MEMORANDUM
RM-5072-PR
SEPTEMBER 1966

AD639163

POWER OF THE F-TEST FOR NONNORMAL DISTRIBUTIONS AND UNEQUAL ERROR VARIANCES

T. S. Donaldson

PREPARED FOR:
UNITED STATES AIR FORCE PROJECT RAND

The RAND Corporation
SANTA MONICA • CALIFORNIA

MEMORANDUM
RM-5072-PR
SEPTEMBER 1966

POWER OF THE F-TEST FOR NONNORMAL DISTRIBUTIONS AND UNEQUAL ERROR VARIANCES

T. S. Donaldson

This research is sponsored by the United States Air Force under Project RAND—Contract No. AF 49(638)-1700—monitored by the Directorate of Operational Requirements and Development Plans, Deputy Chief of Staff, Research and Development, Hq USAF. Views or conclusions contained in this Memorandum should not be interpreted as representing the official opinion or policy of the United States Air Force.

The RAND Corporation
1700 MAIN ST • SANTA MONICA • CALIFORNIA • 90406

PREFACE

An intensive study of base maintenance management has been in progress at The RAND Corporation for several years under the designation Laboratory Problem IV (LP-IV). One portion of this project has concerned the statistical analysis of maintenance management data and the resolution of methodological problems associated with them. This Memorandum attempts to resolve some of the problems arising when maintenance data have extremely nonnormal distributions. While the specific motivation for this effort grew out of problems associated with the statistical analysis of maintenance data, the study is far more general, since it applies to the use of analysis of variance techniques when the assumptions necessary for the analysis are violated.

This Memorandum will interest specialists in statistical data analysis, specifically those concerned with analysis of variance techniques. At present, this audience may be small, but as more refined and computerized data systems become available, a greater effort in the direction of statistical analysis will be made, and many methodological problems can be expected to arise. As a result, this Memorandum will undoubtedly be only one of many aimed at resolving applied problems in statistical data analysis.

SUMMARY

The purpose of this study was to investigate empirically the power of the F-test when the underlying distribution is nonnormal and the within cell treatment variances are not homogeneous. The nonnormal distributions used in this study were the exponential and the lognormal. These distributions were used because they allow one to determine the effects of extreme nonnormality on F-test power. The study was limited to the single classification analysis of variance in which the F-test is used to test for differences among the means of k cells.

A computer was programmed to draw n numbers randomly for each of k cells. The type of distribution and its mean and variance were specified for each cell, then the F-test was calculated and the process repeated 10,000 times. The computer listed the empirical distribution of F, MST (mean square between cells) and MSE (mean square within cells), and the correlation between MST and MSE. The value of F defined by the lower bound of the critical region (of size α) was determined for appropriate values of n and k (denoted as F_α). The proportion of F's larger than F_α approximate the probability of rejecting H_0 (the null hypothesis of no difference between means) and accepting H_1 (the alternate hypothesis). When H_0 is true, this proportion approximates α; when H_1 is true, it approximates the power of the test for a specific value of ϕ, where

$$\phi = \sqrt{\frac{n\sum_{j}^{k}(\mu_j-\mu)^2}{k\sigma_\epsilon^2}} .$$

The power of F (including Type I error when $\phi = 0$) was determined for: (1) nonnormal distributions and homogeneous error variances, (2) nonnormal distributions and heterogeneous error variances, and (3) normal distributions and heterogeneous error variances. In the case of heterogeneous error variances, the within cell error variance was equal to the mean squared and $\sigma_\epsilon^2 = (1/k)\sum_j^k \sigma_j^2$. In all cases, the means were equally spaced. Under H_0, the distributions had a mean of 10 and a variance of 100. This led to skewness (γ_1) and kurtosis (γ_2) parameters equal to 2 and 6, respectively, for the exponential distribution, and 4 and 38 for the lognormal. Selected combinations of n and k were used with k = 2, 4 and 8, and n = 4, 8, 16 and 32.

The results indicate that the F-test is conservative with respect to Type I error for the nonnormal distributions that were considered (i.e., the probability of falsely rejecting H_0 is smaller than α). In the case of equal within cell variances, the power curves based on the nonnormal distributions show greater power than in the normal case over a large range of ϕ (corresponding to a power range of about 10 to 90 percent). For heterogeneous within cell variances, the differences in power are small until ϕ becomes large (corresponding to a power of about 40 percent), at which point the nonnormal distributions show the greater power. In general, the F-test is conservative with respect to both Type I error and power when based on nonnormal distributions of the type investigated. Of the two nonnormal distributions, the lognormal is the more conservative. As either n or k increase, the power curves seem to converge to that based on the normal distribution.

An attempt is made to understand these results in terms of the distributions of the numerator (MST) and the denominator (MSE) of F, and in terms of the correlation between them. The correlation between MST and MSE was derived under H_0, and was found to depend on n, k and γ_2. By using an approximation for the variance of F it was shown that the correlation (or γ_2) reduces the variance of F based on nonnormal distributions to approximately the value it has when the underlying distribution is normal. The implications these results have for experimental design, data analysis, and future studies are discussed.

ACKNOWLEDGMENTS

The author is indebted to a number of RAND people for their many helpful comments. Particularly helpful were George S. Fishman, Allen J. Gross and Ernest M. Scheuer. The author would like to thank Miss Paula M. Oldfather for writing the computer programs used in this study, and Mrs. Nancy Hausner for her aid in tabulating the results.

CONTENTS

BLANK PAGE

1. INTRODUCTION

The F-test in the analysis of variance provides a powerful test for significance of differences among means* under certain restricting assumptions. Two of these assumptions are important for the present study; they are: (1) samples are drawn from normally distributed populations; and (2) the populations have equal variance. The effect of using the F-test when these assumptions are violated has been a topic of interest and research for some time. A review and summary of much of the earlier work is presented in Scheffé (1959, Chap. 10) and will not be discussed here. The general conclusion, however, is that F-test analysis in the case of nonnormality has little effect on inferences about means. The same is true concerning inequality of error variances when the samples are of equal size. The insensitivity of a statistical procedure to the underlying assumptions is referred to as robustness. Most studies of F-test robustness have been concerned only with the effects of departures from assumptions on the Type I error.** The Type II error*** or power (1 - Type II error) has received relatively little attention. There are a few notable exceptions.

---

*In the case of only two means, the t and F tests yield equivalent results, and numerically, $t^2 = F$.

**The Type I error is the error of rejecting the null hypothesis of no difference between means, when it is "actually" true. When the probability of the Type I error is α, the test is said to be of size α.

***The Type II error is the error of not rejecting the null hypothesis (or of not accepting the alternative) when it is false, i.e., true differences exist between means and one concludes that no differences exist.

A recent study by Srivastava (1958) investigated the power of the t-test under deviations from normality. Both skewness (γ_1) and kurtosis (γ_2) of the parent distributions were allowed; however, because of the distributions used (Edgeworth population) the range of γ_1 and γ_2 was limited to a rather narrow range. His results indicate that γ_1 has a larger effect on power than γ_2, but in general, the effect of either one is small. Horsnell (1953)* analytically determined the power of the F-test for four groups under the condition of unequal variance, unequal and equal sample size, and normal distributions. The results indicate that when the samples are of equal size (n=10) the effect on power is small.

In general, the F-test appears to be robust for both Type I error and power. However, there are some serious limitations in these few studies of power. The degree of nonnormality is seriously limited, and in Horsnell's study the distributions were normal. The effect of sample size and the number of groups has not been adequately determined. The combined effect on power of nonnormality and unequal variance is unknown. This latter point is important. The sample mean and variance for any nonsymmetric distribution are correlated, and unequal variances within means are expected when the null hypothesis is false.

The purpose of the present study is to determine the effects of extreme departures from underlying assumptions (two listed above) on the resultant power of the F-test. Extreme departure from the assumption of normality is indicated by high values of γ_1 and γ_2. Extreme departure from the assumption of equal within cell variance is indicated

---

*Reviewed in Scheffé, 1959, pp. 356-358.

by a high ratio of largest to smallest within cell variance. The two nonnormal distributions used in this study are the exponential and lognormal. These distributions were used because they fit the requirements discussed above--not to generate power curves specific to these distributions. The problem is to determine the degree of error that exists if the F-test is used to test for differences among sample means when the assumptions of normality and homogeneity of error are violated to the degree specified by these theoretical distributions.

F-test power curves are generated in this study by actually sampling from the specified distributions (normal, exponential, lognormal). Comparisons in F-test power for the three distributions are made for the case of both equal and unequal variance. This investigation will deal only with the single classification (or one-way layout) analysis of variance with equal observations in each cell,in which the F-test is used to test for significant differences in the treatment groups.

2. THE SINGLE CLASSIFICATION ANALYSIS OF VARIANCE*

The layout for the single classification analysis of variance is shown in Fig. 1. There are k treatment cells, n observations per cell,

	Treatment (cell)				
	1	2	3	j	k
Observation	x_{11}	x_{12}	x_{13}	x_{1j}	x_{1k}
	x_{21}	x_{22}	x_{23}	x_{2j}	x_{2k}
	⋮	⋮	⋮	⋮	⋮
	x_{i1}	x_{i2}	x_{i3}	x_{ij}	x_{ik}
	⋮	⋮	⋮	⋮	⋮
	x_{n1}	x_{n2}	x_{n3}	x_{nj}	x_{nk}
Mean	$\bar{x}_1$	$\bar{x}_2$	$\bar{x}_3$	$\bar{x}_j$	$\bar{x}_k$
Variance	s_1^2	s_2^2	s_3^2	s_j^2	s_k^2

Fig. 1 -- The Single Classification Layout

and the i^{th} observation in the j^{th} cell is denoted as $x_{ij}(1 \le i \le n, 1 \le j \le k)$. The sample mean and sample variance for the j^{th} cell are

(2.1) $$\bar{x}_j = \frac{1}{n}\sum_{i}^{n} x_{ij},$$

---

*For a detailed discussion, see Winer (1962, Chap. 3).

(2.2) $$s_j^2 = \frac{1}{n-1}\sum_i^n (x_{ij}-\bar{x}_j)^2.$$

Under the assumptions of the F-test, the x_{ij} are observations on a normally distributed random variable with mean μ_j and variance σ_j^2. Moreover, $\sigma_j^2 = \sigma_{j'}^2$, $(j \neq j')$ which states that the population distributions in each cell have the same variance (σ_ϵ^2). Under these conditions, the F-test is appropriate* in testing the null hypothesis, $H_0:\mu_1 = \dots = \mu_k$. The details of the F-test are discussed next.

Define MST as the mean square between treatments (T), and MSE as the mean square error (the unbiased estimate of σ_ϵ^2). The F-test for the null hypothesis (H_0) of no difference among the population means is

(2.3) $$F = \frac{MST}{MSE},$$

and F has k-1 and k(n-1) degrees of freedom. That is, under H_0, MST has a chi-square distribution with k-1 degrees of freedom, MSE has a chi-square distribution with k(n-1) degrees of freedom, and the two chi-squares are independent. These mean squares are computed from the between and within cell variations in x_{ij}, and are

(2.4) $$MST = \frac{n}{k-1}\sum_j^k (\bar{x}_j-\bar{x})^2,$$

and

$$MSE = \frac{1}{k(n-1)}\sum_i^n\sum_j^k (x_{ij}-\bar{x}_j)^2 = \frac{1}{k}\sum_j^k s_j^2,$$

*It is assumed here that all other assumptions for the F-test are also met.

where ($\bar{x}$) (without any subscript) denotes the grand mean

$$\bar{x} = \frac{1}{nk}\sum_{i}^{n}\sum_{j}^{k}x_{ij}.$$

Under H_0, the statistic F has a distribution determined by k-1 and k(n-1). Let $F_\alpha (0 < \alpha < 1)$ define a point on the abscissa of the distribution such that the proportion $1-\alpha$ of the cumulative distribution of F lies to the left of F_α, and the proportion α lies to the right and defines the values of F for which H_0 is rejected. The probability that F will lie in this critical region (reject H_0) when H_0 is true is α, and defines the size of the Type I error (the probability that H_0 is falsely rejected). When H_0 is true, 100α percent of the resultant F-tests should be significant (leading to rejection of H_0) by chance. When H_0 is not true, a greater percentage of F's are significant (in the critical region) depending upon the magnitude of the true difference between the means (under the alternate hypothesis H_1: $\mu_1 \neq \mu_2 \neq \ldots \neq \mu_k$.)

The power of the F-test is the probability of declaring F significant when H_1 is true. However, when H_0 is not true, the ratio MST/MSE is not distributed according to the ordinary F distribution but according to the distribution of the random variable F'. The distribution of F' (generally referred to as the noncentral F) depends upon the parameter ϕ, a function of the difference between means,* where

*The noncentrality parameter is generally denoted δ, where

$$\phi = \frac{\delta}{\sqrt{k}} .$$

(2.6)
$$\phi = \sqrt{\frac{n\sum_{j}^{k}(\mu_j - \mu)^2}{k\sigma_\epsilon^2}}\ ,$$

where

$$\mu = \frac{1}{k}\sum_{j}^{k}\mu_j\ .$$

In this expression, n is the number of observations in each cell and σ_ϵ^2 is the within cell error variance estimated by MSE. When the within cell variances are assumed equal, σ_ϵ^2 is identical in each cell. When they are assumed unequal,

$$\sigma_\epsilon^2 = \frac{1}{k}\sum_{j}^{k}\sigma_j^2\ .$$

The power of the F-test depends on the parameter ϕ; that is, the greater the difference between means, the greater the chance of declaring F significant, thus correctly rejecting H_0 and accepting H_1. (H_1 is accepted whenever F falls in the critical region.) The probability of falsely rejecting H_1 (Type II error) is the complement of power. Therefore, the size of the Type II error (β) is $\beta = 1 -$ Power.* Power curves for the F statistic as a function of ϕ and α have been calculated by Pearson and Hartley (1951) and Fox (1956) for various values of k-1 and k(n-1).

---

*Specifying β is identical to specifying power, and both terms are used throughout this Memorandum.

