POWER OF ANALYSIS OF VARIANCE TEST PROCEDURES FOR CERTAIN INCOMPLETELY SPECIFIED MODELS, I[1, 2]

By Helen Bozivich,[3] T. A. Bancroft, and H. O. Hartley

Statistical Laboratory, Iowa State College

TABLE OF CONTENTS

	PAGE
1. Introduction	1017
1.1 Description of pooling procedures	1017
1.2 More precise formulation	1019
1.3 Reduction of mixed models to random models	1021
1.4 Related papers and objectives of present investigation	1021
2. Exact and approximate formulas for power. Component of variance model	1022
2.1 Mathematical formulation of the pooling procedure	1022
2.2 Integral expressions for the power	1022
2.3 Exact formulas	1024
2.31 Series formulas	1024
2.32 Recurrence formulas	1025
2.4 Approximate formulas	1026
2.5 Theory of reduction of mixed model to random model	1027
2.6 Application of derived formulas	1028
3. Discussion of power and size and comparison of test procedures	1029
3.1 Type of recommendations attempted	1029
3.2 Size	1029
3.3 Frequency of pooling	1036
3.4 Power	1037
References cited	1042
Appendix: Figures and tables	1041

1. Introduction.

1.1 *Description of pooling procedures.* The simplest situation of a pooling procedure for testing hypotheses using analysis of variance procedures may be described as follows: We are given three mean squares, V_1, V_2, V_3, based on n_1, n_2, and n_3 degrees of freedom, respectively, and designated as treatment mean square (V_3), the error mean square (V_2), and the doubtful error mean square (V_1). It is desired to test a null hypothesis involving V_3, which can be tested by comparing V_3 with V_2 by the F-test. It is now suspected that V_1 is also a measure of the error variance, that is, has the same expectation as V_2. It is decided, therefore, to first perform a preliminary test of significance by comparing

Received July 12, 1955; revised May 21, 1956.

[1] Journal paper No. J-2962 of the Iowa Agricultural Experiment Station, Ames, Iowa, Project 169.

[2] This paper is based on research sponsored by the Iowa State College Agricultural Experiment Station and the Wright Air Development Center.

[3] Part of the results presented here are contained in a thesis prepared by Miss Helen Bozivich and submitted to the Graduate Faculty of Iowa State College in partial fulfillment of the requirements for the Ph.D. in Statistics, June, 1955.

V_2 against V_1 by the F-test, and, if this turns out to be nonsignificant, to use the pooled mean square $V = (n_1V_1 + n_2V_2)/(n_1 + n_2)$ as error for comparison with V_3 in the final F-test. In the case that V_2 is significantly different from V_1, however, use V_2 as error in the final F-test. This test procedure is usually referred to as the sometimes-pool procedure. In corresponding terminology the single test V_3/V_2 is called the never-pool test, and the procedure of employing $V = (n_1V_1 + n_2V_2)/(n_1 + n_2)$ as error and always testing only V_3/V is called the always-pool test. If the level of significance for the preliminary test is 100%, the sometimes-pool test becomes the never-pool test; if, on the other hand, the level is 0%, the sometimes-pool test becomes the always-pool test. With the sometimes-pool test, the precise nature of the final F-test is, therefore, not determined in advance, but it depends on the relative magnitude of the observed mean squares V_2 and V_1.

When the analysis of variance and associated tests of significance were first developed by R. A. Fisher, such procedures were not advocated. Indeed, Fisher's original description of analysis of variance tests clearly stipulated that for every well-designed experiment there can only be one correct analysis and the test(s) of significance are completely determined before the experimental results are available. With Fisher the appropriate test of significance is determined by a specification of the population from which the experimental data were sampled. We may speak in this case of an analysis determined by a completely specified model.[4] However, in experimental design, situations frequently arise in which the model is not completely specified. Furthermore, with the wider application of analysis of variance to operational research and to the study of routine data, statisticians are often faced with analyzing data which have not resulted from a designed experiment, and in these situations the model is often incompletely specified. In such cases preliminary tests of significance have been used, in practice, as an aid in choosing an appropriate specification from which valid subsequent inferences may be drawn. In particular, preliminary tests of significance procedures have been used in the past in an attempt to increase the number of degrees of freedom associated with the error in a final F-test, thereby apparently increasing the sensitivity of the final F-test. Justification for the use of such methods has been made apparently on intuitive grounds.

The procedures described above and similar pooling procedures can be regarded as dealing with these situations sequentially in two stages: the preliminary stage in which inferences are drawn about the model, and the final stage in which inferences are drawn about the parameter(s) involved in the main hypothesis.

The purpose of the present study is to critically examine the consequences of certain pooling procedures with regard to the resulting errors of the first and second kind, for certain random and mixed models. Finally, on the basis of these results, we shall attempt some general recommendations on the advisability or otherwise of using them.

---

[4] For the general formulation of model specification, see Section 1.2.

TABLE 1

Component of variance model with $\sigma_b^2 > 0$—Analysis of variance

Source of variation	df	Mean square	Exp. mean square
Between A	$n_3 = q - 1$	V_3	$\sigma_z^2 + s\sigma_b^2 + rs\sigma_a^2 = \sigma_3^2$
Between B within A	$n_2 = q(r - 1)$,	V_2	$\sigma_z^2 + s\sigma_b^2 = \sigma_2^2$
Within B	$n_1 = qr(s - 1)$	V_1	$\sigma_z^2 = \sigma_1^2$

1.2 *More precise formulation.* Let us assume the component of variance model

$$x_{ijk} = \mu + a_i + b_{ij} + z_{ijk}, \tag{1}$$

where $i = 1, 2, \cdots, q$; $j = 1, 2, \cdots, r$; $k = 1, 2, \cdots, s$; a_i is $N(0, \sigma_a^2)$, b_{ij} is $N(0, \sigma_b^2)$, and z_{ijk} is $N(0, \sigma_z^2)$. We wish to test a hypothesis concerning a_i. If $\sigma_b^2 \geqq 0$ and $\sigma_z^2 > 0$, then

$$\begin{aligned} x_{ijk} &= \mu + a_i + b_{ij} + z_{ijk} \qquad &\text{for } \sigma_b^2 > 0, \\ x_{ijk} &= \mu + a_i + z_{ijk} \qquad &\text{for } \sigma_b^2 = 0. \end{aligned} \tag{2}$$

In this case (1) is said to be an incompletely specified model. If, however, $\sigma_b^2 > 0$,

$$x_{ijk} = \mu + a_i + b_{ij} + z_{ijk}, \tag{3}$$

and (1) is completely specified. Similarly, if $\sigma_b^2 = 0$,

$$x_{ijk} = \mu + a_i + z_{ijk}, \tag{4}$$

and again (1) is completely specified.

We wish to test the hypothesis $H_0\colon\sigma_a^2 = 0$ against the alternative $H_1\colon\sigma_a^2 > 0$. Now let us assume we have the completely specified model given by (3). Then $\sigma_b^2 > 0$, and we obtain the analysis of variance given in Table 1. Then it follows from the likelihood ratio principle that the appropriate test procedure is to calculate the test statistic

$$F_0 = \frac{V_3}{V_2} \tag{5}$$

and to reject H_0 if $F_0 \geqq F_\alpha(n_3, n_2)$, where α is the prescribed level of significance. This test is the never-pool test.

Next let us assume the completely specified model given by (4). Now the expected mean squares of Table 1 no longer include the σ_b^2 component, since $\sigma_b^2 = 0$. Application of the likelihood ratio test procedure to this model for the test of H_0 gives us the test criterion

$$F_0 = \frac{(n_1 + n_2)V_3}{n_1 V_1 + n_2 V_2} \tag{6}$$

and the rule to reject H_0 if $F_0 \geqq F_\alpha(n_3, n_1 + n_2)$. This gives us the always-pool test.

Finally, we assume $\sigma_b^2 \geqq 0$ and, hence, are confronted with the incompletely specified model given by (2). Ordinarily this model (2) might be assumed when there exists some uncertainty as to whether $\sigma_b^2 = 0$ or $\sigma_b^2 > 0$. In such cases of incomplete specification, attempts are often made to resolve the uncertainty by first testing the hypothesis that $\sigma_b^2 = 0$. The model finally chosen and, hence, the final test (test of H_0) depend upon the outcome of this original test. When the original and final tests are performed on the same set of data, the original test is referred to as a preliminary test of significance. In our example the preliminary test becomes the test of $H_0':\sigma_b^2 = 0$ against $H_1':\sigma_b^2 > 0$. Again, a likelihood ratio test procedure is available for this preliminary test. The statistic $F_0 = V_2/V_1$ is calculated and H_0' rejected at the level α_1 (usually different from α) if $F_0 \geqq F_{\alpha_1}(n_2, n_1)$. If H_0' is rejected, the non-pooling test procedure indicated by (5) is used for the final test. If H_0' is not rejected, the pooling procedure indicated by (6) is used for the final test.

It should be noted that when the final test is carried out, the model is assumed to be completely specified, that is, to be either model (3) or model (4), according as the preliminary test is found to be significant or not significant, respectively.

The essential features of the sometimes-pool procedure as applied to the component of variance model described may be summarized as follows:

(i) The three mean square V_i are independently distributed as $\chi_i^2\sigma_i^2/n_i$, where χ_i^2 is the (central) χ^2 statistic for n_i degrees of freedom;

(ii) The main purpose of the analysis is to test the null hypothesis $\sigma_3^2 = \sigma_2^2$ against the alternative $\sigma_3^2 > \sigma_2^2$;

(iii) The error mean square V_2 has an expectation σ_2^2 which is greater than or equal to the expectation, σ_1^2, of the doubtful error mean square V_1, which may or may not be pooled.

