M = Nz = kvz$, where, referring again to equation (4.14) in which we set the term in brackets to unity, kvz is approximately distributed as a chi-square with $k-1$ d.f.

Under the assumption of an underlying normal distribution and the null hypothesis of equality of variances $\sigma_1^2, \sigma_2^2 = \dots = \sigma_k^2$, it can easily be shown that the s_i^2 have gamma distributions. For, if these two assumptions hold, then

$$v_i s_i^2 / \sigma^2 \sim \chi_{v_i}^2$$

so that

$$s_i^2 \sim \sigma^2 \chi_{v_i}^2 / v_i.$$

Since the moment generating function of a chi-square random variable with ν_i d.f. is given by

$$M_t(x_{\nu_i}^2) = (1 - 2t)^{-\nu_i/2},$$

this implies

$$M_t(s_i^2) = \left(1 - \frac{2\sigma^2 t}{\nu_i}\right)^{-\nu_i/2}$$

since, in general for the chi-square distribution

$$M_t(bx) = E[e^{bxt}] = (1 - 2bt)^{-\nu/2},$$

where b is an arbitrary constant, equal in the above application to σ^2/ν_i .

Furthermore, it is clear that if $\nu_1 = \nu_2 = \dots = \nu_k = \nu$, then each s_i^2 will have the same gamma distribution.

If the ν_i , $i = 1, 2, \dots, k$ are not all equal, then we may still write $M = Nz$, but $\bar{x}$ and $\overline{\ln x}$ are now weighted means of the s_i^2 and $\ln s_i^2$, respectively, where the s_i^2 are each drawn from different chi-square distributions. The distribution of z in this case is apparently unknown.

Some confusion may be avoided by noting that Bartlett's test is based on data obtained from repeated samplings from each of k distributions (usually n observations from each) whereas the application of z in section (4.6) was to n draws from one distribution (under the null hypothesis). Note, too, that Bartlett's test is for equality of variances while the z -test is concerned with the parameter of a gamma distribution represented here by β .

There is yet another relationship between M and z that is of interest. Pearson and Hartley (1958:57) note Bartlett's use of a modification to M which results in a better approximation to the chi-square distribution if some v_i are small. In particular, if

$$C = 1 + \frac{1}{3(k-1)} \left[\sum_{i=1}^k \frac{1}{v_i} - \frac{1}{N} \right] ,$$

then M/C is more closely approximated by a chi-square with $k-1$ d.f. than is M . Now for the case that $v_i = v$ for all i , we have that $N = kv$ so that

$$C = 1 + (k+1)/3vk ,$$

which is just the term in equation (4.14) which was previously set to unity to obtain the result that $kvz \dot{\sim} \chi_{k-1}^2$. Hence, we see that

$$M/C = kvz / [1 + (k+1)/3vk] \dot{\sim} \chi_{k-1}^2$$

or

$$z \dot{\sim} \frac{1}{kv} [1 + (k+1)/3vk] \chi_{k-1}^2 .$$

Note that the test utilizing equation (4.14) considered in section (4.6) involves the unknown parameter β from a gamma distribution, whereas in the above formulation β is replaced by v , the d.f. of each estimate s_i^2 computed from a normal distribution with variance σ^2 .

Under the assumption of a lognormal distribution, one might go about testing the equality of variances computed from log-transformed data by using Bartlett's test. What happens if the underlying distribution is in fact of the gamma form, is checked by a simulation in section 5.2.

4.8 Scheffé's Test for Equality of Variances

Since Bartlett's test is sensitive to departures from normality, Scheffé (1959:83) proposed a test based on an analysis of variance of the logarithms of sample variances applicable no matter what the underlying distribution. Bartlett and Kendall (1946) proposed a test similar to Scheffé's based on logarithms of sample variances, but only for the case of normal data. Gartside (1972) has compared the power and robustness of several tests for homogeneity of variance (including those of Bartlett (1937) and Bartlett and Kendall (1946)) for data from normal or Weibull distributions. He concluded that Bartlett's test was indeed not robust for the non-normal distributions considered, while that of Bartlett and Kendall were quite robust (in agreement with the results of our simulations in sections 5.2 and 5.3). However, the latter test had the least power of any statistic considered by Gartside (1972) when the data were normal (no power results were given for non-normal data).

Scheffé (1959) requires that two or more sample variances be obtained from each population in the set of populations to be compared. He thus transformed the problem to one of a comparison of means rather than variances. A practical problem, of course, comes up in trying to secure more than one sample from a given population. One solution, suggested by Scheffé, is just to use randomly selected subsamples (from a given population) for computing sample variances.

Scheffé's test statistic is an F-ratio. For a one-way analysis of variance of k sets of populations, where the i^{th} set contains J_i subpopulations, and samples of the same size are drawn from each subpopulation, the test statistic is

$$F_s = \frac{v_e}{k-1} \frac{\sum_{i=1}^k J_i (y_{i.} - \bar{y})^2}{\sum_{i=1}^k \sum_{j=1}^{J_i} (y_{ij} - y_{i.})^2} \quad (4.16)$$

where

$$y_{ij} = \ln s_{ij}^2, \quad y_{i.} = J_i^{-1} \sum_{j=1}^{J_i} y_{ij}$$

$$\bar{y} = \frac{\sum_{i=1}^k J_i y_{i.}}{\sum_{i=1}^k J_i} \quad \text{and} \quad v_e = \sum_{i=1}^k (J_i - 1).$$

It is assumed that $E(s_{ij}^2) = \sigma_i^2$ for $j = 1, 2, \dots, J_i$ where s_{ij} is the sample variance from the j^{th} subpopulation of the i^{th} set. Under the null hypothesis that $\sigma_1^2 = \sigma_2^2 = \dots = \sigma_k^2 = \sigma^2$, F has the F-distribution with $k-1$ and v_e degrees of freedom. Scheffé also gives a test statistic for the case where the sample sizes in each subpopulation are not all equal.

Scheffé's test depends on the approximations (obtained by Scheffé using the "delta method")

$$E(y_{ij}) \approx \eta_i = \ln \sigma_i^2$$

and

$$\text{Var}(y_{ij}) \approx 2/(n_{ij}-1) + \gamma_{2,ij}/n_{ij}, \quad (4.17)$$

where $\gamma_{2,ij}$ is the coefficient of kurtosis for the $(ij)^{th}$ subpopulation defined to be $\gamma_2 = \mu_4/\sigma^4 - 3$. The further assumption is made that $\gamma_{2,ij}$ is the same for all populations.

As noted before, one of the main concerns in this report is with the behavior of log-transformed random variables from a gamma distribution. Hence, Scheffé's test is concerned here with the random variable

$$y = \ln s^2 = \ln \left[\frac{1}{n-1} \sum_{i=1}^n (\ln x_i - \overline{\ln x})^2 \right] \quad (4.18)$$

where $x_i \sim \text{gamma}$. (The fact that s^2 is calculated on logarithms of random variables makes for some confusion, but is relevant only back at the point where we obtained the moment generating function (equation (4.5)), and not here). We limit our attention here to the case of equal sample sizes ($n_{ij} = n$ for all i and j). Estimates of the power of Scheffé's test for this situation are obtained from computer simulations using the statistic F_s given above (equation (4.16)) as well as from tables of the power of F tests (Kastenbaum, Hoel and Bowman (1969)) entered using a non-centrality parameter τ or an approximate non-centrality parameter τ' . In the remainder of this section we define τ and derive approximations to τ and τ' for our distributional assumptions. In sections (5.3) and (5.4) we discuss in more detail the procedures used in our simulations and compare the estimates of power obtained using simulations and the above power tables.

The above power tables by Kastenbaum, et al. (1969) are conveniently arranged in terms of

$$\tau = \frac{|\mu_{\max} - \mu_{\min}|}{[\text{Var}(y)]^{1/2}}, \quad (4.19)$$

where $\mu_{\max}$ and $\mu_{\min}$ are the largest and smallest means, respectively, of the k populations under study, and $\text{Var}(y)$ is the (assumed) common variance of these populations. We need to determine τ for our distributional assumptions (log-transforms of gamma variables). We have that the means in equation (4.19) are given approximately by

$$\mu = E(y) \approx \ln \sigma^2$$

where σ^2 is given by equation (4.12). Also, by utilizing the approximation of $\text{Var}(y)$ given in equation (4.17), we have that

$$\text{Var}(y) = 2/(n-1) + k_4/n\sigma^4 \quad . \quad (4.20)$$

Hence, we need to determine k_4 (the fourth cumulant as defined in section (4.3)) and $\sigma^4 = \mu_2^2$.

Equation (4.12) provides an estimate of $\mu_2 = \sigma^2$, but for use in our simulations an additional term is utilized. The result, expressed in terms of the coefficient of variation ($c^2 = 2/\beta$ for equation (2.5)) is

$$\sigma^2 \doteq c^2 + \frac{c^4}{2} + \frac{c^6}{6} - \frac{c^{10}}{30} \quad . \quad (4.21)$$

The cumulant k_4 is obtainable from the cumulant generating function (equation (4.6)). We find that expressing the series expansions of equation (4.6) in terms of c , we have

$$k_4 = \psi'''(\beta/2) \doteq 2c^6 + 3c^8 + 2c^{10} \quad . \quad (4.22)$$

Hence, the approximate power of Scheffé's test when the $x_i \sim \text{gamma}$ and $y = \ln s^2$ as defined in equation (4.18) can be obtained by entering the power tables of Bowman (1971) and Kastenbaum, et.al. (1969) with the non-centrality parameter

$$\tau = \frac{\left| \ln \left(c_1^2 + \frac{c_1^4}{2} + \frac{c_1^6}{6} - \frac{c_1^{10}}{30} \right) - \ln \left(c^2 + \frac{c^4}{2} + \frac{c^6}{6} - \frac{c^{10}}{30} \right) \right|}{\left[2/(n-1) + k_4/n \left(c^2 + \frac{c^4}{2} + \frac{c^6}{6} - \frac{c^{10}}{30} \right)^2 \right]^{1/2}}. \quad (4.23)$$

Power calculations are somewhat simplified by an approximation achieved by retaining only the first term of equations (4.21) and (4.22) in computing (4.20). Doing this we obtain

$$\text{Var}(y) \doteq 2/(n-1) + 2 c^2/n \quad (4.24)$$

so that equation (4.23) is approximately given by

$$\tau' = \frac{\left| \ln c_1^2 - \ln c^2 \right|}{\left[2/(n-1) + 2 c^2/n \right]^{1/2}}, \quad (4.25)$$

where c_1 and c are the largest and smallest coefficients of variation, respectively, in the k populations under consideration. The usual calculation of the non-centrality parameter assumes a constant variance, whereas here the variance and mean both depend on the same parameter (β). In this study the denominator is arbitrarily based on the smaller value of the coefficient of variation. Perhaps a preferable alternative would be to use an average value of c in the denominator inasmuch as exact power statements in terms of τ require that one assume all of the populations other than those with $\mu_{\max}$ and $\mu_{\min}$ fall at the midpoint between those two.

