LIMITS FOR A VARIANCE COMPONENT WITH AN EXACT CONFIDENCE COEFFICIENT

By W. C. Healy, Jr.

Ethyl Corporation, Detroit

1. Introduction and summary. This paper deals with confidence interval estimation and hypothesis testing for components of variance, in analysis-of-variance situations embraced by the Model II of Eisenhart [1], including also the so-called nested classifications. Several authors have treated the problem of setting confidence limits for variance components, and several approximate methods have been proposed. Four approximate methods are described in Anderson and Bancroft [2] and briefly in Crump [3], and references to original sources and extensive bibliographies are given in both [2] and [3]. In [4], Green gives more refined approximations which are, however, not presently in a form for practical use. Huitson [5], Welch [6] and Cochran [7] discuss related problems involving linear combinations of variances, and offer approximate methods for these problems. The many references on approximate tests and confidence limits in variance component problems emphasize the absence of exact methods. The present paper points out that, using a randomization device, exact confidence limits and tests for a variance component become available in a simple way. These exact confidence limits will usually but not always define a single confidence interval; in the exceptional case (having small probability in practice) the exact limits may define an interval with a gap in it. Numerical illustrations are given, together with comparisons with results using some of the available approximate methods. Also, asymptotic power comparisons between the exact test and two approximate tests are discussed.

There are at least three notions of confidence that can be associated with the statement: "$a(x) \leqq \theta \leqq b(x)$ is a $100(1 - \alpha)\%$ confidence interval" for the parameter θ with possible nuisance parameters η, based on observations x. They are

(a) $\Pr\{a(x) \leqq \theta \leqq b(x)\} \equiv 1 - \alpha$ for all θ, η

(b) $\Pr\{a(x) \leqq \theta \leqq b(x)\} \geqq 1 - \alpha$ for all θ, η with equality for some θ, η.

(c) $\Pr\{a(x) \leqq \theta \leqq b(x)\} \geqq 1 - \alpha$ for all θ, η.

The phrase "exact confidence" we shall interpret in the sense of (a) above. So far as the author is aware, for a variance component no confidence limits satisfying either of the notions (a) or (b) have previously been constructed. An interval satisfying (c) has been constructed in [8], using a two-stage sampling procedure.

In the present approach, the mathematical difficulties ordinarily caused in

Received June 13, 1957; revised October 21, 1960.

variance component analysis by the presence of nuisance parameters are circumvented by a process of randomization. The resulting confidence limits depend not only on the mean squares of the analysis of variance table, but also on auxiliary observations on a random variable with known normal distribution. A consequence is that two statisticians confronted with the same analysis of variance table will in general construct different confidence limits. While practically this may afford some discomfiture, it remains true nevertheless that these exact confidence limits meet the ordinary claim (a) as to probability of containing the true variance component. In the examples tried, the limits are plausible and they are not difficult to compute. Moreover, the agreement between numerical results using the method proposed herein and the usual approximate methods may serve to increase one's faith in the approximate methods in small samples.

2. Statement of the problem. Specifically, the kind of problem dealt with here can be described in terms of two observed mean squares U and V such that nV/σ^2 and $mU/(\sigma^2 + r\sigma_0^2)$ are independently distributed in chi-square distributions with n and m degrees of freedom respectively. The variance components σ^2 and σ_0^2 are unknown, and r is a known constant depending on the experimental design. It is required to find confidence limits for σ_0^2 having confidence coefficient exactly $1 - \alpha$. This problem arises from balanced Model II variance analyses, from certain Mixed Model analyses, and from analyses of nested classifications, when suitable normality assumptions are made. For full discussions of these analyses and models, the reader is referred to [1], [2], and [3], and the accompanying bibliographies.

Any problem of the above type, concerned with estimating or testing hypotheses on a variance component, can always be reduced to a corresponding problem concerning the difference between unknown variances of two normal distributions with known zero means. In the following section the latter formulation will be used, later converting the results into the usual terms of variance component analysis.