It is clear that the above hierarchical classification is not the only analysis of variance situation giving rise to the above conditions. As an example we may quote the two-way classification with both factors random and cell repetition. Here V_1 would play the part of the within-cell mean square, while V_2 would be represented by the residual in the two-way analysis.

Aside from the preliminary test for the complete specification of the model, it is to be noted that we have made the assumptions usually made in the customary analysis of variance, namely those associated with an additive analysis of variance model. It is sometimes correctly argued that these latter assumptions may not be justified in certain situations, and in others may represent only an approximation to the actual mechanisms generating the data. This issue is, of course, one affecting the analysis of variance tests in general, and has led to extensive studies of the validity of these tests when the basic assumptions are not completely satisfied. If there is some doubt regarding the detailed assumptions for the analysis of variance model, it should be possible also to formulate the problem as being incompletely specified in these other respects. We are not concerned with these issues here. In extending the analysis of variance theory based on the assumption of linear models, our results are, strictly speaking, limited to situations in which these other assumptions are satisfied. However, the classical

TABLE 2

Mixed model example—Analysis of variance

Source of variation	df	Mean square	Exp. mean square
Between rations	$n_3 = k - 1$	V_3	$\sigma_3^2 = \sigma_z^2 + m\sigma_d^2 + mn\theta_{(a)}$
Reps x rations	$n_2 = (k-1)(n-1)$	V_2	$\sigma_2^2 = \sigma_z^2 + m\sigma_d^2$
Within pens	$n_1 = nk(m-1)$	V_1	$\sigma_1^2 = \sigma_z^2$

analysis of variance tests have been found to be remarkably robust, that is, not sensitive to certain deviations from the basic assumptions.[5] We expect, therefore, that our present results will likewise be applicable as useful approximations to a wider class of situations.

1.3 *Reduction of mixed models to random models.* The preceding section has been devoted to random models only. Another frequently occurring type of model is the mixed model, in which one of the factors is fixed and the other factors are random, and the hypothesis of interest is concerned with the fixed factor. A typical example of an experiment giving rise to this type of model is a randomized block experiment in which k rations are fed to each of m animals of a pen in each of n replicates. Then a suitable model for these data is given by

$$x_{tij} = \mu + a_t + b_i + d_{ti} + z_{tij},$$

where the replicate variates b_i, error variates d_{ti}, and within-pen error variates z_{tij} are assumed to be random samples from the respective normal populations $N(0, \sigma_b^2)$, $N(0, \sigma_d^2)$, and $N(0, \sigma_z^2)$, while the ration means a_t are fixed parameters. The analysis of variance based on this model is shown in Table 2. Here $\theta_{(a)} = \sum (a_t - \bar{a})^2/(k - 1)$. Following the same general consideration of Section 1.2, it is shown in Section 2.5 how the sometimes-pool procedure for this model can be reduced to that of the random model.

1.4 *Related papers and objectives of present investigation.* The problem to be discussed here is from a general area of preliminary tests of significance. Work in this area includes studies by Bancroft [1], [2]; Mosteller [12]; Paull [14], [15]; Kitagawa [10]; Bechhofer [4]; and Bennett [5]. Paull [14], [15] studied the size and the power for the component of variance model described in Section 1.2. However, he was able to express the size and power in closed form for the case $n_3 = 2$ only, so that all comparisons made by him are restricted to that value of n_3.

The object of the present study is to provide the necessary extension of Paull's investigation to cover all of the important degrees of freedom combinations occurring in the analyses of variance under discussion. This extension was made possible by

(i) the development of the power integrals as series formulas for even values of the degrees of freedom n_1, n_2, and n_3;

[5] See, e.g., Box ([6], [7]) and the numerous references to earlier work given there.

TABLE 3

Component of variance model—Analysis of variance

Source of variation	Mean square	df	Exp. mean square
Treatments	V_3	n_3	σ_3^2
Error	V_2	n_2	σ_2^2
Doubtful error	V_1	n_1	σ_1^2

(ii) the derivation of recurrence formulas for the power for even values of n_1, n_2, and n_3;

(iii) the development of approximate formulas valid for large degrees of freedom for even values of n_1, n_2, and n_3.

2. Exact and approximate formulas for power. Component of variance model.

2.1 *Mathematical formulation of the pooling procedure.* We now derive formulas for the power and size of the pooling procedure applied to the component of variance model described in Section 1. Let us first state the procedure in mathematical terms. We are given an analysis of variance as shown in Table 3.

We are interested in testing the hypothesis $H_0:\sigma_3^2 = \sigma_2^2$ against the alternative $H_1:\sigma_3^2 > \sigma_2^2$ when it is known that $\sigma_3^2 \geqq \sigma_2^2 \geqq \sigma_1^2$. We assume the sums of squares n_iV_i are independently distributed as $\chi_i^2\sigma_i^2$, where χ_i^2 is the central χ^2 statistic based on n_i degrees of freedom. The test procedure with sometimes pooling V_2 and V_1 is then as follows: Reject H_0 if

$$(7) \qquad \begin{cases} \text{either} & \{V_2/V_1 \geqq F_{n_2,n_1}(\alpha_1) \text{ and } V_3/V_2 \geqq F_{n_3,n_2}(\alpha_2)\} \\ \text{or} & \{V_2/V_1 \leqq F_{n_2,n_1}(\alpha_1) \text{ and } V_3/V \geqq F_{n_3,n_1+n_2}(\alpha_3)\}, \end{cases}$$

where $V = (n_1V_1 + n_2V_2)/(n_1 + n_2)$ and $F_{n_i,n_j}(\alpha)$ is the upper 100α % point of the F-distribution with numerator df $= n_i$ and denominator df $= n_j$.

The probability, P, of rejecting H_0, which in general is the power of the test procedure, is a function of the degrees of freedom, n_1, n_2, and n_3, the ratios, $\theta_{32} = \sigma_3^2/\sigma_2^2$ and $\theta_{21} = \sigma_2^2/\sigma_1^2$, and the levels of significance employed, α_1, α_2, and α_3. In the special case when $\theta_{32} = 1$, this power is equal to the size of the test, i.e., the probability of type one error. In general the power P is obtained as the sum of two components corresponding to the mutually exclusive alternatives headed by either, and/or in its definition above, namely,

$$(8) \qquad P_1 = \Pr\{V_2/V_1 \geqq F_{n_2,n_1}(\alpha_1) \text{ and } V_3/V_2 \geqq F_{n_3,n_2}(\alpha_2)\},$$

$$(9) \qquad P_2 = \Pr\{V_2/V_1 \leqq F_{n_2,n_1}(\alpha_1) \text{ and } V_3/V \geqq F_{n_3,n_1+n_2}(\alpha_3)\}.$$

2.2 *Integral expressions for the power.* Definitie integrals for P_1 and P_2 will now be derived. The joint density of V_1, V_2, and V_3 is given by

$$c_1 V_1^{\frac{1}{2}n_1-1}V_2^{\frac{1}{2}n_2-1}V_3^{\frac{1}{2}n_3-1} \exp\left\{-\frac{1}{2}\left(\frac{n_1V_1}{\sigma_1^2} + \frac{n_2V_2}{\sigma_2^2} + \frac{n_3V_3}{\sigma_3^2}\right)\right\},$$

where c_1 is a constant independent of V_1, V_2, and V_3. By introducing new variates,

$$u = \frac{n_2 V_2}{\theta_{21} n_1 V_1}, \qquad v = \frac{n_3 V_3}{\theta_{32} n_2 V_2}, \qquad w = \frac{n_1 V_1}{n_3},$$

and integrating out w, we obtain for the joint distribution of u and v

$$f(u, v) = \frac{k u^{\frac{1}{2}(n_2+n_3)-1} v^{\frac{1}{2}n_3-1}}{(1 + u + uv)^{\frac{1}{2}(n_1+n_2+n_3)}}$$

where

$$k = \frac{1}{B(n_1/2, n_2/2) B(n_3/2, (n_1 + n_2)/2)}$$

The probability of rejecting the hypothesis H_0 is obtained by integrating $f(u, v)$ over the two ranges of variation of u and v which correspond to the two alternatives either and/or of definition (7). These ranges are respectively given by either

$$\frac{u_1^0}{\theta_{21}} \leqq u < \infty, \frac{u_2^0}{\theta_{32}} \leqq v < \infty$$

or

$$0 \leqq u \leqq \frac{u_1^0}{\theta_{21}}, \qquad \frac{u_3^0(1 + \theta_{21} u)}{\theta_{32}\, \theta_{21}\, u} \leqq v < \infty.$$

where

(10) $$u_1^0 = \frac{n_2}{n_1} F_{n_2, n_1}(\alpha_1), \qquad u_2^0 = \frac{n_3}{n_2} F_{n_3, n_2}(\alpha_2).$$

and

$$u_3^0 = \frac{n_3}{n_1 + n_2} F_{n_3, n_1+n_2}(\alpha_3).$$

Hence the formulas for the two power components become

$$P_1 = \int_a^\infty \int_d^\infty f(u, v)\, dv\, du$$

and

$$P_2 = \int_0^a \int_{c(1+\theta_{21}u)/u}^\infty f(u, v)\, dv\, du.$$

where

(11) $$a = \frac{u_1^0}{\theta_{21}}, \qquad c = \frac{u_3^0}{\theta_{21}\, \theta_{32}}, \qquad d = \frac{u_2^0}{\theta_{32}}, \qquad \theta_{21} \geqq 1,$$

$$\theta_{32} \geqq 1, \qquad a \leqq u_1^0, \qquad d \leqq u_2^0$$

2.3 *Exact formulas.*

2.3.1 Series formulas.

$$P_1 = k \int_a^\infty \int_d^\infty \frac{u^{\frac{1}{2}(n_2+n_3)-1} v^{\frac{1}{2}n_3-1}}{(1 + u + uv)^{\frac{1}{2}(n_1+n_2+n_3)}} \, dv \, du.$$