We have checked on the accuracy of τ' in equation (4.25) as compared with τ in equation (4.23). The results (given in Table 4.1) indicates τ' tends to underestimate τ , particularly as n and c_1 become large.

4.9 Confidence Limits on the Coefficient of Variation

This section is included for convenience in future work. We do not use the results given here in this report.

Koopmans, Owens and Rosenblatt (1964) derive confidence limits on σ^2 and c of the lognormal distribution (equation (2.1)) as follows: Let $x_1, x_2, \dots, x_n$ be a random sample from the lognormal distribution, and let $y_i = \ln x_i$ for $i = 1, 2, \dots, n$. Then

$$\frac{(n-1)s^2}{\sigma^2} \sim \chi_{n-1}^2$$

where $s^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2$. Hence, if $\chi_{1,n}^2$ and $\chi_{2,n}^2$ are solutions of the equations

$$\Pr[\chi^2 \leq \chi_{1,n}^2] = \alpha_1 \quad \text{and} \quad \Pr[\chi^2 \geq \chi_{2,n}^2] = \alpha_2$$

and if

$$\sigma_L^2 = (n-1) s^2 / \chi_{2,n}^2$$

$$\sigma_U^2 = (n-1) s^2 / \chi_{1,n}^2,$$

then

$$\sigma_L^2 \leq \sigma^2 \leq \sigma_U^2$$

is a $1-\alpha$ confidence interval for σ^2 . Since $[\exp(\sigma^2) - 1]^{1/2}$ is a strictly increasing function of σ ,

TABLE 4.1

Comparison of τ and τ' Computed Using
Equations (4.23) and (4.25) When $c = 0.20$

c_1	$n = 6$		$n = 21$	
	τ	τ'	τ	τ'
0.20	0.0000	0.0000	0.0000	0.0000
0.25	0.7116	0.6942	1.4198	1.3851
0.30	1.3001	1.2613	2.5943	2.5169
0.35	1.8052	1.7409	3.6020	3.4738
0.40	2.2502	2.1563	4.4898	4.3027
0.45	2.6502	2.5227	5.2880	5.0338
0.50	3.0156	2.8504	6.0172	5.6878
0.55	3.3538	3.1469	6.6920	6.2794
0.60	3.6702	3.4176	7.3232	6.8196
0.65	3.9687	3.6666	7.9188	7.3164

$$\left[\exp(\sigma_L^2) - 1 \right]^{1/2} \leq c \leq \left[\exp(\sigma_U^2) - 1 \right]^{1/2}$$

is a $1-\alpha$ confidence interval for c .

Linhart (1965) provides confidence limits for c in gamma distributions based on the z statistic.

5.0 SIMULATION RESULTS FOR COMPARING COEFFICIENTS OF VARIATION

In consequence of the various approximations in the preceeding results, it is worthwhile to do some checking via simulation. One can seldom narrow down the range of possibilities enough to use simulation results for a general confirmation, but the results given below may help support the plausibility of approximate theoretical results. Since the programs (see Appendix) are mostly quite simple, one can readily generate the necessary test for particular practical problems as they come up.

5.1 Tests for Evaluation of Coefficients of Variation

In this section we describe our simulation results concerning the behavior of the tests

$$F_{\lambda n} = s_1^2 / s_2^2$$

and

$$F_g = \frac{n_1 \beta_1 z_1 / (n_1 - 1)}{n_2 \beta_2 z_2 / (n_2 - 1)} \quad .$$

We know that if s_1^2 and s_2^2 are computed on log-transformed samples of size n_1 and n_2 from two lognormal distributions for which the coefficients of variation are equal, then $F_{\lambda n}$ has the F distribution with $n_1 - 1$ and

$n_2 - 1$ degrees of freedom. We also recall that F_g will be approximately distributed as an F with $n_1 - 1$ and $n_2 - 1$ d.f. if samples of size n_1 and n_2 are drawn from two gamma distributions with equal coefficients of variation. Results are obtained here for the case of equal sample sizes, i.e., $n_1 = n_2 = n$.

The behavior of these tests are studied here in terms of their power, i.e., what proportion of the time do we get a significant test result using F_{ln} and F_g for a given difference in coefficient of variation between the two distributions. When the two coefficients of variation are equal, the power should approximate the chosen level (α) for Type I error.

F_{ln} was applied to log-transformed random variables generated from both lognormal and gamma distributions having both a standard value of the coefficient of variation (c) and some larger value. Similarly F_g was computed on samples from both lognormal and gamma distributions. The required pseudo-random variates were generated as described in section (2.2). The simulation results were obtained using the computer programs LØGCV and GAMACV listed in the Appendix.

A summary of the simulation results appears in Table 5.1. Each entry in the table was obtained by generating 1000 samples of size $n = 6$ or 21 from two gamma or lognormal distributions with equal or unequal coefficients of variation c and computing the statistic F_{ln} or F_g . The upper part of the table shows the increase in proportion of significant F-ratios (increasing power) as the coefficient of variation of the second set of data is increased.

The lower part of the table shows the same process, but the "reference" value c of the coefficient of variation (0.200) is lower than in the upper part. The right and left halves of the table do not quite correspond in consequence of the necessity to generate gamma-distributed random variables by using an integral number of draws from an exponential distribution (see section (2.2)). The lack of correspondence is an oversight in that the simulated lognormal distributions could have been matched to have the same set of coefficients of variation as the gamma distributions, but enough computing time has been invested in the results to argue against repeating the simulation.

The first row of each part of the table corresponds to a test with identical coefficients of variation, so an $\alpha = 0.05$ significance level test utilizing $F_{\ell n}$ and F_g should result in values of 0.05 in each such entry (aside from sampling variability) if the tests are robust. The only statistically significant differences from the 0.05 nominal level were obtained when generating samples of size 21 from gamma distributions with $c = 0.447$ and 0.200. We see that the actual Type I errors of the test $F_{\ell n}$ utilizing s^2 are significantly greater than 0.05 for these values of c . One possible explanation for this result is that the test $F_{\ell n}$ is not robust when sampling from gamma rather than lognormal distributions when $c = 0.447$ and 0.200 for $n = 21$. The result for F_g utilizing z is also significant for $c = 0.200$. This latter result suggests that the distribution of $n\beta z$ is perhaps not closely approximated by the chi-square distribution with $n-1$ d.f., at least for $c = 0.200$ and $n = 21$. Since no significant results

TABLE 5.1

Simulated Power for Tests F_{ln} and F_g Using s^2 and z
as Criteria on Gamma and Lognormal Variables

Gamma Distribution					Lognormal Distribution				
Comparisons Against $c = 0.447$					Comparisons Against $c = 0.400$				
c	$n = 6$		$n = 21$		c	$n = 6$		$n = 21$	
	s^2	z	s^2	z		s^2	z	s^2	z
0.447	0.049	0.043	0.074*	0.055	0.400	0.046	0.047	0.041	0.043
0.500	0.091	0.077	0.143	0.123	0.450	0.070	0.074	0.111	0.117
0.577	0.150	0.141	0.355	0.309	0.600	0.179	0.174	0.494	0.486
0.707	0.300	0.253	0.696	0.659	0.700	0.274	0.255	0.701	0.699
1.000	0.607	0.543	0.978	0.071	0.800	0.341	0.323	0.815	0.812

Comparisons Against $c = 0.200$					Comparisons Against $c = 0.200$				
c	$n = 6$		$n = 21$		c	$n = 6$		$n = 21$	
	s^2	z	s^2	z		s^2	z	s^2	z
0.200	0.049	0.048	0.074*	0.069*	0.200	0.046	0.046	0.041	0.042
0.258	0.141	0.135	0.320	0.308	0.250	0.112	0.111	0.214	0.213
0.301	0.221	0.215	0.589	0.582	0.300	0.198	0.187	0.545	0.544
0.447	0.529	0.520	0.740	0.720	0.400	0.393	0.384	0.900	0.902
0.707	0.864	0.852	0.900	0.900	0.600	0.695	0.696	0.996	0.995

* Statistically different from nominal level $\alpha = 0.05$ using 2-tailed large sample normal test.

were obtained for F_g when sampling from lognormal distributions, this suggests the distribution of F_g is fairly robust, i.e., it is not greatly affected if sampling occurs from lognormal rather than gamma distributions.

