BU - 321 - M

Best Quadratic Unbiased Estimation of Variance Components from Unbalanced Data in the 1-way Classification[1,2]

E. C. Townsend[3] and S. R. Searle

Biometrics Unit, Cornell University, Ithaca, N.Y.

Abstract

Best quadratic unbiased estimators (BQUE's) of variance components from unbalanced data in the one-way classification random model are derived under zero mean and normality assumptions. An estimator of the between-class variance is also developed for the non-zero mean case from analogy with the zero mean situation. These estimators are functions of the ratio of the population variances, $\rho = \sigma_a^2/\sigma_e^2$. Numerical studies indicate that for badly unbalanced data and for values of ρ larger than one, estimators of σ_a^2 having variance less than that of the analysis of variance estimator can be obtained by substituting even a rather inaccurately predetermined value of ρ into the BQUE of σ_a^2 .

1/ Paper Number BU-215 in the Biometrics Unit and Number 599 in the Department of Plant Breeding and Biometry, Cornell University.

2/ The results in this paper are from a thesis submitted by the senior author in partial fulfillment of the Ph.D. degree

3/ Present address: Dept. of Statistics and Computer Science, University of West Virginia, Morgantown, West Virginia.

1. Introduction

Graybill [1961] has shown that the usual analysis of variance procedures produce best quadratic unbiased estimators (BQUE's) of variance components when data are balanced, i.e., have the same number of observations in each subclass. Choosing among unbiased estimators when data are unbalanced is more difficult. Several estimators have been suggested but there is relatively little information upon which to base a decision to use one rather than another. This paper attempts to alleviate this situation by developing BQUE's of variance components for the 1-way classification random model, with unbalanced data. Two cases are considered: when the overall mean in the model is assumed zero, and when it is assumed non-zero.

Although the practical value of the BQUE's is limited by the fact that they are functions of the unknown variance components, investigation of their behavior as various limits are approached permits evaluation of different procedures. In addition, if an experimenter has some prior knowledge, or even a reasonable guess, about the relative magnitudes of the unknown components, approximate BQUE procedures may be useful in yielding estimators that have smaller variances than the usual analysis of variance (ANOVA) estimators. Evidence of this is available from numerical studies.

2. Zero Mean Model

The simplest variance component model for data in a one-way classification is where the linear model for an observation, y_{ij}, is

$$y_{ij} = \alpha_i + e_{ij} , \qquad (1)$$

with $i = 1,2,\ldots,c$ and $j = 1,2,\ldots,n_i$. The α_i and e_{ij} are assumed to be independent random samples from two normal populations with zero means and variances σ_a^2 and σ_e^2 , respectively. It is also assumed that no relationship exists between the number of observations in a subclass (n_i) and the subclass effects (α_i), a form of the model considered by Harville [1967].

Suppose the vector of observations is written as

$$\underline{y}' = (y_{11} \;\; y_{12}\cdots y_{1n_1}\cdots y_{i1}\cdots y_{in_i}\cdots y_{c1}\cdots y_{cn_c}) \;\; .$$

Then $E(y) = \underline{0}$, where E denotes expectation over repeated sampling with the same values of the n_i . The variance-covariance matrix of $\underline{y}$ shall be denoted by $\underline{V}$. It is a diagonal matrix of submatrices $\underline{V}_i$, of order n_i , and can be written as

$$\underline{V} = \mathrm{diag}\{\underline{V}_1 \;\; \underline{V}_2\cdots\underline{V}_c\} \equiv \sum_{i=1}^{c}{}^{+} \underline{V}_i \tag{2}$$

where Σ^+ denotes the operation of taking the direct sum of matrices. The form of $\underline{V}_i$ is

$$\underline{V}_i = \sigma_a^2\underline{J}_i + \sigma_e^2\underline{I} \tag{3}$$

with $\underline{I}_i$ being an identity matrix of order n_i and $\underline{J}_i$ being a matrix of order $n_i \times n_i$, with every element unity.

2a. Estimators of σ_a^2 and σ_e^2

A best quadratic unbiased estimator (BQUE) of a variance σ^2 is defined as that quadratic form of the observations which is an unbiased estimator of σ^2, which from among all such quadratic forms has minimum variance. Consider

$$\hat{\sigma}^2 = \underset{\sim}{y}'\underset{\sim}{A}\underset{\sim}{y} \tag{4}$$

where, without loss of generality, $\underset{\sim}{A}$ is symmetric. It can be shown (e.g., Graybill [1961, problems 17.6 and 17.7]), that the mean and variance of $\underset{\sim}{y}'\underset{\sim}{A}\underset{\sim}{y}$ are respectively $\mathrm{tr}(\underset{\sim}{V}\underset{\sim}{A})$ and $2\mathrm{tr}(\underset{\sim}{V}\underset{\sim}{A})^2$. The estimator in (4) will therefore be a BQUE of σ^2 provided $\underset{\sim}{A}$ is chosen so that $2\mathrm{tr}(\underset{\sim}{V}\underset{\sim}{A})^2$ is minimized subject to $\sigma^2 = \mathrm{tr}(\underset{\sim}{V}\underset{\sim}{A})$. Derivation in this manner, of the BQUE's (and their variances) of the variance components σ_a^2 and σ_e^2 of the model (1), based on $\underset{\sim}{V}$ given in (2) and (3), is shown in the Appendix. On defining

$$\rho = \sigma_a^2/\sigma_e^2 \quad \text{and } N = n_{.} = \sum_{i=1}^{c} n_i$$

and

$$r = \sum_{i=1}^{c} 1/(1+n_i\rho)^2 + N-c\ , \quad s = \sum_{i=1}^{c} n_i^2/(1+n_i\rho)^2 \text{ and } t = \sum_{i=1}^{c} n_i/(1+n_i\rho)^2 \tag{5}$$

we have, from (A17) - (A20) of the appendix, the BQUE of σ_e^2 as

$${}_B\hat{\sigma}_e^2 = (rs-t^2)^{-1}\left\{\sum_{i=1}^{c} (s-tn_i)(1+n_i\rho)^{-2}(y_{i.}^2/n_i) + s\left(\sum_{i=1}^{c}\sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{c} y_{i.}^2/n_i\right)\right\}, \tag{6}$$

with variance

$$\operatorname{var}({}_B\hat{\sigma}_e^2) = 2s\sigma_e^4/(rs-t^2) \quad ; \tag{7}$$

and the BQUE of σ_a^2 is

$${}_B\hat{\sigma}_a^2 = (rs-t^2)^{-1}\left\{\sum_{i=1}^{c}(rn_i-t)(1+n_i\rho)^{-2}(y_{i.}^2/n_i) - t\left(\sum_{i=1}^{c}\sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{c} y_{i.}^2/n_i\right)\right\} \tag{8}$$

with variance

$$\operatorname{var}({}_B\hat{\sigma}_a^2) = 2r\sigma_e^4/(rs-t^2) \quad . \tag{9}$$

It should be noted that (7) and (9) can be shown equal to the equivalent expressions given by Crump [1951] and Searle [1956] for the large sample variances of maximum likelihood estimators of σ_e^2 and σ_a^2 when a general mean μ is included in the model.