When the underlying assumptions of the F-test are met, power may be read off from power curves.* If the underlying distributions are nonnormal, or if the within cell variances are not equal, the F-test (central or noncentral) is invalid. Also, when the underlying within cell populations are normally distributed, the expected value of the correlation between MST and MSE is zero, and in fact, MST and MSE are statistically independent. The correlation is not zero for many non-normal distributions.**

---

*For example, see Fox (1956) or Pearson and Hartley (1951).

**A correlation between the sample mean and sample variance exists for any nonsymmetric distribution ($\gamma_1 \neq 0$). As will be proven in this investigation, a correlation between MST and MSE exists whenever $\gamma_2 \neq 0$; it does not depend on γ_1. Thus MST and MSE based on some nonnormal distributions (whenever $\gamma_2 = 0$) are not correlated.

3. THE SAMPLING PROCEDURE

A computer was programmed to sample n numbers from the distribution specified in each cell of the single classification layout. The type of parent distribution and its mean and variance for each of the k cells was specified. An F-test was computed for the k random samples, and this operation was replicated 10,000 times. The computer listed the distributions of F, MST, and MSE. Thus, when the null hypothesis was true (when the means of the parent distribution were equal and normality and homogeneity held), approximately 10,000α of the obtained F's exceeded F_α, where F_α was the critical value of F tabulated* for specific values of n and k. The tabulated F had degrees of freedom given by k-1, k(n-1). For example, let k = 4, n = 16. Then, by referring to tables of the F distribution, one found that for 3 and 60 degrees of freedom, $F_{.05} = 2.76$. Therefore, approximately 500 (10,000 x .05) of the obtained empirical F's should be larger than 2.76. Conversely, by counting the number of F's that actually exceeded 2.76 (or any value of F_α), the empirical Type I error (within the limits of sampling error) for any sampled distribution was obtained. The similarity between this empirical probability of Type I error and the true α level indicated the robustness of F with respect to probability of Type I error.

The same procedures were used to compute the power (1-β). In this case, the parent distributions have unequal means (and/or variances), and ϕ had some value depending upon the true population differences

---

*See tables of the F distribution.

between means. The critical region was determined from the value of F_α under the null hypothesis, and the power of the test was approximated by counting the number of F's in the critical region.

A computer program generated the random numbers. The numbers drawn were determined by an input control number. Several sets of random numbers (for several control numbers) were found that gave small sampling error. For all practical purposes, the numbers were drawn from continuous distributions, and errors due to discreteness are negligible.

Cumulative distributions of F were obtained for the 10,000 replications in intervals of 0.1 from 0 to 100. A linear interpolation was used to read F values to two decimal places. MST and MSE were accumulated in intervals of 1 from 0 to 250, and in intervals of 10 from 250 to 2750. The critical values of F_α were read off from the F interval scale of the cumulative distribution for $\alpha = .01, .05$, and .10.

The correlation between MST and MSE for the nonnormal distributions was calculated and printed out. Also printed out were the sample variance of MST and the sample variance of MSE (based on the sample size of 10,000).

Under the null hypothesis, all distributions (nonnormal and normal) had a mean of 10 and a variance of 100. These distributions are shown in Fig. 2. Each distribution and the method for sampling from it will be discussed.

THE NORMAL DISTRIBUTION

Normally distributed random variables with mean μ and variance σ^2 were generated by the method of Box and Muller (1958).

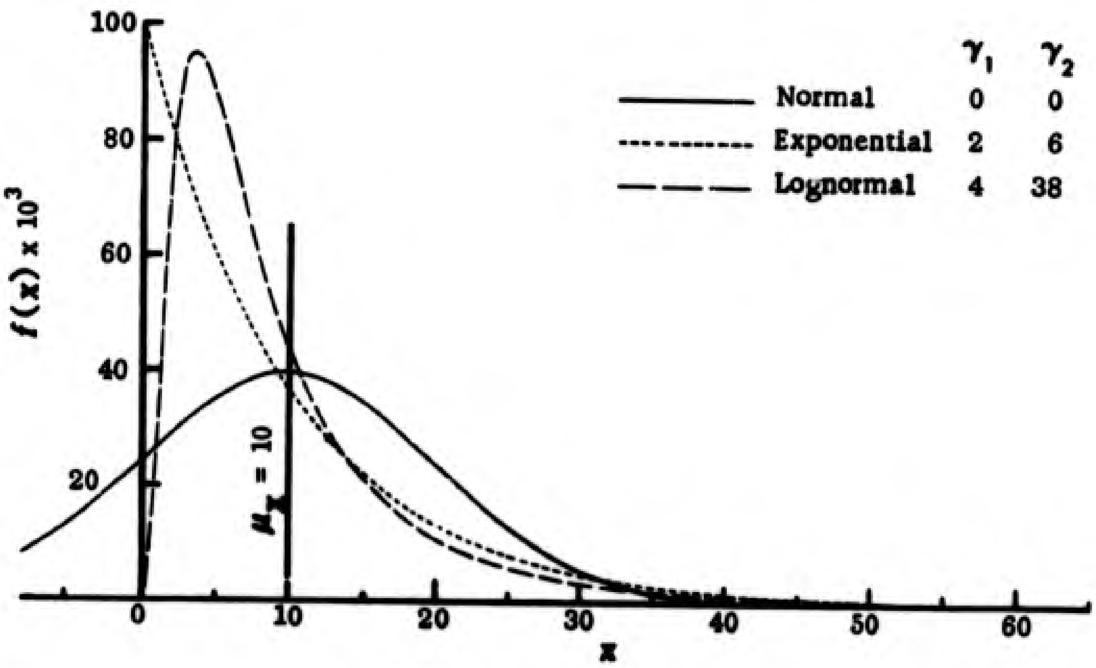


Fig. 2 -- Theoretical Distributions

THE EXPONENTIAL DISTRIBUTION

The probability density function for an exponentially distributed random variable x is

(3.1) $$f(x) = \lambda e^{-\lambda x}, \quad x > 0$$
$$= 0, \quad x \leq 0.$$

Taking expected values (or the moment generating function)* it is straightforward to obtain the mean and the higher moments about the mean. From the first four moments it follows that:

*See Cramér (1951).

(i) Mean $\mu = 1/\lambda$,

(ii) Variance $\sigma^2 = \mu^2$,

(iii) Skewness $\gamma_1 = \frac{2\mu^3}{\sigma^3} = 2$,

(iv) Kurtosis $\gamma_2 = \frac{9\mu^4}{\sigma^4} - 3 = 6$.

It is seen that the coefficients of skewness and kurtosis are constant for all exponential distributions and that the mean is equal to the standard deviation (σ).

In order to generate exponentially distributed random numbers, the computer was programmed to sample from the rectangular distribution of the random variable r in the 0 to 1 interval. Let

$$x = -\mu \log(r), \tag{3.2}$$

where log is the natural logarithm; then x is exponentially distributed with mean μ.

THE LOGNORMAL DISTRIBUTION

Let y be $N(\mu,\sigma^2)^*$ and set $x = e^y$, then x has the (two parameter) lognormal distribution

$$f(x) = \frac{1}{\sigma x\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{\log x-\mu}{\sigma}\right)^2}, \quad x > 0 \tag{3.3}$$

$$= 0, \qquad x \le 0.$$

---

*This notation indicates that x is normally distributed with mean μ and variance σ^2.

In this study it is necessary to find values of μ and σ^2 for specific values of the mean and variance of x (not y), since it is the random variable x that is sampled for the k cells. Denote the mean and variance of x by M and V^2 respectively. Aitchison and Brown (1957, p. 8) show that

$$M = e^{\mu+(\frac{1}{2})\sigma^2}, \tag{3.4}$$

$$V^2 = M^2(e^{\sigma^2} - 1). \tag{3.5}$$

Solving Eqs. (3.4) and (3.5) for μ and σ^2 yields

$$\mu = \log\left(\frac{M^2}{\sqrt{M^2 + V^2}}\right), \tag{3.6}$$

$$\sigma^2 = \log\left(1 + \frac{V^2}{M^2}\right). \tag{3.7}$$

For example, in this study, under H_0, M = 10 and $V^2 = 100$. Thus, from Eqs. (3.6) and (3.7), $\mu = 1.95601$ and $\sigma^2 = 0.69351$.

The skewness and kurtosis are (Aitchison and Brown, p. 8),

$$\gamma_1 = z^3 + 3z, \tag{3.8}$$

$$\gamma_2 = z^8 + 6z^6 + 15z^4 + 16z^2, \tag{3.9}$$

where $z^2 = e^{\sigma^2} - 1$. However, when $V^2 = M^2$, Eq. (3.7) reduces to

$$\sigma^2 = \log(2),$$

so that

$$z^2 = e^{\log(2)} - 1 = 1.$$

Thus, under the restriction that $V^2 = M^2$ (used in this study)

$$\gamma_1 = 4,$$

$$\gamma_2 = 38.$$

While γ_1 and γ_2 are not constant for all lognormal distributions, they are constant for all $V^2 = M^2$.

To generate lognormally distributed numbers, the computer was programmed to sample the variable y which has $N(\mu,\sigma^2)$ distribution. It then sets $x = e^y$, and the x's are entered into the cells of the single classification layout.

PARAMETER VALUES USED IN THIS STUDY

The values of n and k used in the study are shown in Table 1. In the case of equal within cell variance ($\sigma_\epsilon^2 = 100$), power was investigated for samples of size n = 4, 8, 16 and 32, and for k = 2 and 4. The case of eight treatment cells was investigated for a sample of size 4. In the case of unequal variances ($\sigma_j^2 = \mu_j^2$), three combinations of n and k were used; these were n = 16, k = 2; n = 4, k = 4; and n = 16, k = 4.

In one case the assumption of homogeneity of within cell error variances was true. Under this condition, when the true population cell means varied (H_0 not true), the within cell population variance remained constant for all cells. In order to hold the error variances constant for all cells when sampling from the exponential and lognormal distributions, the random numbers were always drawn from population distributions with a mean of 10. To obtain unequal means, a constant equal to μ_j - 10 (M_j - 10 for the lognormal) was added to each number. In effect, this shifts the distribution along the abscissa, holding its variance constant while varying the mean.

Table 1

PARAMETER VALUES USED IN THE STUDY

	n			
k	4	8	16	32
2	*	*	* **	*
4	* **	*	* **	*
8	*			

* {Normal, Exponential, Lognormal} $\sigma_j^2 = \sigma_{j'}^2 = 100$

** {Normal, Exponential, Lognormal} $\sigma_j^2 = \mu_j^2$

In the case of unequal within cell variance, the cell variance was proportional to the square of the mean. In this case, both assumptions of normality and homogeneity of error are violated for the two nonnormal distributions. For the normal one, only the assumption of homogeneity of error is violated.

4. RESULTS AND DISCUSSION

The effect of nonnormality when H_0 is true (Type I error) will be presented first. Then a presentation will follow describing how nonnormality affects F-test power when the within cell variances are equal. The effects of nonnormality on F as a function of the correlation between the numerator and the denominator of F will be discussed and some empirical and analytical results presented. Finally, the F-test power in the case of unequal variances will be presented along with some empirical results concerning the correlation between MST and MSE.

TYPE I ERROR

The observed Type I error (α') for each of the three distributions is shown in Table 2 for three values of α (where α is the true Type I error for the F-test when the assumptions underlying the test are met). The difference between α' and α for the normal distribution is due to sampling error and the linear interpolation. The difference is small in all cases. It may be observed in Table 2 that the nonnormal distributions lead to conservative Type I errors, i.e., the observed values are always smaller than the theoretical α level. Thus, if a test is designed with α level protection against a Type I error under the assumption of a normal distribution, even more protection against a Type I error exists if the distribution is of the nonnormal type specified here. It may be noted that the difference between the normal and the nonnormal distributions decreases as the sample size increases. Further, the size of $\alpha' - \alpha$ increases as the size of

Table 2

OBSERVED TYPE I ERROR (α')

k	Distribution	n: 4	3	16	32
			α = .10		
2	Normal	.101	.098	.100	.098
	Exponential	.087	.098	.099	.099
	Lognormal	.074	.085	.094	.096
4	Normal	.099	.096	.096	.102
	Exponential	.083	.086	.099	.092
	Lognormal	.076	.076	.085	.092
8	Normal	.100			
	Exponential	.086			
	Lognormal	.079			
			α = .05		
2	Normal	.048	.049	.050	.049
	Exponential	.042	.044	.046	.049
	Lognormal	.031	.035	.038	.046
4	Normal	.048	.047	.048	.051
	Exponential	.041	.040	.047	.045
	Lognormal	.035	.034	.037	.044
8	Normal	.053			
	Exponential	.044			
	Lognormal	.040			
			α = .01		
2	Normal	.009	.009	.009	.010
	Exponential	.007	.007	.007	.009
	Lognormal	.005	.005	.004	.007
4	Normal	.010	.009	.011	.010
	Exponential	.010	.007	.009	.007
	Lognormal	.008	.006	.006	.008
8	Normal	.011			
	Exponential	.009			
	Lognormal	.009			

γ_1 (skewness) and γ_2 (kurtosis) increases. All of these results agree with previous findings.

POWER: THE CASE OF EQUAL WITHIN CELL VARIANCE

The actual values of the means and the resultant values of ϕ are shown in Appendix A for the various combinations of n and k used in the study. Also shown is the observed power of the F-test for three values of α (.10, .05, .01) for each of the distributions. These data are shown in graphic form in Figs. 3, 4, and 5 for two levels of α (.05 and .01).