The simulation results for the criterion s^2 can be cross-checked inasmuch as $F_{\ell n}$ when computed on log-transformed observations from two lognormal distributions follows the incomplete beta function. Following Pearson and Hartley (1958:33) when the same size sample (n) is drawn from each normal distribution the probability that $F_{\ell n}$ exceeds a specified value, F_0 (the critical value of $F_{\ell n}$ for testing at the $\alpha = 0.05$ level of significance) is

$$\int_{F_0}^{\infty} f(F/v, v) dF = I_{x_0} \left(\frac{v}{2}, \frac{v}{2} \right)$$

where

$$v = n-1 \quad \text{and}$$

$$x_0 = v/(v + v F_0)$$

and I_x represents the incomplete beta-function defined as

$$I_x(a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \int_0^x u^{a-1} (1-u)^{b-1} du$$

The values of $I_{x_0} \left(\frac{v}{2}, \frac{v}{2} \right)$ for $v = 5$ and 20 were obtained from Pearson's tables (Pearson (1934)) and are given in Table 5.2 along with the simulated results for $F_{\ell n}$ from Table 5.1. The simulated and exact power are seen to agree quite well.

Returning to Table 5.1 it appears that the power of F_{ln} and F_g is almost identical when samples are drawn from lognormal distributions. For samples from gamma distributions it appears that the power is consistently smaller for the z-criterion (test F_g), suggesting use of tests involving s^2 rather than z. It is possible, however, that this apparent difference in power is a result of the non-robustness of F_g so that the power should be computed by reference to an F value different than the nominal value.

We also note that there is a considerable degree of similarity between results in the two halves of the table, i.e., the powers are of roughly the same magnitude for similar comparisons regardless of the underlying distribution. However, the main conclusion from these results is that it seems that either criterion, s^2 or z, gives about the same results against either distribution.

5.2 Robustness of Bartlett's Test

The program BART (see Appendix) was written to check the consequences of using Bartlett's test (section 4.7 above) on log-transformed data from a gamma distribution. Scheffé (1959:83) and Pearson and Hartley (1958:57) warn the user to beware of the test's sensitivity to non-normality. The program computes 10 variances, each based on 6 log-transformed random variables from a simulated gamma distribution for a given coefficient of variation. Each set of 10 variances is subjected to Bartlett's test and the process is replicated 100 times for each of 6 different values of the coefficient of variation. The number of significant (0.05 level) chi-square values is recorded, and would be expected to be 5 for each lot of 100. Four separate runs were performed in order to achieve a total sample size of 400.

TABLE 5.2

Simulated and Expected[†] Power of $F_{\ell n}$ on Log-transformed
Variables From the Lognormal Distribution

<u>Comparison Against $c = 0.400$</u>				
<u>c</u>	<u>n = 6</u>		<u>n = 21</u>	
	<u>Simulated</u>	<u>Expected</u>	<u>Simulated</u>	<u>Expected</u>
0.400	0.046	0.050	0.041	0.050
0.450	0.070	0.075	0.111	0.119
0.600	0.179	0.175	0.494	0.479
0.700	0.274	0.253	0.701	0.699
0.800	0.341	0.330	0.815	0.839

<u>Comparison Against $c = 0.200$</u>				
<u>c</u>	<u>n = 6</u>		<u>n = 21</u>	
	<u>Simulated</u>	<u>Expected</u>	<u>Simulated</u>	<u>Expected</u>
0.200	0.046	0.050	0.041	0.050
0.250	0.112	0.110	0.214	0.244
0.300	0.198	0.191	0.545	0.531
0.400	0.393	0.379	0.900	0.898
0.600	0.695	0.679	0.996	0.997

[†] Using Pearson's (1934) tables.

The results (which appear in Table 5.3, along with those for Scheffé's test of equality of variances) indicate a larger number of significant chi-square values than desirable, particularly for the larger values of the coefficient of variation.

A rough guide to significance of the observed results is given by the binomial approximation to "normal" 95% confidence limits on p , i.e., we reject the hypothesis that $p = 0.05$ if $\hat{p}$ falls outside the limits

$$p \pm 2 \left(\frac{pq}{n} \right)^{1/2}$$

or, for $p = 0.05$, whenever

$$0.072 \leq \hat{p} \leq 0.028 \quad .$$

Hence, when $n = 400$ trials, we reject whenever the number of significant chi-square values falls outside the interval from 11 to 29. The results in Table 5.3 support the supposition that it may be hazardous to employ Bartlett's test on log-transformed data from a gamma distribution.

5.3 Robustness of Scheffé's Test

In this section we check on the robustness of Scheffé's test F_s when the underlying data are gamma distributed. Scheffé's test (section 4.8 above) constitutes a one-way analysis of variance on logarithms of variances (see equation (4.18)). It is thus necessary to have replicate sets of data from each of the populations being compared (Scheffé suggests random subdivision of samples when necessary).

TABLE 5.3

Tests of Equality of Variances for Log-transformed Random
Variables From a Simulated Gamma Distribution[†]

c	Bartlett's Test					Scheffé's Test				
	1	2	3	4	Total ^{††}	1	2	3	4	Total ^{††}
0.100	8	8	5	8	29	8	5	6	4	23
0.200	4	5	5	4	18	5	8	2	3	18
0.302	8	11	3	5	27	7	8	4	3	22
0.408	4	10	8	11	33	5	6	6	3	20
0.500	11	10	13	21	55	5	7	9	10	31
0.577	11	13	9	9	42	4	9	5	8	26

[†] Entries are number of significant chi-square values in 100 replications.

^{††} Based on 400 replications.

The simulated test was performed on the same general basis as for Bartlett's test (100 replicates of 10 variances each based on 6 log-transformed random variables from a simulated gamma distribution) except that it is necessary to use successive pairs of variances as entries in "cells" of the one-way analysis of variance. In the present arrangement there are thus 5 "treatments" each having 2 "observations", and the degrees of freedom are thus 4 and 5. In the notation of section (4.2) we have $k = 5$, $J_i = 2$ for $i = 1, 2, \dots, k$ and $n = 6$. The program used here appears in the Appendix (SCHEF).

Results of the simulation (Table 5.3) are encouraging, being mostly near to the expected 20 significant chi-squares in 400 replications. Two "observations" per "treatment" is the minimum possible and one might intuitively suppose that the test should not be very sensitive to the number of "treatments" (as also evidenced by examination of power tables), so that the range for further simulation study presumably depends mainly on varying the sample size on which each variance is based (now 6). Appraisal of the power of the test (sections 4.8 and 5.4) is also important in judging the need for further study by simulations.

5.4 Power of Scheffé's Test

Power of the Scheffé test F_s for equality of variances was studied using computer simulations by modifying the program described above (SCHEF) so that the first pair of variances was based on a fixed coefficient of variation ($c = 0.20$) and the last pair ranged over a selected set of values of c . In order to make the results compatible with the available power

tables (Kastenbaum, et al. (1969)), the intermediate pairs of variances were computed to have mean values approximately midway between the initial and final pair.

The basis for calculating a coefficient of variation for the intermediate pairs of variances depends on Scheffé's approximation for the mean of the random variables used in the analysis of variance (i.e., logarithms of sample variances). This is just

$$E(y_{ij}) \doteq \ln \sigma_i^2 ,$$

where we replace σ_i^2 by the first term of its definition in equation (4.21), i.e., $\sigma_i^2 = c_i^2$. We want to find the mean M of the two extremes in a particular simulation, i.e.,

$$\begin{aligned} M &= (\ln c^2 + \ln c_1^2) / 2 \\ &= \ln c + \ln c_1 \\ &= -\frac{1}{2} (\ln r + \ln r_1) , \end{aligned}$$

where

$$r = 1/c^2 \quad \text{and} \quad r_1 = 1/c_1^2 ,$$

and c and c_1 were chosen so that r and r_1 were integers. The integral number of draws from the exponential distribution for the mid-value coefficient of variation was taken to be the nearest integer less than or equal to r_M , where

$$r_M = e^{-M} = \exp \left[\frac{1}{2} (\ln r + \ln r_1) \right] .$$

Since r_M must be expressed as an integer, the resulting coefficient of variation will not be exactly equal to the mean of the two values of c being compared. The effect for the principal coefficients of variation used here is shown below for $c = 0.200$ (for which $r = 25$).

r_1	$c_1 = (1/r_1)^{1/2}$	M	r_M (as an integer)	M using r_M
6	0.408	-2.505	12	-2.485
5	0.447	-2.414	11	-2.398
4	0.500	-2.302	9	-2.197
3	0.577	-2.159	8	-2.079
2	0.707	-1.956	7	-1.946

Thus, the means of the "intermediate" sets of simulated data are not exactly midway between the two extreme values, as required for an exact comparison with power tables. Presumably the effect will be to inflate the numerator variance in the F-ratio of Scheffé's test slightly.

The only other change in this program from SCHEF is a provision for recording and printing out the average variance of the logarithms of the sample variances (denominator of F_S). We have seen that this variance is approximated by equation (4.20) which can be further approximated by equation (4.24). Note that equations (4.20) and (4.24) do not depend on the number of sets of variances used, but only on the number of random variables used to generate each variance. Hence, τ and τ' , with which we enter the power tables of Kastenbaum, et al. (1969), are not influenced by the number of "treatments" (number of sets of variances under test). This differs from the usual power calculation, where $\text{Var}(y)$ is estimated from the error mean square and thus decreases with increasing number of sets of observations.

Four simulations were conducted with the program, and the results appear in Table 5.4 along with tabular power data obtained using both τ and τ' (equations (4.23) and (4.25)). The first two simulation runs employed 10 sets of variances per replicate (so pairing gives 5 groups, or $k = 5$ "treatments" in the power tables by Bowman (1971)) and were based on 6 and 21 random variables per variance. The second pair of simulation runs used 20 sets of variances ($k = 10$) and also 6 and 21 random variables per variance. All of the runs used 100 replicates per coefficient of variation. Some notion of the sampling variability in simulated power results in Table 5.4 is obtained from the approximate 95% confidence intervals given by

$$\hat{p} \pm 2\sqrt{\frac{\hat{p}\hat{q}}{n}}.$$

For $n = 100$ the values of $2(\hat{p}\hat{q}/n)^{1/2}$ for simulated estimates of power $\hat{p} = 0.2, 0.50$ and 0.90 are $0.08, 0.10$ and 0.06 , respectively, indicating it may be desirable in future studies to increase the number of replicates for better precision.