The transformation $z = (1 + u)/(1 + u + uv)$ yields

$$P_1 = k \int_a^\infty \int_0^{x_1} \frac{z^{\frac{1}{2}(n_1+n_2)-1}(1 - z)^{\frac{1}{2}n_3-1} u^{\frac{1}{2}n_2-1}}{(1 + u)^{\frac{1}{2}(n_1+n_2)}} \, dz \, du,$$

where

$$x_1 = \frac{1 + u}{1 + u(1 + d)}.$$

The binomial expansion of $(1 - z)^{(n_3/2)-1}$ gives us

$$P_1 = k \int_a^\infty \int_0^{x_1} \frac{u^{\frac{1}{2}n_2-1} f(z)}{(1 + u)^{\frac{1}{2}(n_1+n_2)}} \, dz \, du,$$

where

$$f(z) = \sum_{j=0}^{\frac{1}{2}n_3-1} (-1)^j \binom{n_3/2 - 1}{j} z^{\frac{1}{2}(n_1+n_2)+j-1}$$

Upon performing the integrations with respect to u and z, we obtain

$$P_1 = \sum_j \left[\frac{(-1)^{j-1} \binom{n_3/2 - 1}{j - 1}}{[(n_1 + n_2)/2 + j - 1] B(n_1/2, n_2/2) B(n_3/2, (n_1 + n_2)/2)(1 + d)^{n_2/2}} \right]$$
$$\times \left[\sum_{r=0}^{j-1} \frac{\binom{j - 1}{r} B(n_1/2 + j - 1 - r, n_2/2 + r) \cdot I_{x_2}(n_1/2 + j - 1 - r, n_2/2 + r)}{(1 + d)^r} \right],$$

where

(12) $$x_2 = \frac{1}{1 + a + ad}.$$

We now consider

$$P_2 = k \int_0^a \int_{x_3}^\infty \frac{u^{(n_2+n_3)/2-1} v^{n_3/2-1}}{(1 + u + uv)^{(n_1+n_2+n_3)/2}} \, dv \, du,$$

where

$$x_3 = [c(1 + \theta_{21} u)]/u = (c + bu)/u,$$

and

(13) $$b = \frac{u_3^0}{\theta_{32}}.$$

Using procedures similar to those used in deriving P_1, we obtain

$$P_2 = \sum_j \left\{ \left[\frac{(-1)^{j-1}\binom{n_3/2-1}{j-1}}{((n_1+n_2)/2+j-1)B(n_1/2, n_2/2) \cdot B(n_3/2, (n_1+n_2)/2)(1+b)^{n_2/2}(1+c)^{n_1/2}} \right] \times \left[\sum_{r=0}^{j-1} \frac{\binom{j-1}{r} B[(n_2/2)+r, (n_1/2)+j-1-r] \cdot I_{x_4}[(n_2/2)+r, (n_1/2)+j-1-r]}{(1+b)^r(1+c)^{j-1-r}} \right] \right\},$$

where

$$x_4 = \frac{a(1+b)}{1+c+a(1+b)}. \tag{14}$$

2.3.2 Recurrence formulas. Integrating P_1 (as originally given in section 2.2) by parts with respect to v, we obtain

$$P_1(n_3) = \frac{(d)^{n_3/2-1}}{(n_1+n_2+n_3)/2-1} k \int_a^\infty \frac{u^{(n_2+n_3)/2-2}}{(1+u(1+d))^{(n_1+n_2+n_3)/2-1}} du + \frac{(n_3/2-1)k}{(n_1+n_2+n_3)/2-1} P_1(n_3-2).$$

Upon integrating with respect to u, we obtain

$$P_1(n_3) = \frac{(d)^{n_3/2-1} I_{x_2}(n_1/2, (n_2+n_3)/2-1)}{(n_3/2-1)B(n_3/2-1, n_2/2)(1+d)^{(n_2+n_3)/2-1}} + P_1(n_3-2), \tag{15}$$

where x_2 is given by (12). For the set of initial values at $n_3 = 2$ it is found that

$$P_1(2) = \frac{I_{x_2}(n_1/2, n_2/2)}{(1+d)^{n_2/2}}. \tag{16}$$

The recurrence development for P_2 is similar but more cumbersome. We obtain the relation

$$P_2(n_1, n_3) = \frac{1}{1+c} \frac{(1/a)^{n_1/2-1} I_{x_5}((n_1+n_2)/2-1, n_3/2)}{((n_1+n_2)/2-1)B(n_1/2, n_2/2)(1+1/a)^{(n_1+n_2)/2-1}} + c \cdot P_2(n_1, n_3-2) + P_2(n_1-2, n_3), \tag{17}$$

where

$$x_5 = \frac{1+(1/a)}{1+(1/a)+b+(c/a)}. \tag{18}$$

The formulas for the initial values are found to be

$$P_2(n_1, 2) = \frac{I_{x_4}[(n_2/2), (n_1/2)]}{(1+b)^{n_2/2}(1+c)^{n_1/2}} \tag{19}$$

and

$$P_2(2, n_3) = \frac{1}{1+c}\left\{\frac{I_{x_2}[(n_2/2), (n_3/2)]}{[1+(1/a)]^{n_2/2}} + c\cdot P_2(2, n_3 - 2)\right\}, \tag{20}$$

where x_2 and x_4 are given by (12) and (14), respectively.

2.4 *Approximate formulas.* We now derive simpler approximate formulas. We first consider P_2. Writing $F_1 = F_{n_2,n_1}(\alpha_1)$, $F_2 = F_{n_3,n_2}(\alpha_2)$, $F_3 = F_{n_3,n_1+n_2}(\alpha_3)$, we have

$$P_2 = \Pr\{V_2/V_1 \leqq F_1 \text{ and } V_3/V \geqq F_3\}.$$

As $n_1 \to \infty$ both $V_1 \to \sigma_1^2$ and $V \to \sigma_1^2$ and, in the limit, the two ratios V_2/V_1 and V_3/V are independently distributed. It is therefore suggested that for large n_1 we use the approximation

$$\begin{aligned} P_2 &\doteqdot \Pr\{V_2/V_1 \leqq F_1\} \Pr\{V_3/V \geqq F_3\} \\ &\doteqdot [1 - I_{x_6}(\tfrac{1}{2}n_1, \tfrac{1}{2}n_2)] I_{x_7}(\tfrac{1}{2}(n_1+n_2), \tfrac{1}{2}n_3), \end{aligned} \tag{21}$$

where

$$x_6 = 1\Big/\left(1 + \frac{1 - x(\alpha_1)}{\theta_{21}\, x(\alpha_1)}\right),$$

$$x_7 = (n_1+n_2)\Big/\left(n_1 + n_2 + \frac{(n_2\theta_{21} + n_1)(1 - x(\alpha_3))}{\theta_{21}\theta_{32}\, x(\alpha_3)}\right),$$

and $x(\alpha_1)$, $x(\alpha_3)$ are respectively the roots, x, of $I_x(\frac{1}{2}n_1, \frac{1}{2}n_2) = \alpha_1$ and $I_x(\frac{1}{2}(n_1+n_2), \frac{1}{2}n_3) = \alpha_3$. Here we have used the well-known relation between the incomplete Beta function $I_x(a, b)$ and the F-integral, viz,

$$\Pr\{F_{\nu_1,\nu_2} \leqq F_0\} = I_x(\tfrac{1}{2}\nu_1, \tfrac{1}{2}\nu_2), \text{ with } x = \nu_1 F_0/(\nu_2 + \nu_1 F_0).$$

We have also used the approximation that for large n_1, V is approximately distributed as $(n_1\sigma_1^2 + n_2\sigma_2^2)\chi^2_{n_1+n_2}/(n_1+n_2)^2$.

Next we turn to

$$P_1 = \Pr\{V_2/V_1 \geqq F_1 \text{ and } V_3/V_2 \geqq F_2\}.$$

Here we could use a similar argument if we were to let $n_2 \to \infty$. This limit would however, not yield useful results. The important situation in pooling procedures, is one in which n_2 is moderate or small. Instead we use the well-known normal approximation to $\log V_i$. M. S. Bartlett and D. G. Kendall [3] have shown that $\log V_i$ is approximately $N(\log \sigma_i^2, 2/(n_i - 1))$, provided that n_i is not too small. Writing

$$u = \log V_2 - \log V_1 \text{ and } z = \log V_3 - \log V_2,$$

it follows that the joint distribution of u and z is approximately bivariate normal with correlation coefficient

$$\rho = -1\Big/\left\{\left(1 + \frac{n_2 - 1}{n_3 - 1}\right)\left(1 + \frac{n_2 - 1}{n_1 - 1}\right)\right\}^{\frac{1}{2}}. \tag{22}$$

We may therefore employ the tables of the double probability integral of a bivariate normal surface of K. Pearson [16]. Tables VIII and IX. If x and y follow a bivariate normal distribution with both means equal to 0, correlation coefficient ρ, and both standard deviations equal to unity, then these tables give the probabilities $P_\xi(h, k)$, for $x \geqq h$ and $y \geqq k$. In our case, ρ is given by (22) and h and k by

$$h = \frac{2z_{n_2,n_1}(\alpha_1) - \log \theta_{21}}{[2/(n_1 - 1) + 2/(n_2 - 1)]^{\frac{1}{2}}}, \qquad k = \frac{2z_{n_3,n_2}(\alpha_2) - \log \theta_{32}}{[2/(n_2 - 1) + 2/(n_3 - 1)]^{\frac{1}{2}}},$$

where $z_{n_i,n_j}(\alpha)$ is the upper 100α per cent point of Fisher's z distribution with numerator degrees of freedom n_i and denominator degrees of freedom n_j.[6]

2.5 *Theory of reduction of mixed model to random model.* Certain mixed models of analysis of variance were described in Section 1.3. No new formulas are required for these models, as we shall show that the joint distribution of the three mean squares is, at least approximately, equal to that of the component of variance model. The exact specifications of the distribution for the mixed model being considered are as follows. (Primed parameters will be used to specify the parameters for the mixed model.)