3. Exact confidence limits for the difference between two variances. Suppose there are two independent samples $(x_1, x_2, \cdots, x_n)$ and $(y_1, y_2, \cdots, y_m)$ from $N(0, \sigma_1^2)$ and $N(0, \sigma_1^2 + \sigma_2^2)$ respectively. It is required to construct exact two-sided confidence limits for σ_2^2, having confidence coefficient $1 - \alpha$. (Corresponding one-sided limits are easily obtained from the two-sided case.) Now an equivalent problem is that of constructing a similar, size-α test of

$$H: \sigma_2^2 = \delta^2$$

$$A: \sigma_2^2 \neq \delta^2;$$

that is, a test satisfying $\Pr\{\text{reject } H \text{ when } \sigma_2^2 = \delta^2\} \equiv \alpha$ for all σ_1^2. Such a test will yield exact confidence limits for σ_2^2.

It is, however, easy to construct such a test by the adjunction of an independent sample $(z_1, z_2, \cdots, z_n)$ from $N(0, 1)$; the statistic defined by

$$w = \frac{\sum_1^m y^2/m}{\sum_1^n (x + \delta z)^2/n}$$

has, when H is true, the F distribution with (m, n) degrees of freedom and accordingly provides similar tests of H. For example, the "equal tail" acceptance region for H is defined by $F_1 \leqq w \leqq F_2$, where F_1 and F_2 are respectively the 100 $(\frac{1}{2}\alpha)$ and 100 $(1 - \frac{1}{2}\alpha)$ per cent points of the F distribution with (m, n) degrees of freedom. Corresponding one-sided tests can be described in the obvious way. We shall agree to define w for $\delta \geqq 0$ only, corresponding to the positive root of σ_2^2; the sign of δ does not affect the validity of the significance test, but a consistent sign for δ is necessary in deriving the confidence limits.

To obtain exact confidence limits for σ_2^2 starting with the equal-tail acceptance region based on w, we have

$$\{F_1 \leqq w \leqq F_2\}$$

$$= \left\{\frac{1}{F_2} \leqq \frac{1}{w} \leqq \frac{1}{F_1}\right\} \tag{1}$$

$$= \left\{\frac{n \sum y^2}{mF_2} \leqq \sum (x + \delta z)^2 \leqq \frac{n \sum y^2}{mF_1}\right\}$$

$$= \left\{\frac{n \sum y^2}{mF_2} - \sum x^2 \leqq 2\delta \sum xz + \delta^2 \sum z^2 \leqq \frac{n \sum y^2}{mF_1} - \sum x^2\right\}. \tag{2}$$

From (2) it is seen that an undesirable feature has crept in; if we proceed in the obvious way from (2), the resulting limits will involve terms in $\sum xz$. Such terms would prove inconvenient in applications to variance component analysis, since independent quantities $x_1, x_2, \cdots, x_n$ having distribution $N(0, \sigma_1^2)$ are not observed there, and would become available only after suitable transformation of the original data. It would be better to have confidence limits which depend on the original data only through the mean squares (e.g., U and V) usually computed in the analysis of variance. The argument which follows serves to construct such computationally more convenient limits.

Divide the inequalities (2) by $2\delta(\sum x^2)^{\frac{1}{2}}$, obtaining

$$\left\{\frac{1}{2\delta(\sum x^2)^{\frac{1}{2}}}\left[\frac{n \sum y^2}{mF_2} - \sum x^2\right] \leqq \frac{\sum xz}{(\sum x^2)^{\frac{1}{2}}} + \frac{\delta \sum z^2}{2(\sum x^2)^{\frac{1}{2}}} \right.$$
$$\left. \leqq \frac{1}{2\delta(\sum x^2)^{\frac{1}{2}}}\left[\frac{n \sum y^2}{mF_1} - \sum x^2\right]\right\}, \tag{3}$$

and set

$$t = \frac{\sum xz}{(\sum x^2)^{\frac{1}{2}}} + \frac{\delta \sum z^2}{2(\sum x^2)^{\frac{1}{2}}}.$$