The power curves for the F-test for k = 2 are shown in Fig. 3. The power curve when the underlying distribution is normal is shown by the solid line. Over most of the range of ϕ, the power curve in the normal case is lower by a substantial amount compared to the case of either the exponential (dashed-line) or the lognormal (long-dashed-line) distributions. Further, the power curve based on the lognormal distribution is higher than it is when based on either the normal or the exponential. As the sample size increases, the curves based on the nonnormal distributions approach the one based on the normal distribution. It may be observed that small values of n result in larger differences in power. In general, the nonnormal curves at $\phi = 0$ (α level) are below the normal curve; they rise above it very quickly (before $\phi = \frac{1}{2}$) and remain above until ϕ is quite large. As α gets smaller, the point at which the normally based curves rise above the nonnormal ones occurs at a smaller value of ϕ. It is inferred that for extremely small values of α, the F-test would be conservative for only a limited range of ϕ or for only small values

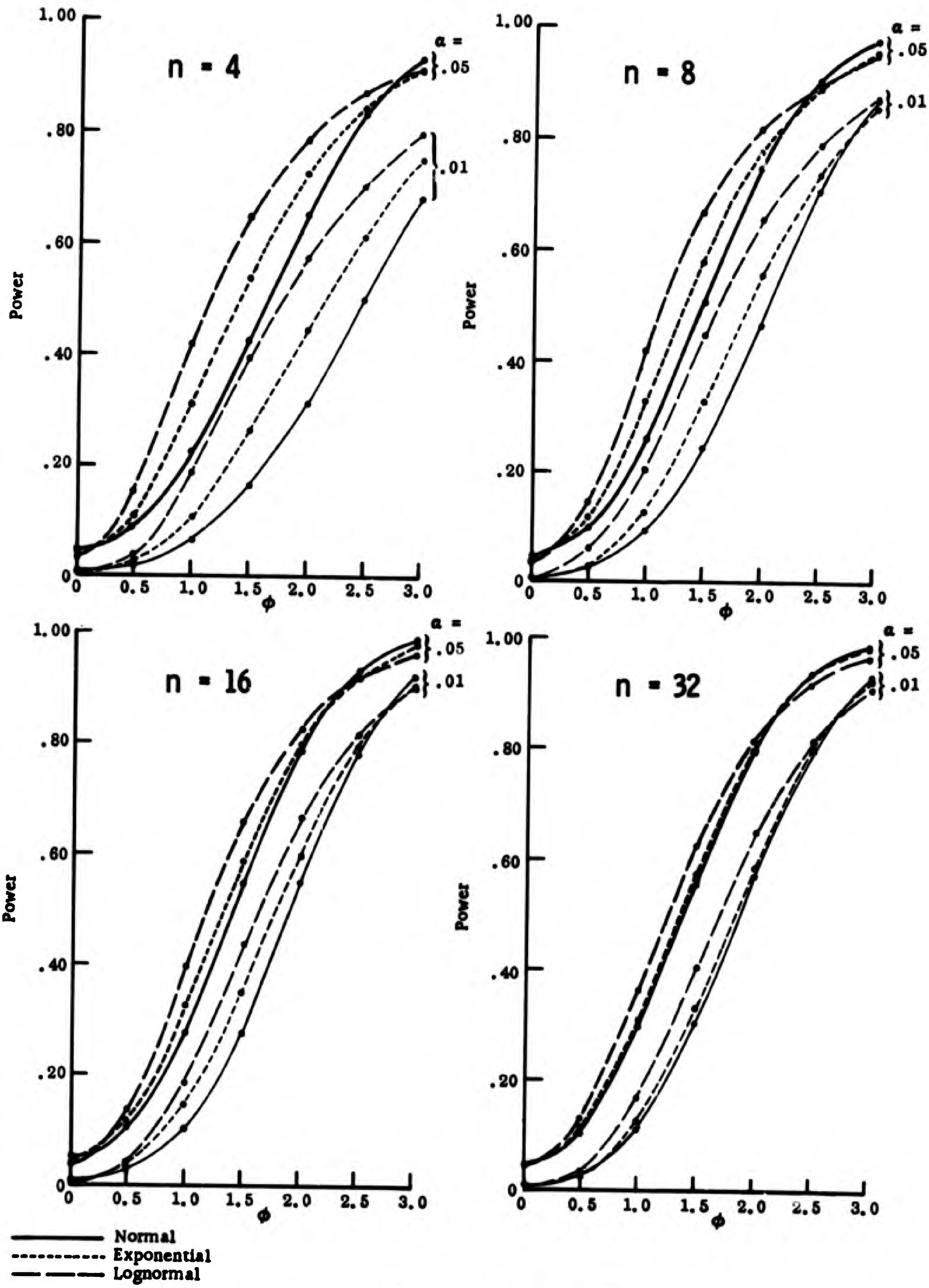


Fig. 3 -- Power Curves for k = 2

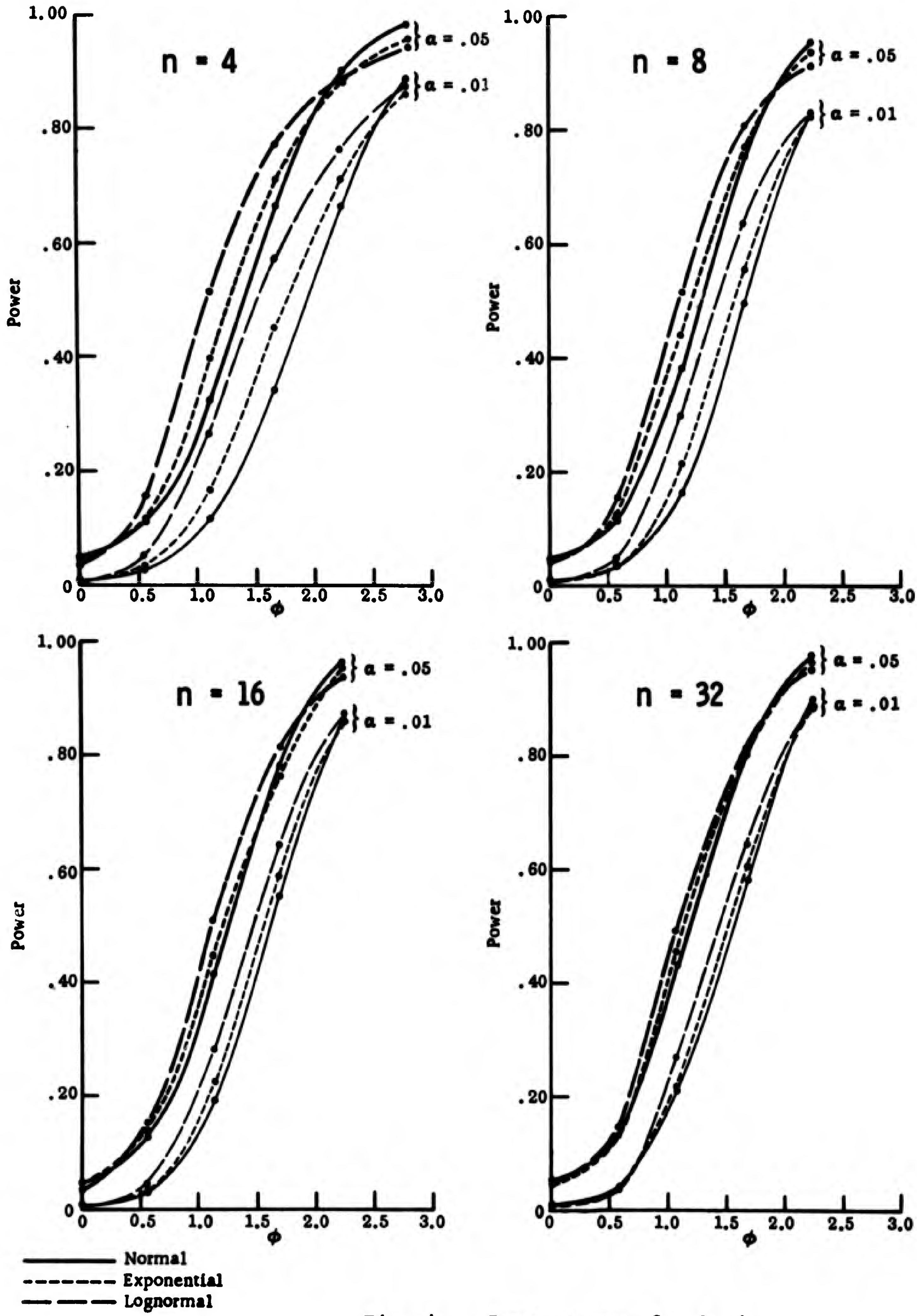


Fig. 4 -- Power Curves for k = 4

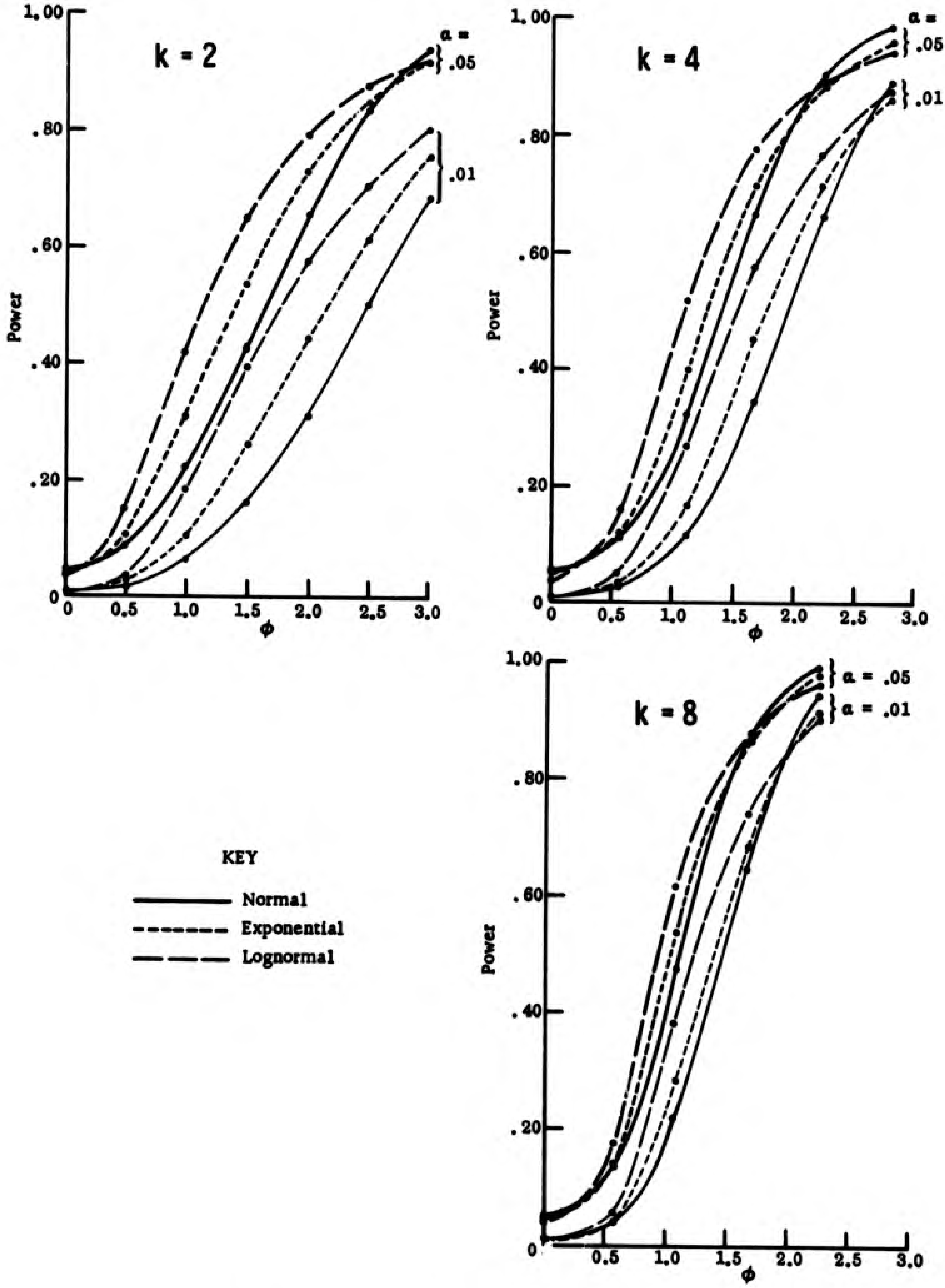


Fig. 5 -- Power Curves for k = 2, 4, 8 at n = 4

of absolute power. Practically speaking, this is of little consequence. Experiments are rarely designed for $\alpha < .01$, and when they are, appreciable power can be obtained only by the use of very large samples, in which case the test would be highly robust.

Figure 4 shows results for k = 4. Compared with k = 2, these curves are less disparate, and power for the nonnormal distributions approaches the normal case more rapidly.

The effect of the size of k is shown clearly in Fig. 5. In this figure, the power curves for three values of k (2,4,8) at n = 4 are shown. As with the Type I error, increasing either n or k results in F-tests that more nearly approximate the normal case. As also with the Type I error, the tests are more conservative with respect to power for those distributions with the highest γ_1 and γ_2 parameters.

These results are based on equal differences between successive means (see values of μ_j in Appendix A). As long as the within cell variances are equal, however, the relative differences between successive means is of no consequence. This is apparent from Eq. (2.6). Since the variances are equal, σ_ϵ^2 is independent of the location of μ_j, and ϕ is proportional to $\sum(\mu_j - \mu)^2$.

To summarize, not only is the F-test robust, but for small samples it is generally conservative with respect to both Type I error and power. As the sample size (or number of samples) increases, the test is less conservative for nonnormal distributions, and the power curves approach that of the F-test based on normality. As n increases, the F-test becomes truly robust (distributed as F regardless of the underlying distribution). These results are similar to those reported by Srivastava (1958) for positive values of γ_1 and γ_2, although the difference in power between the normal and nonnormal distributions is greater in the present study than it was in his. This is due to the higher values of γ_1 and γ_2 that were possible in this study. Srivastava's results indicated that both γ_1 and γ_2 contribute to the effect of nonnormality on F. Further, he found that positive values of γ_1 and γ_2 led to conservative F-tests, while negative values lead to an opposite effect. The characteristics of the nonnormal distributions used in the present study did not allow manipulation of the sign of γ_1 and γ_2.

DISTRIBUTION OF MST AND MSE

The effect of the underlying distribution on the F-test originates from the combined influence of the numerator (MST) and denominator (MSE). When the underlying distribution is normal, MST and MSE are distributed as chi-square and they are independent. If the distributions are nonnormal, MST and MSE are not distributed as chi-square, nor are they independent. Thus, the effect of nonnormality on F may be due to deviations in the distributions of MST and MSE from that of chi-square and to the correlation between MST and MSE.