(a) The error mean square V_2 and the doubtful error mean square V_1 are distributed as $\chi_i^2\sigma_i^2/n_i'$ $(i = 1, 2)$, where χ_i^2 is the central χ^2 statistic with n_i' degrees of freedom. On the other hand, the treatment mean square V_3 is distributed as $\chi_{n_3}'^2\sigma_2^2/n_3'$, where $\chi_{n_3}'^2$ is the noncentral χ^2 statistic with n_3' degrees of freedom and noncentrality parameter

$$\lambda = \frac{n_3'\sigma_3^2 - n_3'\sigma_2^2}{2\sigma_2^2} = \frac{n_3'}{2}(\theta_{32}' - 1),$$

where $\theta_{32}' = \sigma_3^2/\sigma_2^2$. V_1, V_2, and V_3 are independent.

(b) The main purpose of the analysis is to test the hypothesis $H_0 : \sigma_3^2 \equiv \sigma_2^2$ against the alternative $H_1 : \sigma_3^2 > \sigma_2^2$.

(c) The true error mean square, V_2, has an expectation σ_2^2 which is greater than or equal to the expectation, σ_1^2, of the doubtful error mean square.

The probability P of rejecting H_0 is obtained as the sum of the two components,

$$P_1 = \Pr\{V_2/V_1 \geqq F_{n_2',n_1'}(\alpha_1') \text{ and } V_3/V_2 \geqq F_{n_3',n_2'}(\alpha_2')\} \tag{23}$$

and

$$P_2 = \Pr\{V_2/V_1 < F_{n_2',n_1'}(\alpha_1') \text{ and } V_3/V \geqq F_{n_3',n_1'+n_2'}(\alpha_3')\}. \tag{24}$$

In evaluating these probabilities we use, the approximation first used by Patnaik [13]. We replace $\chi_{n_3}'^2$ by $C\chi_{\nu_3}^2$, $\chi_{\nu_3}^2$ being the central χ^2 statistic based upon ν_3 degrees of freedom, where

$$\nu_3 = n_3' + \frac{4\lambda^2}{n_3' + 4\lambda} \text{ df}$$

[6] See Appendix Table 4 for illustrations of the nature of the approximation to the integral P_1.

TABLE 4

Modified parameters for random model corresponding to specified parameters for mixed model

Specified parameters for mixed model	Modified parameters for random model
n_1'	$n_1 = n_1'$
n_2'	$n_2 = n_2'$
n_3'	$n_3 = \nu_3 = n_3' + \dfrac{4\lambda^2}{n_3' + 4\lambda}$
α_1'	$\alpha_1 = \alpha_1'$
α_2'	$\alpha_2 = \text{Root of } F_{n_3', n_2'}(\alpha_2') = F_{\nu_3 n_2'}(\alpha_2)$
α_3'	$\alpha_3 = \text{Root of } F_{n_3', n_1'+n_2'}(\alpha_3') = F_{\nu_3, n_1'+n_2'}(\alpha_3)$
$\theta_{21}' = \sigma_2^2/\sigma_1^2$	$\theta_{21} = \theta_{21}'$
$\theta_{32}' = \sigma_3^2/\sigma_2^2$	$\theta_{32} = (2\lambda + n_3')/n_3'$

and

$$C = 1 + \frac{2\lambda}{n_3' + 2\lambda}.$$

Since the use of this approximation reduces the noncentral χ^2 statistic to a central χ^2 statistic, all three statistics are now central.

We now compare the power for the mixed model as defined by (23) and (24) with the corresponding formulas (8) and (9), for modified values of the eight parameters as indicated in Table 4. Entering the random model tables with these altered parameters we obtain the mixed model power. It will be seen that when we deal with the size for the mixed model we have $\lambda = 0$ and hence $\nu_3 = n_3'$, so that all primed parameters agree with those without primes. Thus our entire size discussion to follow is directly applicable to the mixed model. On the other hand, the power evaluations, which refer to $\alpha_2 = \alpha_3 = .05$, will in general provide answers for larger values of α_2' and α_3', and these levels α_2' and α_3' will vary with λ. For a proper evaluation of power corresponding to a given pair of significance levels α_2' and α_3', say, $\alpha_2' = \alpha_3' = .05$, a more extensive tabulation of (8) and (9) as described in the subsequent section, would be required.

2.6 *Application of derived formulas.* The recurrence formulas derived in Section 2.3.2 were used to construct master tables of P_1 and P_2. These master tables were constructed for

$$\frac{n_2}{2} = 5, \qquad \frac{n_1}{2} = 3(1)\ 10, \qquad \text{and } \frac{n_3}{2} = 1(1)\ 6.$$

Also, tables were constructed for $\frac{1}{2}n_2 = 3$, in order that the effect of small error degrees of freedom could be better studied. However, the latter tables were confined to the values

$$\tfrac{1}{2}n_3 = 1 \quad \text{and} \quad \tfrac{1}{2}n_1 = 4, 7, \quad \text{and} \quad 10.$$

To compute the power component P_1 in a given problem from these master tables, for specified degrees of freedom n_1, n_2, and n_3 and levels of significance α_1, α_2, the values of the parameters u_1^0 and u_2^0 are computed from (10). From these values and those specified for θ_{21} and θ_{32}, the corresponding values of a and d from (11) and hence the value of x_2 from (12) are computed. P_1 is then obtained by interpolation in the appropriate master table.

The procedure used to compute the component P_2 was similar, and required evaluation of an interpolation for the parameters a, b, c, , x_1 and x_2 ; but interpolation with respect to a was avoided by choosing values of θ_{21} which would result in tabular values of a. This accounts for the decimal values of θ_{21} found in our tables.

The approximate formula (21) for P_2 derived in Section 2.5 is exact for $n_1 = \infty$ and was found to be very effective for large n_1, yielding either P_1 values directly to sufficient accuracy, or facilitating extrapolation of the master tables.

3. Discussion of power and size curves and comparison of test procedures.

3.1 *Type of recommendations attempted.* We have seen that the power of our test procedures depends upon the following eight parameters: the degrees of freedom n_1, n_2, and n_3 ; the variance ratios $\theta_{21} = \sigma_2^2/\sigma_1^2$, $\theta_{32} = \sigma_3^2/\sigma_2^2$; and the levels of significance α_1, α_2, α_3. Of these, the degrees of freedom n_1, n_2, and n_3 are completely determined by the analysis of variance table, while the variance ratios are generally unknown (except in the case of the size of the procedure, when $\theta_{32} = 1$). Any recommendations that are to be made must therefore be confined to the levels of significance, α_1, α_2, and α_3. We shall here be primarily concerned with the size of the procedure being in the vicinity of .05. It will be apparent from what follows that a convenient way to achieve this is to choose $\alpha_2 = \alpha_3 = .05$, that is, to choose procedures in which the significance levels of both final tests are .05. However, the remaining parameter, α_1, the level of significance for the preliminary test, is entirely at our disposal. In attempting recommendations, therefore, we shall be concerned with the choice of the level of α_1. Should α_1 be, say, .05, .25, .50, or should we use what Paull ([14], p. 4; [15]) has called the borderline test, where α_1 will be near .70 to .80? In choosing the level of α_1, we shall consider

(i) the variation in the size of our test procedure as a function of the parameter θ_{21}, and

(ii) a comparison of the power of our test procedure with that of the never-pool test of the same size.

3.2 *Size.* The size of our test procedure does not equal the nominal level of .05, but varies about this level as α_1 and θ_{21} vary. Figures 1 to 10[7] give us examples of size curves, illustrating the variations in type one error with variation in θ_{21} for fixed values of the remaining parameters.

---

[7] A selection of figures and tables has been assembled in the Appendix. Additional size and power curves and tables, illustrating the points to be made in the ensuing discussion will be found in [8].

Note that as θ_{21} becomes large, the size approaches .05; for, as $\theta_{21} \to \infty$, the preliminary test will almost certainly be significant, pooling will almost certainly not occur, and hence the final test will almost certainly be that of V_3/V_2, having a size of .05.

At the lower extreme, that is, at $\theta_{21} = 1$, the size is at its minimum, which is less than .05. This minimum, and even more so the size peak, are points of particular interest.