In Table 5.4 we see that the approximate tabular power of Scheffé's test using τ' is somewhat smaller than that obtained using τ . The difference appears to be larger for $n = 6$ than when $n = 21$, although only the result for $c_1 = 1.000$ is available for $n = 6$. We also note that in nearly all cases considered, the simulated power of Scheffé's test is closer to the tabular power obtained using τ' than that found using τ . With the results in hand, it appears there is little reason to

TABLE 5.4

Comparison of Simulated and Tabular Power of Scheffé's Test for Homogeneity of Variances

5 "Treatments"							5 "Treatments"						
n = 6							n = 21						
c_1	τ	Tabular Power	τ'	Tabular Power	Simulated Power	Simulated Variance	c_1	τ	Tabular Power	τ'	Tabular Power	Simulated Power	Simulated Variance
0.200	0.0000	0.050	0.0000	0.050	0.01	0.510	0.200	0.0000	0.050	0.0000	0.050	0.04	0.111
0.408	2.3169	†	2.2179	†	--	---	0.408	4.6229	0.648	4.4256	0.611	0.61	0.119
0.447	2.6273	†	2.5019	†	0.17	0.526	0.447	5.2422	0.759	4.9923	0.717	0.70	0.116
0.500	3.0156	†	2.8504	†	0.21	0.601	0.500	6.0172	0.863	5.6878	0.823	0.79	0.116
0.577	3.5271	†	3.2960	†	0.32	0.573	0.577	7.0377	0.950	6.5769	0.917	0.91	0.132
0.707	4.2910	†	3.9281	†	0.45	0.566	0.707	8.5619	0.992	7.8382	0.979	1.00	0.123
1.000	5.7369	0.829	5.0067	0.720	0.79	0.528	1.000	11.4468	1.000	9.9905	1.000	--	---
10 "Treatments"							10 "Treatments"						
0.200	0.0000	0.050	0.0000	0.050	0.03	0.497	0.200	0.0000	0.050	0.0000	0.050	0.04	0.108
0.408	2.3169	†	2.2179	†	--	---	0.408	4.6229	†	4.4256	†	--	---
0.447	2.6273	†	2.5019	†	0.18	0.565	0.447	5.2422	0.725	4.9923	0.677	0.59	0.134
0.500	3.0156	†	2.8504	†	0.20	0.547	0.500	6.0172	0.847	5.6878	0.804	0.79	0.123
0.577	3.5271	†	3.2960	†	0.18	0.570	0.577	7.0377	0.949	6.5769	0.913	0.92	0.123
0.707	4.2910	†	3.9281	†	0.36	0.555	0.707	8.5619	0.994	7.8382	0.981	0.95	0.133
1.000	5.7369	0.810	5.0067	0.680	0.70	0.553	1.000	11.4468	1.000	9.9905	1.000	1.00	0.126

†Not tabulated in power tables.

bother computing τ when the simpler expression τ' is available. Since there are approximations involved in all three cases (simulated, τ and τ'), it is difficult to decide whether to recommend the use τ or τ' to enter the power tables. It is perhaps well to recall that $\text{Var}(y)$ (equation (4.20)) is itself an approximation derived by Scheffé using the first term of a Taylor series expansion. In addition, of course, we have the errors introduced by the truncated series expansions of k_4 and σ^2 utilized in computing equation (4.20).

The "simulated variances" of Table 5.4 may be compared with $\text{Var}(y)$ (equation (4.24)). For $n = 6$ and 21 when $c = 0.20$ we find $\text{Var}(y) = 0.413$ and 0.104 , respectively. The simulated values are clearly somewhat larger than these "expected values" suggesting that one should perhaps use some intermediate value of c when computing $\text{Var}(y)$ in τ and τ' . However, any such decision needs to be preceded by an appraisal of other factors, including the fact that these simulated variances are averages of sample variances each based on two observations. Also, the simulated values, while somewhat inflated, do not change appreciably with increasing coefficients of variation.

In general, the simulated power calculations seem in sufficiently good accord with tabular values to encourage the use of τ or τ' (equations (4.23) and (4.25)) to obtain rough guides to the power of Scheffé's test for equality of variances using log-transformed data from a gamma distribution.

It may be noted that, if one assumes lognormality, equation (4.24) becomes

$$\text{Var}(y) = \frac{2}{n-1}$$

since $\gamma_2 = 0$ for the normal distribution resulting from a logarithmic transformation. From equation (2.1) the variance (after transformation) is σ^2 , so the equation corresponding to (4.25) becomes

$$\tau' = \frac{|\ln \sigma_1^2 - \ln \sigma^2|}{\left(\frac{2}{n-1}\right)^{1/2}}.$$

Recalling from equation (2.2) that $c^2 = e^{\sigma^2} - 1$ for the lognormal distribution, one can approximate σ on the basis of guesses as to c and the range of c that it is considered desirable to detect.

6.0 CONCLUSIONS

Our general conclusion from this study is that it appears reasonably safe to neglect the differences between the lognormal and the gamma distributions for many practical purposes. Although the two criteria (z and s^2) examined here seem to behave reasonably well for either distribution, the simplest procedure for statistical analyses is just to use a logarithmic transformation and to proceed as though the data were thus "normalized". Scheffé's test of equality of variances, which is simply a one-way analysis of variance on logarithms of the sample variances, is recommended over that of Bartlett's as a test of equality of variances

when the data are non-normal. An approximation of the noncentrality parameter for the power of Scheffé's test is given by equations (4.23) or (4.25).

Our simulation studies of Jackson's "separate families" tests for distinguishing between lognormal and gamma distributions indicate their power to be rather low for the cases studied unless the sample size is at least 100, particularly if the density functions being compared are not widely spaced. Also, it appears in general that the null distribution of the tests is not well approximated by the $N(0,1)$ unless n is at least 100. It appears that the power and null distribution depend to some extent on the particular lognormal or gamma distribution present, as determined in our case by the value of c .

Several of the results obtained above appear either to need more study to support the tentative conclusions reached here, or to be of broader interest and thus worth treatment in greater detail than was attempted here. These are summarized in the following paragraphs.

- (1) It might be useful to conduct a detailed study of the sampling distribution of $\ln \bar{x}$, where $\bar{x}$ is the mean of random variables from a lognormal distribution. This result is necessary for determining the sampling distribution of z . Since, from Tables 5.1 and 5.2 we have that the simulated power of F_g (utilizing the z criterion) calculated on lognormal data is nearly identical to that calculated from the variance s^2 of log-transformed data utilizing the statistic $F_{\lambda n}$, it would seem intuitively plausible

that the distribution of z might be well approximated by the chi-square. Such results might also throw some additional light on the behavior of Bartlett's test with other than normally distributed random variables. Naus (1969) gives some results on the distribution of the logarithm of the sum of two lognormal variates.

- (2) Simulation studies of the power of Scheffé's test F_S (when applied to log-transformed data from a gamma distribution) suggest an approximation to the non-centrality parameter, equations (4.23) or (4.25), gives a reasonable indication of the power. These results might usefully be extended, but such an extension should be preceded by an effort to look into the theoretical aspects in further detail (at least to limit the amount of simulation needed, when possible).
- (3) The results obtained here should shed some light on the effects of pooling and compositing on the subsequent analysis of data. In particular, the approximations to expected values of z (e.g., equations (4.4) and (4.11)) permit comparison of the expected results of two kinds of sampling: (i) compositing (or pooling), as represented by $\bar{x}$; and (ii) taking a logarithmic average of individual determinations, as represented by $\overline{\ln x}$. The expected values of $z = \ln \bar{x} - \overline{\ln x}$, are thus approximated by $c^2/2$ (section 4.5), giving an indication of the magnitude of the difference in results obtained by the two procedures. The problem of transforming back to the original scale of measurement has not been mentioned above but obviously needs also to be considered in practice.

7.0 REFERENCES

- Abramowitz, Milton and I. A. Stegun (Editors). (1964). Handbook of Mathematical Functions. National Bureau of Standards, Applied Mathematics Series, 55. U. S. Government Printing Office, Washington, D.C.
- Aitchison, J. and J. A. C. Brown. (1969). The Lognormal Distribution. Cambridge University Press.
- Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. Proc. Royal Soc. A, 160:268-282.
- Bartlett, M. S. and D. G. Kendall. (1946). The statistical analysis of variance-heterogeneity and the logarithmic transformation. J. Roy. Statist. Soc. Suppl. 8:128-138.
- Bowman, K. O. (1971). Sample size requirement: Continuation. ORNL-4712, Oak Ridge National Laboratory, Oak Ridge, Tenn.
- Box, G. E. P. and M. E. Muller (1958). A note on the generation of normal deviates. Ann. Math. Stat. 29:610-611.
- Cox, D. R. (1962). Further results on tests of separate families of hypothesis. J. R. Statist. Soc. B. 24:406-424.
- Cox, D. R. (1961). Tests of separate families of hypothesis. Proc. 4th Berkeley Symp., 1:105-123.
- Eberhardt, L. L. (1964). Variability of the strontium-90 and caesium-137 burden of native plants and animals. Nature 204(4955):238-240.
- Ellett, W. H. and G. L. Brownell (1964). The time analysis and frequency distribution of caesium-137 fall-out in muscle samples. pp. 155-166 in Assessment of Radioactivity in Man, Vol. II, International Atomic Energy Agency, Vienna.
- Gartside, P. S. (1972). A study of methods for comparing several variances. Jour. of the Amer. Statistical Assoc., 67(338):342-346.
- Gilbert, R. O. (1973). An evaluation of four pseudo-random number generators. AEC Research and Development report BNWL-1743, PNL, BMI, May 1973.
- Hahn, G. J. and S. S. Shapiro, (1967). Statistical Models in Engineering, J. Wiley, New York.
- Jackson, O. A. Y. (1969). Fitting a gamma or log-normal distribution to fibre-diameter measurements on wool tops. Appl. Statistics 18:70-75.