Now for fixed $(x_1, x_2, \cdots, x_n)$, an orthogonal transformation from $(z_1, z_2, \cdots, z_n)$ to $(z_1', z_2', \cdots, z_n')$ with $z_1' = (\sum xz)/(\sum x^2)^{\frac{1}{2}}$ yields

$$t' = z_1' + \frac{\delta}{2(\sum x^2)^{\frac{1}{2}}} \sum_1^n z'^2, \tag{4}$$

where $(z_1', z_2', \cdots, z_n')$ are each distributed $N(0, 1)$, are mutually independent are are independent of $(x_1, x_2, \cdots, x_n)$. The statistics t and t' have the same distribution and since t' is computable directly from $\sum x^2$ and $(z_1', z_2', \cdots, z_n')$ it will serve as the desired replacement for t. From this discussion, a third possible substitute for t is seen to be

$$t'' = z + \frac{\delta}{2(\sum x^2)^{\frac{1}{2}}} z^2 + \frac{\delta}{2(\sum x^2)^{\frac{1}{2}}} \chi^2_{n-1},$$

where z is $N(0, 1)$, χ^2_{n-1} is a chi-square variate with $n - 1$ degrees of freedom, and z and χ^2_{n-1} are independent.

Replacing t by t' in (3), and dropping primes, we have

$$\left\{\frac{1}{2\delta(\sum x^2)^{\frac{1}{2}}}\left[\frac{n \sum y^2}{mF_2} - \sum x^2\right] \leqq z_1 + \frac{\delta}{2(\sum x^2)^{\frac{1}{2}}} \sum_1^n z^2 \leqq \frac{1}{2\delta(\sum x^2)^{\frac{1}{2}}}\left[\frac{n \sum y^2}{mF_1} - \sum x^2\right]\right\}$$

which yields, upon completing the square in δ,

$$\left\{\frac{1}{\sum z^2}\left[\frac{n \sum y^2}{mF_2} - \sum x^2\right] + \left(\frac{z_1(\sum x^2)^{\frac{1}{2}}}{\sum z^2}\right)^2 \leqq \left(\delta + \frac{z_1(\sum x^2)^{\frac{1}{2}}}{\sum z^2}\right)^2 \leqq \frac{1}{\sum z^2}\left[\frac{n \sum y^2}{mF_1} - \sum x^2\right] + \left(\frac{z_1(\sum x^2)^{\frac{1}{2}}}{\sum z^2}\right)^2\right\}. \tag{5}$$

An exact confidence region for σ_2 is then defined as the set of values δ satisfying both (5) and $\delta \geqq 0$. These are the values of σ_2 for which the null hypothesis $\sigma_2 = \delta$ would be accepted, and since the acceptance region (1) has probability identically $1 - \alpha$ when $\sigma_2 = \delta$, the confidence region defined by (5) and $\delta \geqq 0$ has confidence coefficient exactly $1 - \alpha$.

It remains to determine the limits defining this region, and the discussion will be simplified if we let

$$k = \frac{1}{\sum z^2}\left[\frac{n \sum y^2}{mF_2} - \sum x^2\right],$$

$$l = \frac{1}{\sum z^2}\left[\frac{n \sum y^2}{mF_1} - \sum x^2\right],$$

and

$$b = -\frac{z_1(\sum x^2)^{\frac{1}{2}}}{\sum z^2}.$$

In these terms the inequalities (5) become

$$k + b^2 \leqq (\delta - b)^2 \leqq l + b^2,$$

which constrain δ to satisfy either

$$b - (l + b^2)^{\frac{1}{2}} \leqq \delta \leqq b - \{\max[0, k + b^2]\}^{\frac{1}{2}} \tag{6}$$

or else

$$b + \{\max[0, k + b^2]\}^{\frac{1}{2}} \leqq \delta \leqq b + \{l + b^2\}^{\frac{1}{2}}. \tag{7}$$