We first consider the size peak. Referring to the size curves for a preliminary test carried out at the 5% level (see Figs. 1 and 2), we note that the peak is usually very high. Clearly, a preliminary test carried out at this level will in many cases admit an unacceptable size disturbance. This is due to the fact that at this level, the preliminary test will frequently admit pooling V_2 and V_1 when σ_1^2 is smaller than the true error mean square σ_2^2, and thereby increase the probability of type one error. We therefore seek a preliminary test in which pooling is admitted less readily; we next investigate the level $\alpha_1 = .25$. At this level (compare Figs. 2 and 3), size control is considerably better, and in many

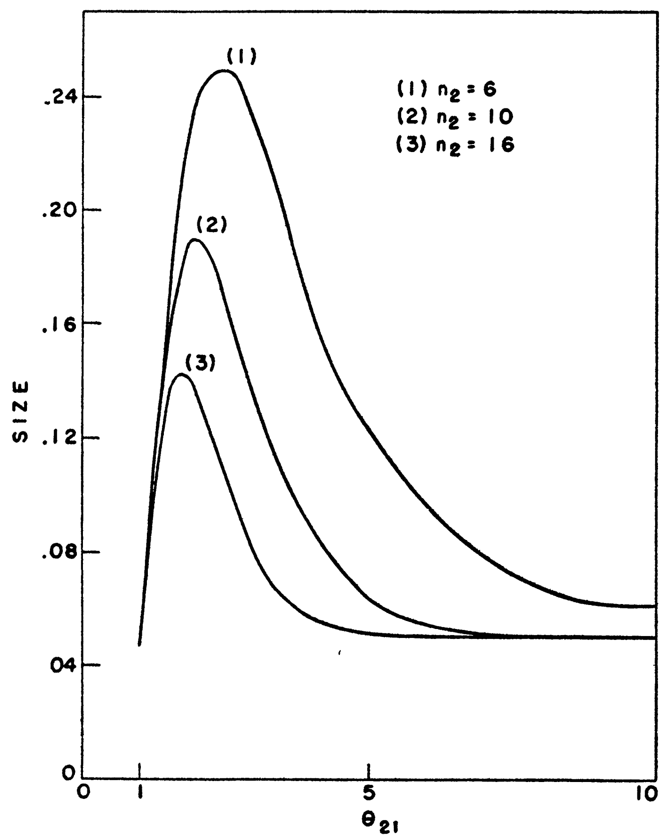

FIG. 1. Size curves for $n_1 = \infty$, $n_3 = 6$, $\alpha_1 = \alpha_2 = \alpha_3 = .05$

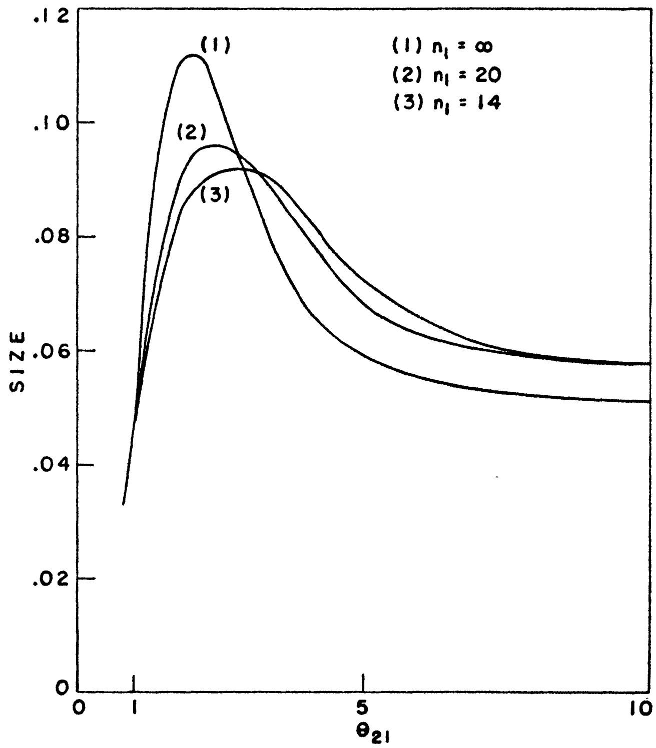

FIG. 2. Size curves for $n_3 = 2$, $n_2 = 10$, $\alpha_1 = \alpha_2 = \alpha_3 = .05$

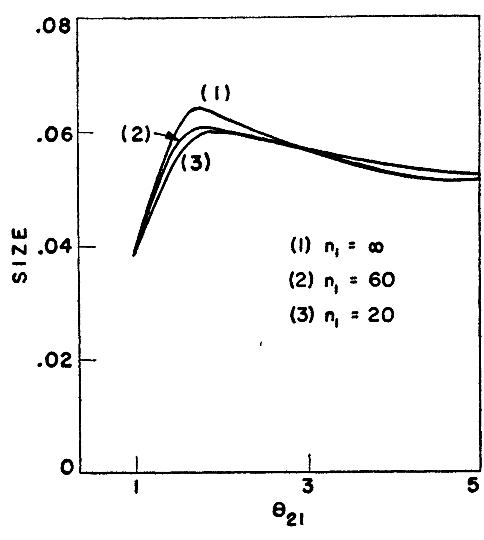

FIG. 3. Size curves for $n_3 = 2$, $n_2 = 10$, $\alpha_1 = .25$, $\alpha_2 = \alpha_3 = .05$

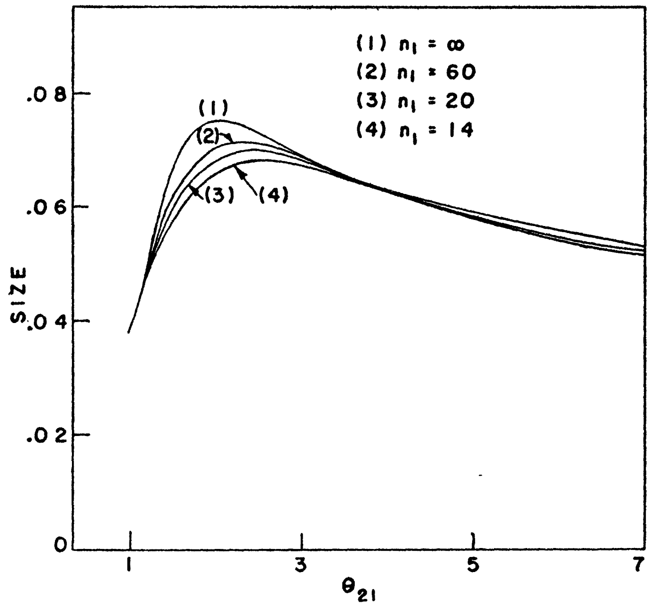

FIG. 4. Size curves for $n_3 = 2$, $n_2 = 6$, $\alpha_1 = .25$, $\alpha_2 = \alpha_3 = .05$

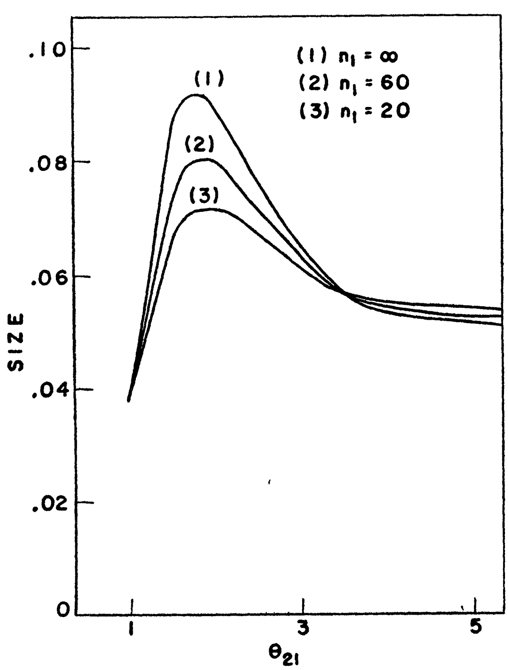

FIG. 5. Size curves for $n_3 = 6$, $n_2 = 10$, $\alpha_1 = .25$, $\alpha_2 = \alpha_3 = .05$

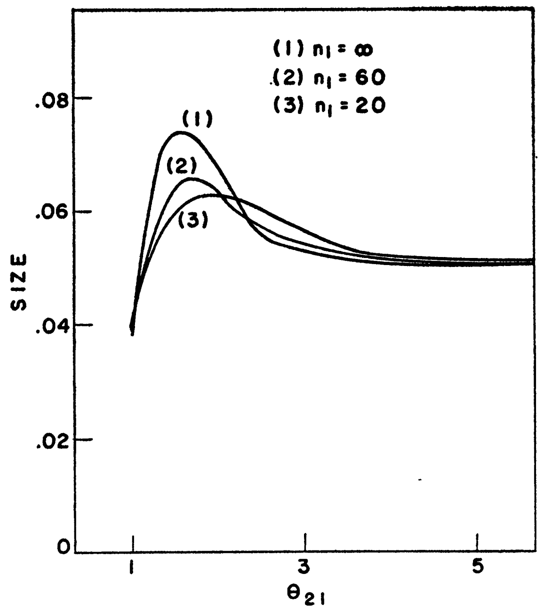

FIG. 6. Size curves for $n_3 = 6$, $n_2 = 16$, $\alpha_1 = .25$, $\alpha_2 = \alpha_3 = .05$

cases the peaks do not go beyond .08. (See, e.g., Figs. 3, 4, 5, and 6.) It is observed that, in general, the size peak increases as n_1 or n_3 increases or as n_2 decreases. (See Figs. 1 through 6.)