2b. Comparisons of BQUE's with other estimators

The BQUE of σ_e^2 is first compared to the ANOVA estimator $\hat{\sigma}_e^2$, where

$$\hat{\sigma}_e^2 = \left(\sum_{i=1}^{c}\sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{c} y_{i.}^2/n_i\right)/(N-c) \quad . \tag{10}$$

The most obvious relationship is that

$${}_B\hat{\sigma}_e^2 = (rs-t^2)^{-1}\left\{\sum_{i=1}^{c}(s-tn_i)(1+n_i\rho)^{-2}(y_{i.}^2/n_i) + s(N-c)\hat{\sigma}_e^2\right\} .$$

Therefore, in some sense it can be said that the first term of ${}_B\hat{\sigma}_e^2$ provides information about σ_e^2 which is not utilized in the ANOVA estimator. When this additional information is important, and to what extent its inclusion in ${}_B\hat{\sigma}_e^2$ affects the variance of the estimator can be partially answered by noting that

$$\lim_{\rho\to\infty} ({}_B\hat{\sigma}_e^2) = \hat{\sigma}_e^2 ,$$

as shown in Townsend [1968]. Hence the difference between $\hat{\sigma}_e^2$ and ${}_B\hat{\sigma}_e^2$ diminishes as σ_a^2 becomes large compared to σ_e^2 .

Further comparison of ${}_B\hat{\sigma}_e^2$ and $\hat{\sigma}_e^2$ is made by looking at their variances. That for ${}_B\hat{\sigma}_e^2$ is given in (7) and

$$\mathrm{var}(\hat{\sigma}_e^2) = 2\sigma_e^4/(N-c) \tag{11}$$

in the usual way. To facilitate discussion we define

$$p_e = \frac{\mathrm{var}(\hat{\sigma}_e^2) - \mathrm{var}({}_B\hat{\sigma}_e^2)}{\mathrm{var}({}_B\hat{\sigma}_e^2)} \; . \tag{12}$$

This is a measure of the extent to which the variance of an estimator of σ_e^2 is increased by using the ANOVA estimator rather than the BQUE, this increase in variance being measured as a fraction of the minimum variance, namely that of the BQUE. p_e therefore represents the penalty incurred, in this variance sense, by using the ANOVA estimator rather than the BQUE .

Since the BQUE tends to the ANOVA estimator when $\rho \to \infty$ we look at the other extreme, namely when $\rho = 0$. In this case, putting $\rho = 0$ in (5) gives $r = N$,

$s = \sum_{i=1}^{c} n_i^2 = S_2$ and $t = N$, thus allowing (6) to be rewritten as

$$({}_B\hat{\sigma}_e^2|\rho = 0) = (S_2 \sum_{i=1}^{c} \sum_{j=1}^{n_i} y_{ij}^2 - N \sum_{i=1}^{c} y_{i.}^2)/N(S_2-N) \tag{13}$$

with its variance, from (7), being

$$\text{var}({}_B\hat{\sigma}_e^2 \,|\rho = 0) = 2S_2\sigma_e^4 \,/\, N(S_2-N) \quad . \tag{14}$$

Since $\hat{\sigma}_e^2$ is unaffected by differing values of ρ its variance is as in (11). Using that and (14) in (12) gives

$$p_e(\rho=0) = \frac{cS_2 - N^2}{(N-c)S_2} \quad . \tag{15}$$

For given c and N, (15) is maximum when S_2 is maximum which, it can be proved (using integer algebra), is maximum when all groups except one have only one observation, i.e., when the n_i's are $(1,1,\ldots,1,N-c+1)$. This maximum value of S_2 is $(c-1) + (N-c+1)^2 = (N-c)^2 + 2N-c$, in which case the maximum value of p_e reduces to

$$\max_{n_i} p_e(\rho=0) = \frac{(c-1)(N-c)}{(N-c)^2 + (2N-c)} \quad . \tag{16}$$

For given N, it can be readily shown that (16) is maximized over c when $c = N + 1 - \sqrt{N}$, which makes

$$\max_{c,n_i} p_e(\rho=0) = \frac{(\sqrt{N} - 1)^2}{2\sqrt{N-1}} \tag{17}$$

It should be emphasized that ${}_B\hat{\sigma}_e^2$ of (6) is a function of ρ, which is unknown in most practical cases. Therefore ${}_B\hat{\sigma}_e^2$ cannot be used directly but only approximated, by using a value for ρ which is somehow felt to be reasonable. Of course this increases the variance of the procedure because it is no longer best. Since in practice ρ is neither zero nor known, and because sets of n_i-values as extreme as used in (16) and (17) rarely occur, the usual ANOVA procedure would seem to be a fairly good choice for estimating σ_e^2.

"There is no page 9."

Returning to the case when $\rho = 0$ it can be noted from (13) that $({}_B\hat{\sigma}_e^2 \mid \rho = 0)$ is the same as an estimator suggested by Koch [1967]. The same is true for σ_a^2. Putting $\rho = 0$ in (8) gives

$$({}_B\hat{\sigma}_a^2 \mid \rho = 0) = \left[\sum_{i=1}^{c} (n_i^2 - n_i)\bar{y}_{i.}^2 - (N-c)\hat{\sigma}_e^2\right] / (S_2 - N) \quad . \tag{18}$$

which will be found identical to the estimator of σ_a^2 put forward by Koch [1967] for the zero mean model.

The usual ANOVA estimator of σ_a^2 for the one-way classification does not utilize the zero mean assumption being used here. When this assumption <u>is</u> made, the ANOVA estimator of σ_a^2 is

$$\hat{\sigma}_a^2 = \left(\sum_{i=1}^{c} n_i\bar{y}_{i.}^2 - c\hat{\sigma}_e^2\right) / N \tag{19}$$

with variance

$$\text{var}(\hat{\sigma}_a^2) = 2\sigma_e^4[\rho^2 S_2 + 2\rho N + cN/(N-c)]/N^2 \; . \tag{20}$$

Although, as noted previously, the limit of the BQUE of σ_e^2 as $\rho \to \infty$ is the ANOVA estimator of σ_e^2, the same kind of limiting result does not hold for the BQUE of σ_a^2. Indeed

$$\lim_{\rho \to \infty} {}_B\hat{\sigma}_a^2 = \left(\sum_{i=1}^{c} \bar{y}_{i.}^2 - \sum_{i=1}^{c} n_i^{-1}\hat{\sigma}_e^2\right) / c \tag{21}$$

as shown in Townsend [1968]. Furthermore, because the terms $(1+n_i\rho)^2$ occur in demonimators of summations in ${}_B\hat{\sigma}_a^2$, and because both $\text{var}(\hat{\sigma}_a^2)$ and $\text{var}({}_B\hat{\sigma}_a^2 \mid \rho = 0)$

bear no simple relation to one another as they do in the case of $\hat{\sigma}_e^2$ [see (15), for example], it is difficult to examine the behaviour of ${}_B\hat{\sigma}_a^2$ for ρ varying between zero and infinity. However, $({}_B\hat{\sigma}_a^2|\rho = 0)$ of (18), $\hat{\sigma}_a^2$ of (19) and the limit of ${}_B\hat{\sigma}_a^2$ as $\rho\to\infty$ of (21) can be compared to some extent, by considering the terms in $\bar{y}_{i.}^2$ in each - i.e., the terms in $\bar{y}_{i.}^2$ over and above their occurrence in $\hat{\sigma}_e^2$ in each case.