(We have assumed $l + b^2 \geqq 0$; the confidence region is empty otherwise.) Consideration of (6) and (7), together with the requirement $\delta \geqq 0$, shows that the confidence interval will be defined by (7) alone, unless $-b^2 \leqq k < 0$ and $b > 0$ both obtain. In the latter case the confidence region consists of both (7) and the non-overlapping interval

$$\max(b - \{l + b^2\}^{\frac{1}{2}}, 0) \leqq \delta \leqq b - \{\max[0, k + b^2]\}^{\frac{1}{2}}. \tag{8}$$

There thus exists a possibility that the exact confidence limits may define a region consisting of two separate intervals, which would in practice be a disconcerting event, (though of course the "gap" could be included in the confidence region at the sacrifice of the exactness property.) It is of interest to examine the chances of getting two intervals, and we note first that b^2 converges to zero in probability so that asymptotically the chance of two intervals is negligible. Also, $\Pr\{b > 0\} = \frac{1}{2}$. Note too that $k < 0$ is the condition that the variance ratio $(\sum y^2/m)/(\sum x^2/n)$ be not significant at the $\frac{1}{2}\alpha$ level for testing $\sigma_2 = 0$.

Some simple computations can give a rough bound on the probability that two intervals will result. Using the definition for k, the condition $-b^2 \leqq k < 0$ is readily found equivalent to the condition.

$$\left(F_2 - \frac{z_1^2}{\sum z^2}\right)\left(\frac{\sigma_1^2 + \delta^2}{\sigma_1^2 + \sigma_2^2}\right) \leqq F \leqq F_2\left(\frac{\sigma_1^2 + \delta^2}{\sigma_1^2 + \sigma_2^2}\right),$$

where F has the F-distribution with (m, n) degrees of freedom. The probability of this event will be greatest when $F_2[(\sigma_1^2 + \delta^2)/(\sigma_1^2 + \sigma_2^2)]$ is close to the mode $(mn - 2n)/(mn + 2m)$ of this F distribution, or roughly when $F_2 \sim (\sigma_1^2 + \sigma_2^2)/(\sigma_1^2 + \delta^2)$. Thus we can get a rough upper bound to the probability of two intervals by computing the probability in an interval of length $z_1^2/F_2\sum z^2$ at the mode of this F distribution. In the neighborhood of the mode, a normal approximation should suffice for our present purposes. Taking F to be approximately normal $N[1, 2(1 + a)/(na)]$, where $a = m/n$, the desired probability bound can then be approximated by

$$E\left\{\frac{z_1^2}{F_2\sum z^2}\frac{n^{\frac{1}{2}}}{(2\pi)^{\frac{1}{2}}}\frac{1}{\left(\dfrac{2(1+a)}{a}\right)^{\frac{1}{2}}}\right\} \sim \frac{1}{(2\pi n)^{\frac{1}{2}}F_2}\left(\frac{a}{2(1+a)}\right)^{\frac{1}{2}}.$$

Numerically, for $m = 5$, $n = 24$ as an illustration, we obtain $\sim .01$ as an approximate maximum to the probability of $-b^2 \leqq k < 0$. Since for two intervals to result, we must also have $b > 0$, we can say that roughly speaking the probability of two intervals will not exceed $\sim .005$ for these values of m and n, no matter what the configuration of σ_1^2 and σ_2^2. We conclude from this type of investigation that the possibility of a confidence region consisting of two intervals by the exact method of this paper is remote for practical values of m and n, and hence should not cause difficulty in practice.

4. Exact confidence limits for a variance component. It remains now to convert the result of Section 3 to the variance component problem of Section 2. The complete procedure can be stated as follows. Let nV/σ^2 and $mU/(\sigma^2 + r\sigma_0^2)$ have independent chi-square distributions with n and m degrees of freedom respectively, V and U being observed mean squares in a variance component analysis and r being a known constant. Let $(z_1, z_2, \cdots, z_n)$ be a sample of size n from $N(0, 1)$, which can be obtained from tables of random normal deviates, eg. [10]. Then exact two-sided $100(1 - \alpha)\%$ confidence limits for σ_0^2 are given (for the usual case of a single interval) by