It is of course quite arbitrary to specify any rules for maintaining an acceptable upper tolerance for the size peak, since what is considered acceptable is a matter of opinion. In using a nominal size of .05, if we stipulate that our size peak should not go much beyond 10 per cent, then we find that even with the 25 per cent preliminary level, there are situations in which this upper limit is exceeded. Generally speaking, these unacceptable size peaks occur when

$$n_3 \geqq n_2 \quad \text{and} \quad n_1 \geqq 5n_2 . \tag{25}$$

(It should be noted that the occurrence of $n_3 > n_2$ is clearly rare.) This means that when the treatment degrees of freedom are greater than or equal to the error degrees of freedom, we must be careful if at the same time the doubtful error degrees of freedom are greater than or equal to five times the true error degrees of freedom; or, briefly, we must be careful when pooling promises a large gain in the precision of the error estimate. This rule has been established by an empirical study of an extensive number of size curves, and is not based on any analytic study. See, for example, Fig. 7. Here the situation represented by Curve 1 would be excluded by our rule. See also Figs. 8 and 9, in which the situations represented by Curves 1 and 2 would be excluded. If our rule is followed, size disturbances such as are represented in Figs. 3, 4, 5, and 6 occur; and also disturbances such as are represented in Figs. 7, 8, and 9 occur, with the

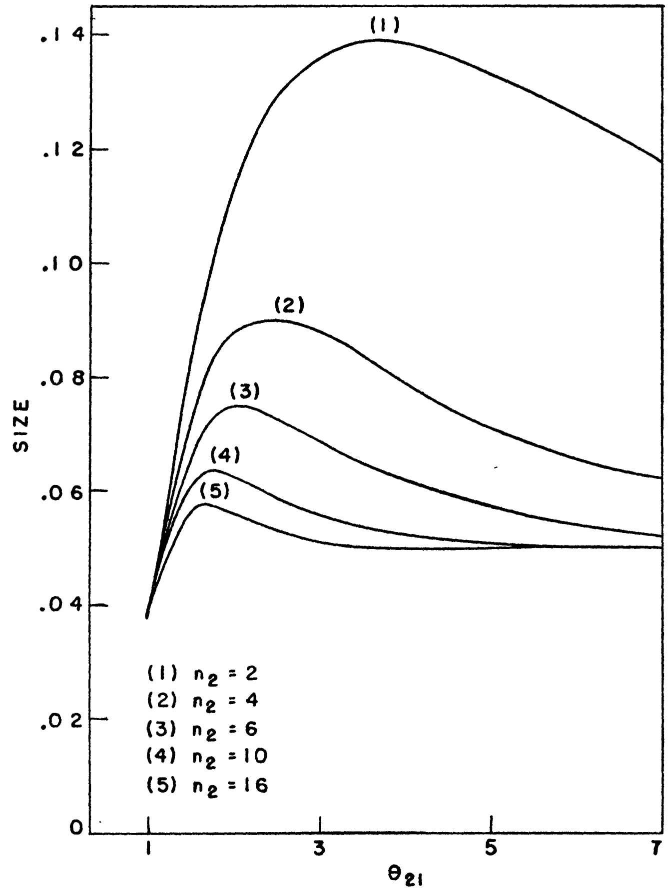

FIG. 7. Size curves for $n_3 = 2$, $n_1 = \infty$, $\alpha_1 = .25$, $\alpha_2 = \alpha_3 = .05$

exception of the excluded cases mentioned above. In the situations represented by (25), a more conservative level of α_1 would be appropriate. From a study of a number of size curves it appears that a preliminary test at the 50 per cent level will ensure adequate control of the size peak in these cases. (See Fig. 9.)

Not only the size peak, but also the size minimum is affected by the level of the preliminary test. From theorems proved by Paull ([14], Chap. 4; [15]), we know that the size of our test procedures is a minimum with respect to θ_{21} at θ_{21} equal to one, and that a lower bound for the size for this value of θ_{21} is

$$(1 - \alpha_1)(.05).$$

These lower bounds are .0475, .0375, and .025 for $\alpha_1 = .05$, .25, and .50, respectively. For some of our curves the plotted minimum sizes are situated very close to these lower bounds. For the borderline test, where, as proved by Paull ([14], [15]) the size is always less than .05, this lower bound varies in magnitude from approximately .01 to approximately .015. We have computed actual

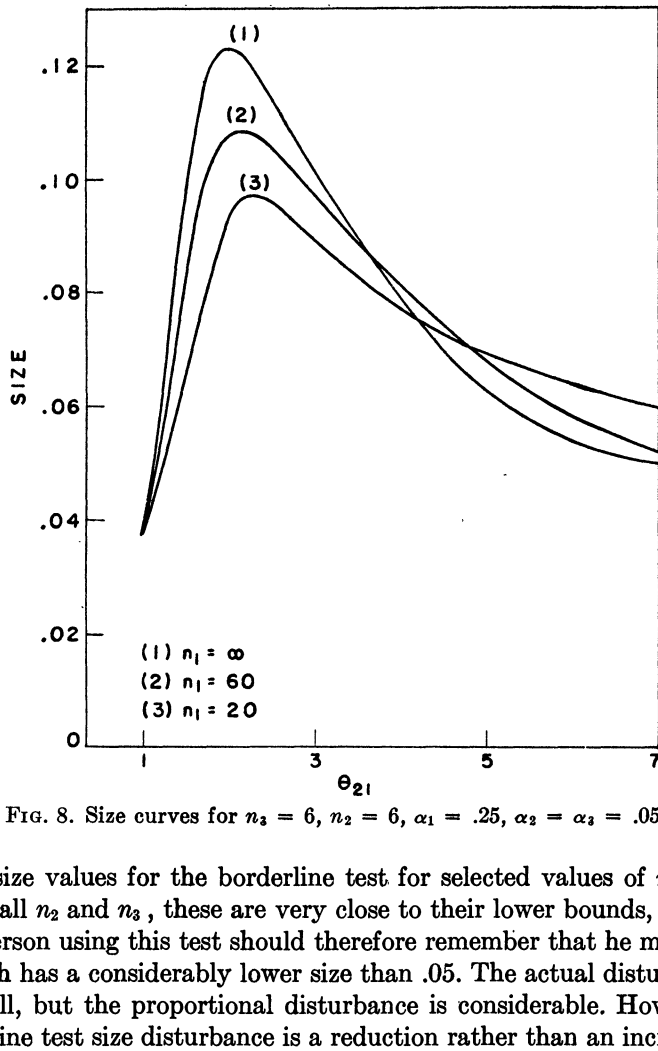

FIG. 8. Size curves for $n_3 = 6$, $n_2 = 6$, $\alpha_1 = .25$, $\alpha_2 = \alpha_3 = .05$

minimum size values for the borderline test for selected values of n_1, n_2, and n_3. For small n_2 and n_3, these are very close to their lower bounds, irrespective of n_1. A person using this test should therefore remember that he may be using a test which has a considerably lower size than .05. The actual disturbance is of course small, but the proportional disturbance is considerable. However, since the borderline test size disturbance is a reduction rather than an increase in size and is therefore on the conservative side, we are not attempting to make any definite rules as to when the experimenter should avoid the use of this test, but merely to remind him that large proportional size disturbances occur when n_2 and n_3 are both small ($\leqq 6$).

Summarizing our considerations of size control, therefore, we have narrowed down our recommendable range of α_1 to $\alpha_1 \geqq .25$, with the reservations that in certain cases characterized by inequalities (25), $\alpha_1 = .25$ would not be desirable, as it would admit too large a peak in the size curve; and that for very small values of n_2 and n_3 the experimenter may not wish to use the borderline test, as this would admit too low a size minimum.

The discussion thus far has been concerned with test procedures in which

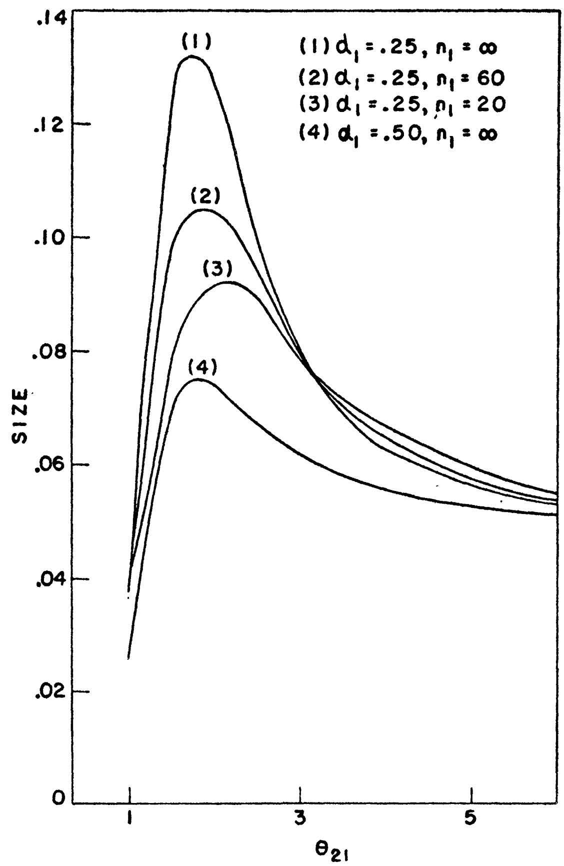

FIG. 9. Size curves for $n_3 = 12$, $n_2 = 10$, $\alpha_2 = \alpha_3 = .05$

$\alpha_2 = \alpha_3 = .05$. A few special cases for $\alpha_2 = \alpha_3 = .01$ and $\alpha_1 = .25$ have also been investigated. In all these situations, larger proportional size disturbances than those found for $\alpha_2 = \alpha_3 = .05$ were experienced, even for cases which our rule would accept. (See Fig. 10.)