The weight given to $\bar{y}_{i.}^2$ when ρ is zero is seen in (18) to be proportional to $(n_i^2-n_i)$ and as ρ approaches infinity, in (21), equal weights are given. The ANOVA estimator (19) always gives weights proportional to subclass size. Therefore, $\hat{\sigma}_a^2$ gives too much weight to small groups and not enough to large groups when ρ is small whereas the converse is true when ρ is large. These results correspond closely to the conclusions drawn by Robertson [1962] for small ρ and Tukey [1957] for large ρ .

Because the largest difference between the variance of $\hat{\sigma}_e^2$ and ${}_B\hat{\sigma}_e^2$ occurred at $\rho = 0$, it was readily established that in most practical situations the difference is rather small. This type of comparison between $\hat{\sigma}_a^2$ and ${}_B\hat{\sigma}_a^2$ is more difficult because the maximum difference does not occur at the same value of ρ for all sets of n_i-values. Therefore the practical value of using the estimator ${}_B\hat{\sigma}_a^2$ on actual data has yet to be considered. However, in order to use ${}_B\hat{\sigma}_a^2$, a numerical value is needed for ρ . Since ρ is unknown, an estimate is required, implying that any practical procedure based on ${}_B\hat{\sigma}_a^2$ will not actually be a BQUE of $\hat{\sigma}_a^2$. Hence one asks how accurate must the estimate of ρ be to give an estimator whose variance is less than that of the ANOVA procedure and how much less will it be? Some empirical answers to these questions follow.

2c. An approximate BQUE of σ_a^2

A prior estimate of ρ , say ρ_o, is substituted into ${}_B\hat{\sigma}_a^2$ of (8) and the resulting estimator designated as ${}_B\tilde{\sigma}_a^2$, with variance

$$\text{var}({}_B\tilde{\sigma}_a^2) = 2(r_o s_o - t_o^2)^{-2}\sigma_e^4\left[\sum_{i=1}^{c}(r_o n_i - t_o)^2(1+n_i\rho)^2(1+n_i\rho_o)^{-4}+(N-c)t_o^2\right], \quad (22)$$

where r_o , s_o and t_o are r,s and t of (5) computed using ρ_o . Thus the variance of the approximate BQUE is a function of both the true ρ and the prior estimate, ρ_o . The intractability of (22) forces numerical comparisons between $\text{var}(\hat{\sigma}_a^2)$ and $\text{var}({}_B\tilde{\sigma}_a^2)$.

Analogous to p_e of (12) the ratio

$$p_a = \left[\text{var}(\hat{\sigma}_a^2)-\text{var}({}_B\tilde{\sigma}_a^2)\right] / \text{var}({}_B\tilde{\sigma}_a^2) \quad (23)$$

is defined to aid the comparisons. It represents the extent to which the variance of $\hat{\sigma}_a^2$ differs from that of ${}_B\tilde{\sigma}_a^2$, relative to the variance of ${}_B\tilde{\sigma}_a^2$. When $p_a = 0$, both estimators have the same variance, and when p_a is positive, ${}_B\tilde{\sigma}_a^2$ is better than $\hat{\sigma}_a^2$ where "better" means smaller variance. The converse is true when p_a is negative, because p_a negative means ${}_B\tilde{\sigma}_a^2$ is poorer than $\hat{\sigma}_a^2$.

Comparisons made over a range of ρ and ρ_o for selected n-patterns (sets of n_i-values) are summarized in Figures 1-5 . Figures 1, 2, and 4 are for cases of N = 25 and c = 5, Figures 1a and 2a being extensions of Figures 1 and 2 . Figure 3 has n_i's that are double those of figures 1 and 2, Figure 4 has N = 25 and c = 10, and Figure 5 is a case of N = 380 and c = 50 . Each figure is for a particular n-pattern with the abscissa showing values of $\rho = \sigma_a^2/\sigma_e^2$ and the ordinates giving values of ρ_o, the prior value of ρ. Each curve is for a given value of p_a, e.g., a curve labeled $p_a = 0$ connects those points (ρ, ρ_o) at which $p_a = 0$. These points were determined by solving

$$\mathrm{var}(\hat{\sigma}_a^2) - (p_a - 1)\mathrm{var}({}_B\tilde{\sigma}_a^2) = 0 \tag{24}$$

for ρ_o, given the n-pattern and ρ. As the expression for $\mathrm{var}({}_B\tilde{\sigma}_a^2)$ in (22) is particularly intractable, (24) was solved numerically for ρ_o using the method of reguli falsi (Conte, 1965).

In each figure, two curves for $p_a = 0$ are given, since for a given n-pattern and ρ, two values of ρ_o satisfy (24). In general, one of the $p_a = 0$ curves lies above the line $\rho_o = \rho$ and the other below it. This is because ${}_B\tilde{\sigma}_a^2$ is BQUE when $\rho_o = \rho$ and, as ρ_o deviates from ρ in either direction, the variance of ${}_B\tilde{\sigma}_a^2$ increases until it equals that of $\hat{\sigma}_a^2$, at which time $p_a = 0$. Thus the upper $p_a = 0$ boundary indicates how much ρ_o can be larger than ρ without making ${}_B\tilde{\sigma}_a^2$ poorer than $\hat{\sigma}_a^2$. Similarly the lower $p_a = 0$ boundary shows how much ρ_o can be smaller than ρ without making ${}_B\tilde{\sigma}_a^2$ poorer than $\hat{\sigma}_a^2$. For example, in Figure 1, with $\rho = 0.4$ we may overestimate ρ by as much as 0.349 ($\rho_o = 0.749$) or under-estimate it by as much as 0.167 ($\rho_o = 0.233$) without causing the approximate BQUE to be poorer (have a larger variance) than the ANOVA estimator.

The region between the two $p_a = 0$ curves in any figure is the set of points (ρ, ρ_o) for which ${}_B\tilde{\sigma}_a^2$ is better than $\hat{\sigma}_a^2$. (Remaining parts of the quadrant are where ${}_B\tilde{\sigma}_a^2$ is poorer than $\hat{\sigma}_a^2$. In all cases this region has approximately the same form. Its lower boundary first rises sharply, until it reaches a ρ_o

corresponding to the value of ρ at which the ANOVA estimator comes nearest to being the BQUE of σ_a^2, at which point the lower boundary levels off. Conversely, as ρ increases from zero, the upper boundary of the region remains nearly constand, at approximately the same value of ρ_o as the level portion of the lower boundary, until that value of ρ at which the lower boundary levels off is reached. The upper boundary then climbs sharply. Although these two boundaries (the $p_a = 0$ curves) do not touch in Figure 1, they do so in some situations (for other n-patterns) as is discussed subsequently.

The region between the two $p_a = 0$ curves in Figure 1 is very large. This means that when ρ is larger than approximately 0.7 a very inaccurate ρ_o (so far as being close to ρ is concerned) used in ${}_B\tilde{\sigma}_a^2$ gives an estimator whose variance is smaller than that of $\hat{\sigma}_a^2$. For example, with $\rho > 0.6$, any $\rho_o > 0.24$ makes the variance of ${}_B\tilde{\sigma}_a^2$ less than that of the ANOVA estimator. It is not much less, however, as can be seen in Figure 1a which is an extension of Figure 1 for ρ and ρ_o extending (on a different scale) up to 10 and 13, well beyond their limits of 1.0 and 1.3 on Figure 1. And in Figure 1a the presence of the $p_a = 0.05$ curve when $\rho \geq 2$ indicates that the penalty incurred by using the ANOVA estimator rather than the approximate BQUE can be only 5% or a little greater for large values of ρ , and the line $p_a = 0.10$ does not appear. In many instances, for data as moderately unbalanced as is the n-pattern of Figures 1 and 1a, this 5% penalty does not seem sufficiently large to warrant using ${}_B\tilde{\sigma}_a^2$ rather than $\hat{\sigma}_a^2$.