$$(9) \quad \text{Lower limit:} \quad \frac{1}{r}\left\{\left[\max 0, \frac{n}{\sum z^2}\left(\frac{U}{F_2} - V\right) + \left(\frac{z_1(nV)^{\frac{1}{2}}}{\sum z^2}\right)^2\right]^{\frac{1}{2}} - \frac{z_1(nV)^{\frac{1}{2}}}{\sum z^2}\right\}^2$$

$$(10) \quad \text{Upper limit:} \quad \frac{1}{r}\left\{\left[\frac{n}{\sum z^2}\left(\frac{U}{F_1} - V\right) + \left(\frac{z_1(nV)^{\frac{1}{2}}}{\sum z^2}\right)^2\right]^{\frac{1}{2}} - \frac{z_1(nV)^{\frac{1}{2}}}{\sum z^2}\right\}^2$$

where F_1 and F_2 are respectively the $100\ (\frac{1}{2}\alpha)$ and $100\ (1 - \frac{1}{2}\alpha)$ per cent points of the F distribution with (m, n) degrees of freedom. Furthermore, if

$$-\left(\frac{z_1(nV)^{\frac{1}{2}}}{\sum z^2}\right)^2 \leqq \frac{n}{\sum z^2}\left(\frac{U}{F_2} - V\right) < 0 \quad \text{and} \quad z_1 < 0,$$

the interval having

$$(11) \quad \text{Lower limit:} \quad \frac{1}{r}\left\{\max 0, -\left[\frac{n}{\sum z^2}\left(\frac{U}{F_1} - V\right) + \left(\frac{z_1(nV)^{\frac{1}{2}}}{\sum z^2}\right)^2\right]^{\frac{1}{2}} - \frac{z_1(nV)^{\frac{1}{2}}}{\sum z^2}\right\}^2$$

$$(12) \quad \text{Upper limit:} \quad \frac{1}{r}\left\{-\left[\max 0, \frac{n}{\sum z^2}\left(\frac{U}{F_2} - V\right) + \left(\frac{z_1(nV)^{\frac{1}{2}}}{\sum z^2}\right)^2\right]^{\frac{1}{2}} - \frac{z_1(nV)^{\frac{1}{2}}}{\sum z^2}\right\}^2$$

is to be included as well.

An equivalent alternate procedure is to observe z_1 and χ^2_{n-1} and replace $\sum_1^n z^2$ by $(z_1^2 + \chi^2_{n-1})$ in the preceding limits. However, tables of the chi-square distribution are not sufficiently complete to permit simulation of sampling from a chi-square distribution and there are no tables of random χ^2 variates. While one could employ tables of random numbers and a table of the Incomplete Gamma Function, it may be that the best way to obtain a χ^2_{n-1} variate is as the sum of the squares of $n - 1$ N(0, 1) variates, which brings us back to the first procedure.

5. Numerical illustrations and comparisons. We shall take as an illustration the example based on the analysis of variance table on page 323 of [2], from which the authors give, on page 324, 90% confidence limits for a variance component by several approximate methods. Three of these approximate methods may be described briefly as

(i) normal approximation to the distribution of $(1/r)[U - V] = \hat{\sigma}_0^2$

(ii) χ^2 approximation to the distribution of $\hat{\sigma}_0^2$

(iii) replacement of σ^2 by V in exact confidence limits for σ_0^2/σ^2.

The pertinent data are

$$U = 46{,}659 \qquad m = 3$$
$$V = 459 \qquad n = 72$$
$$r = 300$$

Based on sampled values $z_1 = 0.628$, $\sum_1^n z^2 = 62.72$, exact confidence limits for σ_0^2 are computed from (9) and (10) as

Lower limit: 62

Upper limit: 1514.

Application of the three approximate methods gave the results in the following table, from page 324 of [2].