3.3 *Frequency of pooling.* We have been discussing the effect of increasing α_1 in order to achieve size control. It is obvious that for $\alpha_1 = 1$, our preliminary F per cent point would be zero, and pooling would never occur. The question arises as to the relative frequency of pooling for the intermediate values of α_1 that we have been considering. When $\theta_{21} = 1$, the probability that V_2/V_1 exceeds $F_{\alpha_1}(n_2, n_1)$ is α_1, so that pooling occurs with relative frequency $1 - \alpha_1$. As θ_{21} increases, this frequency rapidly decreases, approaching the limit zero as θ_{21} becomes infinite. Evaluations of these frequencies of pooling show that, while for $\alpha_1 = .25$ and small values of θ_{21}, pooling will occur in the majority of experiments, when $\alpha_1 = .50$ the frequency is usually well below .50. This frequency of pooling is of course even smaller for the borderline test, where α_1

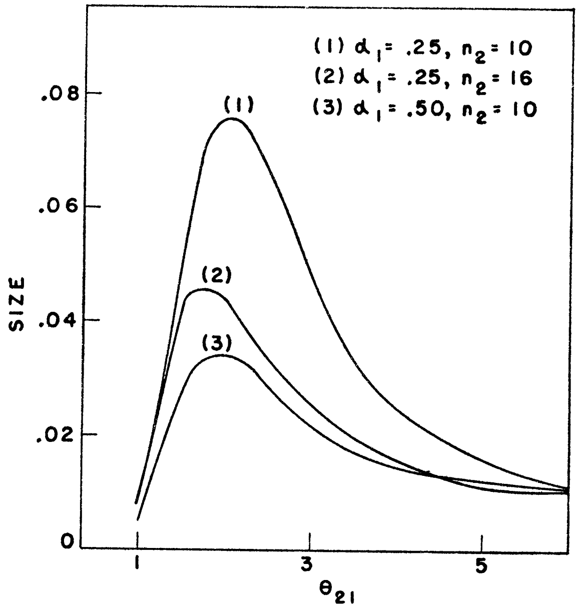

FIG. 10. Size curves for $n_3 = 12$, $n_1 = \infty$, $\alpha_2 = \alpha_3 = .01$

usually takes on values in the neighborhood of .7 to .8; when such large values of α_1 are employed, pooling occurs in only about 25 per cent of all situations for which $\theta_{21} = 1$, and this pooling percentage rapidly decreases as θ_{21} increases. While this property by itself cannot be regarded as a disadvantage of the borderline test, it is clear that, if this test were the only one recommended to the experimenter, he would hardly ever pool.

3.4 *Power*. We now attempt a comparison of the power of our sometimes-pool procedure with that of the never-pool test. As is well known, any comparison of power of any two test procedures is a fair comparison if the two test procedures have the same size. We have seen that the size of our sometimes-pool procedures is not at the constant level of .05, but varies about this, depending upon the parameter θ_{21}. The method of power comparison we have therefore adopted is as follows:

(i) Assume a fixed value of the parameter θ_{21}.

(ii) For this value of θ_{21}, evaluate the size of the sometimes-pool test.

(iii) For this level of size, evaluate the power curve of the never-pool test; this power is then directly comparable with that of the sometimes-pool test corresponding to the chosen value of θ_{21}.

For an illustration of such comparisons, see Table 1. Here we have $n_1 = 20$, $n_2 = 6$, $n_3 = 2$, $\alpha_1 = .25$ and $\alpha_2 = \alpha_3 = .05$. For $\theta_{21} = 1$, the sometimes-pool procedure is always more powerful than the never-pool test of the same size. For $\theta_{21} = 1.5$, the powers are very similar; on the other hand, for $\theta_{21} = 2$, the never-pool test is always more powerful. See also Tables 2 and 3, which again

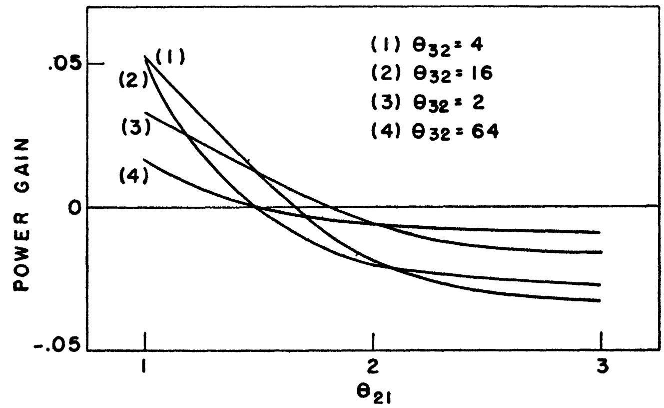

FIG. 11. Power gain of the sometimes-pool procedure over the never-pool test of the same size for $n_1 = 20$, $n_3 = 2$, $n_2 = 6$, $\alpha_1 = \alpha_2 = \alpha_3 = .05$.

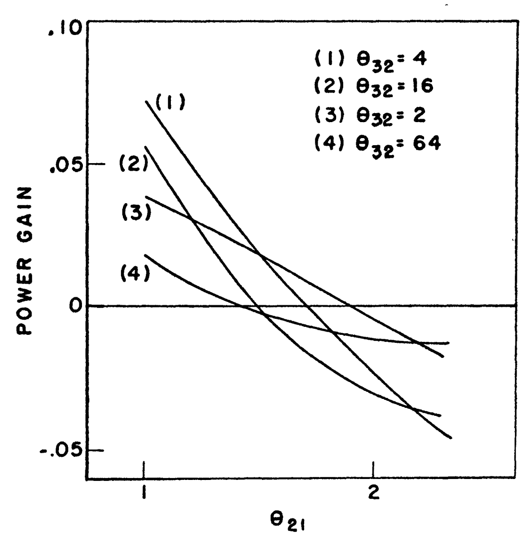

FIG. 12. Power gain of the sometimes-pool procedure over the never-pool test of the same size for $n_1 = 20$, $n_3 = 2$, $n_2 = 6$, $F_1 = 2\,F_{.50}$, $\alpha_2 = \alpha_3 = .05$.

illustrate the fact that the sometimes-pool procedure is more powerful for small θ_{21} but less powerful for large θ_{21}.

In order to show more clearly the dependence of these power differences on θ_{21}, we have plotted in Figs. 11, 12 and 13 the difference between two corresponding power points against θ_{21}. Here each curve corresponds to a fixed value of θ_{32}—i.e., that value of θ_{32} at which the difference between the power ordinates of the power curves was taken. It will be seen, again, that for small θ_{21} the differences are positive (the sometimes-pool procedure is more powerful than the never-pool test), while for larger θ_{21} the position is reversed. As $\theta_{21} \to \infty$, the difference tends to 0, since both procedures tend to the never-pool test at the

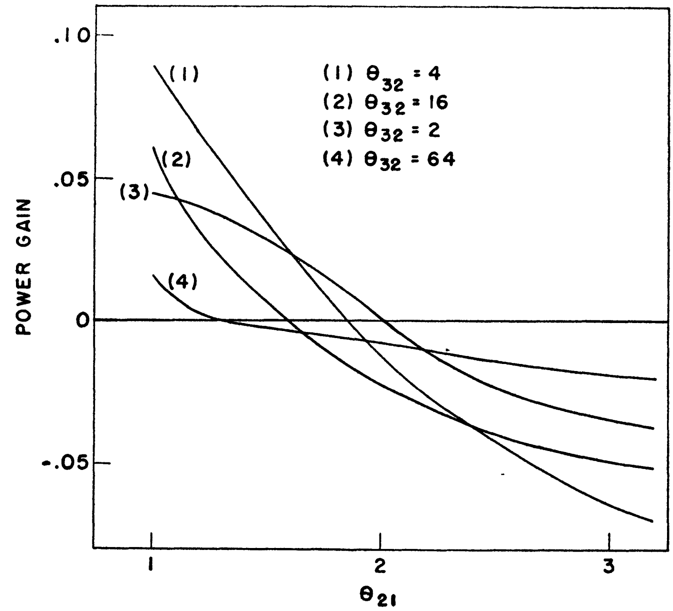

FIG. 13. Power gain of the sometimes-pool procedure over the never-pool test of the same size for $n_1 = 20$, $n_3 = 2$, $n_2 = 6$, $\alpha_1 = .25$, $\alpha_2 = \alpha_3 = .05$.

.05 level of significance. The transition from favorable to unfavorable power conditions generally occurs between $\theta_{21} = 1.5$ and $\theta_{21} = 2.0$. From other similar curves not shown here, it is seen that the magnitude of these power gains and losses increases with increasing n_3, or decreasing n_2, or increasing n_1.

Figures 11, 12, and 13 also illustrate the effect of decreasing the per cent point of F for the preliminary test. In these figures, corresponding power comparisons are given, respectively, for $\alpha_1 = .05$, $F_1 = 2F_{.50}(n_2, n_1)$,[8] and $\alpha_1 = .25$. There is a general tendency for both power gains and power losses to diminish as the per cent point of F decreases—i.e., as α_1 increases from values such as .05 through intermediate values such as .25 to the level of the borderline test (approximately $\alpha = .70$ to .80). Here the gain in power has diminished further, but the power losses have completely disappeared. In fact, a theorem by Paull ([14], p. 61; [15]) proves that the borderline test is always more powerful than the corresponding never-pool test of the same size, although the power gain is small for large θ_{21}. However, as stated in Section 3.2, this size is below the nominal level of .05. If we compare the borderline test power with that of the never-pool test at the nominal level of .05, the former is always less powerful. It is likewise less powerful than our sometimes-pool procedures for $\alpha_1 = .50$ and $\alpha_1 = .25$, which have, of course, a larger size.

We now attempt recommendations, considering the relative merits of the procedures at $\alpha_1 = .25$, $\alpha_1 = .50$, and $\alpha_1 = .7$ to .8 (the borderline level). These recommendations are somewhat subjective, since they are contingent upon what

[8] This means we use a preliminary test $V_2/V_1 \geqq 2\ F_{.50}(n_2, n_1)$.

the experimenter may regard as a reasonable assumption concerning the parameter θ_{21}.

(i) If the experimenter is reasonably certain that only small values of θ_{21} can be envisaged as a possibility, he is advised to use $\alpha_1 = .25$ except in the cases (25) when he should use $\alpha_1 = .50$, in order to ensure size control. Our figures show that the range of small values of θ_{21}, when the sometimes-pool procedure gives a gain in power, is approximately between 1 and 1.5 to 2. An experimenter about to adopt this recommendation but not quite certain about his assumptions may wish to know the consequences which result from his adopting this procedure when, in fact, unknown to him, θ_{21} is large. It is seen from Figs. 1 to 10 that in such a situation he will still have control of the size of his test; in fact the size will be near .05 for large θ_{21}. All he loses (as is illustrated by our power figures and tables) is the power of his test; this is a risk that he may well be prepared to take.