The cumulative empirical distributions of MST and MSE under H_0 are shown in Fig. 6 for k = 4 and n = 4 and 32. These distributions are shown in terms of the percentile points of the cumulative distribution of MST and MSE when the underlying distribution is normal. In the normal case

(4.1) $$\chi^2_\nu = \frac{\nu s^2}{\sigma^2},$$

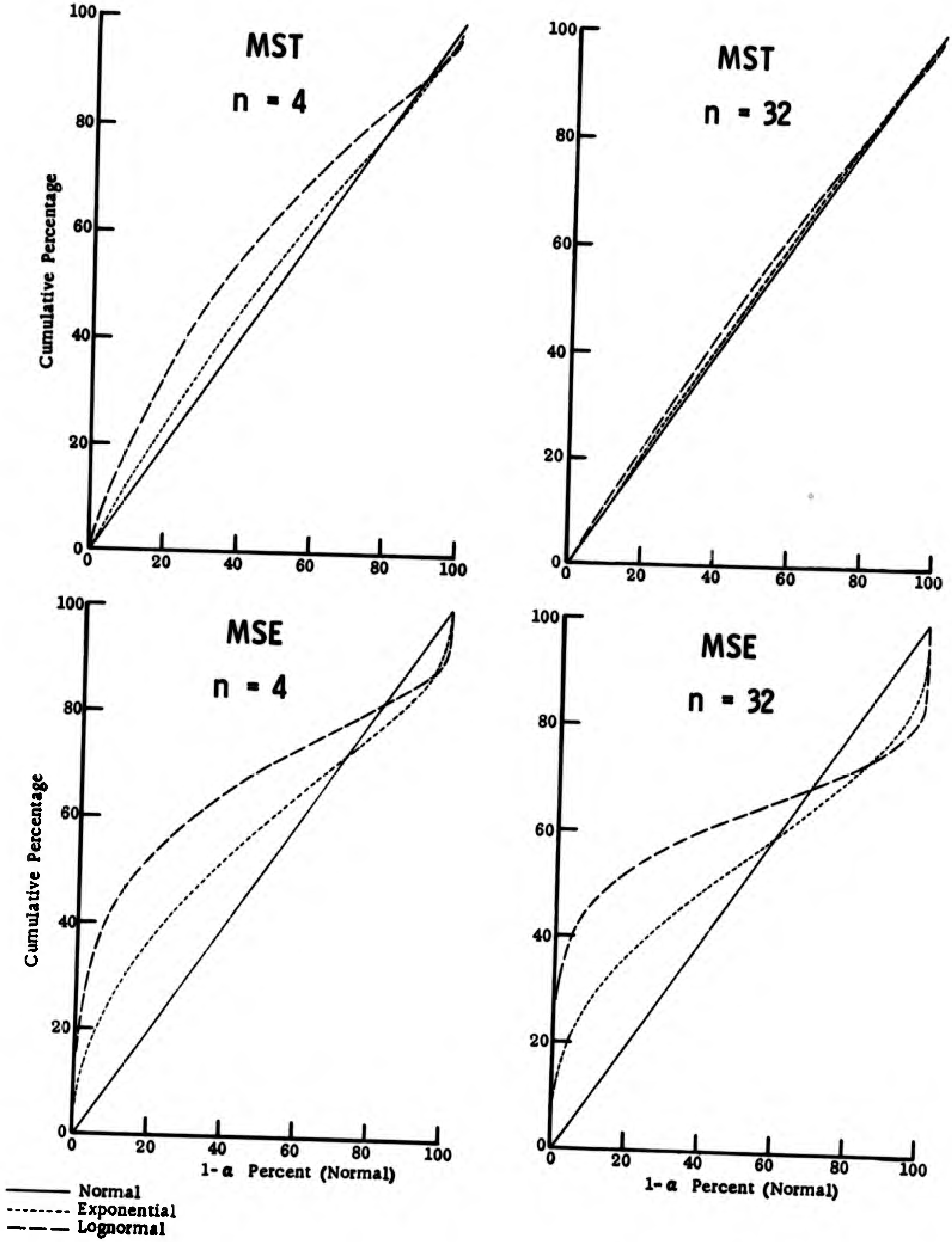


Fig. 6 -- Distributions of MST ($\phi = 0$) and MSE

where s^2 is any variance with v degrees of freedom. From tables of χ^2 one can readily determine the value of s^2 below which is $100(1-\alpha)$ percent of the cumulative distribution of s^2 (and χ^2). Setting s^2 = MST (with k-1 degrees of freedom), and rearranging terms, the $100(1-\alpha)$ percentile points of MST are,

(4.2) $$MST_{1-\alpha} = \left(\frac{\sigma^2}{k-1}\right) \chi^2_{1-\alpha}.$$

For example, for k = 4 and $\sigma^2 = 100$, $MST_{1-\alpha} = 33.333\chi^2_{1-\alpha,v}$. Below this value lies $100(1-\alpha)$ percent of the distribution of MSE when the assumption of normality is true. From tables of chi-square, $\chi^2_{0.50,3} = 2.366$, so that $MST_{.50} = 78.86$, and 50 percent of the MST's for a normally distributed variable have an expected value of less than, or equal to 78.86. It may be observed in Fig. 6 that 63.69 percent of the MST's for the lognormal distribution lie below 78.86 (and 50 percent of MST's for the normal). Similar calculations were used for the MSE curves.

Inspection of Fig. 6 indicates that the effect of nonnormality on MST is slight compared to its effect on MSE; and further, as n increases, the effect of MST decreases, while it gets worse for MSE. As n increases, a greater proportion of MSE's for the nonnormal distributions lie in the tails of the region defined by the distribution of the MSE's for the normal distribution. This is easier to observe in Fig. 7, which shows the actual frequency distributions of MSE for n = 4, 8, 16, and 32. These results are expected from theoretical considerations.

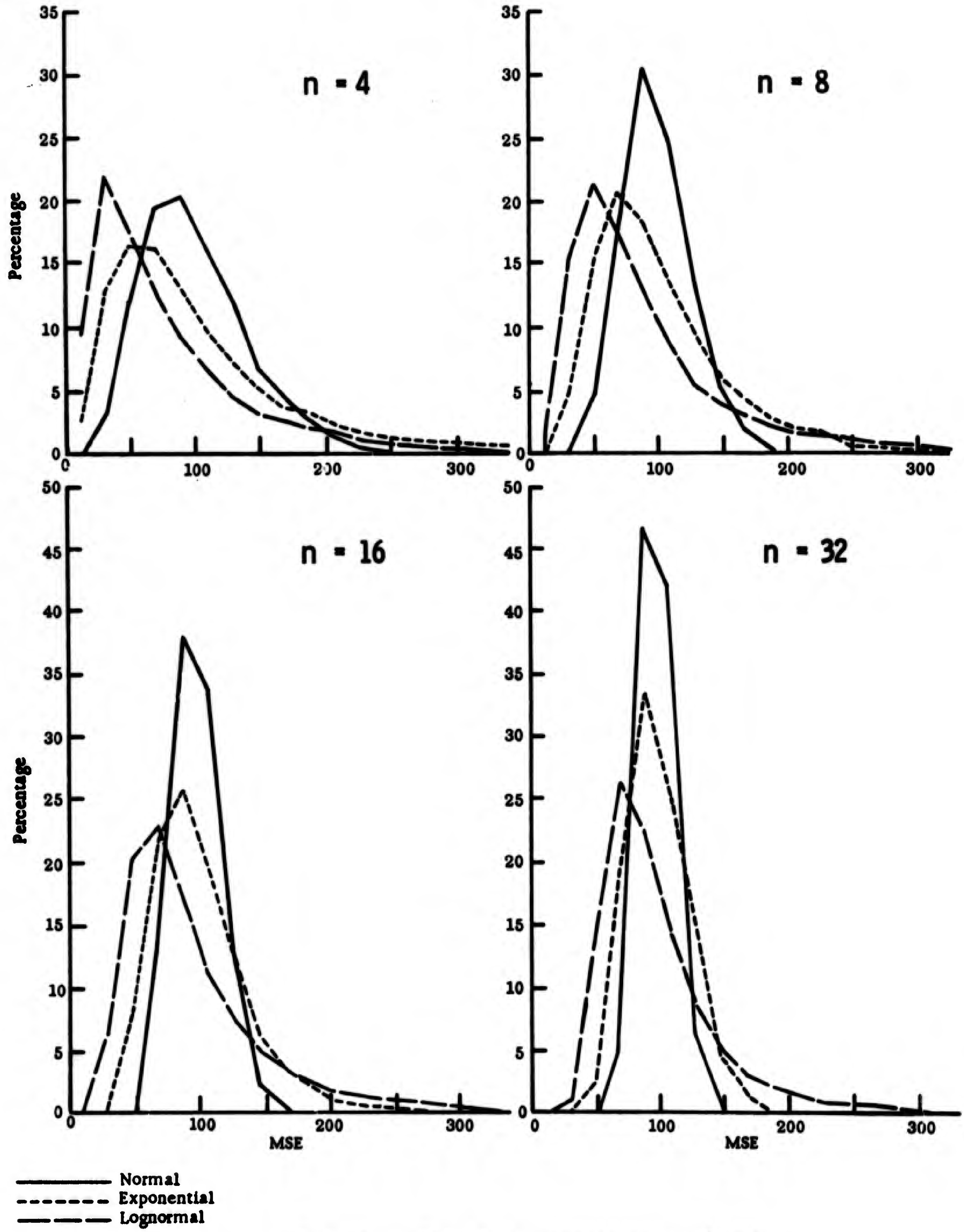


Fig. 7 -- Frequency Distribution of MSE

The calculation of MST (see Eq. 2.4) depends only upon averages. Thus, by the Central Limit Theorem,* as n increases, MST becomes less sensitive to nonnormality. Further, as n or k increases, the effect of nonnormality on the variance of MST is rapidly reduced. From Appendix C, the variance of MST is

(4.3) $$\mathrm{Var(MST)} = \frac{2\sigma^4}{k-1}\left[1 + \tfrac{1}{2}\gamma_2\left(\frac{k-1}{nk}\right)\right],$$

where Var = variance. The actual rate at which Var(MST) converges to its normal theory value for the distributions used in this study is shown in Table 3 as a function of n and k.

For MSE the situation is very different. Atiquallah (1962)** shows that

(4.4) $$\mathrm{Var(MSE)} = \frac{2\sigma^4}{k(n-1)}\left[1 + \tfrac{1}{2}\gamma_2\left(\frac{(n-1)}{n}\right)\right].$$

When $\gamma_2 = 0$,

(4.5) $$\mathrm{Var(MSE)} = \frac{2\sigma^4}{k(n-1)},$$

which is the case for the normal distribution. If $\gamma_2 \neq 0$, however, it is apparent that the variance of MSE in the nonnormal case is larger than in the normal. Further, as n increases, the variance

---

*The Central Limit Theorem states that if the variance of x exists, $\bar{x}$ is normally distributed for large n ($n \to \infty$) regardless of the distribution of x.

**This may also be derived from a lemma in Scheffé (1959, p. 255) or from the methods discussed in Appendix C of this paper. Atiquallah's approach is somewhat different and further shows that s_ϵ^2 is a minimum unbiased estimate of σ_ϵ^2 for any quadratically balanced design. The conditions of quadratic design are met in the present study.

Table 3

VAR(MST) FOR NONNORMAL RELATIVE TO NORMAL DISTRIBUTIONS

n	k: 2	4	8	16	∞
			$\gamma_2 = 6$		
2	1.7500	2.1250	2.3125	2.4063	2.5000
4	1.3750	1.5625	1.6563	1.7031	1.7500
8	1.1875	1.2813	1.3281	1.3516	1.3750
16	1.0938	1.1406	1.1641	1.1758	1.1875
32	1.0469	1.0703	1.0820	1.0879	1.0938
64	1.0234	1.0352	1.0410	1.0439	1.0469
128	1.0117	1.0176	1.0205	1.0220	1.0234
256	1.0059	1.0088	1.0103	1.0110	1.0117
∞	1.0000	1.0000	1.0000	1.0000	1.0000
			$\gamma_2 = 38$		
2	5.7500	8.1250	9.3125	9.9063	10.5000
4	3.3750	4.5625	5.1563	5.4531	5.7500
8	2.1875	2.7813	3.0781	3.2266	3.3750
16	1.5938	1.8906	2.0391	2.1133	2.1875
32	1.2969	1.4453	1.5995	1.5566	1.5938
64	1.1484	1.2227	1.2598	1.2783	1.2969
128	1.0742	1.1113	1.1299	1.1392	1.1484
256	1.0371	1.0557	1.0649	1.0696	1.0742
∞	1.0000	1.0000	1.0000	1.0000	1.0000

of MSE for nonnormal distributions increases relative to MSE for normal distributions. This is apparent from the ratio of Eq. (4.4) to Eq. (4.5) which is

(4.6) $$1 + \tfrac{1}{2}\gamma_2\left(\frac{n-1}{n}\right).$$

As n gets large: the quantity (n-1)/n approaches 1, the effect of γ_2 is maximum,* and the variance of MSE for nonnormal distributions

*See Scheffé (1959, p. 336) for a nearly identical development.

relative to the normal is maximum. This limit is approached rapidly. For n = 2, the expression is $1 + \frac{1}{4}\gamma_2$, while at n = 8, it is $1 + (7/16)\gamma_2$, which is already quite close to its limit (i.e., the relative distribution of normally and nonnormally based MSE's is fixed at small values of n).

Of course, as n becomes large, the absolute value of Var(MSE) approaches zero and the denominator of F approaches a constant (σ_ϵ^2). Thus, for large n, the distribution of MSE has no effect on the distribution of F. Further, MST is distributed as chi-square (as under normal theory) by the Central Limit Theorem, so that in the limit, the distribution of F is unaffected by the underlying distribution of the random variable. Bradley (1964) has discussed this ultimate ($n \to \infty$) robustness of F in detail. Further, Bradley shows that F is ultimately robust for either heterogeneous within cell variances or unequal sample size, but not for both.