Figure 1 also illustrates situations in which $p_a < 0$, meaning that the variance of $\hat{\sigma}_a^2$ is less than that of an approximate BQUE. For example, at $(\rho,\rho_o) = (0.4, 0.065)$, $p_a = -0.05$, indicating that the ANOVA estimator has a variance equal to 95% of the approximate BQUE at that point.

Figures 2 and 2a are for an n-pattern more unbalanced (n_i = 1,1,1,11,11) than is that of Figures 1 and 1a (n_i = 3,4,5,6,7). A noticeable difference between these two pairs of figures is that in Figures 2 and 2a the two $p_a = 0$ curves appear to cross instead of just approaching one another as in Figures 1 and 1a. The point at which the curves appear to cross is $(\rho,\rho_o) = (0.337, 0.337)$. At this point, the approximate BQUE is the exact BQUE (because $\rho_o = \rho$) and for this n-pattern has variance equal to that of the ANOVA estimator when $\rho = 0.337$. Therefore, for $\rho = 0.337$, there are no values of ρ_o for which the approximate BQUE is better than the ANOVA estimator. Hence, the two boundaries of the region in which ${}_B\tilde{\sigma}_a^2$ is better than $\hat{\sigma}_a^2$, touch at the point $(\rho,\rho_o) = (0.337, 0.337)$. Although in Figures 2 and 2a the boundaries appear to be crossing at this point they are, in effect, just touching. Their behaviour can be envisaged by thinking of the two $p_a = 0$ curves in Figures 1 and 1a as being moved toward each other until they have a point in common. This occurrs for some n-patterns and not for others.

Additional to the tangential property just discussed, Figures 2 and 2a display considerably larger differences between the variances of ${}_B\tilde{\sigma}_a^2$ and $\hat{\sigma}_a^2$ than are to be found in Figures 1 and 1a. This is evidenced by the curves $p_a = 0.25$ and $p_a = 0.50$ in Figure 2a. The $p_a = 0.25$ curve of Figure 2a is considerably closer to the $p_a = 0$ curves than is the $p_a = 0.05$ curve of Figure 1a, indicating that the penalty incurred by using the ANOVA estimator rather than the approximate BQUE is not only greater but is more quickly encountered as one moves away from the $p_a = 0$ boundaries. For example, at the point $(\rho,\rho_o) = (2.25, 1.0)$ the penalty in Figure 1a is approximately 5% but in Figure 2a it exceeds 25%. Also, in Figure 2a, when ρ is large, 4.5 say, and ρ_o is chosen as 1.5, the penalty for using the ANOVA estimator rather than the approximate BQUE is more than 50%. [This is not to be confused with the upper limit of 50% indicated in (17). That is for estimating σ_e^2 and pertains to using the exact BQUE and not an approximate BQUE as is being discussed here for σ_a^2.]

An example of the effect on p_a of increasing the number of observations in the data while leaving the relative values of the n_i undisturbed is illustrated in the comparison of Figures 2 and 3. The n-pattern in Figure 3 is that of Figure 2 but with every n_i doubled. In Figure 3 the p_a curves have shifted down and to the left compared to those in Figure 2. Consequently the region in which the approximate BQUE is better than the ANOVA estimator now extends over a wider range of (ρ,ρ_o) values. Also, the penalty for not using it is larger for any given point in the region. For example, at $(\rho,\rho_o) = (0.9, 0.6)$ the penalty is 10% in Figure 2 but it exceeds 25% in Figure 3.

In contrast to Figure 3, the p_a-curves of Figure 4 are shifted up and to the right, compared to Figure 2. The change in the n-pattern has been one of holding N constant, N = 25 and increasing the number of classes, to have n_i-values = 1,1,1,1,1,4,4,4,4,4. The $p_a = 0$ curves now touch at $(\rho,\rho_o) = (0.64, 0.64)$ approximately, compared to (0.16, 0.16) in Figure 3 and (0.337, 0.337) in Figure 2. This means that the lower left portion of the region bounded by the $p_a = 0$ curves is larger in Figure 4 than in Figures 2 or 3, whereas the upper right portion is smaller. Consequently, for small values of ρ, a less accurate ρ_o can be used without causing p_a to be negative, but for large values of ρ a more accurate ρ_o is needed if the approximate BQUE is to be better than the ANOVA estimator.

An n-pattern considerably larger than those of Figures 1-4 is used in Figure 5; it is for N = 380 and c = 50. The principal consequence is that greater gains and losses can occur when using ${}_B\tilde{\sigma}_a^2$ rather than $\hat{\sigma}_a^2$. For example, the penalty incurred through using $\hat{\sigma}_a^2$ can be as much as 100%, as evidenced by the presence of a $p_a = 1.00$ curve. Conversely, the presence of $p_a = -0.50$ curves indicates that the variance of the ANOVA estimator can be 50% smaller that that of the approximate BQUE. However, the $p_a > 0$ region is very large when ρ is greater than .3 or .4 indicating that quite inaccurate values of ρ_0 may be used in ${}_B\tilde{\sigma}_a^2$ with its variance still being less than the variance of $\hat{\sigma}_a^2$. In view of the substantial reduction which may be obtained when ρ is moderately large and the fact that ρ_0 need not be very accurate to obtain these reductions, ${}_B\tilde{\sigma}_a^2$ would seem to be useful as a practical procedure for this type of n-pattern if ρ is moderately large.

Interesting conclusions regarding the practicality of ${}_B\tilde{\sigma}_a^2$ may be drawn from Figures 1-5. Despite the fact that ${}_B\tilde{\sigma}_a^2$ requires a prior estimate of $\rho = \sigma_a^2/\sigma_e^2$, it is clear that in many situations its use can provide estimates of ρ having smaller variance than the ANOVA estimator, even when the prior estimate, ρ_0, is not close to ρ. This is especially true when ρ is very small or very large, and the n-pattern is badly unbalanced. Under these conditions the variances of ${}_B\tilde{\sigma}_a^2$ can not only be less, but substantially less, than the variance of $\hat{\sigma}_a^2$. Furthermore, the presence of the $p_a = 1.0$ curve in Figure 5 shows that having many groups and a large total number of observations does not automatically imply that the choice of an estimator is unimportant.

Although we have not specifically investigated iterative techniques using successive approximate BQUE's, some inference about this class of procedures can also be made from figures 1-5. Figure 2, for example, clearly indicates that when ρ = .337 an iterative procedure cannot be as good as the ANOVA procedure. This is so true because ${}_B\tilde{\sigma}_a^2$ is the BQUE of σ_a^2 when ρ_o has this value and therefore has the smallest variance possible for unbiased quadratic estimators. On the other hand, when ρ is very small or very large, it seems likely that iterative procedures might lead to estimators whose variances would be smaller than the ANOVA estimator.