90% Confidence Limits

Method	Lower Limit	Upper Limit
(i)	0	316
(ii)	59	1313
(iii)	55	1331

The extent of variations to be expected between conventional approximate methods and the present method may be studied by constructing an approximation based on the present method. A simple approximation can be obtained from (9) and (10) by replacing functions of $(z_1, z_2, \cdots z_n)$ by their exact or approximate expected values. Using

$$E\left(\frac{z_1(nV)^{\frac{1}{2}}}{\sum z^2}\right) = 0, E\left(\sum z^2\right) = n,$$

and

$$E\left(\frac{z_1(nV)^{\frac{1}{2}}}{\sum z^2}\right)^2 \sim \frac{EnVz_1^2}{E(\sum z^2)^2} = \frac{V}{n+2},$$

we have the slightly new approximate limits

$$\text{(13)} \qquad \text{Lower limit: } \frac{1}{r}\left[\frac{U}{F_2} - V\left(\frac{n+1}{n+2}\right)\right]$$

$$\text{Lower limit: } \left[\frac{U}{F_1} - V\left(\frac{n+1}{n+2}\right)\right]. \tag{14}$$

Replacing $(n + 1)/(n + 2)$ in the above limits by unity yields the approximate limits from method (iii) above. It is clear that the ratio between (14) and (10) is essentially the ratio between n and $\sum z^2$, or χ_n^2/n. That is, the exact limits will tend to deviate from this conventional approximation (iii) proportionally to variations in χ_n^2/n, being more variable than the approximate limits because of this extra element of random variation. The variance of χ^2/n being $2/n$, one obtains an idea of the size of discrepancies to be encountered between exact and approximate limits.

6. Power comparisons. As remarked earlier, the power of the similar test from which exact confidence limits have been derived herein can be computed directly from the F distribution. The one-sided tests are clearly unbiased. However, no investigation of the standing of these tests (one-sided or two-sided) in the class of all similar tests has been attempted. This would seem to be a difficult problem due to difficulties in characterizing similar tests of the hypothesis $\sigma_2^2 = \delta^2$. (The similar test herein derived does not have Neyman structure, which means that $(\sum x^2, \sum y^2)$ is not boundedly complete, which means that the methods depending on boundedly complete sufficient statistics do not apply.)

One might suspect that the element of randomization introduced to achieve similarity could result in a serious impairment of power. For this reason, comparisons of asymptotic power functions have been made among the similar test of Section 3 and the tests corresponding to approximate confidence intervals (ii) and (iii) of the preceding section. We consider only one-sided tests

$$H: \sigma_2^2 = \delta^2$$

$$A: \sigma_2^2 > \delta^2;$$

we do find that the exact test is somewhat inferior in large-sample power, but that the amount of power impairment is not likely to be serious.

Corresponding to the approximate confidence interval (ii) based on a chi-square distribution, the one-sided rejection region is defined by

$$\frac{1}{\delta^2}\left(\frac{\sum y^2}{m} - \frac{\sum x^2}{n}\right) \geqq \chi_f^2/f,$$

where χ_f^2 is the 100 $(1 - \alpha)\%$ point of the chi-square distribution with f degrees of freedom, and f is determined here by

$$f = \frac{\left(\frac{\sum y^2}{m} - \frac{\sum x^2}{n}\right)^2}{\frac{1}{m}\left(\frac{\sum y^2}{m}\right)^2 + \frac{1}{n}\left(\frac{\sum x^2}{n}\right)^2}.$$

In these comparisons of large-sample power, we will consider that m and n approach infinity in a fixed ratio, say $m = na$. Then $f = ng$, where

$$g = \frac{a\left(\frac{\sum y^2}{m} - \frac{\sum x^2}{n}\right)^2}{\left(\frac{\sum y^2}{m}\right)^2 + \left(\frac{\sum x^2}{n}\right)^2}.$$