- Jackson, O. A. Y. (1968). Some results on tests of separate families of hypothesis. Biometrika 55:355-363.
- Kastenbaum, M. A., D. G. Hoel, and K. O. Bowman (1969). Sample size requirements: tests of equality of several Gaussian means. ORNL-4468, Oak Ridge National Laboratory, Oak Ridge, Tenn.
- Koopmans, L. H., D. B. Owen, and J. I. Rosenblatt. (1964). Confidence intervals for the coefficient of variation for the normal and log-normal distributions. Biometrika 51 (1 and 2):25-32.
- Linhart, H. (1965). Approximate confidence limits for the coefficient of variation of gamma distributions. Biometrics 21:733-738.
- Monfort, F. (1964). Fitting gamma distributions to fibre-diameter measurements on wool tops. J. Text. Inst. 55:T185-T191.
- Naus, J. I. (1969). The distribution of the logarithm of the sum of two log-normal variates. J. Amer. Stat. Assn. 64(326):655-659.
- Olmstead, J. M. H. (1959). Real Variables. Appleton - Century - Crofts, N.Y.
- Pearson, E. S. and H. O. Hartley (1958). Biometrika Tables for Statisticians. Cambridge University Press.
- Pearson, K. (1934). Tables of the Incomplete Beta-Function. Cambridge University Press.
- Remmenga, E. E. and F. W. Whicker (1967). Sampling variability in radio-nuclide concentrations in plants native to the Colorado Front Range. Health Physics. 13:977-983.
- Scheffé, H. (1959). The Analysis of Variance. J. Wiley, New York.

8.0 APPENDIX: SIMULATION PROGRAMS USED IN THE STUDY

1. NORM

Produces pair of random selections from unit normal curve (mean zero). Only the first number was used in this study. Starting number contained in program.

2. RN2

Produces random number on interval (0,1). Starting number input from main program.

3. RN4

Produces two random numbers on interval (0,1). Starting number input from main program.

4. LOGN

Simulates samples from lognormal distribution and computes both sample and expected coefficients of skewness and kurtosis along with variances and a significance test. Running time 72 seconds (UNIVAC 1108, with constants as listed).

5. SKEW

Same process as LOGN but with gamma distribution. Running time 320 seconds.

6. LOGCV

Runs series of pairwise F-tests over 5 values of c . Based on lognormal distribution. Running time for 6 random variables per set 33 seconds; 100 seconds for $N1 = 21$.

7. GAMA

Runs series of pairwise F-tests (based on z-criterion) in similar manner to LOGCV, but using gamma distribution. Running time ranges from 30 to 300 seconds depending on $N1$ and c (small values of c require many draws from RN2).

8. BART

Simulates Bartlett's test for equality of variances on log-transformed gamma distribution. Running time 260 seconds.

9. SCHEF

Simulates Scheffé's test for equality of variances on log-transformed gamma distribution. Running time 280 seconds.

10. SCH

Simulation to calculate power of Scheffé's test. Based on SCHEF. Running time 40 to 240 seconds for simulations discussed in section 4.4 (140 seconds for parameters as listed in program).

11. PCALC

Calculates mean value of logarithm of coefficient of variation (to show difference between means actually used in power calculations and expected values).

12. TRANS

Transforms the power tables by Kastenbaum et. al. (1969) into ratios of maximum to minimum values of c that can be detected with given probability.

NORM

```

890 @E   FR5  NORM,NORM
900      SUBROUTINE NORM (X,Y)
910 C---PSEUDORANDOM NORMAL NUMBER GENERATOR USING UNIFORM GENERATOR RN4
920      DATA U1,U2 / 0233362477003, 0212312312323 /
930      CALL RN4 (U1,U2,X1,X2)
940      SQ = SQRT(-2. * ALOG(X1))
950      AA = 6.283185306 * X2
960      X = SQ * COS(AA)
970      Y = SQ * SIN(AA)
980      RETURN
990      END

```

RN2

RN2 17:47 08/17/72

```

10 @N   ASM  RN2,RN2
20 B11  EQU  11
30 A0    EQU  12
40 A1    EQU  13
      N (0,1)
50 A2    EQU  14
      AFTER
60 XM    EQU  15
      2
70 .
80 .
90 RN2*  LUF  A0,*0,B11
100      SSC  A1,27
110      JP   A1,$+3
      2**35))
120      AND  A1,(0377777777777777)
130      LA   A1,A2
140      AA   A1,*0,B11
150      JP   A1,$+3
      2**35))
160      AND  A1,(0377777777777777)
170      LA   A1,A2
180      A,XM A1,1
190      JP   A1,$+3
      2**35))
200      AND  A1,(0377777777777777)
210      LA   A1,A2
220      SA   A1,*0,B11
230      AND  A1,(0777777777777777)
240      AA   A2,TW0
250      FA   A2,TW0
260      SA   A2,*1,B11
270      J    3,B11
280 TW0  +    0200000000000000
290      END

```

. SUBROUTINE RN2(U,FU)
 . U IS A RANDOM 35 DIGIT INTEGER
 . FU FLOATING POINT NUMBER RANDOM 0

 . INITIALLY U MUST BE FURNISHED

 . THAT IT IS KEPT CURRENT BY RN

 . SUGGESTED INITIAL VALUE IS
 . U=011060471625 (OCTAL)
 . LOAD U
 . MULTIPLY BY 2**9
 . TEST FOR OVERFLOW AND RETAIN A1(MODC

 . ADD U
 . TEST FOR OVERFLOW AND RETAIN A1(MODC

 . ADD 1
 . TEST FOR OVERFLOW AND RETAIN A1(MODC

 . SAVE NEW U
 . FLOAT U

 . RETURN FU

END-

READY

RN4

```

120 @N   ASM   RN4,RN4
130 . REFERENCE   RICHARD KRØNMAL, EVALUATION OF A PSEUDORANDOM NORMAL
      NUMBER
140 .             GENERATOR JACM VOL 11, NO 3 (JULY 1964) PP 357-363
150 B11   EQU    11      . SUBROUTINE RN4(U1,U2,FU1,FU2)
160 A0    EQU    12      .   U-S ARE PANDØM 35 DIGIT INTEGER
      S
170 A1    EQU    13      .   FU-S FLOATING POINT NUMBERS PAND
      ØM ØN (0,1)
180 A2    EQU    14      .   INITIALLY U-S MUST BE FURNIS
      HED AFTER
190 XM    EQU    15      .   THAT THEY ARE KEPT CUPRENT B
      Y RN4
200 .             . SUGGESTED INITIAL VALUES IF NONE ARE
      AVAILABLE
210 .             . U1=233362477003 AND U2=212312312323
      (ØCTAL)
220 RN4*  LA     A1,*0,B11 .   LOAD U1
230      LSSL   A1,12      .   MULTIPLY BY 2**12
240      JP     A1,$+3      .   TEST FOR ØVERFLOW AND RETAIN A1(MØDC
      2**35))
250      AND    A1,(Ø377777777777) .
260      LA     A1,A2      .
270      AA     A1,*0,B11   .   ADD U1
280      JP     A1,$+3      .   TEST FOR ØVERFLOW AND RETAIN A1(MØDC
      2**35))
290      AND    A1,(Ø377777777777) .
300      LA     A1,A2      .
310      A,XM   A1,1       .   ADD 1
320      JP     A1,$+3      .   TEST FOR ØVERFLOW AND RETAIN A1(MØDC
      2**35))
330      AND    A1,(Ø377777777777) .
340      LA     A1,A2      .
350      SA     A1,*0,B11   .   SAVE NEW U1
360      JP     A1,$+3      .   TEST FOR ØVERFLOW AND RETAIN A1(MØDC
      2**35))
370      AND    A1,(Ø377777777777) .
380      LA     A1,A2      .
390      AND    A1,(Ø7777777777) .   FLOAT U1
400      AA     A2,TWØ      .
410      FA     A2,TWØ      .
420      SA     A2,*2,B11   .   RETURN FU1
430      LA     A1,*1,B11   .   LOAD U2
440      LSSL   A1,7        .   MULTIPLY BY 2**7
450      JP     A1,$+3      .   TEST FOR ØVERFLOW AND RETAIN A1(MØDC
      2**35))

```


RN4
(Continued)

```

460      AND      A1,(03777777777777).
470      LA       A1,A2
480      AA       A1,*1,B11      .   ADD U2
490      JP       A1,$+3        .   TEST FOR OVERFLOW AND RETAIN A1(MODC
      2**35))
500      AND      A1,(03777777777777).
510      LA       A1,A2
520      AA       A1,C          .   ADD C
530      JP       A1,$+3        .   TEST FOR OVERFLOW AND RETAIN A1(MODC
      2**35))
540      AND      A1,(03777777777777).
550      LA       A1,A2
560      SA       A1,*1,B11      .   SAVE NEW U2 AND REPEAT AGAIN
570      LSSL     A1,7          .   MULTIPLY BY 2**7
580      JP       A1,$+3        .   TEST FOR OVERFLOW AND RETAIN A1(MODC
      2**35))
590      AND      A1,(03777777777777).
600      LA       A1,A2
610      AA       A1,*1,B11      .   ADD U2
620      JP       A1,$+3        .   TEST FOR OVERFLOW AND RETAIN A1(MODC
      2**35))
630      AND      A1,(03777777777777).
640      LA       A1,A2
650      AA       A1,C          .   ADD C
660      JP       A1,$+3        .   TEST FOR OVERFLOW AND RETAIN A1(MODC
      2**35))
670      AND      A1,(03777777777777).
680      LA       A1,A2
690      SA       A1,*1,B11      .   SAVE NEW U2
700      AND      A1,(077777777777) .   FLOAT U2
710      AA       A2,TW0
720      FA       A2,TW0
730      SA       A2,*3,B11      .   RETURN FU2
740      J        5,B11
750 TV0      +      02000000000000
760 C        +      0311564570652   .   =.788*2**35 (DECIMAL)
770      END
READY

