The accuracy of four approximations to noncentral F

DRAKE R. BRADLEY and RONALD L. RUSSELL

Bates College, Lewiston, Maine

and

CHARLES P. REEVE

Westinghouse Savannah River Company, Aiken, South Carolina

The accuracy of the two-moment, three-moment, square root, and cube root approximations to the noncentral F distribution was assessed using 7,920 entries from Tiku's (1967) power tables. Tiku's tables list exact values of β for $\alpha = .005-.05$, $v_1 = 1-12$, $v_2 = 2-120$, and $\phi = 0.5-3.0$. Analysis of the errors showed generally satisfactory performance for all four approximations. The three-moment approximation was most accurate, registering a maximum error of only .009. The other three approximations had maximum errors of $\pm .02$, except for the square root approximation at $v_2 = 2$, where maximum errors of .05 occur. Approximation error increased with decreases in v_1 and, less consistently, with increases in v_2 . Error was nonmonotonically related to ϕ . A second investigation explored the accuracy of the approximations at values of α ranging from .10 to .90. All four approximations degraded substantially in this situation, with maximum errors ranging from $-.09$ to .05. If the analysis is restricted to cases where $v_1 > 1$ and $v_2 > 2$, maximum errors drop to roughly $\pm .03$. We conclude that the approximations perform reasonably well for small α and moderately well for larger values, if certain restrictions are imposed. From a computational standpoint, however, there is little advantage to using approximate as opposed to exact methods unless exact values of F_α are known in advance.

The last several decades have seen a growing interest in issues surrounding statistical power, ignited in large part by Jacob Cohen's book *Statistical Power Analysis for the Behavioral Sciences* (1969, 1977, 1988). Cohen's book was partly a response to an earlier power survey (Cohen, 1962), which showed that the average power of studies published in the 1960 issue of the *Journal of Abnormal and Social Psychology* was only .48 for detecting medium-size effects. Subsequent research has demonstrated similar low power levels for studies published in other journals and disciplines (see, e.g., Brewer, 1972; Chase & Chase, 1976; Chase & Tucker, 1975, 1976; Katzer & Sadt, 1973; Rossi, 1982, 1986, 1990; Sedlmeier & Gigerenzer, 1989). Inadequate attention to power has motivated the development of both general purpose and specialized programs for doing power analysis. A sampling of such programs is reviewed by Goldstein (1989) and by Thomas and Krebs (1997). Some recent additions include software described by Meyer (1995), Erdfelder, Faul, and Buchner (1996), and Bradley, Russell, and Reeve (1996).¹

A number of approximations to the noncentral F distribution have been developed to facilitate the computation of power for tests employing the F statistic. Illustrations may be found in Johnson and Kotz (1970), Winer (1971), Guenther (1979), and Winer, Brown, and Michels (1991). Approximations have been used in the construction of power tables (see, e.g., Cohen, 1969, 1977, 1988; Rotton & Schönemann, 1978) and to perform power and sample size calculations in power analysis programs (see, e.g., Borenstein & Cohen, 1988; Bradley, 1988; Erdfelder et al., 1996; Parker & Borich, 1980).² Given the increasing reliance on approximations, it is necessary to revisit the question of the relative accuracy of approximate and exact methods for computing power. We first describe the analytical basis for each approximation and then review previous research assessing the accuracy of the approximations. Finally, we report the results of our own comprehensive tests of approximation accuracy.

The Noncentral F Distribution

The random variable $F = (x_1/v_1)/(x_2/v_2)$ is the ratio of two independent chi-square variables, x_1 and x_2 , each divided by its degrees of freedom, v_1 and v_2 , respectively. The numerator chi-square is distributed as $\chi_1^2(v_1, \lambda_1)$ and the denominator chi-square as $\chi_2^2(v_2, \lambda_2)$. With respect to the noncentrality parameter, x_1 is a central chi-square variable if $\lambda_1 = 0$ and a noncentral chi-square variable if $\lambda_1 > 0$. The same distinction holds for x_2 , although in practice one rarely encounters applications where $\lambda_2 > 0$ (but see Scheffé, 1959, pp. 134-136). The distributions of x_1 ,

Investigations 1 and 2 in this paper were conducted by R.L.R. in partial fulfillment of the requirements for a senior honors thesis at Bates College (advisor, D.R.B.). The authors thank Martin Richter (Lehigh University) and Melinda Harder (Bates College) for helpful comments and suggestions on an initial draft of the paper. Any errors that remain are the responsibility of the authors. Reprint requests should be sent to D. R. Bradley, Department of Psychology, Bates College, Lewiston, ME 04240 (e-mail: dbradley@abacus.bates.edu).

x_2 , and F are related as follows. If x_1 and x_2 are central chi-square variables, F follows a central F distribution, $F(v_1, v_2)$. If x_1 and x_2 are distributed as noncentral and central chi-square variables, respectively, then F follows a singly (topside) noncentral F distribution, $F(v_1, v_2; \lambda)$. In this context, λ is understood to be the noncentrality parameter for x_1 . Finally, if both x_1 and x_2 are noncentral chi-square variables, F follows a doubly noncentral F distribution, $F(v_1, v_2; \lambda_1, \lambda_2)$. We follow the convention of denoting central, singly noncentral, and doubly noncentral F variables as F , F' , and F'' , where

$$F = \frac{\chi_1^2/v_1}{\chi_2^2/v_2}, F' = \frac{\chi_1'^2/v_1}{\chi_2^2/v_2}, \text{ and } F'' = \frac{\chi_1'^2/v_1}{\chi_2'^2/v_2}.$$

Likewise, we denote the distributions of these variables as the F distribution, F' distribution, and F'' distribution, respectively. (The central and noncentral chi-squares are denoted χ_1^2 and $\chi_1'^2$, respectively.) No special designation is reserved for the case where $\lambda_1 = 0$ and $\lambda_2 > 0$, as this corresponds to the reciprocal of F' (Johnson & Kotz, 1970, p. 189).

In computing power, the value of F that cuts off α in the upper tail of the F distribution, $F_\alpha(v_1, v_2)$, is referenced to an F' or F'' distribution, as appropriate, and the upper tail area ($1 - \beta$) determined. In most cases, the reference distribution is a singly noncentral F distribution with parameters v_1 , v_2 , and λ . For example, in a one-way analysis of variance (ANOVA) with k groups and n observations per group, the noncentrality parameter is

$$\lambda = \frac{n \sum_{j=1}^k (\mu_j - \mu)^2}{\sigma_e^2} = \frac{n \sum_{j=1}^k \tau_j^2}{\sigma_e^2}.$$

To simplify table organization, most tables and charts of the power of the F test employ the following alternative measure of noncentrality:

$$\phi = \sqrt{\frac{n \sum_{j=1}^k \tau_j^2}{k \sigma_e^2}} = \sqrt{\frac{\lambda}{k}}.$$

The researcher enters the power table (or chart) with the parameters α , v_1 , v_2 , and ϕ and reads the value of $1 - \beta$. Interpolation is generally required if precise estimates of power are desired. Note that some tables present $1 - \beta$ as a function of sample size, n , and effect size, f , where

$$f = \sqrt{\frac{\sum_{j=1}^k \tau_j^2}{k \sigma_e^2}} = \frac{\phi}{\sqrt{n}} = \sqrt{\frac{\lambda}{nk}}.$$

In this case, power is determined by entering the tables with α , v_1 , n , and f . Most researchers find such tables easier to use because n and f relate to the structure of the research design in a more direct way than do v_2 and ϕ .³

Below, we illustrate the computation of power using an exact method and four different approximations. To fa-

cilitate comparison, it will be helpful to develop an example that can be used with each method. Suppose a researcher is interested in computing the power of the F test for a one-way ANOVA. The researcher assumes the following: $k = 4$, $n = 16$, $\sigma^2 = 225$, $\mu_1 = 41$, $\mu_2 = 47$, $\mu_3 = 53$, $\mu_4 = 59$, and $\mu = 50$. In this case, $\Sigma \tau^2 = (-9)^2 + (-3)^2 + (3)^2 + 9^2 = 180$, and the noncentrality parameters are $\lambda = 16(180)/225 = 12.8$ and $\phi = \sqrt{12.8/4} = 1.79$. Power tables in Tiku (1967, p. 535) and power charts in Pearson and Hartley (1976, p. 252) show that for $\alpha = .05$, $v_1 = k - 1 = 3$, $v_2 = k(n - 1) = 60$, and $\phi = 1.79$, power is $1 - \beta \approx .84$. Effect size tables in Cohen (1988, p. 315) produce the same result for $\alpha = .05$, $v_1 = 3$, $n = 16$, and $f = 1.79/\sqrt{16} = 0.447$. The power solution is depicted in Figure 1, which plots the probability density functions of the corresponding F and F' distributions. The critical value, $F_\alpha(3, 60) = 2.758$, cuts off $\alpha = .05$ in the upper tail of the F distribution and $1 - \beta \approx .84$ in the upper tail of the F' distribution.

Tabled values of $1 - \beta$ are based on exact solutions to the integral of F' . A series representation of the cumulative distribution function of F (Reeve, 1986, p. 1) can be used to find exact areas under the F , F' , and F'' distributions:

$$F_Y(x) = \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} A_i B_j I(u, \frac{v_1}{2} + i, \frac{v_2}{2} + j). \quad (1)$$

$F_Y(x)$ is the area to the left of $F = x$ in the distribution, $u = v_1 x / (v_1 x + v_2)$ is the value of F transformed to a beta variable, and $I(u, a, b)$ is the distribution function of the beta variable (also called the incomplete beta ratio). The Poisson probabilities, A_i and B_j , are given by

$$A_i = \frac{\left(\frac{\lambda_1}{2}\right)^i e^{-\lambda_1/2}}{\Gamma(i+1)}, \text{ and } B_j = \frac{\left(\frac{\lambda_2}{2}\right)^j e^{-\lambda_2/2}}{\Gamma(j+1)},$$

where λ_1 and λ_2 are the noncentrality parameters for the numerator and denominator chi-square variables of the F ratio. Since these expressions can be evaluated for $\lambda_1 \geq 0$ and $\lambda_2 \geq 0$, the series representation in Equation 1 can be used to compute the cumulative distribution function (CDF) of F , F' , or F'' .

Evaluation of Equation 1 requires subroutines to calculate the gamma function $\Gamma()$ and the incomplete beta ratio $I()$. Indices for summing the Poisson probabilities also have to be determined since, for computational purposes, the infinite series must be truncated. FORTRAN subroutines written by Reeve (1986) while at the National Bureau of Standards (now the National Institute of Science and Technology) were designed to do this. Reeve notes that, by selecting values for I' , I'' , J' , and J'' that result in

$$\sum_{i=I'}^{I''} A_i \geq 1 - \frac{\epsilon}{2} \text{ and } \sum_{j=J'}^{J''} B_j \geq 1 - \frac{\epsilon}{2},$$

ϵ serves as an absolute error bound on $F_Y(x)$, with ϵ being specified by the researcher as the maximum tolerable error in the solution. The CDF of F , F' , or F'' can therefore be computed with an accuracy that is limited only by the

Figure 1. Probability density functions for central ($\lambda = 0$) and noncentral ($\lambda = 12.8$) F distributions with 3 and 60 df . The power of the F test is shown on the basis of the $\alpha = .05$ critical value, $F_\alpha(3,60) = 2.758$.

available machine precision. To reduce computation time, recurrence relations (Abramowitz & Stegun, 1964, p. 944) are used to limit the number of direct evaluations of the beta CDF. These algorithms may be used to evaluate Equation 1 for the example in Figure 1. The parameters input to Reeve's (1986) program are $F' = F_\alpha = 2.758$, $v_1 = 3$, $v_2 = 60$, $\lambda_1 = 12.8$, and $\lambda_2 = 0$. The exact solution output by the program, rounded to four places, is $P(F' \leq 2.758) = \beta = .1601$. The power is therefore $1 - \beta = .8399$. This figure agrees with the value obtained earlier by interpolation in tables and charts of noncentral F .

Reeve's (1986) algorithms yield up to 12 digits of accuracy in the computed CDF of F' . However, converging on a solution for $\epsilon \leq 10^{-12}$ can require significant amounts of processing and memory, particularly for large values of λ_1 and λ_2 . Approximations to F' are computationally simpler and potentially more efficient in code size and execution time than exact solutions.⁴ These advantages must be weighed against the loss in accuracy that accompanies the use of an approximation. In order to make an informed decision regarding this tradeoff, it is necessary to consider both the overall accuracy and the worst-case performance of the approximation. Below, we review four commonly employed approximations to F' and assess the accuracy of each with respect to the example in Figure 1.