3. Non-zero mean

The mean and variance of $\underline{y}'\underline{A}\underline{y}$ when $\underline{y}$ is normally distributed with vector of means $\underline{\mu}$ and variance-covariance matrix $\underline{V}$ are

$$E(\underline{y}'\underline{A}\underline{y}) = tr(\underline{A}\underline{V}) + \underline{\mu}'\underline{A}\underline{\mu} \tag{25}$$

and

$$var(\underline{y}'\underline{A}\underline{y}) = 2tr(\underline{A}\underline{V})^2 + 4\underline{\mu}'\underline{A}\underline{V}\underline{A}\underline{\mu} .$$

Utilization of these results in the procedure for deriving BQUE's of σ_a^2 and σ_e^2 have so far proven intractable in the case of the non-zero mean model $y_{ij} = \mu + \alpha_i + e_{ij}$. This is so despite the fact that $\underline{\mu}$ of (25) takes the slightly simpler form $\mu\underline{1}$. Nevertheless, the preceding study of the zero mean model provides information which can be applied to the general class of unbiased estimators suggested by Tukey [1957] to investigate an approximate BQUE of σ_a^2 along the lines of that for the zero mean case.

Comparison of ${}_B\hat{\sigma}_e^2$ in (6) to Tukey's [1957] class of estimators of σ_e^2 indicates that the latter is not sufficiently general to include ${}_B\hat{\sigma}_e^2$. However, since for the zero mean case we have seen that little reduction in the variance of an estimator of σ_e^2 can be anticipated by using ${}_B\hat{\sigma}_e^2$ rather than $\hat{\sigma}_e^2$, it is not unreasonable to conjecture that the gain may be even less in the non-zero mean case. We therefore accept the ANOVA estimator $\hat{\sigma}_e^2$ in the non-zero mean case and proceed to investigate an estimator alternative to $\hat{\sigma}_a^2$ for σ_a^2 .

3a. An alternative estimator for σ_a^2

The general class of estimators for σ_a^2 given by Tukey [1957] is

$$ {}_T\hat{\sigma}_a^2 = \left[\sum_{i=1}^{c} w_i \left(\bar{y}_{i.} - \sum_{i=1}^{c} w_i \bar{y}_{i.} \right)^2 - \hat{\sigma}_e^2 \sum_{i=1}^{c} w_i (1-w_i)/n_i \right] / \left(1 - \sum_{i=1}^{c} w_i^2\right) \tag{26} $$

where the w_i are a set of weights which can be assumed, without loss of generality, to sum to unity. Recalling $\hat{\sigma}_e^2$ of (10) we now notice that the BQUE of σ_a^2 given in (8) can be rewritten as

$$ {}_B\hat{\sigma}_a^2 = \left[\sum_{i=1}^{c} (rn_i^2 - tn_i)(1+n_i\rho)^{-2} \bar{y}_{i.}^2 - t(N-c)\hat{\sigma}_e^2 \right] / (rs-t^2) \tag{27} $$

In comparing (26) and (27) we see that, except for subtracting a mean in (26), the first term of both estimators is a weighted sum of squared group means. Also, the second term in each is a multiple $\hat{\sigma}_e^2$, whose purpose is to remove σ_e^2 from the expectation of the first term; and a denominator occurs in both expressions to make the estimators unbiased with respect to σ_a^2 . Therefore Tukey's estimator with

$$ w_i = (rn_i^2 - tn_i)(rs-t^2)^{-1}(1+n_i\rho)^{-2} \quad , \tag{28} $$

is considered as an estimator of σ_a^2 .

Further inspection of the first term of (26) reveals that the mean subtracted from each group mean is a weighted mean, $\bar{y}_w$ say. This is effectively an estimator of the mean μ and as such can be improved upon, i.e., values w_i can be found such that the variance of $\bar{y}_w$ is smaller than when (28) is used for w_i. There appears to be no reason, other than simplicity, for using the same weighting when estimating the mean as is used when summing the squared deviations of the group means from that estimated mean. Therefore in the place of $\bar{y}_w$ we use

$$\bar{y}_u = \sum_{i=1}^{c} u_i \bar{y}_{i.} \quad , \tag{29}$$

with $\sum_{i=1}^{c} u_i = 1$, and define the u_i's so as to minimize $\text{var}(\bar{y}_u)$. Since

$$\text{var}(\bar{y}_{i.}) = (1+n_i\rho)/n_i \sigma_e^2$$

we take

$$u_i = \frac{\text{var}(\bar{y}_{i.})}{\sum_{i=1}^{c} \text{var}(\bar{y}_{i.})} = \frac{n_i}{1+n_i\rho} \Big/ \sum_{i=1}^{c} \frac{n_i}{1+n_i\rho} \tag{30}$$

in (29) and use that value of $\bar{y}_u$ in place of $\sum_{i=1}^{c} w_i \bar{y}_{i.}$ in (26). The resulting estimator we suggest is therefore

$$\hat{\hat{\sigma}}_a^2 = \left[\sum_{i=1}^{c} w_i (\bar{y}_{i.} - \bar{y}_u)^2 - C_2 \hat{\sigma}_e^2 \right] / C_1 \tag{31}$$

where C_1 and C_2 are yet to be chosen, to make $\hat{\hat{\sigma}}_a^2$ an unbiased estimator. It is not difficult to show that

$$E\left[\sum_{i=1}^{c} w_i(\bar{y}_{i.} - \bar{y}_u)^2\right] = \sigma_a^2 \sum_{i=1}^{c}(w_i-2u_iw_i+u_i^2) + \sigma_e^2 \sum_{i=1}^{c}(w_i-2u_iw_i+u_i^2)/n_i$$

so that C_1 and C_2 are taken as

$$C_1 = \sum_{i=1}^{c}(w_i-2u_iw_i+u_i^2) \quad \text{and} \quad C_2 = \sum_{i=1}^{c}(w_i-2u_iw_i+u_i^2)/n_i \quad . \tag{32}$$

The estimator we suggest is therefore (31) using w_i of (28), $\bar{y}_u$ of (29), u_i of (30) and C_1 and C_2 of (32). With these substitutions its variance can, after a little algebraic simplification (as in Townsend [1968]), be written as follows. Define

$$\theta_i = (w_i - 2u_iw_i + u_i^2)/n_i^2 \, ,$$

$$k_{ij} = (u_iu_j - u_iw_j - u_jw_i)/n_in_j \qquad \text{for } i\neq j = k,2,\ldots,c,$$

and

$$\gamma = C_2/(N-c) \quad .$$

Then

$$\mathrm{var}(\hat{\hat{\sigma}}_a^2) = 2\sigma_a^4\left(\sum_{i=1}^{c} n_i^4\theta_i^2 + \sum_{i=1}^{c}\sum_{j\neq i}^{c} n_i^2n_j^2k_{ij}^2\right)$$

$$+ 2\sigma_a^2\sigma_e^2\left[2\sum_{i=1}^{c} n_i^3\theta_i^2 + \sum_{i=1}^{c}\sum_{j\neq i}^{c} n_in_j(n_i+n_j)k_{ij}^2\right] \tag{33}$$

$$+ 2\sigma_e^4\left[\sum_{i=1}^{c} n_i^2\theta_i^2 + (N-c)\gamma^2 + \sum_{i=1}^{c}\sum_{j\neq i}^{c} n_in_jk_{ij}^2\right] \quad .$$

It seems clear that (33) is intractable for analytic study for comparison with the ANOVA estimator, which in this case of the non-zero mean is the familiar

$$\hat{\sigma}_a^2 = \left[\left(\sum_{i=1}^{c} n_i \bar{y}_{i.}^2 - N\bar{y}_{..}^2 \right) / (c-1) - \hat{\sigma}_e^2 \right] / [N-S_2/N)/(c-1)] \quad .$$

Its variance is

$$\mathrm{var}(\hat{\sigma}_a^2) = \left[\frac{2\sigma_e^4(N-1)}{(c-1)(N-c)} + \frac{4\sigma_e^2\sigma_a^2(N^2-S_2)}{N(c-1)^2} + \frac{2\sigma_a^4(N^2S_2 + S_2^2 - 2NS_3)}{N^2(c-1)^2} \right] / f^2$$

where $S_2 = \sum_{i=1}^{c} n_i^2$ as before, $S_3 = \sum_{i=1}^{c} n_i^3$ and $f = (N-S_2/N)/(c-1)$ as, for example, in Searle [1956].