```

LOGN

LOGN 17:21 07/31/72

```

100 @0 RUN EC0005,CUB438, ,120,500
110 @E FR5 LOGN,LOGN
120 C T0 SIMULATE SAMPLES FROM LOG-NORMAL
130 C AND COMPUTE SKEWNESS (G1) AND KURTOSIS (G2)
140     DIMENSION Y(1000),G1S(50),G2S(50)
150     REAL MU,M2,M3,M4,MU1
160 C N1=NUMBER OF RANDOM VARIABLES SIMULATED IN EACH SET
170     N1=1000
180 C N2=NUMBER OF SETS (OF N1 EACH)
190     N2=30
200     WRITE(6,60)
210 60     FORMAT(4X,'CV',6X,'AG1',5X,'EG1',5X,'T1',6X,'AG2',5X,'EG2',5X
220     &,'T2',6X,'VG1',5X,'VG2')
230 C D0 LOOP TO RUN THROUGH 6 VALUES OF COEF. OF VARIATION
240     D0 10 I=1,6
250         AG1=0.
260         AG2=0.
270         VG1=0.
280         VG2=0.
290         C=C+.1
300         CSQ=C*C
310         SSQ=ALOG(1.+CSQ)
320         S=SQRT(SSQ)
330         MU=S/C
340         WRITE(6,70) C
350 70     FORMAT(/5X,'C=',F10.5)
360 C D0 LOOP TO GENERATE N2 SETS (N1 RANDOM VARIABLES IN EACH)
370     D0 50         M=1,N2
380 C GENERATES N1 RANDOM VARIABLES FROM LOGNORMAL DISTRIBUTION
390     D0 20         J=1,N1
400         Y(J)=0.
410         CALL NORM(Z1,Z2)
420         Y(J)=EXP(MU+Z1*S)
430 20     CONTINUE
440         MU1=0.
450 C CALCULATES MEAN VALUE
460     D0 30         K=1,N1
470         MU1=MU1+Y(K)/N1
480 30     CONTINUE
490         M2=0.
500 C CALCULATES MOMENTS
510         M3=0.
520         M4=0.
530         S2=0.

```

LOGN
(Continued)

```

540      DO 40      L=1,N1
550      M2=M2+((Y(L)-MU1)**2)/N1
560      M3=M3+((Y(L)-MU1)**3)/N1
570      M4=M4+((Y(L)-MU1)**4)/N1
580      S2=S2+((Y(L)-MU1)**2)/(N1-1.)
590 40      CONTINUE
600 C SAMPLE COEF. OF VARIATION
610      CVSQ=S2/(MU1**2)
620      CV=SQRT(CVSQ)
630 C COEF. OF SKEWNESS
640      G1=M3/(M2*SQRT(M2))
650      AG1=AG1+G1/N2
660 C COEF. OF KURTOSIS
670      G2=(M4/(M2*M2))-3.
680      AG2=AG2+G2/N2
690      G1S(M)=G1
700      G2S(M)=G2
710 50      CONTINUE
720 C COMPUTES VARIANCES OF G1 AND G2
730      DO 90 II=1,N2
740      VG1=VG1+((G1S(II)-AG1)**2)/(N2-1)
750      VG2=VG2+((G2S(II)-AG2)**2)/(N2-1)
760 C EXPECTED VALUES AND NORMALIZED VALUES
770 90      CONTINUE
780      EG1=3.*C+C**3
790      EG2=16.*(C**2)+15.*(C**4)+6.*(C**6)+C**8
800      T1=(AG1-EG1)/SQRT(VG1)
810      T2=(AG2-EG2)/SQRT(VG2)
820      WRITE (6,80) CV,AG1,EG1, T1,AG2,EG2,T2,VG1,VG2
830 80      FORMAT(9F8.3)
840 10      CONTINUE
850      CALL RN4(U1,U2,X1,X2)
860      WRITE(6,100) U1,U2
870 100      FORMAT(/5X,'U1,U2',5X,2016)
880      END

```

SKEW

```

100 @0 RUN EC0001,CUB438, ,240,500
110 @E FR5 SKEW,SKEW
120 C T0 SIMULATE SAMPLES FROM GAMMA DISTRIBUTION (LAMBDA =1)
130 C AND COMPUTE SKEWNESS (G1) AND KURTOSIS (G2)
140 DIMENSION Y(1000),RS(6),G1S(50),G2S(50)
150 REAL MU,M2,M3,M4,MU1
160 C N1=NUMBER OF RANDOM VARIABLES SIMULATED IN EACH SET
170 N1=1000
180 C N2=NUMBER OF SETS (N1 IN EACH)
190 N2=30
200 C INITIAL VALUE FOR RANDOM NUMBER GENERATOR
210 DATA U /011060471625 /
220 C SET OF COEF. OF VARIATION (C=1/RS)
230 DATA RS /100.,25.,11.,6.,4.,3./
240 WRITE(6,90)
250 90 FORMAT(4X,'CV',6X,'AG1',5X,'EG1',5X,'T1',6X,'AG2',5X,'EG2',5X
260 &,'T2',6X,'VG1',5X,'VG2')
270 C D0 LOOP TO RUN THROUGH 6 VALUES OF CV
280 D0 10 I=1,6
290 AG1=0.
300 AG2=0.
310 VG1=0.
320 VG2=0.
330 R=RS(I)
340 C=1./SQRT(R)
350 WRITE(6,80) C
360 80 FORMAT(/5X,'C=',F10.5)
370 C D0 LOOP TO GENERATE N2 SETS (OF N1 EACH)
380 D0 50 M=1,N2
390 C GENERATES N1 RANDOM VARIABLES FROM GAMMA DISTRIBUTION
400 D0 20 J=1,N1
410 Y(J)=0.
420 N=IFIX(R)
430 Z1=0.
440 D0 60 IJ=1,N
450 CALL RN2(U,RND)
460 Z1=Z1-ALOG(1.-RND)
470 60 CONTINUE
480 Y(J)=Z1
490 20 CONTINUE
500 MU1=0.
510 C CALCULATES MEAN VALUE
520 D0 30 K=1,N1
530 MU1=MU1+Y(K)/N1
540 30 CONTINUE
550 M2=0.

```

SKEW
(Continued)

```

560      M3=0.
570      M4=0.
580      S2=0.
590 C COMPUTES MOMENTS
600      DO 40      L=1,N1
610      M2=M2+((Y(L)-MU1)**2)/N1
620      M3=M3+((Y(L)-MU1)**3)/N1
630      M4=M4+((Y(L)-MU1)**4)/N1
640      S2=S2+((Y(L)-MU1)**2)/(N1-1.)
650 40      CONTINUE
660 C SAMPLE COEF. OF VARIATION
670      CVSQ=S2/(MU1**2)
680      CV=SQRT(CVSQ)
690 C COEF. OF SKEWNESS
700      G1=M3/(M2*SQRT(M2))
710      AG1=AG1+G1/N2
720 C COEF. OF KURTOSIS
730      G2=(M4/(M2*M2))-3.
740      AG2=AG2+G2/N2
750      G1S(M)=G1
760      G2S(M)=G2
770 50      CONTINUE
780 C COMPUTES VARIANCES OF G1 AND G2
790      DO 100 II=1,N2
800      VG1=VG1+((G1S(II)-AG1)**2)/(N2-1)
810      VG2=VG2+((G2S(II)-AG2)**2)/(N2-1)
820 100      CONTINUE
830 C EXPECTED VALUES OF G1 AND G2
840      EG1=2.*C
850      EG2=6.*(C**2)
860      T1=(AG1-EG1)/SQRT(VG1)
870      T2=(AG1-EG1)/SQRT(VG2)
880      WRITE (6,70) CV,AG1,EG1,T1,AG2,EG2,T2,VG1,VG2
890 70      FORMAT(9F8.3)
900 10      CONTINUE
910      CALL RN2(U,RND)
920      WRITE(6,110) U
930 110      FORMAT(/5X,'U',5X,016)
940      END

```

LOGCV

```

100 @X RUN EC0004,CUB438, ,2,500 EBERHARDT
110 @E FR5 LOGCV,LOGCV
120 C T0 SIMULATE SAMPLES FROM A LOGNORMAL DISTRIBUTION
130     DIMENSION Y(100)
140     DIMENSION DS(5)
150     REAL MU,MU1,MU2,MS,MSS(2,100),ZZ(2,100)
160 C INCREMENTS OF CHANGE IN COEF. OF VAR.
170     DATA DS / 0.,.05,.2,.3,.4 /
180 C RUNS THROUGH VALUES OF C.V.
190     D0 170 JL=1,5
200     P11=0.
210     P22=0.
220 C PROGRAM OPERATES IN SETS OF 10 RUNS OF 100 EACH WITH OPTION
230 C (LINES 720,740) OF PRINTING OUT RESULTS FOR EACH OF THE 10 SETS
240     D0 80 JK=1,10
250     P1=0.
260     P2=0.
270 C SAMPLE SIZE FOR EACH TEST
280     N1=6
290 C F-VALUE FOR SAMPLE SIZE
300     F=5.0503
310     D=DS(JL)
320     C=.4
330 C GENERATES PAIRWISE COMPARISONS
340     D0 10 I=1,2
350     IF(I.GT.1)C=C+D
360     CSQ=C*C
370 C FIRST PARAMETER COMPUTED FROM CV
380     SSQ=ALOG(1.+CSQ)
390     S=SQRT(SSQ)
400 C ARBITRARY CHOICE OF SECOND PARAMETER (MEAN OF LOG-TRANS DATA)
410     MU=S/C
420 C RUNS 100 TRIALS OF N1 EACH
430     D0 50 M=1,100
440     D0 20 J=1,N1
450     Y(J)=0.
460     CALL NORM(Z1,Z2)
470 C SIMULATED LOGNORMAL RANDOM VARIABLE
480     Y(J)=EXP(MU+Z1*S)
490 20 CONTINUE
500     MU1=0.
510     MU2=0.
520 C MEANS
530     D0 30 K=1,N1
540     MU1=MU1+Y(K)/N1
550     MU2=MU2+ALOG(Y(K))/N1