Approximations to Noncentral F

Algorithms for approximating the CDF of F' generally fall into one of two categories. Moment approximations transform the F' distribution into a central F distribution (F^*) with equivalent (lower order) moments, thus allowing the use of published tables, charts, or computer sub-

outines to obtain the desired areas. Normal approximations are based on the fact that certain functions of the numerator and denominator χ^2 variables transform F' to a standard normal deviate, z^* . Areas are determined by referencing tables of the standard normal distribution. For example, to compute power, $F' = F_\alpha$ is transformed to either F^* or z^* , and the area to the right is obtained from tables of the central F or standard normal distribution. The CDF of F' is obtained by transforming F' to F^* or z^* and finding the area to the left of the transformed value. Examples of moment and normal approximations are considered next.

Two-moment approximation. Patnaik (1949, p. 221) developed an approximation based on fitting an F distribution with the exact first two moments of the F' distribution of interest. The first and second moments of F and F' are given by:

$$m_1 = \frac{v_2}{v_2 - 2},$$

$$m_2 = \frac{v_2^2(v_1^2 + 2v_1)}{(v_2 - 2)(v_2 - 4)v_1^2},$$

$$m'_1 = \frac{v_2(v_1 + \lambda)}{(v_2 - 2)v_1},$$

$$m'_2 = \frac{v_2^2}{(v_2 - 2)(v_2 - 4)v_1^2} \left[(v_1 + \lambda)^2 + 2(v_1 + 2\lambda) \right].$$

Comparing m_1 and m'_1 shows that the noncentral F distribution is shifted to the right of a corresponding central F distribution by a factor of $k = (v_1 + \lambda)/v_1$. If we wish to employ a central F distribution to approximate F' , the transformation F'/k scales the distribution accordingly.

Note that k is obtained by equating the first moments as follows:

$$\frac{m'_1}{k} = m_1; \frac{v_2(v_1 + \lambda)}{k(v_2 - 2)v_1} = \frac{v_2}{v_2 - 2}.$$

Solving for k verifies that $k = (v_1 + \lambda)/v_1$. To equate the second moments, a modified numerator df , v_1^* , is selected such that:

$$\frac{m'_2}{k^2} = m_2;$$

$$\frac{v_2^2}{k^2(v_2 - 2)(v_2 - 4)v_1^2} \left[(v_1 + \lambda)^2 + 2(v_1 + 2\lambda) \right]$$

$$= \frac{v_2^2(v_1^{*2} + 2v_1^*)}{(v_2 - 2)(v_2 - 4)v_1^{*2}}.$$

Substituting for k and solving for v_1^* gives $v_1^* = (v_1 + \lambda)^2/(v_1 + 2\lambda)$. Hence, the two-moment approximation approximates the noncentral F distribution $F(v_1, v_2; \lambda)$ by a central F distribution, $F^*(v_1^*, v_2)$, where

$$\left\{ F^* = \frac{F'}{k}; k = \frac{(v_1 + \lambda)}{v_1}; v_1^* = \frac{(v_1 + \lambda)^2}{v_1 + 2\lambda} \right\}. \quad (2)$$

For the example in Figure 1 ($v_1 = 3$, $v_2 = 60$, $\lambda = 12.8$) the scaling constant is $k = (3 + 12.8)/3 = 5.2667$ and the modified numerator df is $v_1^* = (3 + 12.8)^2/[3 + 2(12.8)] = 8.728$. To obtain the power of the test, we transform the critical value $F' = F_\alpha = 2.758$ to $F^* = F'/k = 2.758/5.2667 = 0.5237$. Power is the probability that $F^*(8.728, 60) \geq 0.5237$. Using an algorithm from O'Grady (1981) to calculate the upper tail area of central F , we find $1 - \beta = .8472$. The difference between this and the exact value (of .8399) obtained using Equation 1 is approximation error.

Three-moment approximation. Tiku (1965, p. 425) developed an approximation based on fitting an F distribution with the exact first three moments of F' . A modified numerator df , v_1^* , is employed in conjunction with the transformation $F^* = (F' + c)/h$. Equating the first three moments, the equations for F^* , v_1^* , h , and c are found to be

$$\left\{ \begin{aligned} F^* &= \frac{F' + c}{h}, v_1^* = \frac{(v_2 - 2)}{2} \left(\sqrt{\frac{E}{E - 4}} - 1 \right), \\ h &= \frac{v_1^*}{v_1} \left(\frac{1}{2v_1^* + v_2 - 2} \right) \frac{H}{K}, c = \frac{v_2}{(v_2 - 2)} \left(h - \frac{v_1 + \lambda}{v_1} \right) \end{aligned} \right\}, \quad (3)$$

where

$$H = 2(v_1 + \lambda)^3 + 3(v_1 + \lambda)(v_1 + 2\lambda)(v_2 - 2) + (v_1 + 3\lambda)(v_2 - 2)^2,$$

$$K = (v_1 + \lambda)^2 + (v_2 - 2)(v_1 + 2\lambda),$$

and

$$E = \frac{H^2}{K^3}.$$

As above, the noncentral F distribution $F(v_1, v_2; \lambda)$ is approximated by a central F distribution, $F^*(v_1^*, v_2)$. For the example in Figure 1, $v_1^* = 14.01$, $h = 6.423$, and $c = 1.196$. To compute power, we transform $F' = F_\alpha = 2.758$ to $F^* = (F' + c)/h = (2.758 + 1.196)/6.423 = 0.6156$. Referencing a central F distribution, the probability that $F(14.01, 60) \geq 0.6156$ is $1 - \beta = .8417$. This value is somewhat more accurate than the solution provided by the two-moment approximation. This follows from the fact that the three-moment approximation takes into account the skew of the F' distribution.

Square root normal approximation. Patnaik (1949, p. 208) notes that Fisher's transformation may be used to approximate a central chi-square distribution, $\chi^2(v)$. For $v > 30$, the transformed variable $\sqrt{2\chi^2}$ is distributed as a normal variate with mean $\sqrt{2(v - 1)}$ and unit variance. Patnaik (p. 209) developed a similar approximation for the noncentral chi-square distribution, $\chi'^2(v)$, by showing that the transformed variable $\sqrt{2\chi'^2}$ approaches a normal distribution with mean $\sqrt{2(v + \lambda) - (v + 2\lambda)/(v + \lambda)}$ and variance $(v + 2\lambda)/(v + \lambda)$. Laubscher (1960, p. 1109) used these results to develop a square root approximation for the noncentral F distribution. Let $Y = \sqrt{2\chi_1'^2}$ and $X = \sqrt{2\chi_2'^2}$ be the normalizing transformations for the noncentral and central chi-square variables of an F' ratio with v_1 and v_2 df . Then the variable $V = Y/X$ is the ratio of two independent normal random variables. A theorem due to Fieller (1932) states that, if X and Y are normally and independently distributed with means μ_X and μ_Y and standard deviations σ_X and σ_Y , then

$$z^* = \frac{(\mu_X V - \mu_Y)}{(\sigma_X^2 V^2 + \sigma_Y^2)^{1/2}}$$

is approximately normally distributed with zero mean and unit variance. Note that

$$V = \frac{Y}{X} = \frac{\sqrt{2\chi_1'^2}}{\sqrt{2\chi_2'^2}} = \sqrt{\frac{\chi_1'^2}{\chi_2'^2}} = \sqrt{\frac{v_1(\chi_1'^2/v_1)}{v_2(\chi_2'^2/v_2)}} = \sqrt{\frac{v_1 F'}{v_2}}.$$

Substituting for V , μ_X , μ_Y , σ_X^2 , and σ_Y^2 in the equation for z^* gives Laubscher's square root approximation:

$$z^* = \frac{\left((2v_2 - 1) \frac{v_1 F'}{v_2} \right)^{1/2} - \left(2(v_1 + \lambda) - \frac{v_1 + 2\lambda}{v_1 + \lambda} \right)^{1/2}}{\left(\frac{v_1 F'}{v_2} + \frac{v_1 + 2\lambda}{v_1 + \lambda} \right)^{1/2}}. \quad (4)$$

Equation 4 transforms F' to a standard normal deviate, and a table of the standard normal distribution is employed to determine the area to the left or right of z^* . For the example in Figure 1, $F' = F_\alpha = 2.758$ converts to $z^* = -1.008$. An algorithm from Abramowitz and Stegun (1964, p. 932, Equation 26.2.17) shows that the area to the right of this z is $1 - \beta = .8433$.

Cube root normal approximation. Severo and Zelen (1960, p. 416) and Laubscher (1960, p. 1111) derive a cube root approximation using $Y = [\chi_1'^2/(v_1 + \lambda)]^{1/3}$ and

Table 1
Results of Early Investigations of the Accuracy of Several Approximations to the
Noncentral F Distribution: Mean Absolute Error (MAE), Maximum Positive Error (MPE),
Maximum Negative Error (MNE), and Number of Cases Investigated (N)

	Laubscher (1960)			Tiku (1965)		Tiku (1966)		
	cosh ⁻¹	Square Root	Cube Root	Two-Moment	Three-Moment	Cube Root	Two-Moment	Three-Moment
<i>MAE</i>	.02133	.00217	.00283	.00267	.00038	.002867	.000973	.000141
<i>MPE</i>	.078	.008	.007	.007	.000	.0092	.0078	.0013
<i>MNE</i>	-.020	-.005	-.007	-.007	-.001	-.0141	-.0100	-.0017
<i>N</i>	24	24	24	24	24	144	144	120

$X = (\chi_2^2/v_2)^{1/3}$ as the normalizing transformations for the noncentral and central chi-square variables of the F' ratio. Following the same general procedure as the derivation of the square root approximation, Laubscher shows that

$$z^* = \frac{\left(1 - \frac{2}{9v_2}\right) \left(\frac{v_1 F'}{v_1 + \lambda}\right)^{\frac{1}{3}} - \left(1 - \frac{2(v_1 + 2\lambda)}{9(v_1 + \lambda)^2}\right)}{\left[\left(\frac{2}{9v_2}\right) \left(\frac{v_1 F'}{v_1 + \lambda}\right)^{\frac{2}{3}} + \left(\frac{2(v_1 + 2\lambda)}{9(v_1 + \lambda)^2}\right)\right]^{\frac{1}{2}}} \quad (5)$$

is approximately normally distributed with zero mean and unit variance. With respect to Figure 1, $F' = F_\alpha = 2.758$ transforms to $z^* = -1.0273$, and the corresponding power is $1 - \beta = .8479$.

Other approximations. The four approximations reviewed above yield errors (approximate - exact power) of .0073, .0018, .0034, and .0080 for the example in Figure 1. The three-moment approximation is most accurate, followed by the square root, two-moment, and cube root approximations. Other approximations to noncentral F have been proposed, such as the cosh⁻¹ (Laubscher, 1960), log (Barton, David, & O'Neill, 1960), Pearson S_u (Tiku, 1966), Tiku series expansion (Tiku, 1966), and four-moment approximations (Tiku & Yip, 1977). (See also Johnson & Kotz, 1970, pp. 194-196.) These approximations are either less accurate or more difficult to compute than the approximations reviewed above and will not be considered further.

Research on the Accuracy of Approximations to F'

Patnaik (1949) assessed the accuracy of the two-moment approximation at $\alpha = \{.01, .05\}$ for the following: $v_1 = 3$, $v_2 = \{10, 20\}$, $\lambda = \{4, 16\}$; $v_1 = 5$, $v_2 = \{10, 20\}$, $\lambda = \{6, 24\}$; and $v_1 = 8$, $v_2 = \{10, 30\}$, $\lambda = \{9, 36\}$. The resulting 24 combinations of α , v_1 , v_2 , and λ were also used by Laubscher (1960) to compare the accuracy of the cosh⁻¹, square root, and cube root approximations, by Severo and Zelen (1960) to investigate the cube root approximation, and by Tiku (1965) to compare the two-moment and three-moment approximations. In a subsequent investigation, Tiku (1966) compared the accuracy of the two-moment, three-moment, and cube root approximations (as well as Pearson's S_u approximation and Tiku's series expansion). Comparisons were made for the 144 combinations ob-

tained by crossing $\alpha = \{.01, .05\}$, $v_1 = \{3, 8, 15, 30\}$, $v_2 = \{2, 6, 12, 20, 30, 60\}$, and $\phi = \{0.5, 1.0, 2.0\}$. Patnaik (1949), Laubscher (1960), Severo and Zelen (1960), and Tiku (1965) computed solutions to three places, whereas Tiku (1966) computed solutions to four places. Each author displayed the approximation errors (approximate - exact) in tables, but no summary statistics were reported. To facilitate comparison, we determined the mean absolute error (MAE), maximum positive error (MPE), and maximum negative error (MNE) for each of the approximations investigated in the studies. The results are presented in Table 1. Note that Severo and Zelen's (1960) results for the cube root approximation are not included in the table because these same values are available from Laubscher. Likewise, Patnaik's error values for the two-moment approximation are not included, because these values were also computed by Tiku (1965) and our checks show that Tiku's solutions are more accurate. Finally, although Tiku (1965) included six additional cases in his table (for a total of $N = 30$), these are not analyzed in Table 1 in order to permit comparison with Laubscher's results.⁵

The mean absolute errors in Table 1 reflect the same rank order as demonstrated for the example in Figure 1. Taken together, the Laubscher (1960) and Tiku (1965) studies show that the order (from most to least accurate) is: (1) three-moment, (2) square root, (3) two-moment, (4) cube root, and (5) cosh⁻¹. (The cosh⁻¹ approximation is included to illustrate a relatively poor approximation.) Tiku's (1966) more extensive investigation confirms the relative accuracy of the three-moment, two-moment, and cube root approximations. Examination of the MAE , MPE , and MNE trends shows that the three-moment approximation is substantially more accurate than the two-moment, square root, and cube root approximations. Not surprisingly, Tiku's (1966) study, which investigated a larger number of cases ($N = 144$ vs. $N = 24$) over a wider range of parameter values (v_1 , v_2 , ϕ), found larger maximum errors of approximation. This suggests that the accuracy of the approximations should be examined over a much wider range of parameter values than has been done so far, in order to assess the largest errors that can arise.