(ii) If, however, the experimenter can make no such assumption about θ_{21}, and wishes to guard against the possibility of power losses, he may then use the borderline test, which would ensure a power gain, although he must realize

(a) that for large θ_{21} this gain would be very small;

(b) that for small θ_{21} he would use a test procedure of a very much smaller size than $\alpha_1 = .05$ (particularly when n_2 and n_3 are $\leqq 6$) and accordingly a test which is much less powerful than the never-pool test of size .05. In fact, he may in these circumstances prefer not to pool at all.

It may be correctly argued that, in order to control the size peak, to advocate $\alpha_1 = .50$ in the cases characterized by (25), and $\alpha_1 = .25$ otherwise, introduces an artificial discontinuity in our recommendations. It would be quite feasible (although it would require a considerable effort in computation) to evaluate for any given triplet n_1, n_2, and n_3 that value of α_1 which results in a size peak of 0.10 exactly. Since this level of α_1 would depend on the degrees of freedom n_1, n_2, and n_3, it would be necessary to evaluate the associated per cent points of F. For such recommendations to be useful, this table of $F_{\alpha_1}(n_1, n_2)$ (which would be a large 3 parametric table with n_1, n_2, and n_3 as arguments) would have to be published. To encumber the experimenter with special tables for the preliminary F-test in addition to the standard F-tables for the final F-tests appeared to us to be unnecessary, and the use of the published Merrington and Thompson [11] 25% and 50% points of F preferable.

We should note here that a rule favored by Paull ([14], Chap. 6; [15]) advocating testing the ratio V_2/V_1 against $2F_{.50}(n_2, n_1)$ will not ensure adequate control of the size peak, since $2F_{.50} > F_{.25}$ in general, and we have just seen that $F_{.25}$ is sometimes too large and hence not always acceptable as a significance level for the preliminary test. Also, it would appear to us that no rule of the form $V_2/V_1 >$ constant is very satisfactory, for with such a rule the frequency with which pooling occurs, as well as the size, varies considerably with the degrees of freedom n_1 and n_2.

Concerning recommendation ii, the experimenter would require knowledge of the precise level of α_1 for the borderline test, or, better still, the value of F associated with it. Paull ([14], p. 20; [15]) gives a simple formula from which the following is derived: F point for borderline test equals

$$\frac{(n_1 F_{n_3, n_1+n_2}(\alpha_3)}{(n_1 + n_2)(F_{n_3, n_2})(\alpha_2) - n_2 F_{n_3, n_1+n_2}(\alpha_3)},$$

where $F_{n_3, n_2}(\alpha_2)$ represents the 100 α_2 per cent point of F with numerator df n_3 and denominator df n_2. Similar statements can be made for the other symbols.

It has been noted that the above recommendations depend upon some a priori information regarding θ_{21}. It is shown in a number of examples discussed in the Wright-Patterson report how this information can often be obtained from the general conditions under which the experiments were carried out.

APPENDIX

Figures and Tables

TABLE 1

The power of the sometimes-pool procedure and the never-pool test of the same size, for $n_1 = 20$, $n_2 = 6$, $n_3 = 2$, $\alpha_1 = .25$, $\alpha_2 = \alpha_3 = .05$

θ_{21}	Test	θ_{32}				
		1	2	4	16	64
1	s.p.	.038	.161	.368	.757	.930
	n.p.	.038	.127	.314	.705	.913
1.5	s.p.	.060	.189	.385	.756	.930
	n.p.	.060	.178	.373	.757	.930
2	s.p.	.068	.190	.377	.751	.935
	n.p.	.068	.195	.396	.771	.935
3	s.p.	.068	.178	.361	.743	.926
	n.p.	.068	.194	.394	.770	.935
5	s.p.	.058	.164	.348	.738	.924
	n.p.	.058	.175	.369	.754	.930
	n.p.	.050	.156	.343	.737	.924

TABLE 2

The power of the sometimes-pool procedure and the never-pool test of the same size, for $n_1 = 20$, $n_2 = 10$, $n_3 = 12$, $\alpha = .25$, $\alpha_2 = \alpha_3 = .05$

θ_{21}	Test	θ_{32}				
		1	2'	4	10	50
1.066	s.p.	.044	.354	.741	.979	1.000
	n.p.	.044	.259	.679	.973	1.000
1.708	s.p.	.086	.360	.714	.976	1.000
	n.p.	.086	.389	.799	.988	1.000
2.914	s.p.	.079	.292	.701	.975	1.000
	n.p.	.079	.369	.784	.987	1.000
5.399	s.p.	.050	.280	.697	.975	1.000
	n.p.	.050	.280	.702	.976	1.000

TABLE 3

The power of the sometimes-pool procedure and the never-pool test of the same size, for $n_1 = 14$, $n_2 = 10$, $n_3 = 12$, $\alpha_1 = .25$, $\alpha_2 = \alpha_3 = .05$

θ_{21}	Test	θ_{32}				
		1	2	4	10	50
0.714	s.p.	.013	.232	.708	.971	1.000
	n.p.	.013	.111	.448	.911	1.000
1.045	s.p.	.039	.294	.757	.980	1.000
	n.p.	.039	.240	.657	.969	1.000
1.593	s.p.	.075	.360	.727	.977	1.000
	n.p.	.075	.358	.774	.986	1.000
2.552	s.p.	.082	.311	.704	.975	1.000
	n.p.	.082	.377	.790	.988	1.000
8.066	s.p.	.050	.280	.699	.975	1.000
	n.p.	.050	.280	.702	.976	1.000

TABLE 4.

Illustrating the nature of the approximation to the integral P_1 ($n_1 = 20$ throughout)

	θ_{21}	$n_2 = 6$			$n_2 = 10$			$n_2 = 16$		
		Exact	Approx.	Diff.	Exact	Approx.	Diff.	Exact	Approx.	Diff.
$n_3 = 2$	1.0	.0375	.0375	.0000	.0375	.0375	.0000	.0375	.0375	.0000
	1.5	.0574	.0530	.0044	.0484	.0441	.0043	.0400	.0363	.0037
	2.0	.0621	.0528	.0093	.0444	.0364	.0080	.0304	.0245	.0059
	3.0	.0524	.0396	.0128	.0274	.0190	.0084	.0128	.0084	.0044
	5.0	.0288	.0187	.0101	.0088	.0050	.0038	.0020	.0010	.0010
$n_3 = 6$	1.0	.0375	.0375	.0000				.0375	.0375	.0000
	1.5	.0692	.0623	.0069				.0512	.0445	.0067
	2.0	.0932	.0757	.0175				.0441	.0329	.0112
	3.0	.0856	.0618	.0238				.0208	.0122	.0086
	5.0	.0463	.0305	.0158				.0034	.0016	.0018

For exact integral P_1, see Eq. (15).
For approximate integral P_1, see Section 2.4 and Formula (22).

REFERENCES

[1] T. A. BANCROFT, "On biases in estimation due to the use of preliminary tests of significance," *Ann. Math. Stat.*, Vol. 15 (1944), pp. 190–204.

[2] T. A. BANCROFT, "Bias due to the omission of independent variables in ordinary multiple regression analysis," (abstract), *Ann. Math. Stat.*, Vol. 21 (1950), p. 142.

[3] M. S. BARTLETT AND D. G. KENDALL, "The statistical analysis of variance—heterogeneity and the logarithmic transformation," *J. Roy. Stat. Soc.*, *Suppl.*, Vol. 8 (1946), pp. 128–138.

[4] R. E. BECHHOFER, "The effect of preliminary tests of significance on the size and power of certain tests of univariate linear hypotheses," unpublished Ph.D. thesis, Columbia University Library, 1951.

[5] B. M. BENNETT, "Estimation of means on the basis of preliminary tests of significance," *Ann. Math. Stat.*, Vol. 4 (1952), pp. 31–43.

[6] G. E. P. BOX, "Some theorems on quadratic forms applied in the study of analysis of variance problems I," *Ann. Math. Stat.*, Vol. 25 (1954), p. 290.

[7] G. E. P. BOX, "Some theorems on quadratic forms applied in the study of analysis of variance problems II," *Ann. Math. Stat.*, Vol. 25 (1954), pp. 484–498.

[8] H. BOZIVICH, T. A. BANCROFT, AND H. O. HARTLEY, WADC Technical Report (In Press), Wright Air Development Center, Wright-Patterson Air Force Base, Ohio, 1956.

[9] R. A. FISHER, *The Design of Experiments*, Oliver and Boyd, London, 1937.

[10] TOSIO KITAGAWA, "Successive process of statistical inference," *Mem. Faculty of Sci.*, Kyusyu University, Series A, Vol. 6, No. 2, 1950.

[11] M. MERRINGTON AND C. M. THOMPSON, "Tables of percentage points of the inverted Beta (F) distribution," *Biometrika*, Vol. 33 (1943), pp. 73–88.

[12] FREDERICK MOSTELLER, "On pooling data," *J. Amer. Stat. Assn.*, Vol. 43 (1948), pp. 231–242.

[13] P. B. PATNAIK, "The non-central χ^2 and F-distributions and their applications," *Biometrika*, Vol. 36 (1949), pp. 202–232.

[14] A. E. PAULL, "On a preliminary test for pooling mean squares in the analysis of variance," unpublished Ph.D. thesis, University of North Carolina, 1948.

[15] A. E. PAULL, "On a preliminary test for pooling mean squares in the analysis of variance," *Ann. Math. Stat.*, Vol. 21 (1950), pp. 539–556.

[16] K. PEARSON, editor, Tables for statisticians and biometricians, Part 2, London, Biometric Laboratory, 1936.