3b. Comparisons with the ANOVA estimator

The suggested estimator (31), through its dependence on w_i and u_i of (28) and (30), depends on $\rho = \sigma_a^2/\sigma_e^2$ which is unknown. Therefore, to assess its value we use a prior estimate for ρ, ρ_o say, in place of ρ in $\hat{\hat{\sigma}}_a^2$, calling the resulting estimator $\tilde{\tilde{\sigma}}_a^2$. Numeric comparison of $\tilde{\tilde{\sigma}}_a^2$ with the ANOVA estimator $\hat{\sigma}_a^2$ is made by means of variances of the estimators, using

$$p_a = \frac{\mathrm{var}(\hat{\sigma}_a^2) - \mathrm{var}(\tilde{\tilde{\sigma}}_a^2)}{\mathrm{var}(\tilde{\tilde{\sigma}}_a^2)} \quad .$$

in the same manner as previously. Figures 6-8 show some results of these comparisons, portrayed in exactly the same manner as Figures 1-5 .

The n-patterns of Figures 6,7,8 are identical to those of figures 1,2 and 5 respectively. Gross comparison between the two sets of figures indicates great similarity, suggesting that $\tilde{\tilde{\sigma}}_a^2$ is near the true BQUE of σ_a^2 when $\rho_o = \rho$.

One salient difference, illustrated by a comparison of Figures 2 and 7, is that the two $p_a = 0$ curves never touch. This means that the ANOVA estimator is never as good as the suggested procedure when $\rho_o = \rho$ and thus is never the BQUE of σ_a^2. Maybe this is true generally. In the zero mean case when the ANOVA procedure was BQUE, it occurred only for some value of ρ other than $\rho = 0$. Yet in the non-zero mean case, the estimator of the mean used in the ANOVA procedure is $\bar{y}_{..} = \sum_{i=1}^{c} n_i \bar{y}_{i.}/N$, which is the minimum variance unbiased best estimator of μ when $\rho = 0$. Now, when σ_a^2 and σ_e^2 are known, the best estimator of μ is $\underline{1}'\underline{V}^{-1}\underline{y}/\underline{1}'\underline{V}^{-1}\underline{1}$, equal to $\bar{y}_u$ of (29) and (30), and equal to $\bar{y}_{..}$ when $\rho = 0$. Furthermore, the best estimator of μ when simultaneously estimating σ_a^2 and σ_e^2 is unknown, but whatever it is it seems reasonable to conclude that it is not $\bar{y}_{..}$ as used in the ANOVA estimator. If so, it follows that the ANOVA estimator of σ_a^2 for the non-zero mean case is never the BQUE.

4. Conclusions

Figures 1-7 and the preceding discussions thereof lead one to tender the following conclusions.

For the zero mean case:

(a) The BQUE of σ_e^2 has little practical advantage over the ANOVA estimator, except for very unbalanced data and very small (close to zero) values of ρ. [Equation (16)]

(b) The ANOVA estimator of σ_a^2 approaches the BQUE for some value of ρ and actually is BQUE for some designs n-patterns. [Figures 1 and 2.]

(c) When ρ is moderately large a rather inaccurate pre-determined ρ_o may be used in an approximate BQUE of σ_a^2 to yield an estimator with smaller variance than the ANOVA estimator. [Figures 1-5.]

(d) Data which are not badly unbalanced offer little opportunity for reducing the variance of the approximate BQUE of σ_a^2 below that of the ANOVA estimator. [Figure 1.]

(e) Increasing the number of observations, while holding the number of groups constant, extends the region in which the approximate BQUE of σ_a^2 is better than the ANOVA estimator. [Figures 2 and 3.]

(f) Increasing the number of groups while holding the total number of observations constant curtails the region in which the approximate BQUE of σ_a^2 is better than the ANOVA estimator. [Figures 2 and 4.]

(g) An increase in both total number of observations and number of groups does not guarantee that the ANOVA estimator will compare favorably with the BQUE. [Figure 5.]

For the non-zero mean

(h) The suggested estimator of σ_a^2, whilst not a BQUE nor even a direct approximation,thereto appears close to it for $\rho_o = \rho$; and it is better than the ANOVA estimator over a wide range of values. [Figures 6-8.]

References

Conte, S. D. (1965). "Elementary Numerical Analysis". McGraw-Hill Book Co., New York, N. Y.

Crump, S. Lee (1951). The present status of variance component analysis. Biometrics 7:1-16.

Graybill, Franklin A. (1961). "An Introduction to Linear Statistical Models". McGraw-Hill Book Co., New York, N.Y.

Harville, David A. (1967). Statistical dependence between random effects and the numbers of observations on the effects for the unbalanced one-way random classification. J. Amer. Stat. Assoc. 62:1375-1386.

Koch, Gary G. (1967). A general approach to the estimation of variance components. Technometrics 9:93-118.

Lancaster, H. O. (1965). The Helmert Matrices. Ann. Math. Monthly, 72, 4-12.

Robertson, Alan (1962). Weighting in the estimation of variance components in the unbalanced single classification. Biometrics 18:413-417.

Searle, S. R. (1956). "Matrix methods in components of variance and co-variance analysis. Ann. Math. Stat. 27:737-748.

Searle, S. R. (1966). "Matrix Algebra for the Biological Sciences". John Wiley and Sons, Inc., New York, N.Y.

Townsend, E. C. (1968). Unbiased estimators of variance components in single unbalanced designs. Unpublished Ph.D. Thesis, Cornell University, Ithaca, N.Y.

Tukey, J. W. (1957). Variances of variance components: II. The unbalanced single classification. Ann. Math. Stat. 28:43-56.

5. APPENDIX

The equation of the model, (1), can be written as $\mathbf{y} = \mathbf{Z}\boldsymbol{\alpha} + \mathbf{e}$ where $\boldsymbol{\alpha}$ is the $c \times 1$ vector of α_i's and $\mathbf{Z}$ is the corresponding incidence matrix, a diagonal matrix of vectors $\mathbf{1}_i$, for $\mathbf{1}_i$ being a vector of n_i ones. Thus

$$\mathbf{1}_i' = (1\ 1\ 1\ \ldots\ 1), \text{ of } n_i \text{ elements}$$

and

$$\mathbf{Z} = \operatorname{diag}\{\mathbf{1}_1\ \mathbf{1}_2\ \ldots\ \mathbf{1}_c\} = \sum_{i=1}^{c}{}^{+}\, \mathbf{1}_i \quad . \tag{A1}$$

Consequently $\mathbf{V}$, the variance-covariance matrix of $\mathbf{y}$ is

$$\mathbf{V} = \alpha\mathbf{Z}\mathbf{Z}' + e\mathbf{I} = \sum_{i=1}^{c}{}^{+}(\alpha\mathbf{J}_i + e\mathbf{I}_i) \tag{A2}$$

as in (2) and (3), except for writing

$$\alpha \equiv \sigma_a^2 \text{ and } e \equiv \sigma_e^2$$

for notational convenience.