Although permitting comparisons of overall accuracy, Table 1 does not show the relationship between the errors of approximation and the parameters α , v_1 , v_2 , and λ (or ϕ). Examination of the error trends in Tiku's tables (1966, pp. 608-609) shows that the two-moment and three-moment approximations decrease in accuracy with

decreases in v_1 and, less consistently, with increases in v_2 . The trends for the cube root approximation are not well defined, although there is a substantial drop in accuracy at very small values of v_2 . For all three approximations, errors tend to be larger at $\alpha = .05$ than at $\alpha = .01$ and to be nonmonotonically related to ϕ . Subsequent work by Pearson and Tiku (1970, Table 1, p. 178) on the three-moment approximation examined the differences in the moment ratio, $\beta_2 = \mu_4/\mu_2^2$, of the true and approximated distributions for $v_1 = \{1(1)6(2)12, 15, \infty\}$, $v_2 = 40$, and $\phi = \{0.5, 1(1)3\}$. Since the three-moment approximation only ensures that the moment ratio $\beta_1 = \mu_3^2/\mu_2^3$ is correct, differences in β_2 provide an index of the approximation error caused by fitting only the first three (as opposed to four) moments. Pearson and Tiku's results show that the values of β_2 for the true and approximated distributions converge as v_1 increases, and this accounts for the increasing accuracy of the three-moment approximation with increases in v_1 (Tiku, 1966). The relationship between β_2 error and ϕ is nonmonotonic, with error rising to a peak at $\phi \approx 1.0$ and then falling again.⁶

Woodward and Overall (1976) investigated the accuracy of the cube root and two-moment approximations. Power values for the cube root approximation were assessed at $\alpha = \{.01, .05\}$ for the 2,090 combinations of $v_1 = \{2(2)30, 40, 60, 120, \infty\}$, $v_2 = \{1(1)10, 12\}$, and $\phi = \{0.5, 1(0.2)2.2, 2.6, 3.0\}$. Exact values were obtained from Tiku (1967). The mean absolute error of the cube root approximation was .0033. This value is larger than those shown in Table 1 for the Laubscher (1960) and Tiku (1966) studies, which once again reinforces the need to examine a broad range of parameter values. For the two-moment approximation, the same 24 cases investigated by Patnaik (1949) were used. The mean absolute error of approximation was .0027, and the authors note that this compares to a value of .0025 for Patnaik's original data. (Note that six of Patnaik's approximation errors disagree with those computed by Tiku [1965], and the latter source yields a mean absolute error [see Table 1] of .00267.) Woodward and Overall do not report maximum positive and negative approximation errors, measures of dispersion, or other information that would indicate the performance of the approximations in the worst-case scenario. There is also no attempt to indicate the relationships between approximation error and α , v_1 , v_2 , and ϕ . Finally, although this investigation does extend quite considerably the range of parameters investigated, the use of only two α levels continues a limitation inherent in all previous studies.

Rotton and Schönemann (1978) employed a combination of the three-moment and cube root approximations to compute power for a fairly extensive set of power tables. Where comparisons with exact values were possible, they report a maximum error of $-.013$, and a mean absolute error of .003. The latter value is consistent with the value reported by Woodward and Overall (1976) for the cube root approximation.

In his original research on the topic, Laubscher (1960) expressed some surprise that the square root approximation, which is based on Fisher's transformation, $\sqrt{(2\chi^2)}$, proved to be more accurate than the cube root approximation, which is based on the superior Wilson-Hilferty transformation, $(\chi^2/v)^{1/3}$. Fowler (1984) suggested that the greater accuracy of the square root approximation was due to the specific parameter combinations selected by Laubscher. In particular, the values of λ used by Laubscher correspond to effect sizes that are much larger than those typically encountered in the social sciences. The smallest effect size in Laubscher's set was $f = .408$, which constitutes a "large" effect by Cohen's (1988) criteria. In an attempt to sample more representative values, Fowler compared the accuracy of the square root and cube root approximations at $\alpha = \{.50, .30, .20, .10, .05, .01\}$, $v_1 = \{1(1)6, 8, 12, 24\}$, $v_2 = \{6(2)30, 40, 60, 120, 200\}$, and $f = \{0.1, 0.25, 0.4\}$. The results indicated that the cube root approximation was more accurate than the square root approximation at conventional α levels (.01 and .05), except for large effect sizes and large v_2 . The cube root approximation was also superior, with certain exceptions, at $\alpha = \{.10, .20, .30, .50\}$.

Cohen and Nee (1987) compared the accuracy of the square root and cube root approximations for parameter values representative of those arising when computing the power of Rao's F test on $R^2_{Y,X}$ in set correlation analysis. The authors note that set correlation applications are generally associated with larger values of λ and v_2 than are multiple regression and analysis of variance applications. Also, since $v_1 = k_Y k_X$ (where k_Y and k_X are the number of variables in each set), the numerator df for the F test will often exceed those provided in power tables. Tiku's tables (1967), for example, only go to $v_1 = 12$. For these reasons, Cohen and Nee felt that the question of the relative accuracy of the square root and cube root approximations should be considered anew. The cases examined by Cohen and Nee were based on generating samples of $n = \{30, 60, 120, 240\}$ observations from 97 correlation matrices having different values of $R^2_{Y,X}$ and different partitions of k_Y and k_X . The Rao F test was evaluated at $\alpha = \{.01, .05\}$. These parametric variations gave rise to 521 distinct cases, with a mean $R^2_{Y,X}$ of .1679 and a mean f^2 of 0.0914. For each case, power was estimated by empirical simulation ($N = 1,000$ datasets each) and then compared to the values produced by the square root and cube root approximations. The results showed that the MAE for the square root and cube root approximations was .0103 and .0105, respectively. These values seem large in comparison to those reported in Table 1 and by other researchers. However, since the approximation errors are influenced by sampling error in the empirical power estimates on which the differences are based, as well as by the inherent error in the approximations, they overestimate the true approximation error. In fact, the sampling error in the empirical proportion of significant F s is probably larger than the inherent error

in the approximations. If the proportion of significant F s is $p = .5$, for example, the standard error of the proportion is $S_p = \sqrt{p(1-p)/N} = \sqrt{.5(1-.5)/1,000} = .0158$. Or, to consider a less extreme example (in terms of variability), if $p = .05$, $S_p = .0069$. In either case, the sampling error in the empirical power estimate may very well exceed the inherent error in the square root and cube root approximations. Hence, assessments of the accuracy of the two, based on empirical power estimates, are of limited utility.

Fortunately, Cohen and Nee (1987) were able to use Tiku's (1967) power tables to determine exact power for a subset of 205 of the 521 cases (those where $v_1 = 6$). The mean absolute error of the square root and cube root approximations for this subset was .0025 and .0052, respectively.⁷ The substantially smaller MAEs obtained for the subset suggest that the results based on empirical power estimates (above) should be regarded with caution. Note that the subset MAEs are generally in line with those in Table 1 and with the MAE of $\approx .003$ reported by Woodward and Overall (1976) and Rotton and Schönmann (1978) for the cube root approximation. However, the higher MAE obtained for the cube root approximation in the Cohen and Nee study (.0052) suggests that the performance of this approximation does degrade a bit for the larger values of λ and v_2 that characterize set correlation analyses.

Several observations follow from the preceding research review. First, no researcher has explicitly compared the accuracy of all four approximations (two-moment, three-moment, square root, cube root) in the same study. Tentative inferences as to the relative accuracy of the four approximations must therefore be made across studies that employ different ranges of parameters or that investigate only a limited number of cases. Second, previous research indicates that, as the number of cases and the range of parameter values increases, so does the maximum approximation error. To determine the worst-case performance of the approximations requires a larger scale study than has been attempted so far. Third, many researchers have avoided testing the algorithms at very small values of v_2 , such as $v_2 = \{2, 4\}$, either because such values are considered unrealistic, or because the mathematical derivation of the approximation implies that it is not valid for these cases. However, if an approximation is computable (in a purely mechanical sense) for small values of v_2 , we would argue that it is important to assess its accuracy at these values as well (even though adequate mathematical justification may be lacking). Finally, the large majority of studies have examined the performance of the approximations only at the $\alpha = .01$ and $.05$ levels of significance. A few have examined values of α out to $.50$ but no further. This oversight is unfortunate, because one potential use of the approximations is to compute the CDF of F' over its entire range, including values in the lower tail of the F' distribution. The accuracy of the approximations for lower tail values of F' can be evaluated if large values of α (such as $\alpha = .90$) are included in the study.

This will produce small (lower tail) values of F'_α , and these, in turn, will fall in the far left tail of the F' distribution. If the accuracy of the approximation degrades substantially in this situation, the use of the approximation may have to be limited to power (or upper tail CDF) computations.

The purpose of the two investigations reported below is to compare the accuracy of the two-moment, three-moment, square root, and cube root approximations over a wide range of parameter values, including large values of α . This will permit assessment of the worst-case performance of the approximations. On the basis of previous research (Tiku, 1965, 1966), we anticipate that the three-moment approximation will prove to be the most accurate approximation. In addition, we expect that the square root and cube root approximations will yield fairly similar levels of overall accuracy. From Tiku's (1966) work, we would also predict that approximation accuracy will decrease systematically with decreases in v_1 , at least for the two-moment and three-moment approximations. A weak tendency for accuracy to decrease with increases in v_2 is also expected, with the exception that the square root and cube root approximations should perform poorly at very small values of v_2 . Finally, approximation error should vary nonmonotonically with increases in ϕ . This follows from Pearson and Tiku's (1970) research on the three-moment approximation and from the trends in Tiku's (1966) data for the two-moment and cube root approximations.

METHOD

Two separate investigations were conducted. Investigation 1 determined the accuracy of the two-moment, three-moment, square root, and cube root approximations for values of α that are commonly employed in significance testing: $\alpha = .005-.05$. Investigation 2 explored the accuracy of these same approximations for $\alpha = .10-.90$.

Investigation 1: $\alpha = .005-.05$

The specific α values selected for Investigation 1, as well as the values chosen for v_1 , v_2 , and ϕ , were dictated by the availability of tables of exact values of β (or $1 - \beta$) against which to compare the estimates provided by the approximations. Although several sources are available (Lachenbruch, 1966; Lehmer, 1944; Patnaik, 1949; Tang, 1938), the tables published by Tiku (1967, pp. 527-537) are probably the most extensive set of power tables compiled for the noncentral F distribution. Tiku computed Type II error probabilities (β) accurate to four places for the following parameters: $v_1 = \{1(1)10, 12\}$; $v_2 = \{2(2)30, 40, 60, 120\}$; $\alpha = \{.005, .01, .025, .05\}$; and $\phi = \{0.5, 1.0(0.2)2.2, 3.0\}$. This represents a total of $11 \times 18 \times 4 \times 10 = 7,920$ different values of β for use in testing the approximations. The large number of tabled values, as well as the four-place accuracy of the entries, were the primary considerations favoring the use of Tiku's (1967) tables as benchmarks.

Inspection of Equation 3 shows that division by zero will occur for the three-moment approximation when computing the constant c for a denominator df of $v_2 = 2$. In order to avoid this, and to avoid missing values in the final database, we decided to employ the two-moment approximation as a stand-in for the three-moment approximation whenever $v_2 = 2$. Consequently, the three-moment approximation investigated in the present study is a combination algorithm. Since the two-moment approximation is generally less accurate than the three-moment approximation (Table 1), this should serve to provide slightly conservative estimates of the overall accuracy of the latter. (The unique characteristics of the three-moment approximation can still be assessed, however, simply by examining plots of approximation error in which v_2 is included as a parameter in the plot.)