The important point is the speed with which F, based on a nonnormal distribution, approaches its normal theory value as n increases. The results of this study indicate the convergence is very rapid and only small errors in either α or β would be expected at a sample size of 32. This convergence in the distribution of nonnormally based F to the distribution of F in the normal case is shown directly in Figs. 8 and 9. In these figures the cumulative distributions of F are shown for two values of n at $\phi = 0$ and $\phi = 1.68$ for k = 4. For $\phi = 0$ (H_0 true) there is only a slight difference between the normally and nonnormally based F distributions even at n = 4. As ϕ increases, the three distributions appear more unequal. However, these differences

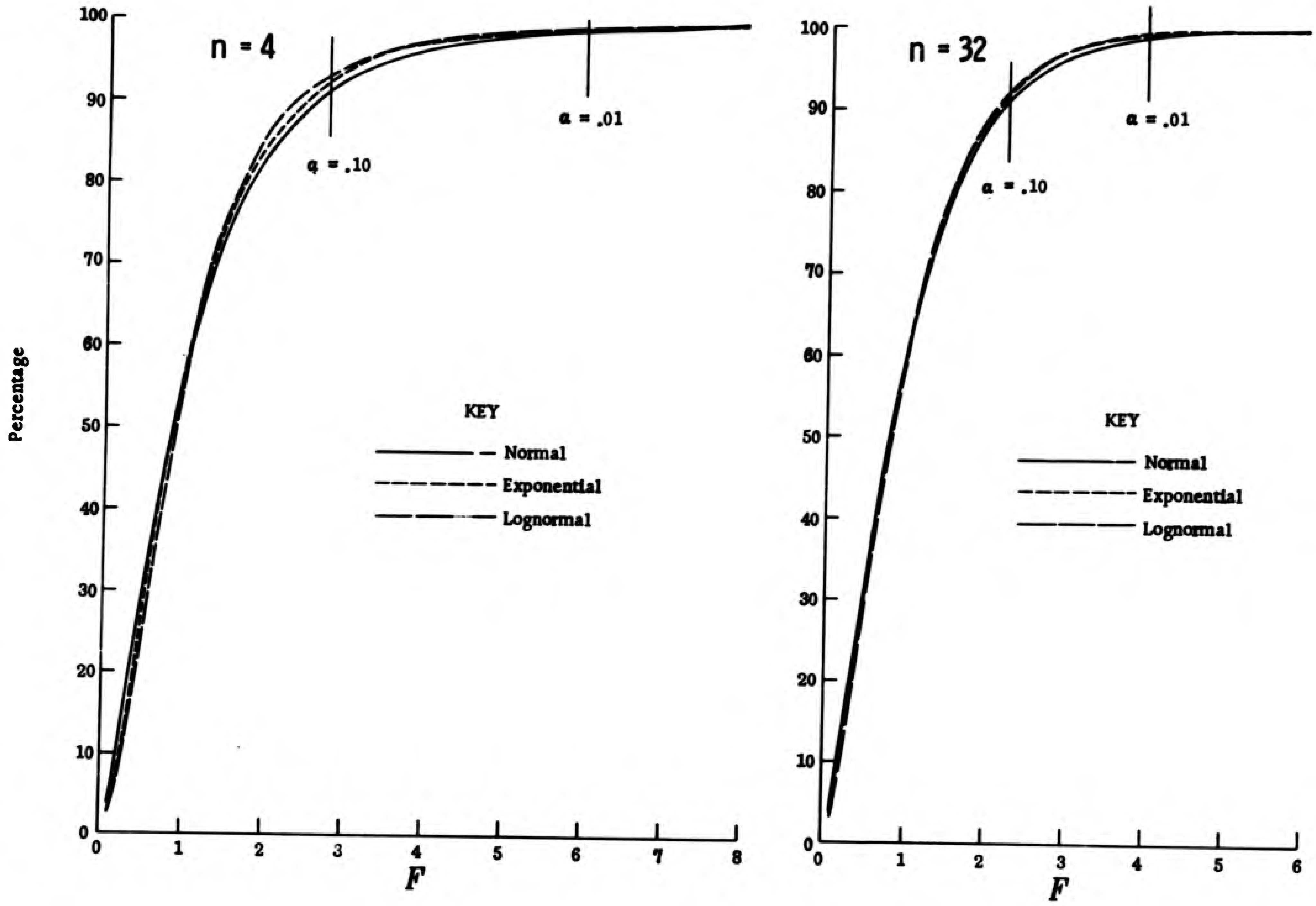


Fig. 8 -- Cumulative Distribution of F fcr $\phi = 0$ and k = 4

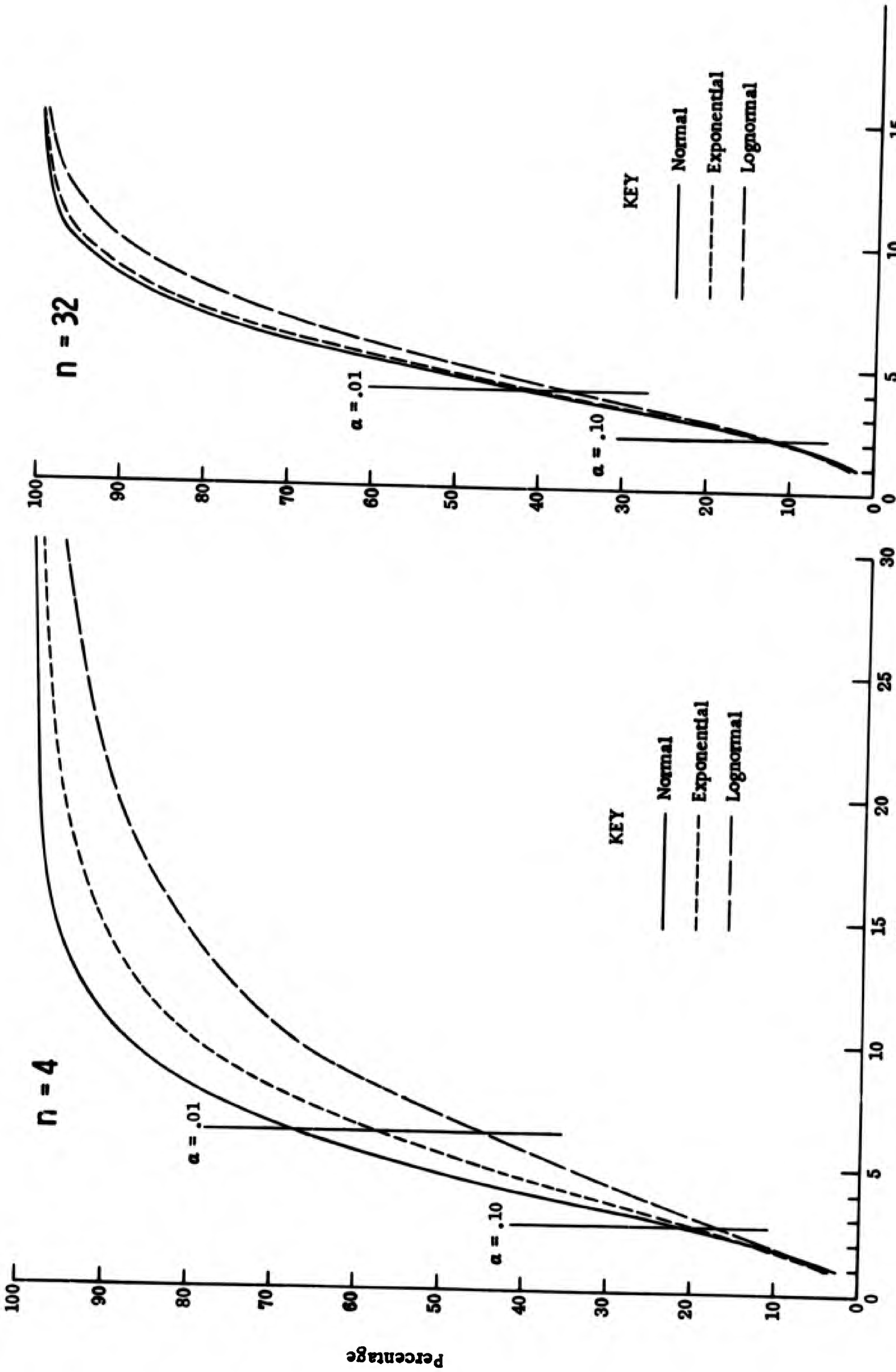


Fig. 9 -- Cumulative Distribution of F for ϕ = 1.68 and k = 4

at $n = 32$ are large only in the region of the distributions corresponding to very small α values. The points on the distributions for which $\alpha = .10$ and $.01$ are shown in the figures.

From Figs. 8 and 9 it is apparent that the distribution of F is much less dependent on the underlying distribution than is the distribution of MSE; that is, F converges to its normal theory distributions for small n, while MSE does not converge for any n. This rapid convergence of F cannot be explained in terms of the distribution of MSE, and certainly, the conservative feature of F for small samples cannot be attributed to any feature of the distribution of MSE. For finite n there are too many nonnormally based MSE's smaller than expected under normal theory and also too many that are larger. Inspection of Figs. 6 and 7 shows that relatively more small MSE's occur than larger ones for small n. This would lead to rejection of H_0 when it is true with a probability higher than α (which does not occur) if MST and MSE are uncorrelated. In order to explain the insensitivity of F to the underlying distribution, it is obviously necessary to consider the correlation between MST and MSE.

It will be proven below that a correlation does exist between MST and MSE for any distribution for which γ_2 exists. Because of this correlation, small MSE's (relative to expected value) have an increased probability of occurring with small MST's. This leads directly to a conservative α level. Similar reasoning explains the conservative feature of F for Type II error. A correlation between MST and MSE would be expected to disappear for large n, and for finite n to at least partially cancel out the effect of nonnormality on F.

CORRELATION BETWEEN MST AND MSE

That a substantial correlation does exist between MST and MSE is shown in Table 4. The values at $\phi = 0$ are theoretical ones; the method to be discussed below. For $\phi \neq 0$, the correlations are empirical ones determined by the computer sampling. It may be noted that the correlation decreases as either n or ϕ increases, or as k decreases.

Table 4

CORRELATION COEFFICIENT BETWEEN MST AND MSE
(Equal Variances)

		Exponential					Lognormal				
		ϕ					ϕ				
k	n	0	.5	1.0	1.5	2.0	0	.5	1.0	1.5	2.0
2	4	.44		.21		.11	.81		.49		.29
	8	.34				.08	.72				.25
	16	.25	.19	.13	.10	.07	.59	.37	.19	.10	.05
	32	.18					.47				
		ϕ					ϕ				
		0	.56	1.12	1.68	2.24	0	.56	1.12	1.68	2.24
4	4	.50	.40	.28	.20	.15	.85	.76	.60	.49	.38
	8	.40		.23		.13	.78		.52		.33
	16	.30		.19		.07	.67		.32		.20
	32	.22				.05	.54				.19
8	4	.52					.87				
	8	.42					.80				

The correlation between MST and MSE under H_0 will be derived for the case of equal within cell variance and equal sample size. Nothing is assumed about the shape of the distribution except that it is the same in all cells.

The correlation is defined as

$$\rho = \frac{\text{Cov(MST,MSE)}}{\sqrt{\text{Var(MST) Var(MSE)}}} \; . \tag{4.7}$$

From the results of Appendix C,

$$\text{Cov(MST,MSE)} = \frac{\gamma_2 \sigma^4}{nk} \; . \tag{4.8}$$

Since Var(MST) and Var(MSE) are known from Eqs. (4.3) and (4.4), the correlation after algebraic simplification is

$$\rho = \frac{\gamma_2}{\sqrt{\frac{4n^2k}{(k-1)(n-1)} + \frac{2n(nk-1)}{(k-1)(n-1)}\gamma_2 + \gamma_2^2}} \; . \tag{4.9}$$

It is apparent from inspection of Eq. (4.9) that as $n \to \infty$, $\rho \to 0$ for any value of k. For finite n and $k \to \infty$,

$$\rho = \frac{\gamma_2}{\sqrt{\frac{4n^2}{(n-1)} + \frac{2n^2}{(n-1)}\gamma_2 + \gamma_2^2}} \; . \tag{4.10}$$

The values of ρ as a function of n and k are shown in Table 5 for the exponential and lognormal distributions. It is apparent that k has a small effect on ρ relative to the effect of n. The correlation between the numerator and denominator of F is largely determined by sample size and the kurtosis (but not the skewness) of the underlying distribution.

It is now possible to show how γ_2 leads to robust F-tests under H_0 and it is clear that the same phenomenon operates under H_1.

Hansen, Hurwitz and Madow (1953, Chap. 4, Sec. 18) have developed an approximation for the variance of the ratio of two random variables.

Applying this approximation to the F ratio, under H_0

$$\text{(4.11)} \qquad \text{Var}\left(\frac{\text{MST}}{\text{MSE}}\right) \sim \frac{1}{\sigma^4}\,[\text{Var(MST)} + \text{Var(MSE)} - 2\,\text{Cov(MST,MSE)}].$$

The validity of the approximation is a function of sample size. Because of the algebraic form of MST and MSE it is difficult to determine the exact error in the approximation as a function of n and k. For the moment the approximation will be accepted and the variance of F under H_0 for any underlying distribution will be explored.*

Table 5

VALUES OF ρ

	k							
	Exponential				Lognormal			
n	2	4	8	∞	2	4	8	∞
2	.507	.564	.584	.600	.865	.891	.899	.905
4	.435	.599	.524	.545	.811	.854	.868	.879
8	.338	.339	.423	.444	.716	.777	.798	.815
16	.252	.302	.322	.341	.594	.668	.695	.717
32	.183	.221	.238	.253	.466	.541	.570	.594
64	.131	.159	.172	.183	.350	.416	.442	.466
128	.093	.114	.123	.131	.256	.307	.330	.350
256	.070	.086	.092	.093	.195	.236	.254	.256
∞	0	0	0	0	0	0	0	0

Substituting Eqs. (4.3), (4.4), and (4.8) in Eq. (4.11) and using parentheses to separate Var(MST) and Var(MSE), the variance of F under H_0 is,

*The variance of MST under H_1 is very complex involving γ_1 as well as γ_2, and a number of complex products of means and variances of several powers. Thus, even if the approximation is good under H_1, it is of little value.

$$(4.12)\qquad \mathrm{Var}\left(\frac{MST}{MSE}\right) \sim \frac{1}{\sigma^4}\left[\left(\frac{2\sigma^4}{k-1} + \frac{\gamma_2\sigma^4}{nk}\right) + \left(\frac{2\sigma^4}{k(n-1)} + \frac{\gamma_2\sigma^4}{nk}\right) - \frac{2\gamma_2\sigma^4}{nk}\right].$$

Inspection of Eq. (4.12) reveals that the terms containing γ_2 add to zero. The effect of γ_2 (or correlation) reduces the variance of F based on nonnormal distributions to approximately the variance it would have when the assumption of normality is true. It may be observed in Fig. 8 that the variance of F under H_0 is not much affected by the shape of the underlying distribution. Thus, the approximation given in Eq. (4.11) is good at even small sample sizes.