```

LOGCV
(Continued)

```

560 30      CONTINUE
570          S2=0.
580          Z1=0.
590          MS=0.
600          D0 40 L=1,N1
610 C VARIANCE OF LOG-TRANSFORMED DATA
620          MS=MS+((ALOG(Y(L))-MU2)**2)/(N1-1)
630          Z1=Z1+ALOG(Y(L))/N1
640 C VARIANCE (UNTRANSFORMED)
650          S2=S2+((Y(L)-MU1)**2)/(N1-1)
660 40      CONTINUE
670 C Z-STATISTIC
680          Z=ALOG(MU 1)-Z1
690 C SAMPLE CV (NOT USED)
700          CVSQ=S2/(MU1**2)
710          CV=SQRT(CVSQ)
720 C STORES VALUES
730          MSS(I,M)=MS
740          ZZ(I,M)=Z
750 50      CONTINUE
760 10      CONTINUE
770 C COMPUTES F-RATIOS
780          D0 70 JJ=1,100
790          F1=MSS(2,JJ)/MSS(1,JJ)
800          F2=ZZ(2,JJ)/ZZ(1,JJ)
810 C TALLIES SIGNIFICANT F-RATIOS
820          IF(F.GT.F1)R1=1.
830          IF(F.GT.F2)R2=1.
840          P1=P1+R1
850 C RECORDS OVERALL PERCENTAGE OF NON-SIGNIFICANT F-RATIOS
860          P11=P11+R1/10.
870          R1=0.
880          P2=P2+R2
890          P22=P22+R2/10.
900          R2=0.
910 70      CONTINUE
920 C OPTIONAL PRINT-OUT OF EACH SET OF 10
930 C          WRITE(6,110) P1
940 110      FORMAT(/5X,'P1=',F10.5)
950 C          WRITE(6,120) P2
960 120      FORMAT(/5X,'P2=',F10.5)
970 80      CONTINUE
980          WRITE(6,130) P11
990          WRITE(6,140) P22
1000 130      FORMAT(/5X,'P11=',F10.5)
1010 140      FORMAT(/5X,'P22',F10.5)
1020 C ACTUAL (NOT SAMPLE) CV
1030          WRITE(6,100) C
1040 100      FORMAT(/5X,'COEF. OF VAR.',F10.5)
1050          WRITE(6,150) D
1060 150      FORMAT(/5X,'D=',F10.5)
1070 170      CONTINUE
1080          WRITE(6,180) N1
1090 180      FORMAT(/5X,'N1=',I5)
1100          END

```

GAMA

LIST

GAMA 17:13 08/29/72

```

100 @S RUN EC0003,CUB438, ,100,500 EBERHARDT
110 @E FR5 GAMACV,GAMACV
120 C T0 SIMULATE SAMPLES FROM GAMMA DISTRIBUTION(LAMBDA=1)
130 C CV GENERATED FROM 1/R,R INTEGRER
140 C SAME ARRANGEMENT AS LOGCV
150     DIMENSION Y(100)
160     DIMENSION DS(5)
170     REAL MU,MU1,MU2,MS,MSS(2,100),ZZ(2,100)
180 C INITIAL RANDOM NUMBER
190     DATA U / 011060471625 /
200     DATA DS / 0.,10.,14.,20.,23. /
210     D0 170 JL=1,5
220     P11=0.
230     P22=0.
240     D0 80 JK=1,10
250     P1=0.
260     P2=0.
270     N1=6
280     F=5.0503
290     D=DS(JL)
300     R=25.
310     CALL RN2(U,YY)
320     D0 10 I=1,2
330     IF(I.GT.1)R=R-D
340     C=1./SQRT(R)
350     MU=1./R
360     D0 50 M=1,100
370     D0 20 J=1,N1
380     Y(J)=0.
390 C GENERATES RANDOM VARIABLE FROM GAMMA DISTRIBUTION
400     N=IFIX(R)
410     Z1=0.
420     D0 60 IJ=1,N
430     CALL RN2(U,Z12)
440     Z1=Z1-ALOG(1.-Z12)
450 60    CONTINUE
460     Y(J)=Z1
470 20    CONTINUE
480     MU1=0.
490     MU2=0.
500     D0 30 K=1,N1
510     MU1=MU1+Y(K)/N1
520     MU2=MU2+ALOG(Y(K))/N1
530 30    CONTINUE
540     S2=0.
550     Z2=0.

```


GAMA

(Continued)

```

560      MS=0.
570      D0 40 L=1,N1
580      MS=MS+((ALOG(Y(L))-MU2)**2)/(N1-1)
590      Z2=Z2+ALOG(Y(L))/N1
600      S2=S2+((Y(L)-MU1)**2)/(N1-1)
610 40    CONTINUE
620      Z=ALOG(MU 1)-Z2
630      CVSQ=S2/(MU1**2)
640 C SAMPLE CV (NOT USED)
650      CV=SQRT(CVSQ)
660      MSS(I,M)=MS
670      ZZ(I,M)=Z
680 50    CONTINUE
690 10    CONTINUE
700      D0 70 JJ=1,100
710      F1=MSS(2,JJ)/MSS(1,JJ)
720      F2=ZZ(2,JJ)/ZZ(1,JJ)
730      IF(F.GT.F1)R1=1.
740      IF(F.GT.F2)R2=1.
750      P1=P1+R1
760      P11=P11+R1/10.
770      R1=0.
780      P2=P2+R2
790      P22=P22+R2/10.
800      R2=0.
810 70    CONTINUE
820 C      WRITE(6,110) P1
830 110    FORMAT(/5X,'P1=',F10.5)
840 C      WRITE(6,120) P2
850 120    FORMAT(/5X,'P2=',F10.5)
860 80    CONTINUE
870      WRITE(6,130) P11
880      WRITE(6,140) P22
890 130    FORMAT(/5X,'P11=',F10.5)
900 140    FORMAT(/5X,'P22=',F10.5)
910      WRITE(6,100) C
920 100    FORMAT(/5X,'COEF. OF VAR.',F10.5)
930      WRITE(6,150) D
940 150    FORMAT(/5X,'D=',F10.5)
950      WRITE(6,160) U
960 160    FORMAT(/5X,'U=',012)
970 170    CONTINUE
980      WRITE(6,180) N1
990 180    FORMAT(/5X,'N1=',I5)
1000      END

```

BART

```

100 @S RUN EC0001,CUB438, ,120,500
110 @E FR5 BART,BART
120 C BARTLETT'S TEST ON LOGTRANSFORM OF GAMMA DISTN
130 C COMPONENTS FOLLOW PROGRAM GAMACV
140     DIMENSION Y(6),RS(6)
150     REAL MU,MU1,MS,MU2,MSS(10)
160     NI=6
170 C NK IS NUMBER OF VARIANCES TO BE COMPARED
180     NK=10
190     DATA U /011060471625/
200 C CHI-SQUARE CRITERION
210     CHI=16.919
220     DATA RS / 100.,25.,11.,6.,4.,3. /
230 C TO PRODUCE 4 RUNS AT ONCE
240     DO 120 KJ=1,4
250     WRITE(6,80)
260 80    FORMAT(/10X,' C      NO. OF SIGNIFICANT CHI-SQUARES')
270     DO 10      I=1,6
280     NCS=0
290 C DETERMINES NUMBER OF TRIALS PER CV VALUE
300     DO 100 II=1,100
310     R=RS(I)
320     C=1./SQRT(R)
330     DO 50      M=1,NK
340     DO 20      J=1,NI
350     Y(J)=0.
360     N=IFIX(R)
370     Z1=0.
380     DO 60      IJ=1,N
390     CALL RN2(U,RND)
400     Z1=Z1-ALOG(1.-RND)
410 60    CONTINUE
420     Y(J)=Z1
430 20    CONTINUE
440     MU1=0.
450     MU2=0.
460     DO 30      K=1,NI
470     MU1=MU1+Y(K)/NI
480     MU2=MU2+ALOG(Y(K))/NI
490 30    CONTINUE
500     S2=0.
510     MS=0.
520     DO 40      L=1,NI
530     MS=MS+((ALOG(Y(L))-MU2)**2)/(NI-1)
540     S2=S2+((Y(L)-MU1)**2)/(NI-1.)
550 40    CONTINUE

```

BART

(Continued)

```

560 C SAMPLE CV (NOT USED)
570     CVSQ=S2/(MUI**2)
580     CV=SQRT(CVSQ)
590     MSS(M)=MS
600 50    CONTINUE
610 C BARTLETT'S TEST
620     S=0.
630     ST=0.
640     NT=NK*(N1-1)
650     D0 70      JJ=1,NK
660     S=S+(N1-1)*MSS(JJ)
670     ST=ST+(N1-1)*ALOG(MSS(JJ))
680 70    CONTINUE
690     B=NT*ALOG(S/NT)-ST
700     C1=1.+((NK/(N1-1))-(1./NT))/(3.*(NK-1))
710     T=B/C1
720     NC=0
730     IF(T.GT.CHI) NC=1
740     NCS=NCS + NC
750 100    CONTINUE
760     WRITE(6,90) C,NCS
770 90    FORMAT(/5X,F10.5,I5)
780 10    CONTINUE
790 120    CONTINUE
810     WRITE (6,110) U
820 110    FORMAT(/5X,'U',5X,016)
825     END