The first step in the present investigation required the entire contents of Tiku's (1967) tables to be manually entered into a computer database.⁸ Each of the 7,920 rows of the database stored the values of v_1 , v_2 , α , ϕ , and β for a particular combination. The order of entry followed Tiku's tables, cycling first through the levels of ϕ , and then through the levels of v_2 , v_1 , and α (in that order). While compiling the Tiku database, several apparent errors in the values of β were detected. For example, for $\alpha = .005$, $v_1 = 3$, and $\phi = 1.2$, the values in Tiku's table for $v_2 = 22, 24, 26$, and 28 are $\beta = .8714, .8867, .8825$, and $.8588$, respectively. Since β should decrease monotonically with increases in v_2 , something is clearly wrong with this sequence. Inspection of adjacent columns in the table, as well as checks using interpolation, suggested that the sequence should be $.8714, .8667, .8625$, and $.8588$. Apparently, a typesetting error occurred in the second digit of the middle two values of the sequence, producing an "8" rather than a "6." Another nonmonotonic pattern occurred for $\alpha = .01$, $v_1 = 3$, and $\phi = 1.8$: The values in Tiku's table for $v_2 = 8, 10$, and 12 are $\beta = .6835, .5234$, and $.5776$, respectively. Interpolation shows that the middle value in this sequence should be $\beta = .6234$. Careful examination of Tiku's table revealed two other nonmonotonic trends in the values of β , each of which was consistent with the hypothesis of a typesetting error: either of a single digit being mistyped (e.g., 8 vs. 9), or of left-right inversions of adjacent digits (e.g., 86 vs. 68). If such errors were to remain uncorrected, "spurious" errors of approximation would arise, biasing the results. Consequently, the "correct" values of β —that is, those obtained assuming a typesetting error and verified using interpolation—were substituted for the original (incorrect) values in the Tiku database.

To assess the accuracy of the four algorithms, the statistical package DATASIM (Bradley, 1988, 1991b) was modified to compute power using the two-moment, three-moment, square root, and cube root approximations.⁹ Each approximation requires that the critical value, F_{α} , be passed as an input parameter to the corresponding subroutine. In addition, the two-moment and three-moment approximations require computation of p -levels for central F , and the square root and cube root ap-

proximations require p -levels for the standard normal distribution. Since DATASIM employs highly accurate algorithms (O'Grady, 1981) for computing critical values and p -levels of the F distribution, as well as an algorithm (Abramowitz & Stegun, 1964, p. 932, Equation 26.2.17) for computing the upper tail area of z , it was not necessary to code these subroutines separately. A customized command, POWF, was added to DATASIM that permitted the computation of power in Investigation 1 to be done in batch mode—that is, by reading the parameters from an input file and writing the results to an output file. A modified version of the Tiku database, which replaced the values of ϕ with the corresponding values of λ , served as the input file for the POWF command.¹⁰ Entering POWF, followed by the name of the input file, caused DATASIM to do the following: (1) read the values of v_1 , v_2 , α , λ , and β from the first row of the input file; (2) compute the exact power, $1 - \beta_{\text{Tiku}}$, from β ; (3) compute the approximate power for each of the four algorithms; (4) concatenate the input and output values and write the resulting string to an output file; and (5) repeat this procedure until all 7,920 rows of the input file had been processed. Using the relation $\phi = \sqrt{[\lambda/(v_1 + 1)]}$, values of ϕ were computed and inserted into the output file between α and λ . (This was done so that subsequent analyses and plots could be performed as a function of either ϕ or λ .) After this, each row of the output file contained values for the following: v_1 , v_2 , α , ϕ , λ , β , $1 - \beta_{\text{Tiku}}$, $1 - \beta_{2\text{-mom}}$, $1 - \beta_{3\text{-mom}}$, $1 - \beta_{\text{sqr}}$, $1 - \beta_{\text{cube}}$.

Finally, the output file was loaded into DATASIM, and the differences between the approximate power and Tiku's exact power were computed for each of the four algorithms. The difference scores were stored in four additional columns joined to the right side of the database, and the results were saved in a file. This file constituted the *augmented* database—that is, the original database from Tiku, plus the approximate power and difference columns computed by DATASIM. The entire database contained $7,920 \times 15 = 118,800$ entries and required approximately 1 MB of storage on disk.

In preliminary analyses of the data, the approximation errors were plotted as a function of v_1 , v_2 , α , ϕ , or λ . It was found that the errors varied systematically with the values of the parameters. However, there were several cases where one or more data points plotted as outliers with respect to otherwise systematic trends. Consider Figure 2, which plots the errors of approximation for the three-moment algorithm as a function of v_1 . A number of outliers are apparent, such as those at $v_1 = 4$ and $v_1 = 7$, which represent exceptions to the trend for errors to decrease systematically with increases in v_1 . When we examined the associated values of β in Tiku's tables, they appeared to be incorrect. Comparing the "suspect" values of β to values obtained by interpolation showed that the former were not properly spaced within the interval created by the adjacent values in the table. For example, for the combination $\alpha = .025$, $v_1 = 7$, and $\phi = 1.8$, the tabled values for $v_2 = 8, 10$, and 12 are $\beta = .4430, .3655$, and

Figure 2. The 7,920 errors of approximation for the three-moment approximation plotted as a function of v_1 . The outliers in the plot suggest possible errors in Tiku's (1967) exact values of β .

.2949, respectively. Interpolation reveals that the middle value in the sequence should be about .3565, which suggests an inversion error for the middle two digits. This error represents a difference of $.3655 - .3565 = .009$, and accounts for the outlier seen at $v_1 = 7$ in Figure 2.

The newly discovered errors in the Tiku database were corrected, the approximation errors recomputed, and the corrected (augmented) database once again subjected to a preliminary analysis. Although the previous outliers were now no longer present in the plots, the greater detail resulting from the reduced range on the y-axis revealed new (previously undetected) outliers. It soon became apparent that, for each set of outliers corrected, additional outliers were sure to appear in subsequent (rescaled) plots of the errors of approximation. These recurring errors in the values of β brought the integrity of Tiku's tables into question. It was not possible to continue Investigation 1 until the integrity of the database could be ensured.¹¹ This necessitated a brief digression from Investigation 1, as detailed next.

Given questions about the integrity of Tiku's tables, an attempt was made to obtain additional tables of β for use in correcting the "suspect" values in Tiku. Cohen (1988, p. 551) refers to an unpublished computer print-out entitled *Tables of Power Points of Analysis of Variance Tests*, developed at the National Bureau of Standards (now the National Institute of Standards and Technology). Queries to this organization succeeded in locating the needed tables, but, since they tabled the percentile points (x) of the noncentral F rather than β , they were not useful for checking Tiku's tables. As it turned out, however, an even

better solution was available. In 1986, Charles Reeve of the Statistical Engineering Division of NBS published an internal report (N-86-4) describing a program for computing the cumulative distribution function of the doubly noncentral F to a specified accuracy (Reeve, 1986). The algorithms employed by the program, described in the introduction, are capable of computing β accurate to 12 digits. A copy of the program was obtained from the author and used to compute the 7,920 values of β in Tiku's tables. Once the accuracy of the program had been verified, the values of β generated by the program were substituted for those values of β in Tiku's tables that appeared to be inaccurate on the basis of outlier analysis (and confirmed by interpolation). The specific substitutions are shown in Table 2: The column labeled β_{Tiku} shows the original (incorrect) values from Tiku's tables, the column

Table 2
Corrections of Typesetting Errors
in the Original Tiku (1967) Database

v_1	v_2	α	ϕ	β_{Tiku}	β_{Reeve}
1	60	.005	1.0	.9240	.9249
3	24	.005	1.2	.8867	.8667
3	26	.005	1.2	.8825	.8625
12	10	.005	1.0	.9430	.9439
3	10	.010	1.8	.5234	.6234
6	20	.025	0.5	.9378	.9387
6	22	.025	2.0	.0932	.0943
7	10	.025	1.8	.3655	.3565
9	20	.025	1.8	.1133	.1333
8	4	.050	0.5	.9274	.9247
9	28	.050	1.6	.1110	.1100
12	30	.050	3.0	.9000	.9000

Table 3
Overall Accuracy of the Approximations for Investigation 1 ($\alpha = .005-.05$) and Investigation 2 ($\alpha = .005-.90$): Mean Error (ME), Mean Absolute Error (MAE), Maximum Positive Error (MPE), and Maximum Negative Error (MNE)

Investigation		Approximation			
		Two-Moment	Three-Moment	Square Root	Cube Root
$\alpha = .005-.05$ ($N = 7,920$)	ME	-.000690	.000197	.002733	-.000329
	MAE	.002537	.000398	.004856	.003132
	MPE	.018295	.008531	.053673	.018574
	MNE	-.018677	-.003050	-.021881	-.018302
$\alpha = .005-.90$ ($N = 25,740$)	ME	.002199	-.000354	.000992	.002143
	MAE	.003355	.001289	.003272	.003414
	MPE	.054433	.046029	.053673	.047479
	MNE	-.018677	-.089215	-.089724	-.018302
$\alpha = .005-.90$ $v_1 > 1, v_2 > 2$ ($N = 22,100$)	ME	.001378	-.000143	.000278	.001339
	MAE	.002344	.000584	.001990	.002346
	MPE	.024864	.010008	.025131	.023191
	MNE	-.013788	-.019507	-.036083	-.013652

labeled β_{Reeve} shows the exact values computed by Reeve's program that were substituted in the database, and the underlined digits in the two columns identify the presumed typesetting errors.

The present investigators confess to having some concern that a table of exact values has errors. Before being critical of Tiku (1967), however, note that the 12 typesetting errors documented in Table 2 represent an error rate of *only* 1.5 errors per 1,000 entries, which is very low. This error rate should be considered in the context of the present investigators' performance: Since the computer database was compiled by manually entering the 7,920 values of β from Tiku's tables, we can use the data entry errors made when creating that database as a baseline for comparison.¹² Coincidentally, we discovered 12 data entry errors in our own database. The errors showed the same patterns as those illustrated in Table 2—that is, single-digit errors and two-digit inversions. Tiku (1972) subsequently published corrections to his original power tables, but only 5 of the 12 errors noted in Table 2 were corrected in that publication: those in rows 1, 4, 7, 8, and 11.

The typesetting corrections and data entry corrections noted above resulted in a final version of the Tiku database that is, one hopes, error free.¹³ Investigation 1 was conducted one last time, using the corrected database as the input file. The POWF command produced an updated output file, which was subsequently augmented with difference columns, as described above.

Investigation 2: $\alpha = .10-.90$

The fortuitous discovery of Reeve's (1986) program made possible Investigation 2, since it was now possible to generate exact values of power to compare with those produced by the four approximations.¹⁴ The combinations of parameters investigated in this case were: $v_1 = \{1(1)10, 12\}$; $v_2 = \{2(2)30, 40, 60, 120\}$; $\alpha = \{.10(.10).90\}$; and $\phi = \{0.5, 1.0(0.2)2.2, 3.0\}$. Except for α , these are the same parameters as those employed in Investigation 1. Creating a Tiku-style database for the input file for Investigation 2 was simplified by the existence of the

Tiku database for Investigation 1. The combinations of parameters for $\alpha = .005$ (v_1, v_2, ϕ) were copied into a new file, and the values for α changed from .005 to .10, using the search/change capability of a word processor. This process was repeated eight more times, and the values of α edited accordingly (.20-.90). In addition to saving time, this approach ensured that no data entry errors would occur in creating the new database.

The next step was to append the exact values of β to this database. Reeve (1986) developed a specially customized version of his program that could read the distribution parameters from an input file, compute the exact solution, and write the results to an output file.¹⁵ DATASIM's POWF command was used to read this file and produce another output file with the four approximate powers appended to the right side. That file was in turn augmented by adding the four difference columns. The augmented database for Investigation 2 therefore had precisely the same structure as that of Investigation 1. For each of the nine α levels, there were $11 \times 18 \times 10 = 1,980$ combinations of v_1, v_2 , and ϕ . Consequently, the database contained $9 \times 1,980 = 17,820$ rows overall and a total of $17,820 \times 15 = 267,300$ entries. This occupied approximately 2 MB of storage on disk.

Since the *maximum* error of approximation over the entire range of α levels was of primary interest, it was decided to make Investigation 2 inclusive of the results of Investigation 1. This was accomplished by joining the two augmented databases into one master database ($\alpha = .005-.90$). The resulting file contained 25,740 rows and 386,100 entries and occupied approximately 3 MB of disk storage.

RESULTS

As in previous studies, the basic unit of analysis was the difference between the approximate and exact power (approximate - exact). Four measures of accuracy were computed, using the difference scores. The first measure was simply the mean error (ME) of the approximation.

Figure 3. Error magnitude as a function of α and approximation method.