When H_0 is not true the situation is more complex. As ϕ increases, the correlation between MST and MSE decreases, while the cumulative distributions of F are less similar (Fig. 9), especially for n small. However, as ϕ increases the variance of MST also increases, thus contributing more to the total variance of F. For small n the affects of nonnormality are therefore severe. A precise explanation of the highly conservative feature of the F-test would require an expression for the variance of MST and covariance of MST, MSE under H_1.

POWER: THE CASE OF UNEQUAL WITHIN CELL VARIANCE

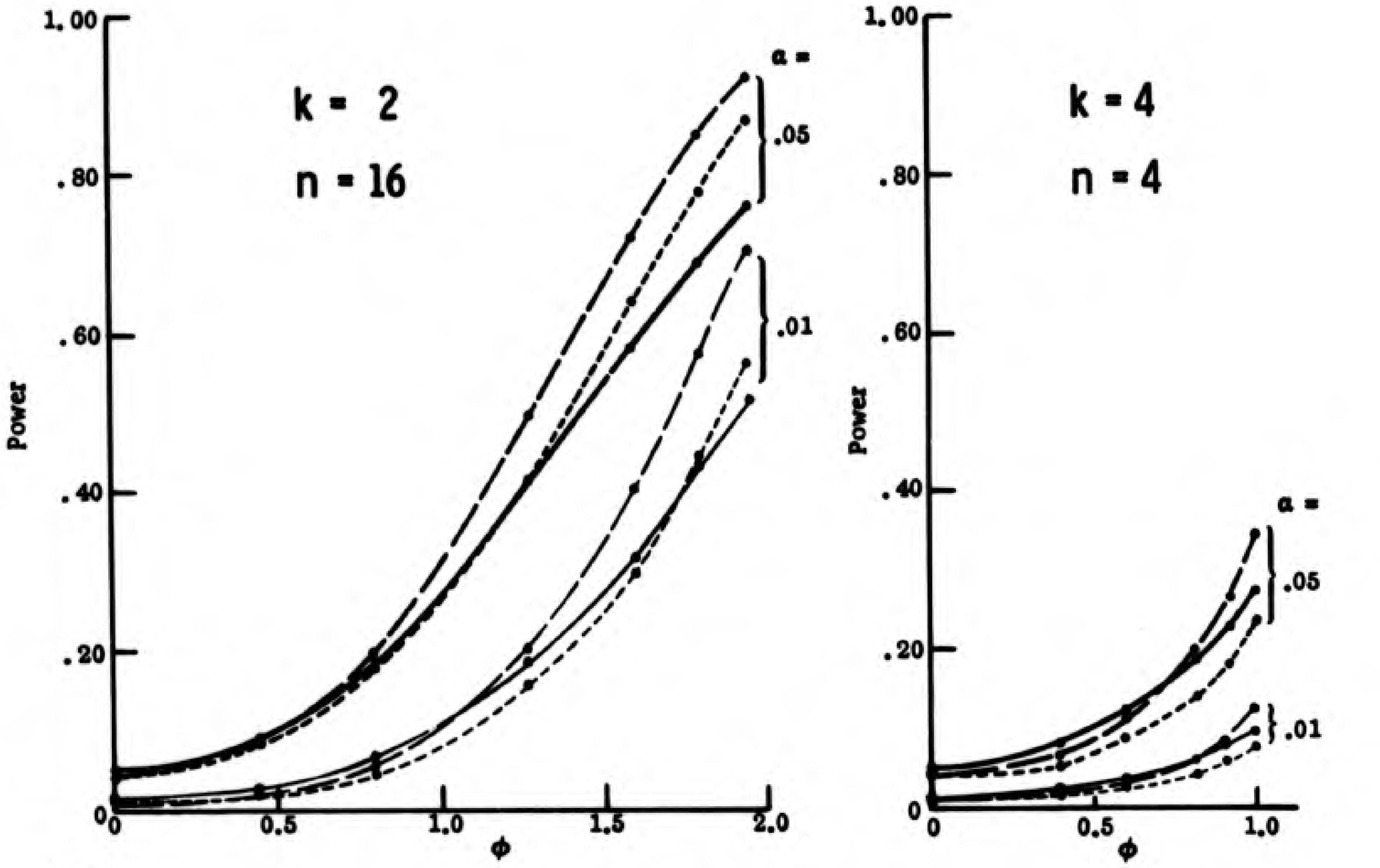
The foregoing results were obtained for the case of equal within cell variance and unequal means. In this section, the within cell variance is a function of the mean. Under H_0, the within cell variances (and means) are equal, and the previous results apply. For the nonnormal distributions used in this study, any other condition raises difficulties. For the exponential distribution, there is only one parameter (λ), and this determines both the mean and variance.

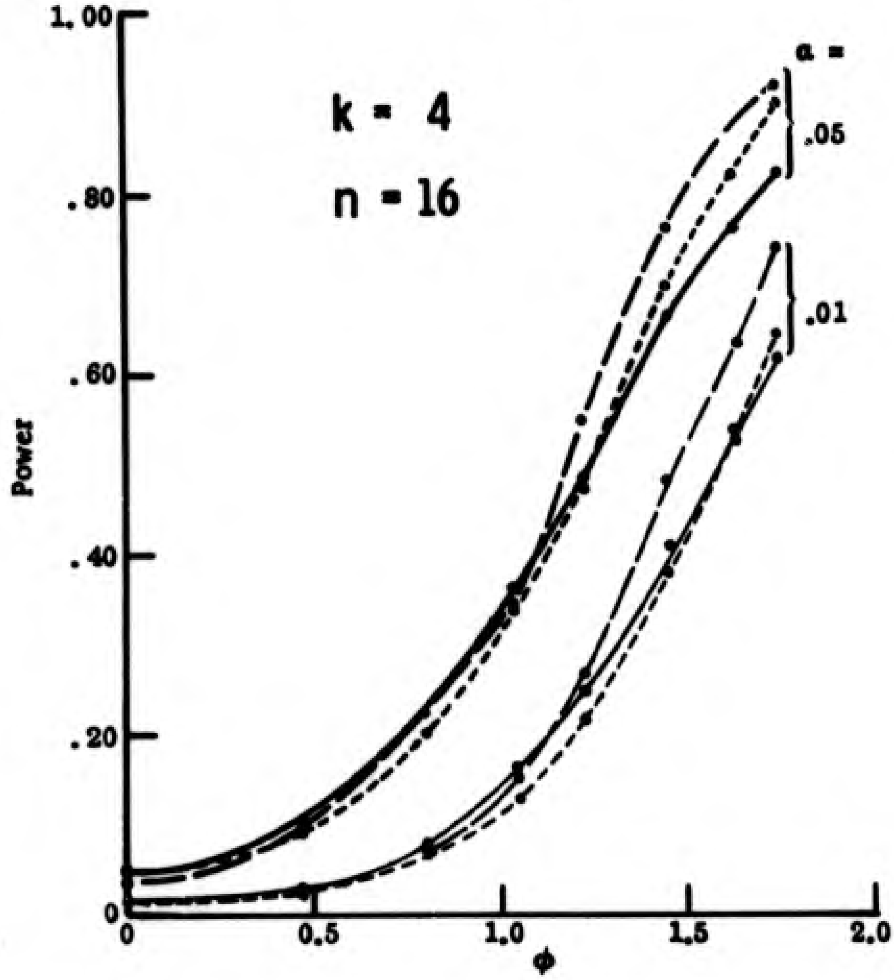
It would be possible to construct lognormal distributions with unequal variances under H_0, but then γ_1 and γ_2 would vary from cell to cell. It appears desirable to hold these constant. Under H_1, the variances are equal to the mean squared.

Under the restriction of equal sample size, MSE is the average of the within cell sample variance (s_j^2). MSE is an unbiased estimate of σ_ϵ^2, the average of the within cell population variances (i.e., $\sigma_\epsilon^2 = \frac{1}{k}(\sigma_1^2 + \ldots + \sigma_k^2)$. The calculation of ϕ (see Eq. 2.6) is based on σ_ϵ^2.

It is possible to obtain any (finite) ratio between the mean and variance under H_0 when the distribution is normal. In this case it is possible to determine the effects of unequal variance alone on the power of the F-test. Previous studies (Horsnell, 1953) indicate that the effect of unequal variance is small if the samples are of equal size. Under the conditions of equally spaced means (used throughout this study), the power curves for equal and unequal variance for a normally distributed variable were nearly identical. The slight differences that existed (generally in the third decimal place) were attributed to sampling error. The normal power curves reported in this section are valid for both equal and unequal variance.

The actual values of μ_j, ϕ, σ_ϵ^2 ($\sigma_j^2 = \mu_j^2$ is not given but is easily obtained) and the resultant power are shown in Appendix B. The power curves for the investigated values of n and k (k = 2, n = 4, k = 4, n = 4, 16) are shown in Fig. 10 for two values of α (.05, .01). The inset table in Fig. 10 shows the ratio of the largest to the smallest of the within cell variances as a function of ϕ. The ratio





KEY
Normal
Exponential
Lognormal

k	n	φ	Max / Min Mean	Max / Min Variance
2	16	.44	1.25	1.56
		.79	1.50	2.25
		1.94	3.50	12.25
4	4	.40	1.75	3.06
		.82	4.00	16.00
		1.01	8.50	72.25
4	16	.47	1.38	1.89
		.80	1.75	3.06
		1.75	4.75	22.56

Fig. 10 -- Power Curves for the Case of Unequal Within-Cell Variance

of the means is also shown. For example, for $k = 2$, $n = 16$, when $\mu_1 = 10$ and $\mu_2 = 30$, the ratio of the means is 3.0, and the ratio of the variances is 9.0 (since $\sigma_1^2 = 100$, $\sigma_2^2 = 900$). The value of σ_ϵ^2 in this example is 500.00. For $k = 4$, $n = 4$, the error variance (σ_ϵ^2) increases very rapidly and large values of ϕ require large differences between the means. This simply shows the low power of a test based on small samples and large error variance. It is observed in Fig. 10 that the normal distribution leads to slightly more powerful tests for small values of ϕ, but as ϕ increases, the power for the normal case falls below that of either the exponential or lognormal distributions. As in the case of equal variance, the lognormal distribution leads to the more powerful tests. Compared to the case of equal within cell variance, these curves show less power for small ϕ and greater power for large ϕ.

The empirical correlations between MST and MSE are shown in Table 6. Unlike the case of equal within cell variance, the correlations tend to increase as ϕ increases. The size of the correlation coefficient is closely associated with the degree to which F is conservative. As the correlation increases, the difference in power for the nonnormal relative to normal distribution increases.

In summary, positive (and equal within cell) kurtosis leads to generally conservative tests when the within cell variances are unequal. This conservative feature of the test is due to the correlation between the numerator and the denominator of F. Obviously, if the within cell variances are unequal and uncorrelated with the means, then these conclusions no longer apply.

Table 6

CORRELATION COEFFICIENTS BETWEEN MST AND MSE
(Unequal Variances)

	Exponential					Lognormal				
	k=2, n=16									
ϕ	0	.44	.79	1.58	1.94	0	.44	.79	1.58	1.94
ρ	.25	.35	.45	.63	.68	.59	.58	.60	.69	.72
	k=4, n=4									
ϕ	0	.40	.61	.82	1.01	0	.40	.61	.82	1.01
ρ	.50	.60	.60	.65	.68	.85	.83	.84	.82	.82
	k=4, n=16									
ϕ	0	.47	1.04	1.45	1.75	0	.47	1.04	1.45	1.75
ρ	.30	.36	.50	.54	.63	.67	.61	.64	.68	.70

5. CONCLUDING COMMENTS

Under the conditions of this study, it is apparent that neither the Type I or II error of the F-test is much affected by nonnormality and heterogeneous errors as long as the sample sizes are equal. For even moderate sample size, the test is robust. This conclusion, however, does not specify the degree of robustness that can be expected under any given set of conditions, and this is the problem the researcher must deal with. There have been several attempts to define robustness quantitatively.

Box and Andersen (1955) developed an approximation for correcting the degrees of freedom in the F-test that indicates the amount of robustness to Type I error. The approximation is primarily a function of γ_2; however, it applies only to H_0, and is based on equal skewness and kurtosis in each cell. One is still faced with the problem of estimating the kurtosis.

In a more recent study, Box and Tiao (1962) demonstrated the use of Bayesian methods for including estimates of γ_2 in the calculation of F-tests. Under certain assumptions, the posterior distribution of μ (the population mean under H_0) is determined as a function of variable kurtosis. The function $\mu(\gamma_2)$ which best estimates the data is then determined from the sample. This procedure does not require the usual χ^2 goodness of fit test nor direct estimation of γ_2. While this approach seems promising, it has not been developed to the point of practical usage.

Bradley (1963, 1964) has extensively studied the factors affecting robustness of F to Type I error. The number of factors and interactions between factors that affect robustness leads one to agree with the author that any statement about a test being "robust" is not generally of much practical value. If one is attempting to specify the degree to which a test is robust under any set of conditions, the task appears enormous.