To compute the asymptotic power of this test we first note that the quantity $n^{\frac{1}{2}}\chi_f^2/f = \chi_{ng}^2/gn^{\frac{1}{2}}$ can be approximated by

$$\chi_{ng}^2/gn^{\frac{1}{2}} \sim [2gn^{\frac{1}{2}}]^{-1}[t_\alpha + (2ng - 1)^{\frac{1}{2}}]^2 = n^{\frac{1}{2}} + 2^{\frac{1}{2}}(t_\alpha/g^{\frac{1}{2}}) + \text{“terms”},$$

where the "terms" approach zero in probability and t_α is the 100 $(1 - \alpha)$ % point of $N(0, 1)$. Also, the quantity $(n^{\frac{1}{2}}/\delta^2)[(\sum y^2/m) - (\sum x^2/n) - \sigma_2^2]$ has asymptotically a normal distribution with mean zero, variance $2\delta^{-4}[\{(\sigma_1^2 + \sigma_2^2)^2/a\} + \sigma_1^4]$. Accordingly the asymptotic power of this test can be written

$$\Pr\left\{\frac{n^{\frac{1}{2}}}{\delta^2}\left[\frac{\sum y^2}{m} - \frac{\sum x^2}{n} - \sigma_2^2\right] + \frac{\sigma_2^2 n^{\frac{1}{2}}}{\delta^2} \geqq n^{\frac{1}{2}} + \left[\frac{2}{a}\right]^{\frac{1}{2}} \frac{t_\alpha}{\sigma_2^2}\left[(\sigma_1^2 + \sigma_2^2)^2 + a\sigma_1^4\right]^{\frac{1}{2}}\right\}$$

which yields, after some reduction and application of standard theorems,

$$\Phi\left(\frac{\rho}{2^{\frac{1}{2}}}\left[\frac{1}{a} + \frac{\sigma_1^4}{(\sigma_1^2 + \sigma_2^2)^2}\right]^{-\frac{1}{2}} - \frac{t_\alpha \delta^2}{\sigma_2^2}\right), \tag{15}$$

where Φ is the standard normal distribution function and

$$\rho = n^{\frac{1}{2}}[(\sigma_2^2 - \delta^2)/(\sigma_1^2 + \sigma_2^2)]$$

measures the divergence from the null hypothesis.

For the one-sided test based on the approximate confidence interval from method (iii) of the preceding section, the rejection region is

$$\frac{\sum y^2/m}{\sum x^2/n + \delta^2} \geqq F_2.$$

Here the quantity

$$n^{\frac{1}{2}}\left(\frac{\sum y^2/n}{\sum x^2/n + \delta^2} - \frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 + \delta^2}\right)$$

has asymptotically a normal distribution with mean zero, variance

$$2\,\frac{(\sigma_1^2 + \sigma_2^2)^2}{(\sigma_1^2 + \delta^2)^2}\left[\frac{1}{a} + \frac{\sigma_1^4}{(\sigma_1^2 + \delta^2)^2}\right],$$

and the asymptotic power function can accordingly be written

$$\Pr\left\{n^{\frac{1}{2}}\left(\frac{\sum y^2/m}{\sum x^2/n + \delta^2} - \frac{(\sigma_1^2 + \sigma_2^2)}{(\sigma_1^2 + \delta^2)}\right) \geqq n^{\frac{1}{2}} F_2 - n^{\frac{1}{2}}\left(\frac{\sigma_1^2 + \sigma_2^2}{\sigma_1^2 + \delta^2}\right)\right\}.$$

Using now the fact that $n^{\frac{1}{2}}(F - 1)$ is asymptotically $N(0, 2[(1/a) + 1])$, we obtain for the power function, after some reduction,

$$\Phi\left(\left[\frac{\rho}{2^{\frac{1}{2}}}-\frac{(\sigma_1^2+\delta^2)}{(\sigma_1^2+\sigma_2^2)}\,t_\alpha\left(\frac{1}{a}+1\right)^{\frac{1}{2}}\right]\left[\frac{1}{a}+\frac{\sigma_1^4}{(\sigma_1^2+\delta^2)^2}\right]^{-\frac{1}{2}}\right). \tag{16}$$

It should be noted that the test statistic $(\sum y^2/m)/(\sum x^2/n+\delta^2)$ and F do not have the same limiting distribution even on the null hypothesis, which incidentally accounts for this more complicated form of the asymptotic power function for approximate method (iii).