This measure reveals the presence of bias in the estimates—that is, whether the estimates are consistently larger (positive bias) or smaller (negative bias) than the exact values. The second measure, the *MAE*, provides an index of the magnitude of the typical error without regard to sign. This measure is useful for comparing the overall accuracy of the four approximations. The third and fourth measures were the *MPE* and the *MNE* of the approximation.¹⁶ The difference between these two measures gives the range ($R = MPE - MNE$). If the range of errors is small, the approximation is regarded as stable. As used here, stability refers to the performance of an algorithm in the worst-case scenario—that is, under that combination of parameters that produces the maximum errors of approximation. The most desirable approximation is one that has an *ME* near zero (unbiased), a small *MAE* (accurate), and an *MPE* and *MNE* that span a narrow range (stable).¹⁷

Table 3 presents the results of Investigation 1 ($\alpha = .005-.05$) and Investigation 2 ($\alpha = .005-.90$), taken over all possible cases for the four measures. Also shown are the results for Investigation 2 when the analysis is restricted to cases having $v_1 > 1$ and $v_2 > 2$. In Investigation 1, the three-moment approximation outperforms the other approximations on all four measures of accuracy.

At its worst, the three-moment approximation gives two-place accuracy, in comparison with only one-place accuracy for the remaining approximations.¹⁸ The two-moment and cube root approximations show very similar levels of accuracy, with a slight edge going to the former on the *MAE* measure. The last-place ranking for the square root approximation may seem surprising, given its second-place ranking in Table 1. The reversal is due to the inclusion of $v_2 = 2$ in the present study, which differentially affects the square root approximation (see below). All four approximations show an increase in *MAE*, *MPE*, and *MNE* relative to those in Table 1, which reflects the continuing degradation of the approximations as more extreme cases are added to the database.

When the database is augmented by the inclusion of large α levels (Investigation 2), the picture changes. Although the three-moment approximation is still the most accurate in terms of *MAE* (.0013), it differs less from the other approximations (*MAEs* $\approx .0033$) than in Investigation 1. The presence of serious degradation in the accuracy of the three-moment approximation at large values of α , which leads to an *MNE* of $-.0892$, is the reason. The square root algorithm shows the same degradation, with an *MNE* of $-.0897$. Large α levels lead to degradation in all four approximations on the *MPE* measure,

Figure 4. Error magnitude as a function of v_1 and approximation method.

with $MPEs$ ranging from .046 to .054. However, if the analysis is restricted to cases for which $v_1 > 1$ and $v_2 > 2$, the maximum errors now range from $-.020$ to .010 and from $-.036$ to .025 for the three-moment and square root approximations, respectively.

The question naturally arises as to the particular combinations of α , v_1 , v_2 , and ϕ that give rise to the MPE and MNE approximation errors in Table 1. Also of interest is the precise functional relationship between the values of the parameters and the accuracy of the approximations. These matters are reviewed in turn for each investigation.

Investigation 1: $\alpha = .005-.05$

The purpose of Investigation 1 was to assess the performance of the approximations at values of α typically encountered in significance testing. Figures 3–6 display the approximation errors as functions of α , v_1 , v_2 , and ϕ , respectively.¹⁹ Each figure consists of four panels that plot the ME (a), the MAE (b), the MPE (c), and the MNE (d) of the approximation. Within each panel, the curves with the open-circle, filled-circle, open-square, and filled-square data points plot the errors for the two-moment, three-moment, square root, and cube root approximations, respectively.²⁰ In panels a and b, each data point represents an average taken across all levels of the re-

maining parameters (other than the one plotted on the x-axis). In panels c and d, on the other hand, each data point represents the maximum error (positive or negative) that occurred across all levels of the remaining parameters.

The relationship between α and the accuracy of approximation is shown in Figure 3. The mean errors plotted in 3a show that the two-moment and cube root approximations are negatively biased, whereas the square root and three-moment approximations are positively biased. Bias decreases with increases in α , except for the three-moment approximation, which shows a very slight increase. The two-moment, three-moment, and cube root approximations show a mild increase in the $MAEs$ (3b) and $MPEs$ (3c) and a mild decrease in the $MNEs$ (3d), as α increases. The square root approximation shows somewhat stronger trends in the reverse direction: The $MAEs$ and $MPEs$ decrease as α increases, whereas the $MNEs$ increase. Overall, the three-moment approximation provides the most accurate estimates of power, with maximum errors of .0085 (3c) and $-.003$ (3d) across the four levels of α . The two-moment and cube root approximations both exhibit maximum positive and negative errors of roughly $\pm .018$. The square root approximation is least accurate, with maximum errors of .054 and $-.022$. These

Figure 5. Error magnitude as a function of v_2 and approximation method.

extreme errors are due to the algorithm's poor performance at $v_2 = 2$. If this condition is excluded, the *MPE* and *MNE* are .023 and -.020, respectively.²¹

Figure 4 shows that approximation error varies inversely with v_1 , as expected. Bias, absolute error, and maximum positive and negative error all decrease as v_1 increases. This is indicated by the error curves approaching $y = 0$ as the numerator *df* goes from $v_1 = 1$ to $v_1 = 12$. An exception is the square root approximation, which shows this trend only weakly. Figure 5 presents the relationship between approximation error and v_2 . The square root approximation shows large errors at small values of v_2 , and a similar but much milder pattern occurs in the cube root approximation. As v_2 increases, the square root and cube root approximation errors rapidly decrease in magnitude and then start to increase again, rising steadily. In contrast, the two-moment and three-moment approximation errors increase steadily throughout the range of v_2 . These trends are easily seen in the error functions for the *MPE* and *MNE* (5c and 5d). They are also present in Figure 5b, but the scale of the y-axis (caused by the large square root errors) obscures the trends. Finally, note that, for $v_2 > 8$, the square root approximation outperforms the two-moment and cube root approximations. The inferior performance of the square root approximation in Figures 3 and 4, relative to those of the two-moment and

cube root approximations, is therefore due to the large errors which occur in this approximation for $v_2 \leq 8$.²²

Figure 6 displays the errors of approximation as a function of ϕ . The trends indicate that the two-moment, three-moment, and cube root approximations are generally most accurate at low and high values of ϕ , with the errors fluctuating systematically throughout the middle range.²³ Although not shown here, the approximation errors can also be plotted as a function of λ . The individual trends are not as easily distinguished as those in Figure 6. This is because λ is a joint function of v_1 and ϕ . Each of the 10 unique values of ϕ in the database therefore gives rise to 11 different values of λ , 1 for each possible value of v_1 . Adjusting for duplications that occur when different pairs of v_1 and ϕ generate (coincidentally) the same value of λ , this produces a total of $10 \times 11 - 4 = 106$ unique values of λ for each error curve. The graphs therefore appear quite cluttered. Nevertheless, they show that the mean errors and mean absolute errors for all four approximations peak somewhere between $\lambda = 0$ and $\lambda = 20$ and decline thereafter. The *MPE* and *MNE* trends are less consistent and vary in form across the four approximations.

Figure 7 clarifies the joint relationship between v_1 , v_2 , and the *MAE* of approximation. Panels a–d plot the errors for the two-moment, three-moment, square root, and

Figure 6. Error magnitude as a function of ϕ and approximation method.

cube root approximations, respectively. Within each panel, errors are plotted as a function of v_1 on the x-axis and v_2 as the parameter defining the different curves. To minimize clutter, large data points are plotted only for the $v_2 = 2$ (open-circle) and $v_2 = 120$ (filled-circle) curves; small data points are employed for all curves with $4 \leq v_2 \leq 60$. The trends in Figure 7 confirm the earlier observations regarding the relationship between the errors of approximation and v_1 and v_2 . For the two-moment and three-moment approximations, MAE increases as v_1 decreases and v_2 increases. These trends are quite systematic and well behaved. The trends for the square root and cube root approximations, on the other hand, are less clear. Relatively large errors are present for small values of v_2 , especially $v_2 = 2$. These errors decrease with increases in v_2 up to some threshold value, after which the errors increase with additional increases in v_2 . A good example of this nonmonotonic relationship is seen in Figure 7c: As v_2 goes from $v_2 = 2$ to $v_2 = 10$, the heights of the error curves decrease steadily.²⁴ This trend subsequently reverses, with the heights of the error curves rising again up to that occupied by the $v_2 = 120$ curve. Although the latter increase seems much less than that observed for the two-moment approximation (7a), this is partly an artifact caused by the different scales employed

for the y-axes. For all four approximations, therefore, it is generally true that errors of approximation increase with decreases in v_1 and increases in v_2 . For the square root and cube root approximations, this generalization is qualified by the existence of relatively large errors at small values of v_2 .

Plots similar to those in Figure 7 were prepared for the MPE s and MNE s. The MPE trends were found to be quite similar to those in Figure 7, and the MNE trends were essentially the mirror image of the MPE trends (with some minor variations). A plot of the ME errors was also made, and the trends were generally consistent with those of Figure 7.

Investigation 2: $\alpha = .005-.90$

The purpose of Investigation 2 was to assess the performance of the approximations when lower tail values of $F' = F_\alpha$ were included for evaluation. Plots of the trends comparable to those in Figures 3-7 were constructed. To conserve space, we present only one of the figures here: Figure 8 plots the approximation errors as a function of α . Note that the figure includes the results of Investigation 1 ($\alpha = .005-.05$), as well as the new data collected for Investigation 2 ($\alpha = .10-.90$). It is evident that large α -levels cause substantial degradation in the

Figure 7. Error magnitude as a function of v_1 and v_2 for the two-moment (a), three-moment (b), square root (c), and cube root (d) approximations.

accuracy of all four approximations, at least for certain combinations of the parameters (see below). The three-moment and square root approximations both show *MNEs* of about $-.089$ at $\alpha = .90$, and all four approximations show *MPEs* of roughly $.045$ – $.055$ over the interval $\alpha = .4$ – $.8$. The three-moment approximation has the smallest bias and *MAE* until $\alpha = .60$ and the smallest *MPE* and *MNE* until $\alpha = .40$ and $\alpha = .20$, respectively. The accuracy of the three-moment approximation decreases steadily after $\alpha = .60$ on the *ME*, *MAE*, and *MNE* measures, until it becomes the least accurate of the four approximations at $\alpha = .90$.

The plot of the approximation error as a function of v_1 showed that bias, *MAE*, *MPE*, and *MNE* all decrease with increases in v_1 , as reflected in the error functions converging on $y = 0$. The only exception is the *MPEs* of the square root approximation, which fail to show improvement with increases in v_1 .²⁵ The plot also revealed that the large errors associated with the three-moment approximation at high values of α (e.g., $\alpha = .90$) are due to its poor performance at $v_1 = 1$. If this condition is excluded, the *MNE* drops from $-.089$ to $-.020$. A plot of the approximation error as a function of v_2 showed unusually large errors for the square root and cube root ap-

proximations at $v_2 = 2$ and $v_2 = 4$. For larger values of v_2 , the bias, the *MAE*, and the *MPE* increase with increases in v_2 for all four algorithms. The *MNEs* do not show consistent trends as a function of v_2 .

Additional plots showed that the square root approximation show a steady decrease in the *MEs*, *MAEs*, and *MNEs* as ϕ increases, whereas the *MPEs* remain relatively unchanged. For the other three approximations, there is a tendency for bias, *MAE*, and *MPE* to be greater at intermediate than at extreme values of ϕ . *MNEs* for the three-moment and square root approximations occur at $\phi = .5$. The fact that all of the error functions approach $y = 0$ as ϕ takes on increasingly large values suggests that the accuracy of the four approximations should continue to hold for $\phi > 3.0$. (The same conclusion is supported by graphs plotting the approximation errors as a function of λ .)

Figures plotting *ME*, *MAE*, *MPE*, and *MNE* as a joint function of v_1 and v_2 (as in Figure 7) were prepared. The *MPE* and *MNE* error plots show that, if $v_1 = 1$ is excluded from the analysis, the maximum errors drop to roughly $\pm .025$ (or less) for all but the square root approximation; for the latter, $v_1 = 2$ and $v_2 = 2$ must also be excluded to reach this level of accuracy. Marginally acceptable per-

Figure 8. Error magnitude as a function of $\alpha = .005-.90$ and approximation method.

formance for the approximations can therefore be attained out to $\alpha = .90$, provided certain restrictions are imposed.