From the above comments, it is apparent that an answer to the question, "How robust is robust?" is not yet available. However, there are a number of findings from studies of F-test robustness that may be used as experimental design criteria. When taken in conjunction with certain features of the data, these findings insure errors no larger than that specified by α and β for testing differences between means. The design criteria are: equal within cell sample size, α levels that are not extremely small,* and within cell sample size as large as possible, especially for negative kurtosis. Inspection of the distribution of even small samples should allow one to determine whether or not the kurtosis is positive or negative, particularly since values near zero are of little consequence regardless of sign. Under these conditions an F-test analysis of differences between means is valid.

Although the overall F-test may be said to be robust, departures from the assumptions of normality and homogeneous errors introduce a number of problems for data analysis. In the case of more than two samples ($k > 2$), and if the sample variances are unequal and correlated

---

*It was noted in this study that small values of α lead to less "robustness." Bradley found serious nonrobust effects for $\alpha < .01$. This does not seem to be a serious criticism. The requirements for such small α levels are rare, and generally when such certainty is required, an analysis of variance design is not the most desirable.

with the mean, differences between the smaller means may be concealed by the large contribution to total error from the larger means. This effect could be severe when multiple comparisons between means are based on the total error. In this case, the researcher might do well to sacrifice the large degrees of freedom of the pooled MSE error in favor of the actual within cell error of the two cells used in the comparison. Otherwise, the multiple comparison method (following a significant overall F) will lead to conservative effects for the smaller means and an opposite bias for the larger means.

Attempts to follow a significant F-test with a trend analysis (such as orthogonal polynomials)* is in serious question as are any F-tests based on regression models. A study by Box and Watson (1962), however, indicates that for certain design criteria, F-tests for regression are robust. One might conclude that a linear trend test under the same criteria would also be robust.

If the sample means and variances are highly correlated, or if there is other information about the true distribution, a data transformation may be used. In many applied situations, however, the true distribution is unknown, and estimating the distribution from small samples is hazardous at best. Even when a transformation is justified, it may lead to undesirable results, and often there is not an appreciable effect on the outcome of the F-test. For a more complete discussion of the advantages and disadvantages of transformations, the reader is referred to Scheffé (1959, pp. 364-368) and Kempthorne (1952, pp. 153-158). Suffice it to say that transformations are not generally an answer to the problems addressed in this Memorandum.

---

*See Winer, (1962).

One alternative open to the researcher is to use a nonparametric test which does not assume anything about the underlying distribution. However, the F-test is very insensitive to distribution alone, and when the variances are heterogeneous, many nonparametric tests are theoretically as invalid as are the parametric ones. Moreover, a study by Pratt (1964) indicates that the t-test is more insensitive to unequal within cell variances ("dispersion" for the nonparametric case) than are the more common nonparametric equivalents.

It is obvious from these concluding comments that there is still much to be learned about "robustness." Further, given robustness, a number of problems remain for data analysis. However, the findings of this study, and others, point out a wide class of conditions under which the F-test is valid even though the underlying assumptions are violated.

Appendix A

DISTRIBUTION PARAMETERS AND POWER IN THE CASE OF EQUAL WITHIN CELL VARIANCE

n	Parameters									Power: Normal			Exponential			Lognormal		
	μ_1	μ_2	μ_3	μ_4	μ_5	μ_6	μ_7	μ_8	ϕ	.10	.05	.01	.10	.05	.01	.10	.05	.01
k = 2																		
4	10	10							0.00	.101	.048	.009	.087	.042	.007	.074	.031	.005
	10	15							0.50	.163	.089	.020	.199	.109	.025	.259	.152	.037
	10	20							1.00	.349	.224	.065	.446	.310	.108	.552	.419	.189
	10	25							1.50	.586	.425	.162	.672	.539	.266	.749	.648	.396
	10	30							2.00	.802	.654	.312	.825	.723	.447	.859	.788	.575
	10	35							2.50	.925	.835	.502	.908	.844	.615	.917	.873	.704
	10	40							3.00	.978	.938	.685	.954	.913	.753	.949	.918	.799
8	10	10							0.00	.098	.049	.009	.098	.044	.007	.085	.035	.005
	10	13.536							0.50	.170	.098	.025	.195	.115	.030	.238	.145	.060
	10	17.072							1.00	.380	.260	.092	.445	.325	.128	.537	.417	.201
	10	20.608							1.50	.647	.505	.244	.687	.573	.328	.751	.663	.449
	10	24.144							2.00	.852	.749	.465	.849	.773	.553	.874	.815	.653
	10	27.680							2.50	.960	.909	.709	.937	.887	.738	.934	.899	.789
	10	31.216							3.00	.993	.978	.875	.975	.951	.856	.966	.949	.877
16	10	10							0.00	.100	.050	.009	.099	.040	.007	.094	.038	.004
	10	12.5							0.50	.174	.102	.028	.199	.117	.030	.217	.132	.037
	10	15.0							1.00	.399	.276	.105	.443	.325	.140	.504	.391	.181
	10	17.5							1.50	.669	.544	.276	.704	.586	.349	.750	.656	.438
	10	20.0							2.00	.869	.787	.548	.871	.797	.595	.882	.824	.667
	10	22.5							2.50	.967	.932	.782	.952	.917	.791	.951	.917	.818
	10	25.0							3.00	.995	.987	.927	.986	.970	.906	.977	.961	.904
32	10	10							0.00	.098	.049	.010	.099	.049	.009	.096	.046	.007
	10	11.768							0.50	.186	.108	.030	.188	.113	.029	.212	.130	.036
	10	13.536							1.00	.408	.289	.114	.426	.307	.129	.479	.361	.172
	10	15.304							1.50	.681	.556	.307	.685	.575	.336	.729	.627	.409
	10	17.072							2.00	.876	.795	.573	.877	.798	.596	.877	.818	.655
	10	18.840							2.50	.967	.934	.803	.964	.933	.807	.952	.918	.820
	10	20.608							3.00	.995	.987	.937	.991	.980	.929	.979	.967	.913
k = 4																		
4	10	10	10	10					0.00	.097	.048	.010	.083	.041	.010	.076	.035	.008
	10	12.5	15.0	17.5					0.56	.190	.110	.028	.204	.117	.032	.265	.158	.049
	10	15.0	20.0	25.0					1.12	.468	.324	.120	.544	.400	.172	.631	.513	.266
	10	17.5	25.0	32.5					1.68	.793	.666	.348	.811	.711	.453	.840	.770	.574
	10	20.0	30.0	40.0					2.24	.958	.902	.669	.936	.885	.714	.931	.887	.769
	10	22.5	35.0	47.5					2.80	.997	.985	.892	.980	.960	.867	.967	.949	.876
8	10	10	10	10					0.00	.096	.047	.009	.086	.040	.007	.076	.034	.006
	10	11.768	13.536	15.304					0.56	.206	.119	.034	.224	.129	.034	.251	.153	.047
	10	13.536	17.072	20.608					1.12	.528	.386	.165	.574	.441	.216	.640	.519	.296
	10	15.304	20.608	25.912					1.68	.851	.753	.496	.854	.770	.555	.868	.808	.637
	10	17.072	24.144	31.216					2.24	.977	.952	.828	.965	.935	.822	.945	.913	.821
16	10	10	10	10					0.00	.095	.048	.011	.099	.047	.009	.085	.037	.006
	10	11.25	12.50	13.75					0.56	.212	.129	.036	.224	.136	.039	.248	.152	.047
	10	12.50	15.0	17.50					1.12	.550	.416	.194	.574	.449	.224	.626	.510	.288
	10	13.75	17.50	21.25					1.68	.870	.786	.559	.869	.764	.590	.879	.817	.647
	10	15.0	20.0	25.0					2.24	.985	.966	.879	.977	.956	.867	.964	.941	.871
32	10	10	10	10					0.00	.102	.051	.010	.092	.045	.007	.092	.044	.008
	10	10.884	11.768	12.652					0.56	.221	.131	.041	.217	.131	.036	.244	.145	.040
	10	11.768	13.536	15.304					1.12	.561	.434	.211	.580	.454	.218	.613	.491	.271
	10	12.652	15.304	17.956					1.68	.882	.804	.584	.878	.804	.607	.882	.816	.645
	10	13.536	17.072	20.608					2.24	.988	.970	.901	.984	.967	.888	.970	.952	.884
k = 8																		
4	10	10	10	10	10	10	10	10	0.00	.100	.053	.013	.086	.044	.009	.079	.040	.009
	10	11.222	12.444	13.666	14.888	16.110	17.332	18.554	0.56	.221	.135	.040	.240	.141	.041	.282	.175	.059
	10	12.444	14.888	17.332	19.776	22.220	24.667	27.108	1.12	.621	.475	.217	.927	.538	.283	.726	.613	.384
	10	13.666	17.332	20.998	24.664	28.330	31.996	35.662	1.68	.941	.878	.647	.988	.868	.687	.920	.876	.741
	10	14.888	19.776	24.664	29.552	34.440	39.328	44.216	2.24	.998	.993	.946	.999	.976	.915	.977	.960	.904

Appendix B

DISTRIBUTION PARAMETERS AND POWER IN THE CASE OF UNEQUAL WITHIN CELL VARIANCE

	Parameters						Power								
							Normal			Exponential			Lognormal		
n	μ_1	μ_2	μ_3	μ_4	ϕ	σ^2_ϵ	.10	.05	.01	.10	.05	.01	.10	.05	.01
k = 2															
16	10	12.5			.44	128.25	.156	.090	.024	.157	.084	.015	.165	.086	.017
	10	15.0			.79	162.50	.289	.186	.063	.287	.176	.042	.324	.200	.054
	10	20.0			1.27	250.00	.548	.414	.185	.582	.415	.153	.653	.498	.201
	10	25.0			1.58	362.50	.706	.586	.317	.789	.643	.297	.846	.726	.405
	10	30.0			1.79	500.00	.802	.692	.432	.891	.786	.443	.935	.856	.572
	10	35.0			1.94	662.50	.846	.760	.512	.948	.872	.566	.972	.924	.703
k = 4															
4	10	12.5	15.0	17.5	.40	196.88	.143	.080	.020	.115	.056	.014	.123	.063	.014
	10	15.0	20.0	25.0	.61	337.50	.209	.122	.032	.162	.089	.022	.205	.112	.029
	10	20.0	30.0	40.0	.82	750.00	.300	.188	.057	.251	.141	.039	.330	.201	.058
	10	25.0	40.0	55.0	.92	1337.50	.351	.228	.074	.313	.181	.053	.420	.265	.085
	10	35.0	60.0	85.0	1.01	3037.50	.405	.274	.045	.388	.235	.073	.523	.348	.123
16	10	11.25	12.50	13.75	.47	142.97	.177	.104	.027	.166	.092	.022	.173	.096	.024
	10	12.50	15.00	17.50	.80	196.88	.338	.221	.079	.318	.204	.064	.354	.225	.073
	10	13.75	17.50	21.25	1.04	261.72	.492	.362	.162	.485	.338	.127	.546	.342	.158
	10	15.00	20.00	25.00	1.22	337.50	.621	.484	.247	.633	.475	.211	.697	.549	.266
	10	17.5	25.0	32.5	1.45	521.88	.773	.664	.406	.827	.698	.379	.874	.761	.480
	10	20.0	30.0	40.0	1.63	750.00	.857	.762	.527	.921	.827	.531	.941	.876	.636
	10	22.5	35.0	47.5	1.75	1021.88	.897	.822	.611	.963	.902	.641	.970	.925	.737

Appendix C*

DERIVATION OF VARIANCE MST AND COVARIANCE MST, MSE

In this Appendix, the variance of MST and the covariance of MST, MSE are derived for any distribution for which the first two moments exist and are finite. MST and MSE are, respectively, the numerator and denominator of the F-test. The following derivation is for a single classification analysis of variance based on k cells and n observations per cell.

The derivations are based on the assumption that $\mu = 0$. This does not result in any loss of generality, and simply represents a shift in the distribution along the x axis. The r^{th} moment about the mean is defined as $\mu_r = E\left[(x-\mu)^r\right]$. Under the assumption $\mu = 0$, $\mu_r = E(x^r)$.

VARIANCE MST

The variance of MST (s_T^2) is defined as,

(1) $$Var(s_T^2) = E\left[(s_T^2)^2\right] - \left[E(s_T^2)\right]^2.$$

Under H_0, $E(s_T^2) = \sigma^2$ and

(2) $$Var(s_T^2) = E\left[(s_T^2)^2\right] - \sigma^4.$$

---

*For a general theory of the method used in this Appendix, see Cramér, (1951, Chapters 15 and 27).

Define $\bar{G} = \frac{1}{k}\sum_{j}^{k}\bar{x}_j$, then

(3)
$$E\left[\left(s_T^2\right)^2\right] = \left(\frac{n}{k-1}\right)^2 E\left[\left(\sum_{j}^{k}\bar{x}_j^2 - k\bar{G}^2\right)^2\right]$$

$$= \left(\frac{n}{k-1}\right)^2\left\{E\left[\left(\sum_{j}^{k}\bar{x}_j^2\right)^2\right] - 2kE\left(\bar{G}^2\sum_{j}^{k}\bar{x}_j^2\right) + k^2E\left(\bar{G}^4\right)\right\}.$$

Each of the terms inside the brackets will be evaluated separately.

TERM 1 (EQUATION 3)

(4)
$$E\left[\left(\sum_{j}^{k}\bar{x}_j^2\right)^2\right] = E\left[\sum_{j}^{k}\bar{x}_j^2\sum_{j}^{k}\bar{x}_j^2\right] = \sum_{j}^{k}\sum_{\ell}^{k}E\left(\bar{x}_j^2\,\bar{x}_\ell^2\right)$$

$$= \underset{(j\neq\ell)}{\sum_{j}^{k}\sum_{\ell}^{k}}E\left(\bar{x}_j^2\right)E\left(\bar{x}_\ell^2\right) + \underset{(j=\ell)}{\sum_{j}^{k}}E\left(\bar{x}_j^4\right) \equiv Y + Z.$$

This follows from the mutual independence of the cell means whenever $j \neq \ell$. The final two terms in Eq. (4) will, for the moment, simply be identified as Y and Z. The parenthetic expressions under the summation sign refer to the subscript values of the original equation.