To cast the asymptotic power function of the one-sided similar test into analogous form we use the limiting distribution of F. The power function is

$$\Pr\left\{F \geqq F_2\,\frac{(\sigma_1^2+\delta^2)}{(\sigma_1^2+\sigma_2^2)}\right\}$$

which asymptotically becomes

$$\Phi\left\{\frac{\rho}{2^{\frac{1}{2}}}\left[\frac{1}{a}+1\right]^{-1}-t_\alpha\,\frac{(\sigma_1^2+\delta^2)}{(\sigma_1^2+\sigma_2^2)}\right\}. \tag{17}$$

Comparison of the power expressions (15), (16), and (17) shows that the approximate chi-square test based on method (ii) is asymptotically most powerful among these three tests. Comparison of the F approximation method (iii) with the similar test shows that for $\delta > 0$, the method (iii) will have superior power for large values of ρ while the similar test will have superior power for small values of ρ.

It is of interest to evaluate the magnitudes of these power differences. Consider a comparison of (15) and (17), which compares the asymptotic power of the similar test with that of the best test among these three. We first note that $n \to \infty$ and the definition of ρ imply that for $\delta > 0$ both $(\sigma_1^2+\delta^2)/(\sigma_1^2+\sigma_2^2)$ and δ^2/σ_2^2 will be close to unity, so that the difference in power resulting from the difference in multipliers of t_α can be neglected. One way to compare tests is on the basis of sample sizes required for equivalent power, and for equivalent power we see from (15) and (17) that sample sizes must be approximately in the ratio

$$R=\left(1+\frac{1}{a}\right)\div\left(\frac{1}{a}+\frac{\sigma_1^4}{(\sigma_1^2+\sigma_2^2)^2}\right)$$

for these two tests. Now R must satisfy

$$1 \leqq R \leqq 1+a, \tag{18}$$

and a is ordinarily a fairly small number, corresponding to the fact that a "between" variance is usually estimated on much fewer degrees of freedom than is a "within" variance. For example, with a one-way classification having r observations per class, $1/a \sim r$. The inequality (18) thus means that with $a = \frac{3}{72}$ as found in the first numerical example of Section 5, the exact test requires less than 4% more observations for equivalent asymptotic power.

These results seem to say, without further numerical investigation, that for most designs yielding fairly small values of a, our use of randomization to achieve exact similarity in small samples does not cost much in terms of large-sample power. Numerical power comparisons would be interesting but have not been computed.

REFERENCES

[1] CHURCHILL EISENHART, "The assumptions underlying the analysis of variance," *Biometrics*, Vol. 3 (1947), pp. 1–21.

[2] R. L. ANDERSON AND T. A. BANCROFT, *Statistical Theory in Research*, McGraw Hill, New York, 1952.

[3] S. L. CRUMP, "The present status of variance component analysis," *Biometrics*, Vol. 7 (1951), pp. 1–16.

[4] J. R. GREEN, "A confidence interval for variance components," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 671–686.

[5] A. HUITSON, "A method of assigning confidence limits to linear combinations of variances," *Biometrika*, Vol. 42 (1955), pp. 471–479.

[6] B. L. WELCH, "On linear combinations of several variances," *J. Amer. Stat. Ass'n.*, Vol. 51 (1956), pp. 132–148.

[7] W. G. COCHRAN, "Testing a linear relation among variances," *Biometrics*, Vol. 7 (1951), pp. 17–32.

[8] A. BIRNBAUM AND WILLIAM C. HEALY, JR., "Estimates with prescribed variance based on two-stage sampling," *Ann. Math. Stat.*, Vol. 31 (1960), pp. 662–676.

[9] The RAND Corporation, *One Million Random Digits and 100,000 Normal Deviates*, The Free Press, Glencoe, Ill., 1955.

[10] H. CRAMÉR, *Mathematical Methods of Statistics*, Princeton University Press, Princeton, 1946.