Worst-case analysis. Examination of the various plots reveal that the three-moment and square root maximum negative errors ($MNEs \approx -.089$) occur at $\alpha = .90$, $v_1 = 1$, $v_2 = 4$, and $\phi = 0.5$. To further explore worst-case performance, approximation accuracy was assessed for this case at $\alpha = .90(.01)1.00$.²⁶ The results show that MNE increases steadily over this interval, with the three-moment and square root approximations giving values of $-.1543$ and $-.1318$, respectively, at $\alpha = 1.00$. It is possible that peaks in the maximum errors exist for values between those tested in Investigation 2 and/or for $\phi < 0.5$. To assess this possibility, we explored worst-case performance for $v_1 = 1$, $v_2 = 4$, and $\alpha = \{.90, 1.00\}$ at 93 levels of ϕ : specifically, $\phi = \{0, 0.001, 0.01, 0.02(0.02)1.0(0.05)3\}$.²⁷ For $\alpha = .90$, the three-moment approximation had a peak at $\phi = 0.68$ ($MNE = -.1003$), and the square root approximation at $\phi = 0.16$ ($MNE = -.1064$). For $\alpha = 1.00$, the three-moment peak was at $\phi = 0.60$ ($MNE = -.1591$), and the square root peak at $\phi = 0$ ($MNE = -.1587$). The MNE of $-.1591$ is the largest error we uncovered in our investigation. To see if the large $MNEs$ would decrease substantially for $v_1 > 1$, the same analyses were conducted again, but with $v_1 = 2$. The peak errors were as follows:

$-.0221$ ($\alpha = .90$, $\phi = 0.96$) and $-.0354$ ($\alpha = 1.00$, $\phi = 0.86$) for the three-moment approximation, and $-.0391$ ($\alpha = .90$, $\phi = 0.30$) and $-.0416$ ($\alpha = 1.00$, $\phi = 0$) for the square root approximation.

The reduction in error observed when $v_1 = 1$ is excluded, along with the improvement in the square root approximation when $v_2 = 2$ is excluded, suggested that the entire Investigation 2 database should be reanalyzed for the subset $\{v_1 > 1, v_2 > 2\}$. The overall results for this analysis were presented earlier in Table 3 and show that the maximum errors of approximation fall to roughly $\pm .03$ in this case.

Criterion analysis. The trends reported above are based on average or maximum values, both of which are influenced by the existence of extreme outliers in the errors of approximation. Since some of these outliers represent poor performance at the margins for v_1 , v_2 , α , and ϕ , many of which may never arise in practice, it is useful to assess accuracy in a way that is not differentially influenced by such outliers. One way to do this is to establish an accuracy criterion, or error tolerance, and compare the performance of the four algorithms in terms of how often they meet the established criterion. For example, we might decide that errors within $\pm .01$ of the true value of $1 - \beta$ are acceptable. We can then determine

Figure 9. Exact and approximated CDFs for F' for the worst-case scenario $v_1 = 1$, $v_2 = 4$, and $\lambda = 0.72$ ($\phi = 0.60$). Figure 9b employs an expanded scale to illustrate the approximation error that occurs for lower tail values of F' .

how often each of the approximations meets this particular criterion. The existence of a few extreme outliers will not distort this type of analysis.

We determined the percent of the errors that are within $\pm .01$ of the exact value for Investigation 1 and Investigation 2 as a function of α and the approximation method. For Investigation 1 ($\alpha = .005-.05$), 100% of the power estimates of the three-moment approximation meet the established criterion. For the two-moment, cube root, and square root approximations, respectively,

95%–97%, 92%–95%, and 87%–92% of the estimates meet the criterion. For Investigation 2 ($\alpha = .005-.90$), the percent of the estimates meeting criterion are as follows: three-moment, 91%–100% ($M = 97.3\%$); square root, 87%–99% ($M = 91.8\%$); two-moment, 86%–97% ($M = 91.8\%$); and cube root, 87%–96% ($M = 91.6\%$).²⁸ The three-moment approximation has at least 97% of its estimates within $\pm .01$ of the exact value out to $\alpha = .6$, where it begins to decline until reaching its lowest percentage, 91%, at $\alpha = .9$.

DISCUSSION

As expected, the three-moment approximation surpasses the accuracy of the other approximations by a fair margin, registering the lowest *ME* and *MAE* from $\alpha = .005$ –.50. The performance of all four approximations becomes less stable for $\alpha > .50$, with the three-moment and square root approximations showing large *MNEs* at $\alpha = .90$. Imposing the restriction $\{v_1 > 1, v_2 > 2\}$ gives moderately acceptable performance for large α , with maximum errors of roughly $\pm .03$. A consistent trend in the data is the decrease in approximation error that occurs with increases in v_1 . The maximum errors are particularly sensitive to this parameter. In addition, error magnitude is a nonmonotonic function of ϕ and an increasing function of v_2 . The latter trend is somewhat weak and inconsistent and is qualified by the presence of large approximation errors in the square root and cube root approximations at small v_2 . Finally, although all four approximations give satisfactory performance at the α levels typically encountered for power calculations, their deterioration at high α levels suggests that caution should be used when employing the approximations to estimate the CDFs of lower tail values of F' .

The systematic fluctuation of the approximate values about the exact values, as reflected in the error functions of Figures 2–8, was apparently first noted by the editor of *Biometrics* in a footnote to Patnaik's two-moment approximation (1949, p. 223). We attempt to account now for some similar characteristics in our own data. For example, why is it that the two-moment and cube root approximations have error functions that are so nearly alike, given that they come from different classes (moment vs. normal)? From Equation 2 in the introduction, we see that the scaling constant for the two-moment approximation is $k = (v_1 + \lambda)/v_1$, and the transformed value of F' is $F^* = F'/k = v_1 F'/(v_1 + \lambda)$. Equation 5 for the cube root transformation has this quantity in both the numerator and the denominator of the expression for z^* . In Laubscher's derivation of the cube root approximation (1960, p. 1111), the quantity shows up explicitly as $V = Y/X = [v_1 F'/(v_1 + \lambda)]^{1/3}$, where V is the ratio of two independent normal variables, $Y = [\chi_1^2/(v_1 + \lambda)]^{1/3}$ and $X = (\chi_2^2/v_2)^{1/3}$. The ratio thus reduces to $V = (F^*)^{1/3}$, and the cube root approximation is seen to be a normalized version of the two-moment approximation. This accounts for the similarity in the error functions of the two approximations.²⁹

A second question concerns the dramatic increase in approximation error for the three-moment and square root approximations at $\alpha \geq .90$. Pearson and Hartley (1976, p. 67) point out that, although the three-moment approximation to F' gives the correct first three moments, it does not give the correct *start* to the distribution. To see this, recall that the three-moment transformation is $F^* = (F' + c)/h$, where F^* is evaluated against a central F distribution with v_1^* and v_2 degrees of freedom. Con-

sequently, $F' = 0$ transforms to $F_{\min}^* = c/h$, and since $c/h > 0$ for $\lambda > 0$,³⁰ the cumulative area approximated by $P\{F^*(v_1^*, v_2) \leq c/h\}$ will overestimate the true area, $P\{F'(v_1, v_2, \lambda) \leq 0\} = 0$. Likewise, $P\{F^*(v_1^*, v_2) \geq c/h\}$ will underestimate the area to the right of $F' = 0$, which is 1.0. As we show below, this distortion is present in varying degree as $F' \rightarrow 0$ ($\alpha \rightarrow 1$), and this is why the three-moment approximation underestimates power at high α (Figure 8d). A similar problem arises for the square root and cube root approximations. To yield correct solutions, $F' = 0$ should transform to $z_{\min}^* = -\infty$, or at least to a value far enough out in the tail that the area to the left of $z_{\min}^*$ is ≈ 0 . However, Equation 4 in the introduction shows that the square root approximation transforms $F' = 0$ to

$$z_{\min}^* = -\left(\frac{2(v_1 + \lambda)^2}{v_1 + 2\lambda} - 1\right)^{\frac{1}{2}}, \quad (6)$$

and Equation 5 shows that the cube root approximation transforms $F' = 0$ to

$$z_{\min}^* = \frac{1-a}{\sqrt{a}}, \text{ where } a = \frac{2(v_1 + 2\lambda)}{9(v_1 + \lambda)^2}. \quad (7)$$

Equations 6 and 7 reveal that the most highly truncated values of $z_{\min}^*$ occur at $v_1 = 1$ and $\lambda = 0$, with $z_{\min}^* = -1$ and $z_{\min}^* = -1.65$, respectively. These limits generate maximum possible errors of $-.1587$ and $-.0495$ for the square root and cube root approximations. Of the four approximations, the two-moment approximation is the only one that transforms $F' = 0$ to a correct starting value, $F^* = F'/k = 0/k = 0$. Hence, the two-moment approximation is not susceptible to the lower tail bias of the other three approximations.

The preceding shows that constraints in the minimum values of F^* and z^* lead to errors in estimating the areas to the left or right of lower tail values of F' . Figures 9a and 9b illustrate the problem for $v_1 = 1$, $v_2 = 4$, and $\lambda = .72$ ($\phi = 0.60$). These parameters correspond to the worst-case scenario for the three-moment approximation (see the Results section). The figures plot the exact CDF of F' computed by Reeve's (1986) algorithm, as well as the CDFs obtained using each of the four approximations. The CDFs were evaluated at $F' = \{0(0.05)1(2)20\}$. For reference, the α levels for $F' = F_\alpha$ are shown below the values of F' on the x -axis. Figure 9a plots the CDFs over the range $0 \leq F' \leq 20$, whereas Figure 9b does so for the lower tail values only ($F' \leq 0.50$). The fit of the approximated CDFs to the exact CDF appears to be excellent in Figure 9a. However, the expanded scale of Figure 9b shows that the three-moment and square root CDFs diverge from the exact CDF at $F' \leq 0.30$, with maximum errors of 0.1591 and 0.1163, respectively, at $F' = 0$. The cube root CDF runs below the exact CDF until $F' = .05$, at which point it crosses over to show an error of .0282 at $F' = 0$. The divergence of the lower tails of the exact and approximate CDFs is due to lower limits

of $F_{\min}^* = 0.1125$, $z_{\min}^* = -1.194$, and $z_{\min}^* = -1.908$, respectively, for the three-moment, square root, and cube root approximations. These limits prevent the CDFs from going to zero at $F' = 0$.³¹ This problem does not arise for the two-moment approximation, since $F_{\min}^* = 0$.

It might be supposed that one reason the three-moment and square root approximations do so poorly in Figure 9b is because $v_2 = 4$. Equations in Patnaik (1949, p. 221) show that computation of the first moment requires $v_2 > 2$, the second moment $v_2 > 4$, and the third moment $v_2 > 6$. Consequently, the use of the three-moment approximation for $v_2 \leq 6$ (or of the two-moment approximation for $v_2 \leq 4$) would seem to lack mathematical justification, even though it is possible (mechanically) to compute solutions for $v_2 > 2$. Likewise, Patnaik (p. 208) notes that Fisher's normalizing function for chi-square, employed by the square root approximation, gives good results if $v > 30$. By implication, relatively poor performance would be expected for $v_2 = 4$. If these factors are responsible for the large approximation errors in Figure 9b as $F' \rightarrow 0$, then a larger value of v_2 should eliminate the errors. However, this is not the case: For $v_2 = 120$, errors of 0.1436 and 0.1163 occur at $F' = 0$ in the three-moment and square root CDFs, respectively. Equation 6 indicates that the minimum transformed value for the square root approximation is independent of v_2 , and this explains why the approximation error is unchanged from that obtained for $v_2 = 4$.

The distortion in the lower tails of the CDFs caused by constraints in the minimum values of F^* and z^* should be greatest for small values of v_1 . The density function of F' for small v_1 is J-shaped, with the peak density falling at or near $F' = 0$. For the three-moment approximation, the area to the left of c/h is potentially much greater in this case. We would therefore expect approximation error to decrease as v_1 increases and the mode of the F' distribution moves to the right (away from $F' = 0$). Consider the example in Figure 9 ($v_2 = 4$, $\lambda = 0.72$): as v_1 goes from 1 to 10, v_1^* goes from 1.471 to 10.126, c/h goes from 0.1125 to 0.0013, and the approximation error at $F' = 0$ goes from 0.1591 to 0.0000. The most substantial reduction in error occurs from $v_1 = 1$ to $v_1 = 2$, with CDF error at $F' = 0$ dropping from 0.1591 ($c/h = 0.1125$) to 0.0275 ($c/h = 0.0432$). These illustrations show another factor contributing to the improvement in accuracy: Increasing v_1 causes c/h to move closer to zero, and this leads to a corresponding decrease in the approximation errors for lower tail values of F' . These conclusions are also consistent with the analysis provided by Pearson and Tiku (1970), which showed that the error arising from not fitting the fourth moment of the F' distribution decreases with increases in v_1 .

The improvement in lower tail accuracy for the square root approximation with increases in v_1 follows from Equation 6. As v_1 goes up, the numerator of the ratio increases at a greater rate (due to the squaring operation) than the denominator, and this gives values of $z_{\min}^*$ further out in the left tail of the normal distribution. For the

example in Figure 9, as v_1 goes from 1 to 10, $z_{\min}^*$ goes from -1.194 to -4.369 , and the approximation error at $F' = 0$ goes from 0.1163 to 0.0000. As with the three-moment approximation, the most substantial reduction in error occurs at $v_1 = 2$, with CDF error at $F' = 0$ dropping to 0.0346 ($z_{\min}^* = -1.817$).