BQUE's of e and α are derived by obtaining $\mathbf{A}$ such that $2\operatorname{tr}(\mathbf{VA})^2$ is minimized subject to $\sigma^2 = \operatorname{tr}(\mathbf{VA})$ for $\sigma^2 = e$ and $\sigma^2 = \alpha$ in turn. This would be easily achieved if $\mathbf{V}$ were diagonal. To attain this form we make the transformation $\mathbf{x} = \mathbf{P}'\mathbf{y}$ where $\mathbf{P}$ is an orthogonal matrix such that $\mathbf{P}'\mathbf{VP} = \mathbf{D}$, the diagonal matrix of latent roots of $\mathbf{V}$. $\hat{\sigma}^2$ of (4) is then $\hat{\sigma}^2 = \mathbf{x}'\mathbf{P}'\mathbf{APx} = \mathbf{x}'\mathbf{Qx}$ for $\mathbf{Q} = \mathbf{P}'\mathbf{AP}$ and we have to minimize

$$2\operatorname{tr}(\mathbf{VA})^2 = 2\operatorname{tr}(\mathbf{PDP}'\mathbf{A})^2 = 2\operatorname{tr}(\mathbf{P}'\mathbf{APD})^2 = 2\operatorname{tr}(\mathbf{QD})^2 \tag{A3}$$

subject to

$$\sigma^2 = \mathrm{tr}(\underline{\underline{V}}\underline{\underline{A}}) = \mathrm{tr}(\underline{\underline{Q}}\underline{\underline{D}}) \quad . \tag{A4}$$

Having found $\underline{Q}$ to achieve this, the BQUE of σ^2 is

$$_B\hat{\sigma}^2 = \underline{x}'\underline{\underline{Q}}\underline{x} = \underline{y}'\underline{\underline{P}}\underline{\underline{Q}}\underline{\underline{P}}'\underline{y} \tag{A5}$$

with variance

$$\mathrm{var}(_B\hat{\sigma}^2) = 2\mathrm{tr}(\underline{\underline{Q}}\underline{\underline{D}})^2 \quad . \tag{A6}$$

This is done in turn for $\sigma^2 = e$ and $\sigma^2 = \alpha$. We first find the latent roots and vectors of $\underline{V}$ and use them to derive $\underline{D} = \underline{\underline{P}}\underline{\underline{V}}\underline{\underline{P}}'$ and $\underline{P}$.

4a. Latent roots of $\underline{V}$

From (A2), the latent roots of $\underline{V}$ are the solutions for λ to

$$|\underline{V}-\lambda\underline{I}| = \prod_{i=1}^{c} |\alpha\underline{J}_i + (e-\lambda)\underline{I}_i| = \prod_{i=1}^{c} (e-\lambda)^{n_i-1}(n_i\alpha + e-\lambda) = 0 ,$$

using, for example, Searle (1966, p. 198) for the expansion of the determinants. Hence

$$(e-\lambda)^{N-c}\prod_{i=1}^{c}(n_i\alpha + e-\lambda) = 0$$

and the latent roots of $\underline{V}$ are e with multiplicity $N-c$, and $n_i\alpha + e$ for $i=1,\ldots,c$. We denote these by

$$\lambda_k = n_k\alpha + e \text{ for } k=1,2,\ldots,c \text{ and } \lambda_k = e \text{ for } k = c+1,\ldots,N , \tag{A7}$$

and have

$$\underline{D} = \mathrm{diag}\{\lambda_k\} \text{ for } k = 1,2,\ldots,c,c+1,\ldots,N \quad . \tag{A8}$$

4b. Latent vectors of $\mathbf{V}$

Let

$$\mathbf{U} = \left(\mathbf{U}_1 \;\vdots\; \mathbf{U}_2\right) = \left(\mathbf{u}_1\ \mathbf{u}_2, \ldots, \mathbf{u}_c \;\vdots\; \mathbf{u}_{c+1}\ \mathbf{u}_{c+2}, \ldots, \mathbf{u}_N\right)$$

be the matrix of N latent vectors corresponding to the latent roots in (A7). We seek a value for $\mathbf{U}$ such that its columns are pair-wise orthogonal; column normalization of $\mathbf{U}$ then yields $\mathbf{P}$, required in (A5) and (A6).

Note first that $\mathbf{Z}$ is a satisfactory value for $\mathbf{U}_1$; i.e., that columns of $\mathbf{Z}$ are latent vectors of $\mathbf{V}$ corresponding to the latent roots $\lambda_k = n_k\alpha + e$ for $k = 1,2,\ldots,c$. This is so because from (A1) and (A2)

$$\mathbf{VZ} = \left[\sum_{i=1}^{c}{}^{+}(\alpha\mathbf{J}_i + e\mathbf{I}_i)\right]\left[\sum_{i=1}^{c}{}^{+}\mathbf{1}_i\right] = \sum_{i=1}^{c}{}^{+}(\alpha n_i\mathbf{1}_i + e\mathbf{1}_i) = \sum_{i=1}^{c}{}^{+}(\alpha n_i + e)\mathbf{1}_i ,$$

which, on comparison with (A2) is $\mathbf{Z}$ with its i'th column multiplied by $\alpha n_i + e$. Furthermore, from (A2), it is clear that $\mathbf{Z}'\mathbf{Z} = \text{diag}\{n_i\}$ for $i = 1,2,\ldots,c$; hence columns of $\mathbf{Z}$ are pair-wise orthogonal. Thus we take $\mathbf{U}_1 = \mathbf{Z}$ and write $\mathbf{U} = (\mathbf{Z}\ \mathbf{U}_2)$. For the columns of $\mathbf{U}$ to be pair-wise orthogonal we want $\mathbf{U}_2'\mathbf{U}_2$ diagonal and $\mathbf{Z}'\mathbf{U}_2 = 0$; and for them to also be latent vectors of $\mathbf{V}$ corresponding to the N-c latent roots e of (A7) we must have $\mathbf{VU}_2 = e\mathbf{U}_2$. Using (A2) for $\mathbf{V}$ this means $\mathbf{ZZ}'\mathbf{U}_2 = \mathbf{0}$, which is satisfied when $\mathbf{Z}'\mathbf{U}_2 = \mathbf{0}$. Hence it is sufficient that $\mathbf{U}_2'\mathbf{U}_2$ be diagonal and $\mathbf{Z}'\mathbf{U}_2 = 0$. We show that

$$\mathbf{U}_2 = \sum_{i=1}^{c}{}^{+}\mathbf{E}_i \text{ with } \mathbf{E}_i = \begin{bmatrix} 1 & 1 & 1 & \cdots & 1 \\ -1 & 1 & 1 & \cdots & 1 \\ 0 & -2 & 1 & \cdots & 1 \\ 0 & 0 & -3 & \cdots & 1 \\ \vdots & \vdots & & \ddots & \vdots \\ 0 & \cdots & & & -(n_i-1) \end{bmatrix}, \quad \text{of order } n_i \times (n_i - 1)$$

satisfies these conditions. We have, by the nature of $\underset{\sim}{E}_i$,

$$\underset{\sim}{U}_2'\underset{\sim}{U}_2 = \overset{c}{\underset{i=1}{\Sigma^+}} \underset{\sim}{E}_i'\underset{\sim}{E}_i = \overset{c}{\underset{i=1}{\Sigma^+}} \left[\mathrm{diag}\{j(j+1)\} \text{ for } j = 1,2,\ldots,n_i-1\right]$$

which is diagonal; and from (A1)

$$\underset{\sim}{Z}_2'\underset{\sim}{U}_2 = \overset{c}{\underset{i=1}{\Sigma^+}} \underset{\sim}{1}_i'\underset{\sim}{E}_i = \underset{\sim}{0} \quad .$$