TERM 2 (EQUATION 3)

(5)
$$2kE\left(\bar{G}^2\sum_{j}^{k}\bar{x}_j^2\right) = \frac{2}{k}\sum_{j}^{k}\sum_{\ell}^{k}\sum_{p}^{k}E\left(\bar{x}_j\bar{x}_\ell\bar{x}_p^2\right)$$

$$= \frac{2}{k}\sum_{\substack{j \\ (j=\ell\neq p)}}^{k}\sum_{p}^{k} E\left(\bar{x}_j^2\right)E\left(\bar{x}_p^2\right) + \frac{2}{k}\sum_{\substack{j \\ (j=\ell=p)}}^{k} E\left(\bar{x}_j^4\right) = \frac{2}{k}\left(Y + Z\right),$$

since whenever $j \neq \ell$, $E(x)$ o[illegible] and this is zero.

TERM 3 (EQUATION 3)

(6) $$k^2 E\left(\bar{G}^4\right) = \frac{1}{k^2}\sum_{j}^{k}\sum_{\ell}^{k}\sum_{p}^{k}\sum_{q}^{k} E\left(\bar{x}_j\bar{x}_\ell\bar{x}_p\bar{x}_q\right)$$

$$= \frac{3}{k^2}\sum_{\substack{j \\ (j=\ell\neq p=q,\text{etc.})}}^{k}\sum_{\ell}^{k} E\left(\bar{x}_j^2\right)E\left(\bar{x}_p^2\right) + \frac{1}{k^2}\sum_{\substack{j \\ (j=l=p=q)}}^{k} E\left(\bar{x}_j^4\right) = \frac{1}{k^2}\left(3Y + Z\right).$$

This follows from $E(x) = \mu = 0$, and that the three conditions defined by $j = \ell \neq p = q$, $j = p \neq \ell = q$ and $j = q \neq \ell = p$ all lead to $\bar{x}_j^2\ \bar{x}_j^2$. Equation 4 may now be written

(7) $$E\left[\left(s_T^2\right)^2\right] = \frac{n^2}{k^2(n-1)^2}\left[Y(k^2-2k+3) + Z(k-1)^2\right].$$

The terms Y and Z will now be evaluated. From the basic definition of moments

(8) $$E(x^2) = \mu_2 = \sigma^2,$$

(9) $$E(\bar{x}^2) = \frac{\mu_2}{n} = \frac{\sigma^2}{n}.$$

Therefore,

(10) $$Y = \sum_{j}^{k}\sum_{\substack{\ell \\ j\neq\ell}}^{k} E\left(\bar{x}_j^2\right)E\left(\bar{x}_\ell\right)^2 = \sum_{j}^{k}\sum_{\substack{\ell \\ j\neq\ell}}^{k}\left(\frac{\sigma^2}{n}\right)^2 = \frac{k(k-1)}{n^2}\sigma^4.$$

(11) $$Z = \sum_{j}^{k} E\left(\bar{x}_j^4\right) = \frac{1}{n^4}\sum_{j}^{k} E\left(\sum_{i}^{n} x_{ij}\right)^4$$

$$= \frac{1}{n^4}\sum_{j}^{k}\sum_{i}^{n}\sum_{u}^{n}\sum_{v}^{n}\sum_{w}^{n} E(x_{ij}x_{uj}x_{vj}x_{wj})$$

$$= \frac{k}{n^3}\left[\mu_4 + 3(n-1)\sigma^4\right],$$

where μ_4 is the fourth moment about the mean. The result follows from a straightforward expansion under the summation with $E(x) = 0$ and $\mu_2^2 = \sigma^4$. See Cramér (1951, p. 348).

By definition $\mu_4 = (\gamma_2 + 3n)\sigma^4$, and

(12) $$Z = \frac{k\sigma^4}{n^3}(\gamma_2 + 3n).$$

Substituting these values of Y and Z in Eq. (10), and simplifying

(13) $$E\left[\left(s_T^2\right)^2\right] = \frac{\sigma^4}{kn(k-1)}\left[nk^2 + nk + \gamma_2(k-1)\right].$$

Finally, from Eq. (2)

(14) $$\mathrm{Var}\left(s_T^2\right) = \frac{2\sigma^4}{k(n-1)}\left[1 + \tfrac{1}{2}\gamma_2\frac{(n-1)}{n}\right].$$

COVARIANCE (MST,MSE)

The covariance of MST, MSE (s_T^2, s_e^2) is defined as

$$\text{(15)} \qquad \operatorname{Cov}(s_T^2, s_e^2) = E\left\{\left[s_T^2 - E(s_T^2)\right]\left[s_e^2 - E(s_e^2)\right]\right\}$$

$$= E(s_T^2 s_e^2) - \sigma^4.$$

Define $s_e^2 = \frac{1}{k}\sum_j^k s_j^2$, then

$$\text{(16)} \qquad E(s_T^2 s_e^2) = \frac{n}{k(k-1)} E\left[\left(\sum_j^k \bar{x}_j^2 - k\bar{G}^2\right)\sum_j^k s_j^2\right]$$

$$= \frac{n}{k(k-1)}\left[E\left(\sum_j^k \bar{x}_j^2 \sum_j^k s_j^2\right) - kE\left(\bar{G}^2\sum_j^k s_j^2\right)\right].$$

The terms inside the brackets of Eq. (16) will be evaluated separately.

TERM 1 (EQUATION 16)

$$\text{(17)} \qquad E\sum_j^k \bar{x}_j^2 \sum_j^k s_j^2 = \sum_j^k \sum_{\ell}^k E(\bar{x}_j^2 s_\ell^2) = \sum_{\substack{j \\ (j=\ell)}}^k E(\bar{x}_j^2 s_j^2)$$

$$+ \sum_{\substack{j \\ (j\neq\ell)}}^k \sum_{\ell}^k E(\bar{x}_j^2)E(s_\ell^2) \equiv U + V,$$

where U and V are used for the moment to identify the two terms on the right side of Eq. (17).

TERM 2 (EQUATION 16)

(18)
$$kE\left(\bar{G}^2\sum_j^k s_j^2\right) = \frac{1}{k}\sum_j^k\sum_{\ell}^k\sum_p^k E\left(\bar{x}_j\bar{x}_{\ell}s_p^2\right)$$

$$= \frac{1}{k}\sum_{\substack{j \\ (j=\ell=p)}}^k E(\bar{x}_j^2 s_j^2) + \frac{1}{k}\sum_{\substack{j \\ (j=\ell\neq p)}}^k\sum_p^k E(\bar{x}_j^2 s_p^2) \equiv \frac{1}{k}(U+V),$$

since all terms of $j \neq \ell$ involve $E(x) = 0$.

Substituting these values in Eq. (16),

(19)
$$E(s_T^2 s_e^2) = \frac{n}{k^2}(U+V).$$

In order to evaluate U and V, write

$$s_j^2 = \frac{1}{n-1}\left(\sum_i^n x_{ij}^2 - n\bar{x}_j^2\right),$$

and it follows that

(20)
$$U \equiv \frac{1}{n-1}\sum_j^k E\left[\bar{x}_j^2\left(\sum_i^n x_{ij}^2 - n\bar{x}_j^2\right)\right]$$

$$= \frac{1}{n-1}\left[\sum_j^k E\left(\bar{x}_j^2\sum_i^n x_{ij}^2\right) - n\sum_j^k E(\bar{x}_j^4)\right].$$

The first term inside the brackets may be written

(21) $$\sum_{j}^{k} E\left(\bar{x}_j^2 \sum_{i}^{n} x_{ij}^2\right) = \frac{1}{n^2}\sum_{j}^{k} E\left[\left(\sum_{i}^{n} x_{ij}\right)^2 \sum_{i}^{n} x_{ij}^2\right]$$

$$= \frac{1}{n^2}\sum_{j}^{k}\sum_{i}^{n}\sum_{u}^{n}\sum_{v}^{n} E(x_{ij}x_{uj}x_{vj}^2)$$

$$= \frac{1}{n^2}\sum_{j}^{k}\left[\sum_{\substack{i \\ (i=u=v)}}^{n} E(x_{ij}^4) + \sum_{\substack{i \\ (i=u\neq v)}}^{n}\sum_{v}^{n} E(x_{ij}^2)E(x_{vj}^2)\right]$$

$$= \frac{1}{n^2}\sum_{j}^{k}\left[n\mu_4 + n(n-1)\sigma^4\right]$$

$$= \frac{k}{n}\left[\mu_4 + (n-1)\sigma^4\right].$$

This result follows from application of $E(x_i^4) = \mu_4$ and $E(x_i^2) = \sigma^2$.

The second term inside the brackets of Eq. (20) is nZ (see Eq. 11). Therefore,

(22) $$U = \frac{1}{n-1}\left\{\sum \frac{k}{n}\left[\mu_4 + (n-1)\sigma^4\right] - \frac{k}{n^2}\left[\mu_4 + 3(n-1)\sigma^4\right]\right\}$$

$$= \frac{k}{n^2}\left[\mu_4 + \sigma^4(n-3)\right];$$

(23) $$V = \sum_{\substack{j \\ (j\neq \ell)}}^{k}\sum_{\ell}^{k} E(\bar{x}_j^2)E(s_\ell^2) = \sum_{\substack{j \\ (j\neq \ell)}}^{k}\sum_{\ell}^{k} \frac{\sigma^2}{n}\sigma^2 = \frac{k(k-1)}{n}\sigma^4.$$

Substituting these values of U and V in Eq. (24) and simplifying,

(24) $$E(s_T^2 s_e^2) = \frac{1}{k}\left[\frac{\mu_4 + \sigma^4(n-3)}{n} + (k-1)\sigma^4\right].$$

Recalling that $\mu_4 = (\gamma_2 + 3)\sigma^4$, and substituting Eq. (24) into (15),

(25) $$\mathrm{Cov}(s_T^2, s_e^2) = \frac{\gamma_2 \sigma^4}{nk}.$$

REFERENCES

1. Aitchison, J., and J.A.C. Brown, The Lognormal Distribution, Cambridge University Press, London, 1957.

2. Atiquallah, M., "The Estimation of Residual Variance in Quadratically Balanced Least-Squares Problems and the Robustness of the F-test," Biometrika, 49, 1962, pp. 83-91.

3. Bradley, J. V., Studies in Research Methodology IV: A Sampling Study of the Central Limit Theorem and the Robustness of One-Sample Parametric Tests, AMRL Technical Document Report, 6570th Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio, 1963, pp. 63-69.

4. -----, Studies in Research Methodology VI: The Central Limit Effect for a Variety of Populations and the Robustness of Z, t, and F, AMRL Technical Document Report, Aerospace Medical Research Laboratories, Wright-Patterson Air Force Base, Ohio, 1964, pp. 64-123.

5. Box, G.E.P., and S. L. Andersen, "Permutation Theory in the Derivation of Robust Criteria and the Study of Departures from Assumptions," J. Roy. Stat. Soc. B., 17, 1955, pp. 1-34.

6. -----, and M. E. Muller, "A Note on the Generation of Normal Deviates," Ann. Math. Stat., 29, 1958, pp. 610-611.

7. -----, and G. C. Tiao, "A Further Look at Robustness via Bayes's Theorem," Biometrika, 49, 1962, pp. 419-432.

8. -----, and G. S. Watson, "Robustness to Non-Normality of Regression Tests," Biometrika, 49, 1962, pp. 93-106.

9. Cramér, Harold, Mathematical Methods of Statistics, Princeton University Press, Princeton, New Jersey, 1951.

10. Fox, M., "Charts of the Powers of the F-tests," Ann. Math. Stat., 27, 1956, pp. 484-497.

11. Hansen, H. H., W. N. Hurwitz, and W. G. Madow, Sample Survey Methods and Theory, John Wiley and Sons, New York, 1953.

12. Horsnell, G., "The Effect of Unequal Group Variance on the F-test for the Homogeneity of Group Means," Biometrika, 40, 1953, pp. 128-136.

13. Kempthorne, Oscar, The Design and Analysis of Experiments, John Wiley and Sons, New York, 1952.

14. Pearson, E. S., and H. O. Hartley, "Charts of the Power Function of the Analysis of Variance Tests, Derived from the Non-central F-Distribution," Biometrika, 38, 1951, pp. 112-130.

15. Pratt, J. W., "Robustness of Some Procedures for the Two-Sample Location Problem," J. Am. Stat. Assoc., 59, 1964, pp. 665-680.

16. Scheffé, Henry, The Analysis of Variance, John Wiley and Sons, New York, 1959.

17. Srivastava, A.B.L., "Effect of Non-Normality on the Power Function of t-Test," Biometrika, 45, 1958, pp. 421-430.

18. Winer, B. J., Statistical Principles in Experimental Design, McGraw-Hill, New York, 1962.

DOCUMENT CONTROL DATA

1 ORIGINATING ACTIVITY THE RAND CORPORATION	2a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b. GROUP

3. REPORT TITLE

POWER OF THE F-TEST FOR NONNORMAL DISTRIBUTIONS AND UNEQUAL ERROR VARIANCES

4. AUTHOR(S) (Last name, first name, initial)

Donaldson, T. S.

5. REPORT DATE September 1966	6a. TOTAL No. OF PAGES 67	6b. No. OF REFS. 18
7. CONTRACT OR GRANT No. AF 49(638)-1700	8. ORIGINATOR'S REPORT No. RM-5072-PR	
9a. AVAILABILITY/LIMITATION NOTICES DDC-1	9b. SPONSORING AGENCY United States Air Force Project RAND	

10. ABSTRACT

An empirical investigation of the power of the F-test when the underlying distribution is nonnormal and the within-cell variances are not homogeneous. It is shown that even for small samples the analysis-of-variance F-test is valid, although the usual assumptions of normal distribution and homogeneous error variances are seriously violated. The test is shown to be valid for both Type I and Type II error. An analysis of the correlation between the numerator and denominator of the null hypothesis F-ratio demonstrates that the distribution of F is little affected by the shape of the underlying distribution.

11. KEY WORDS

Statistical methods and processes
Variational theory
Distribution
Logistics
Maintenance
Laboratory Problem IV