A last issue to consider is the relationship between approximation accuracy and λ . Examination of the three-moment approximation indicates that CDF approximation error at $F' = 0$ is zero for $\lambda = 0$ (since $c/h = 0$), increases to a peak value as λ increases, and then decreases again with further increases in λ . This nonmonotonic trend arises because the peak error is a joint function of v_1^* and c/h . The former increases, and the latter increases and then decreases, with increases in λ . The interaction of these two effects produces a peak error at $\lambda > 0$. The square root approximation, on the other hand, has its maximum CDF error at $F' = 0$ and $\lambda = 0$, and this error decreases systematically with increases in λ . Equation 6 shows why larger values of λ are associated with smaller CDF errors at $F' = 0$: The squaring operation in the numerator causes larger values of λ to produce values of $z_{\min}^*$ that are further out in the left tail, and this reduces the error.

The preceding analysis has attempted to explain some of the factors that contribute to the worst-case (lower tail) performance of the approximations. Excessive approximation error is most likely to arise when solutions are desired for $F' \rightarrow 0$ in distributions having $v_1 = 1$ and $\lambda \rightarrow 0$. Fortunately, this problem does not arise for computing statistical power, because conventional significance levels produce critical values ($F' = F_{\alpha}$) that fall in the upper tail of the F' distribution when λ is small, and if λ is large (so that F_{α} falls in the lower tail of the F' distribution), the problem does not arise for that reason. When the approximation is needed to compute the CDF of the F' distribution, however, the researcher should exercise special care to ensure that the solutions are accurate for small values of F' .

Most researchers investigating the accuracy of approximations conclude by expressing a preference for a particular approximation (Cohen & Nee, 1987; Fowler, 1984; Laubscher, 1960; Tiku, 1965, 1966). Our results suggest that the best approximation depends on the application and the needs of the user. Although the three-moment approximation is clearly superior to the others for $\alpha \leq .50$, it has the poorest worst-case performance at $\alpha \geq .90$. Similarly, although the two-moment approximation is intermediate in overall accuracy, it is the only approximation that yields correct solutions for $F' = 0$, and generally more accurate solutions for $F' \rightarrow 0$. Finally, the square root and cube root approximations are roughly comparable in overall accuracy, but the cube root approximation shows less degradation at very small values of v_2 , whereas the square root approximation performs better at $v_2 \geq 10$.

In addition to accuracy, a second issue of concern in selecting an approximation is ease of use. Equations 2–5

indicate that the two-moment approximation is simplest to use, followed by the square root, cube root, and three-moment approximations. If computations are done manually, the square root and cube root approximations have the advantage that tables of the standard normal distribution provide more detail for determining $1 - \beta$ (or the CDF) than do tables of the central F distribution. If the approximation is being coded for use in a program, such differences are perhaps less important than the overall accuracy of the approximation. If so, this would argue for the three-moment approximation. However, the two- and three-moment approximations require that the transformed value F^* be evaluated against a central F distribution, and algorithms for computing $P\{F^*(v_1^*, v_2) \geq x\}$ (O'Grady, 1981) are considerably more complex than algorithms for computing $P\{z^* \geq x\}$ (Abramowitz & Stegun, 1964). Although the latter fact argues for the square root or cube root approximation, the advantage is more apparent than real, because to compute power all four approximations require evaluation of F_α . Determination of the critical value requires an algorithm for computing upper tail areas of F , and, once this is included, it can be used to compute areas with respect to F^* as well.³²

The need to supplement the approximations with an algorithm for computing upper tail areas of central F raises an interesting paradox. Although the amount of program code required to implement any of the four approximations is trivial, the code required to compute exact values of F_α is not. If this code is included in the overhead, none of the four approximations is substantially more efficient, in terms of program size, than Reeve's (1986) implementation of the series representation in Equation 1, which computes the CDF exactly. This is because the same program code used by Reeve to compute the CDF of F' (or F'') is used to compute the CDF of F (the latter being the special case where $\lambda_1 = \lambda_2 = 0$). Consequently, no additional overhead is required for computing F_α , and Equation 1 is nearly as efficient (in terms of code size) as one of the approximations in Equations 2–5, supplemented by an algorithm for computing central F probabilities (O'Grady, 1981).

It might be argued that the approximations at least offer an advantage in terms of computational speed. Although they do compute their solutions very quickly, as compared with an exact method, this advantage is swamped by the time required to compute F_α . As a result, total execution time shows only a small advantage for the approximations. To illustrate, consider the original example in Figure 1, where $\alpha = .05$, $v_1 = 3$, $v_2 = 60$, and $\lambda = 12.8$. The time required to compute $F_\alpha = 2.758$ and $1 - \beta \approx .84$ for this case was averaged over 100 successive calls to the subroutines for computing the exact and approximate solutions. The average time required to complete these two computations was 0.086, 0.079, 0.080, 0.072, and 0.073 sec for the exact, two-moment, three-moment, square root, and cube root methods, respectively.³³ The approximations are, therefore, only slightly

faster than the exact solution when the overhead for computing F_α is included. It seems doubtful that the loss of accuracy inherent in an approximation can be justified simply to achieve a difference of 0.086 versus 0.072 sec.

Consequently, with respect to both program size and speed, there is little to be gained by using an approximate over an exact method. An exception to this is when critical values of F_α are known in advance, and either the square root or cube root transformation is used to approximate $1 - \beta$. In this case, calls to subroutines for computing critical values and p values of central F are not required, and computations of $1 - \beta$ are performed quickly. This is presumably why Cohen employs the square root and cube root algorithms for approximating power in his well-known book on statistical power analysis (Cohen, 1988, p. 550).³⁴ Another exception is when F_α is approximated rather than computed exactly. Although this approach is substantially more efficient in program size and speed than an exact method, it has potentially serious problems with worst-case performance. Since the error in approximating F_α is now compounded with the error inherent in the four noncentral F approximations, approximation errors much larger than those identified in this report can arise.

In conclusion, the present study shows that the two-moment, three-moment, square root, and cube root approximations can provide satisfactory estimates of power, at least when certain restrictions are imposed. However, the approximations prove to be only marginally more efficient than exact methods for computing the CDF of F' . With the availability of inexpensive, high-performance computers, there would no longer seem to be compelling reasons to rely on approximations. Likewise, the increasing emphasis on statistical power analysis by contemporary researchers (and journal editors) warrants, in our view, the exclusive use of exact methods in statistical power software.

REFERENCES

- ABRAMOWITZ, M., & STEGUN, I. A. (EDS.) (1964). *Handbook of mathematical functions* (Applied Mathematics Series No. 55). Washington, DC: National Bureau of Standards.
- BARTON, D. E., DAVID, F. N., & O'NEILL, A. F. (1960). Some properties of the distribution of the logarithm of non-central F . *Biometrika*, *47*, 417-429.
- BORENSTEIN, M., & COHEN, J. (1988). *Statistical power analysis: A computer program*. Hillsdale, NJ: Erlbaum.
- BRADLEY, D. R. (1988). *DATASIM*. Lewiston, ME: Desktop Press.
- BRADLEY, D. R. (1989a). Computer simulation with DATASIM. *Behavior Research Methods, Instruments, & Computers*, *21*, 99-112.
- BRADLEY, D. R. (1989b). A general purpose simulation program for statistics and research methods. In G. Garson & S. Nagel (Eds.), *Advances in social science and computers* (Vol. 1, pp. 145-186). Greenwich, CT: JAI Press.
- BRADLEY, D. R. (1991a). Anatomy of a DATASIM simulation: The Doob and Gross horn-honking study. *Behavior Research Methods, Instruments, & Computers*, *23*, 190-207.
- BRADLEY, D. R. (1991b). *Datasim for the Macintosh*. Lewiston, ME: Desktop Press.
- BRADLEY, D. R. (1993). Multivariate simulation with DATASIM: The

- Mihal and Barrett study. *Behavior Research Methods, Instruments, & Computers*, **25**, 148-163.
- BRADLEY, D. R., & FLEISHER, C. L. (1994). Generating multivariate data from nonnormal distributions: Mihal and Barrett revisited. *Behavior Research Methods, Instruments, & Computers*, **26**, 156-166.
- BRADLEY, D. R., HEMSTREET, R. L., & ZIEGENHAGEN, S. T. (1992). A simulation laboratory for statistics. *Behavior Research Methods, Instruments, & Computers*, **24**, 190-204.
- BRADLEY, D. R., RUSSELL, R. L., & REEVE, C. P. (1996). Statistical power in complex experimental designs. *Behavior Research Methods, Instruments, & Computers*, **28**, 319-326.
- BRADLEY, D. R., SENKO, M. W., & STEWART, F. A. (1990). Statistical simulation on microcomputers. *Behavior Research Methods, Instruments, & Computers*, **22**, 236-246.
- BREWER, J. (1972). On the power of statistical tests in *American Educational Research Journal*. *American Educational Research Journal*, **9**, 391-401.
- CHASE, L. J., & CHASE, R. B. (1976). A statistical power analysis of applied psychological research. *Journal of Applied Psychology*, **61**, 234-237.
- CHASE, L. J., & TUCKER, R. K. (1975). A power-analytic examination of contemporary communication research. *Speech Monographs*, **42**, 29-41.
- CHASE, L. J., & TUCKER, R. K. (1976). Statistical power: Derivation, development, and data-analytic implications. *Psychological Record*, **26**, 473-486.
- COHEN, J. (1962). The statistical power of abnormal-social psychological research: A review. *Journal of Abnormal & Social Psychology*, **65**, 145-153.
- COHEN, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- COHEN, J. (1977). *Statistical power analysis for the behavioral sciences* (rev. ed.). Hillsdale, NJ: Erlbaum.
- COHEN, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- COHEN, J., & NEE, J. C. M. (1987). A comparison of two noncentral F approximations, with applications to power analysis in set correlation. *Multivariate Behavioral Research*, **22**, 483-490.
- ERDFELDER, E., FAUL, F., & BUCHNER, A. (1996). GPOWER: A general power analysis program. *Behavior Research Methods, Instruments, & Computers*, **28**, 1-11.
- FIELLER, E. C. (1932). The distribution of the index in a normal bivariate population. *Biometrika*, **24**, 428-440.
- FOWLER, R. L. (1984). Approximating probability levels for testing null hypotheses with the noncentral F distributions. *Educational & Psychological Measurement*, **44**, 275-281.
- GOLDSTEIN, R. (1989). Power and sample size via MS/PC-DOS computers. *American Statistician*, **43**, 253-260.
- GUENTHER, W. C. (1979). The use of noncentral F approximations for calculation of power and sample size. *American Statistician*, **33**, 209-210.
- JOHNSON, N. L., & KOTZ, S. (1970). *Continuous univariate distributions* (Vol. 2). New York: Houghton Mifflin.
- KATZER, T., & SODT, J. (1973). An analysis of the use of statistical testing in communication research. *Journal of Communication*, **23**, 251-265.
- KOELE, P. (1982). Calculating power in the analysis of variance. *Psychological Bulletin*, **92**, 513-516.
- KOELE, P., & HOOGSTATEN, J. (1980). *Power and sample size calculations in the analysis of variance* (Révész Berichten No. 12). Amsterdam: Universiteit van Amsterdam, Psychologisch Laboratorium.
- LACHENBRUCH, P. A. (1966). The noncentral F distribution extensions of Tang's tables [Abstract]. *Annals of Mathematical Statistics*, **37**, 774.
- LAUBSCHER, N. F. (1960). Normalizing the noncentral t and F distributions. *Annals of Mathematical Statistics*, **31**, 1105-1112.
- LEHMER, E. (1944). Inverse tables of probabilities of errors of the second kind. *Annals of Mathematical Statistics*, **15**, 388-398.
- MEYER, G. E. (1995). Power & Effect: A statistical utility for Macintosh and Windows systems. *Behavior Research Methods, Instruments, & Computers*, **27**, 134-138.
- O'GRADY, K. E. (1981). Probabilities and critical values for z , chi-square, t , and F . *Behavior Research Methods & Instrumentation*, **13**, 55-56.
- PARKER, R., & BORICH, G. (1980). POWERF: A program that computes the power of F tests in fixed-effects ANOVA designs. *Behavior Research Methods & Instrumentation*, **12**, 76.
- PATNAIK, P. B. (1949). The non-central χ^2 and F distributions and their applications. *Biometrics*, **36**, 202-232.
- PEARSON, E. S., & HARTLEY, H. O. (EDS.) (1976). *Biometrika tables for statisticians* (Vol. II). London: Biometrika Trust.
- PEARSON, E. S., & TIKU, M. L. (1970). Some notes on the relationship between the distributions of central and non-central F . *Biometrika*, **57**, 175-179.
- REEVE, C. P. (1986). *An algorithm for computing the doubly noncentral F C.D.F. to a specified accuracy*. Statistical Engineering Division Note 86-4. Gaithersburg, MD: National Bureau of Standards.
- ROSSI, J. S. (1982, April). *Meta-analysis, power analysis and artifactual controversy: The case of spontaneous recovery of verbal associations*. Paper presented at the 53rd Annual Meeting of the Eastern Psychological Association, Baltimore.
- ROSSI, J. S. (1986, June). *Effects of inadequate statistical power on specific research literatures*. Paper presented at the annual meeting of the Psychometric Society, Toronto.
- ROSSI, J. S. (1990). Statistical power of psychological research: What have we gained in 20 years? *Journal of Consulting & Clinical Psychology*, **58**, 646-656.
- ROTON, J., & SCHÖNEMANN, P. H. (1978). Power tables for analysis of variance. *Educational & Psychological Measurement*, **38**, 213-229.
- SCHEFFÉ, H. (1959). *Analysis of variance*. New York: Wiley.
- SEDLMEIER, P., & GIGERENZER, G. (1989). Do studies of statistical power have an effect on the power of studies? *Psychological Bulletin*, **105**, 309-316.
- SEVERO, N., & ZELEN, M. (1960). Normal approximation to the chi-square and noncentral F probability functions. *Biometrika*, **47**, 411-416.
- TANG, P. C. (1938). The power function of the analysis of variance tests with tables and illustrations of their use. *Statistical Research Memoirs*, **2**, 126-149.
- THOMAS, L., & KREBS, C. J. (1997). A review of statistical power analysis software. *Bulletin of the Ecological Society of America*, **78**, 128-139.
- TIKU, M. L. (1965). Laguerre series forms of non-central χ^2 and F distributions. *Biometrika*, **52**, 415-427.
- TIKU, M. L. (1966). A note on approximating the non-central F distribution. *Biometrika*, **53**, 606-610.
- TIKU, M. L. (1967). Tables of the power of the F -test. *Journal of the American Statistical Association*, **62**, 525-539.
- TIKU, M. L. (1972). More tables of the power of the F -test. *Journal of the American Statistical Association*, **67**, 709-710.
- TIKU, M. L., & YIP, D. Y. N. (1977). A four-moment approximation based on the F distribution. *Australian Journal of Statistics*, **20**, 257-261.
- WINER, B. J. (1971). *Statistical principles in experimental design* (2nd ed.). New York: McGraw-Hill.
- WINER, B. J., BROWN, D. R., & MICHELS, K. M. (1991). *Statistical principles in experimental design* (3rd ed.). New York: McGraw-Hill.
- WOODWARD, A. J., & OVERALL, J. E. (1976). A computer program for calculating the power of the F -test. *Educational & Psychological Measurement*, **36**, 165-168.
- WRIGHT, J. (1993). DATASIM 1.0. *Psychology Software News*, **4**, 15-16.