```

SCHEF

```

100 @0 RUN EC0003,CUB438, ,300,500
110 @E FR5 SCHEF,SCHEF
120 C SIMULATES SCHEFFE TEST FOR EQUALITY OF VARIANCES
130 C BASED ON PROGRAM GAMMA CV
140 C I=5,J=2;F HAS 4,5 D.F.
150     DIMENSION Y(6)
160     REAL MU,MU1,MS,MU2,MSS(10),RS(6)
170     N1=6
180 C NK IS NO. OF VARIANCES COMPUTED
190 C (MUST BE AN EVEN NUMBER)
200     NK=10
210 C FC IS F-CRITERION (MUST AGREE WITH DF)
220     FC=5.19
230     DATA U/ 011060471625 /
240     DATA RS / 100.,25.,11.,6.,4.,3. /
250 C TO PRODUCE 4 RUNS
260     DO 140 KJ=1,4
270     WRITE(6,110)
280 110    FORMAT(/10X,'      C      NO. OF SIGNIFICANT VALUES OF F')
290     DO 10 I=1,6
300     NCF=0
310     DO 130 II=1,100
320     R=RS(I)
330     C=1./SQRT(R)
340     MU=1./R
350     DO 50 M=1,NK
360     DO 20      J=1,N1
370     Y(J)=0.
380     N=IFIX(R)
390     Z1=0.
400     DO 60      IJ=1,N
410     CALL RN2(U,RND)
420     Z1=Z1-ALOG(1.-RND)
430 60    CONTINUE
440     Y(J)=Z1
450 20    CONTINUE
460     MU1=0.
470     MU2=0.
480     DO 30      K=1,N1
490     MU1=MU1+Y(K)/N1
500     MU2=MU2+ALOG(Y(K))/N1
510 30    CONTINUE
520     M2=0.
530     M3=0.
540     M4=0.
550     S2=0.

```

SCHEF

(Continued)

```

560      MS=0.
570      D0 40          L=1,N1
580      MS=MS+((ALOG(Y(L))-MU2)**2)/(N1-1)
590      S2=S2+((Y(L)-MU1)**2)/(N1-1.)
600 40      CONTINUE
610      CVSQ=S2/(MU1**2)
620      CV=SQRT(CVSQ)
630      MSS(M)=MS
640 50      CONTINUE
650 C SCHEFFE'S TEST
660      YB=0.
670      YN=0.
680      DY=0.
690      D0 70          JJ=1,NK
700      YB=YB+ALOG(MSS(JJ))/NK
710 70      CONTINUE
720      NKK=NK-1
730      D0 80          JK=1,NKK,2
740      YD=0.
750      MN=JK+1
760      D0 90          JL=JK,MN
770      YD=YD+ALOG(MSS(JL))/2.
780 90      CONTINUE
790      YN=YN+(YD-YB)**2
800      D0 100         JM=JK,MN
810      DY=DY+(ALOG(MSS(JM))-YD)**2
820 100     CONTINUE
830 80      CONTINUE
840      T=(NK*YN)/(((FLOAT(NK)/2.)-1.)*DY)
850      NC=0
860      IF(T.GT.FC) NC=1
870      NCF=NCF+NC
880 130     CONTINUE
890      WRITE(6,120) C,NCF
900 120     FORMAT(/5X,F10.5,I5)
910 10      CONTINUE
920      WRITE (6,150) U
930 150     FORMAT(/5X,'U',5X,016)
940 140     CONTINUE
950      END

```

SCH

```

100 @S  RUN EC0003,CUB438, ,120,500
110 @E  FR5 SCHEF,SCHEF
120 C  SIMULATES SCHEFFE TEST FOR EQUALITY OF VARIANCES
130 C  BASED ON PROGRAM SCHEF
140 C  I=5,J=2;F HAS 4,5 DF FOR NK=10 (DOESNT DEPEND ON N1)
150      DIMENSION Y(21)
160      REAL MU1,MS,MU2,MSS(10),RS(6)
170      N1=21
180 C  NK IS NO. OF VARIANCES COMPUTED
190 C  (MUST BE AN EVEN NUMBER)
200      NK=10
210 C  FC IS F-CRITERION (MUST AGREE WITH DF)
220      FC=5.19
230 C  NUMBER OF REPLICATIONS
240      NT=100
250      DATA U/ 011060471625 /
260      DATA RS/25.,6.,5.,4.,3.,2./
270      WRITE(6,110)
280 110  FORMAT(/10X,'  C      NO. SIGNIF.F  AVE VAR LOG SSQ')
290      DO 10 I=1,6
300      NCF=0
310      DO 130 II=1,NT
320      DO 50 M=1,NK
330      R=25.
340 C  SELECTS INTERMEDIATE CV FOR MIDDLE SET OF VARIANCES
350      IF(M.GT.2) R=EXP(.5*(LOG(R)+LOG(RS(I))))
360 C  MAKES LAST PAIR OF VARIANCES CORRES TO CV FROM RS(I)
370      N2=NK-2
380      IF(M.GT.N2) R=RS(I)
390      DO 20      J=1,N1
400      Y(J)=0.
410      N=IFIX(R)
420      Z1=0.
430      DO 60      IJ=1,N
440      CALL RN2(U,RND)
450      Z1=Z1-ALOG(1.-RND)
460 60  CONTINUE
470      Y(J)=Z1
480 20  CONTINUE
490      MU1=0.
500      MU2=0.
510      DO 30      K=1,N1
520      MU1=MU1+Y(K)/N1
530      MU2=MU2+ALOG(Y(K))/N1
540 30  CONTINUE
550      S2=0.

```

SCH

(Continued)

```

560      MS=0.
570      D0 40      L=1,N1
580      MS=MS+((ALOG(Y(L))-MU2)**2)/(N1-1)
590      S2=S2+((Y(L)-MU1)**2)/(N1-1.)
600 40      CONTINUE
610 C SAMPLE CV (NOT USED)
620      CVSQ=S2/(MU1**2)
630      CV=SQRT(CVSQ)
640      MSS(M)=MS
650 50      CONTINUE
660 C SCHEFFES TEST
670      YB=0.
680      YN=0.
690      DY=0.
700      D0 70      JJ=1,NK
710      YB=YB+ALOG(MSS(JJ))/NK
720 70      CONTINUE
730      NKK=NK-1
740      D0 80      JK=1,NKK,2
750      YD=0.
760      MN=JK+1
770      D0 90      JL=JK,MN
780      YD=YD+ALOG(MSS(JL))/2.
790 90      CONTINUE
800      YN=YN+(YD-YB)**2
810      D0 100     JM=JK,MN
820      DY=DY+(ALOG(MSS(JM))-YD)**2
830 100     CONTINUE
840 80      CONTINUE
850      T=(NK*YN)/(((FLOAT(NK)/2.)-1.)*DY)
860      N3=(NK*NT)/2
870      SDY=SDY+DY/N3
880      NC=0
890      IF(T.GT.FC) NC=1
900      NCF=NCF+NC
910 130     CONTINUE
920      C=1./SQRT(RS(I))
930      WRITE(6,120) C,NCF,SDY
940 120     FORMAT(/5X,F10.5,5X,I5,5X,F15.5)
950      SDY=0.
960 10      CONTINUE
970      WRITE(6,150) U
980 150     FORMAT(/5X,'U',5X,016)
990      END

```

PCALC

LOAD PCALC
READY
LIST

PCALC 16:46 09/06/72

```
100 REM VALUES OF AVERAGE R
110 R=25
120 PRINT"    MEAN          R1          CALC MEAN    R3"
130 FOR I=1 TO 24
140 R1=R-I+1
150 C=1/SQR(R1)
160 M=LOG(.2)+LOG(C)
170 R2=EXP(-M)
180 R3=INT(R2)
190 M2=-LOG(R3)
200 PRINT M,R1,M2,R3
210 PRINT" "
220 NEXT I
230 END
READY
```


TRANS

TRANS 16:45 09/06/72

```
100 REM T0 TRANSFORM POWER TABLES (ORNL-4468)
110 DIM C(4,10),H(4,10)
120 N=9
130 PRINT "WHAT IS K?"
140 INPUT K
150 PRINT "WHAT IS COEFFICIENT OF VARIATION"
160 INPUT C1
170 FOR S=1 TO 4 STEP 1
180 PRINT "LIST VALUES FOR TAU"
190 FOR R=1 TO N STEP 1
200 INPUT B
210 C(S,R)=B
220 NEXT R
230 FOR I=1 TO N STEP 1
240 H1=LOG(1+C1↑2)
250 H(S,I)=EXP(C(S,I)*SQR(H1))
260 NEXT I
270 NEXT S
280 FOR T=1 TO N STEP 1
290 PRINT T+1,H(1,T),H(2,T),H(3,T),H(4,T)
300 NEXT T
310 END
READY
```

DISTRIBUTION

No. of
Copies

OFFSITE

1	<u>AEC Chicago Patent Group</u> G. H. Lee
3	<u>AEC/Headquarters Division of Biomedical and Environmental Research</u> Washington, D.C. H. H. Hollister J. Swinebroad B. Waters
1	<u>AEC-Nevada Operations Office</u> Las Vegas, NV P. B. Dunaway
3	<u>AEC Technical Information Center</u>
1	<u>Battelle Memorial Institute</u> <u>Battelle-Columbus Laboratory</u> Columbus, OH W. E. Martin
1	<u>Washington State University</u> Pullman, WA V. Schultz

ONSITE-
HANFORD

1	<u>AEC/RL Patent Attorney</u> R. M. Poteat
2	<u>AEC Richland Operations Office</u> W. E. Lotz B. J. Melton

No. of
Copies

68

Battelle-Northwest

J. C. Fox
R. L. Hooper
J. W. Johnston
T. M. Beetle
P. J. Dionne
B. E. Vaughan
W. L. Templeton
L. L. Eberhardt (10)
J. M. Thomas
R. F. Foster
C. L. Simpson
W. J. Bair
A. M. Platt
J. M. Nielsen
D. E. Olesen
R. O. Gilbert (40)
Technical Information (3)
Technical Publications