NOTES

1. Meyer's (1995) program, Power & Effect, offers a calculator and formula-based metaphor for computing effect size. The GPOWER software of Erdfelder et al. (1996) is unique in performing compromise power analysis: With this approach, solutions are computed so as to preserve a prescribed ratio of Type I to II error probabilities (α , β). The enhancements to DATASIM described by Bradley et al. (1996) permit researchers to generate complete power and sample size tables in either conventional format (based on df and δ or λ) or the format used in Cohen's 1988 book (based on n and the effect size, f). DATASIM is also one of the few programs that compute power for the doubly noncentral F distribution.

2. The programs by Erdfelder et al. (1996; GPOWER) and Bradley (1988; DATASIM) have algorithms for computing exact (as well as

approximate) power. Several of the programs reviewed by Goldstein (1989) and Thomas and Krebs (1997) also employ exact methods.

3. It is especially helpful to conceptualize power in terms of f when no basis exists for estimating the population parameters required to compute λ or ϕ . In this case, Cohen's guidelines (1988, pp. 281–284) for defining small, medium, and large effects may be used to select values of f for entry into the power tables. Alternatively, the researcher can compute power for detecting effects that account for a given proportion of the variation in the data, η^2 , where $f = \sqrt{\eta^2/(1 - \eta^2)}$.

4. As we explain in the discussion section, the efficiency of approximation over exact methods is largely lost if values of F_α must also be computed by the program (rather than supplied as input).

5. Note that the N for the three-moment approximation in Tiku (1966) is 120 rather than 144 because the three-moment approximation cannot be computed for $v_2 = 2$ (see v_1^* and c in Equation group 3, above).

6. An apparent exception to this generalization occurs at $v_1 = 1$, where the error in the moment ratio decreases consistently from $\phi = 0.5$ to $\phi = 3.0$. As we show in the Results and Discussion sections, however, a peak does in fact exist in the interval $0.5 \leq \phi \leq 1.0$.

7. Cohen and Nee report that this difference is significant at $p < .01$; however, a significance test is not really necessary, because the errors of approximation are not stochastic, and the mean absolute errors are therefore not subject to sampling error.

8. Optical scanning with OCR software was briefly considered but rejected due to insufficient quality in the copies of Tiku's tables obtained through interlibrary loan. Previous experience with OCR in this situation showed error rates exceeding those associated with manual entry of the data.

9. As documented elsewhere (Bradley, 1989a, 1989b, 1991a, 1993; Bradley & Fleisher, 1994; Bradley, Hemstreet, & Ziegenhagen, 1992; Bradley, Senko, & Stewart, 1990; Wright, 1993), DATASIM is a general purpose data analysis and statistical simulation program.

10. This modification was necessary because the power approximations in Exhibits 3–6 employ λ rather than ϕ as the noncentrality parameter. The change was made by reading the data from the Tiku database into DATASIM and using the LET command to transform ϕ to λ through the relation $\lambda = (v_1 + 1)\phi^2$.

11. It makes little sense to assess the magnitude of the errors of approximation when such errors reflect not only the inaccuracy of the algorithm, but also the inaccuracy of the "exact" values to which they are being compared. This is particularly true when the accuracy of the approximation is sufficiently high that even small typesetting errors in the tabled values of β can seriously bias the results. Since the approximation errors for the four algorithms are usually within $\pm .01$ of the true value (see the Results section), a typesetting error in a four-place decimal number can substantially distort the apparent accuracy of an algorithm, especially for measures based on the MPE s or MNE s, which is a major focus of the present study.

12. All but two of these errors were discovered in the same way as the errors in Tiku's tables—that is, the errors caused data points in plots of the errors of approximation to show up as outliers. The last two errors were detected after a systematic comparison, entry by entry, of the values in the computer database against the values in Tiku's published tables.

13. The 7,920 values in the final database were compared to the 7,920 exact values produced by Reeve's (1986) program. The maximum absolute difference between corresponding values of β was .000145, and 95.2% of the values of β agreed exactly when rounded to four places (the precision of Tiku's values). The option existed at this point to conduct Investigation 1 using Reeve's exact values rather than Tiku's (1967) tables. However, since Investigation 1 was largely complete at this point, we elected to use Reeve's program to identify and correct possible errors in Tiku, rather than as the primary source of exact values for $1 - \beta$.

14. It was at this point that Charles Reeve became actively involved as a coinvestigator in the present research project.

15. The program BATNFI.EXE reads the parameters v_1 , v_2 , α , and λ from an input file, determines the critical value $F_\alpha(v_1, v_2)$, and computes the area to the right of F_α in the noncentral F distribution $F(v_1, v_2; \lambda)$. The input parameters and the exact power are then written to an output file. For consistency with Investigation 1, DATASIM was used to

modify the output file by adding columns including β and ϕ in the appropriate locations.

16. Although the standard deviation (SD) of the errors of approximation would provide information about the range of errors encountered, it was rejected for two reasons. First, the distributions of the errors of approximation are not symmetrical, so the SD is a potentially misleading index of the range of errors above and below the mean. Second, the SD does not directly communicate the maximum error (either positive or negative) that can arise using a particular approximation, and this information was of primary interest to the investigators.

17. Accuracy, stability, and bias are interdependent in that a small MAE tends to be associated with a small ME and a small $MPE - MNE$. An unbiased approximation is not necessarily accurate or stable, however, since large positive and negative approximation errors can exist that average out to zero. Hence, although a small MAE implies a small ME , the reverse is not true.

18. Inspection of the raw data shows that the three-moment approximation gives three-place accuracy or better in many cases.

19. In interpreting these and subsequent figures, the reader is cautioned to pay careful attention to the scale of the y -axis. In order to maximize resolution, the y -axis in each plot is scaled so as to employ the full area of the plotting window. This means that the start, stop, and increment values of the y -axis change from one plot to the next. Comparison of the heights of particular curves across plots is, therefore, not meaningful unless the corresponding y -values on each axis are taken into account.

20. As a point of comparison, perfect accuracy on each measure would be represented by a horizontal line extending out from $y = 0$ on the vertical axis.

21. The reader may wonder why significance tests are not reported for the trends plotted in Figure 4 (and elsewhere)—for example, to see if the four approximations differ significantly in accuracy. The answer is that the values making up the database are not *stochastic* (i.e., they are not random variables following a probability distribution). The errors of approximation are analytic, in that, given the same parameters, the same values will always be obtained. Consequently, the data points plotted in Figure 4 are not subject to sampling error, and the issue of significance does not arise.

22. Interestingly, the error trends for the two-moment and cube root approximations are quite similar in Figures 3–5, as is the overall accuracy of the two algorithms. In several of the panels, the two curves (open circles vs. filled squares) are very nearly superimposed. This is somewhat surprising, since the two algorithms come from different families (i.e., moment vs. normal). This matter is considered further in the discussion section.

23. The two-moment and cube root approximations once again have very similar error functions in Figures 6a, b, and d, and in Figure 6c for $\phi \geq 1.8$.

24. The relative order of the curves in Figure 7c is difficult to discern for the intermediate values of v_2 . The original plot of Figure 7c was displayed on a color monitor, and the different colors used by the plotting software permitted the investigators to unambiguously establish the order of the individual curves.

25. This is due, again, to the unusual performance of the square root algorithm at $v_2 = 2$. When the maximum positive error trends for higher values of v_2 are considered, decreases in the errors do occur with increases in v_1 .

26. The inclusion of $\alpha = 1.00$ means that the performance of the approximations is being evaluated for $F' = F_\alpha = 0$. The upper tail area is therefore $1 - \beta = 1.00$, and this holds for all values of v_1 , v_2 , or ϕ .

27. Including $\phi = 0$ tests the approximations when applied to the central F distribution. Tests at $\alpha = 1.00$, $\phi = 0$, and small v_1 and v_2 assess the accuracy of the approximations at the extreme endpoints of their intended applicability, and this helps to reveal the worst-case performance of the approximations.

28. These results confirm a conclusion arrived at earlier—namely, that the three-moment approximation is noticeably more accurate than the two-moment, square root, and cube root approximations and that the latter three are roughly comparable in (overall) accuracy.

29. We thank Melinda Harder of the Department of Mathematics, Bates College, for suggesting this interpretation.

30. As $\lambda \rightarrow 0$, the constants converge as follows: $v_1^* \rightarrow v_1$, $c \rightarrow 0$, $h \rightarrow 1$, and, therefore, $c/h \rightarrow 0$.

31. The constants for the three-moment approximation are $v_1^* = 1.4707$, $c = 0.205001$, and $h = 1.8225$. Hence, $F^* = c/h = 0.1125$ produces a minimum possible CDF of $P\{F^*(1.4707, 4) \leq .1125\} = .1591$. The CDF errors at $F^* = 0$ for the square root and cube root approximations are found by determining $P\{z^* \leq -1.194\}$ and $P\{z^* \leq -1.908\}$, respectively.

32. F_α is found using an iterative search procedure that systematically varies the value of F until one is found which cuts off α in the upper tail. Of course, tables of the central F distribution can be consulted to find F_α . However, such tables are of limited utility insofar as only certain values of v_1 , v_2 , and α are tabled. Although interpolation can, in some cases, be used to find the desired value of F_α , this defeats the point of automating the calculations of $1 - \beta$ or CDF. Also, inaccuracies arising from linear interpolation will add to the inherent inaccuracy of the approximation, causing potentially unacceptable increases in error.

33. The computations were performed on a Macintosh Centris 650 with a 68040 processor and FPU.

34. However, Bradley et al. (1996) show that Cohen's use of approximate solutions in conjunction with an adjustment procedure for computing power in factorial designs (Cohen, 1988, p. 365) can lead to serious bias in the power and sample size estimates obtained using his tables. They cite an example in which Cohen's approach gives $1 - \beta \approx .70$ for the power of F_A in a $3 \times 4 \times 5$ design (Cohen, 1988, Table 8.3.34), when in fact the actual power is $1 - \beta = .8516$. Errors in the treatment of factorial ANOVA in Cohen's earlier (1969 and 1977) editions were noted by Koele (1982, see footnote 1) and reviewed in detail in an unpublished technical report by Koele and Hoogstraten (1980). Erdfelder et al. (1996) have also addressed this issue.

(Manuscript received March 31, 1997;
revision accepted for publication May 30, 1997.)