Hence

$$\underset{\sim}{U} = (\underset{\sim}{Z}\ \underset{\sim}{U}_2) = \left[\overset{c}{\underset{i=1}{\Sigma^+}} \underset{\sim}{1}_i \quad \overset{c}{\underset{i=1}{\Sigma^+}} \underset{\sim}{E}_i\right]$$

is a matrix of latent roots of $\underset{\sim}{V}$. Normalizing the columns of $\underset{\sim}{U}$ yields $\underset{\sim}{P}$ as

$$\underset{\sim}{P} = \left[\overset{c}{\underset{i=1}{\Sigma^+}} \sqrt{n_i^{-1}}\,\underset{\sim}{1}_i \quad \overset{c}{\underset{i=1}{\Sigma^+}} \underset{\sim}{G}_i\right] \tag{A9}$$

where

$$\underset{\sim}{G}_i = \underset{\sim}{E}_i \left[\mathrm{diag}\{\ \sqrt{2}\ \ \sqrt{6}\ \ldots\ \sqrt{j(j+1)}\ \ldots\ \sqrt{n_i(n_i-1)}\}\right]^{-1} \tag{A10}$$

$$= \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{6} & \ldots & 1/\sqrt{n_i(n_i-1)} \\ -1/\sqrt{2} & 1/\sqrt{6} & \ldots & 1/\sqrt{n_i(n_i-1)} \\ 0 & -2/\sqrt{6} & \ldots & 1/\sqrt{n_i(n_i-1)} \\ 0 & 0 & \ddots & \vdots \\ \vdots & & & \\ 0 & 0 & \ldots & -(n_i-1)/\sqrt{n_i(n_i-1)} \end{bmatrix} \quad .$$

We may note that a rearrangement of the columns of $\underline{P}$ gives a matrix $\sum_{i=1}^{c}{}^{+}\left[\sqrt{n_i^{-1}}\,\underline{1}_i \quad \underline{G}_i\right]$, where $\left[\sqrt{n_i^{-1}}\,\underline{1}_i' \quad \underline{G}_i'\right]$ is a Helmert matrix in the "strict sense," as given by Lancaster [1965] .

<u>BQUE of σ_e^2</u>

Using $\underline{P}$ of (A9), the BQUE of σ_e^2 comes from (A3) and (A4) by choosing $\underline{Q}$ so as to minimize, for $\underline{D}$ of (A8),

$$2\mathrm{tr}(\underline{Q}\underline{D})^2 = 2\sum_{k,k'} \lambda_k \lambda_{k'} q_{kk'}^2$$

subject to

$$e = \mathrm{tr}(\underline{Q}\underline{D}) = \sum_{k=1}^{N} \lambda_k q_{kk} = \alpha \sum_{k=1}^{c} n_k q_{kk} + e \sum_{k=1}^{N} q_{kk} \quad .$$

This last equation implies

$$\sum_{k=1}^{c} n_k q_{kk} = 0 \quad \text{and} \quad \sum_{k=1}^{N} q_{kk} = 1 \quad , \tag{A11}$$

so that we have to minimize

$$\theta = 2\sum_{k,k'=1}^{N} \lambda_k \lambda_{k'} q_{kk'}^2 + 4m_1\left(\sum_{k=1}^{N} q_{kk} - 1\right) + 4m_2 \sum_{k=1}^{c} n_k q_{kk} \quad ,$$

where $4m_1$ and $4m_2$ are Lagrange multipliers. Equating to zero the derivatives of θ with respect to $q_{kk'}$ for $k \neq k'$ gives $8\lambda_k\lambda_{k'}q_{kk'} = 0$. Since none of the λ's are zero [see (A7)] this means $q_{kk'} = 0$ for all $k \neq k'$. Hence $\underline{q}$ is diagonal. Equating to zero the derivatives of θ with respect to m_1 and m_2 yields (A11); and with respect to q_{kk} yields

$$4\lambda_k^2 q_{kk} + 4m_1 + 4m_2 n_k = 0 \quad \text{for } k \leq c \quad ,$$

and (A12)

$$4\lambda_k^2 q_{kk} + 4m_1 = 0 \qquad \text{for } k > c \ .$$

Substituting for q_{kk} from these equations into (A11) then gives

$$um_1 + wm_2 = -1$$

and (A13)

$$wm_1 + vm_2 = 0$$

where

$$u = \sum_{k=1}^{N} 1/\lambda_k^2 \ , \quad v = \sum_{k=1}^{c} n_k^2/\lambda_k^2 \quad \text{and} \quad w = \sum_{k=1}^{c} n_k/\lambda_k^2 \ . \tag{A14}$$

With

$$\Delta = uv - w^2$$

(A13) have solutions

$$m_1 = -v/\Delta \quad \text{and} \quad m_2 = w/\Delta \ . \tag{A15}$$

Hence from (A12)

$$q_{kk} = (v - wn_k)/(\Delta\lambda_k^2) \quad \text{for } k = 1,2,\ldots,c$$

and (A16)

$$= v/(e^2\Delta) \qquad \text{for } k = c+1,\ldots,N \ .$$

Substituting these values for the elements of the (diagonal) matrix $\underline{Q}$ in (A5) it can be shown, after a little algebraic simplification, that the BQUE of σ_e^2 is

$$ {}_B\hat{\sigma}_e^2 = (rs-t^2)^{-1}\left\{\sum_{i=1}^{c}(s-tn_i)(1+n_i\rho)^{-2}(y_{i.}^2/n_i) + s\left(\sum_{i=1}^{c}\sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{c} y_{i.}^2/n_i\right)\right\} \quad (A17)$$

where $\rho = \sigma_a^2/\sigma_e^2$, $N = \sum_{i=1}^{c} n_i$ and r, s, and t are as given in (5), derived from (A14) by multiplying by σ_e^2 and using (A7) and ρ . Also, the variance of this estimator, derived from (A6) using (A15) for $\underline{Q}$ and (A7) for $\underline{D}$ reduces to

$$\mathrm{var}({}_B\hat{\sigma}_e^2) = 2s\sigma_e^4/(rs-t^2) \quad . \qquad (A18)$$

BQUE of σ_a^2

The BQUE of σ_a^2 is derived exactly as is that of σ_e^2 save for minimizing $2\mathrm{tr}(\underline{QD})^2$ subject to

$$\alpha = \mathrm{tr}(\underline{QD}) = \alpha\sum_{k=1}^{c} n_k q_{kk} + \sum_{k=1}^{N} q_{kk} = 1 \quad .$$

The effect of this is to interchange the 0 and 1 on the right-hand sides of the equations in (A11). Consequently the right-hand sides of (A13) get interchanged so that solutions of the resulting equations are

$$m_1 = w/\Delta \quad \text{and} \quad m_2 = -u/\Delta \quad .$$

Compared to (A15) this means replacing -v by w and w by -u , or equivalently -s by t and t by -r, in the numerator of ${}_B\hat{\sigma}_e^2$ to obtain ${}_B\hat{\sigma}_a^2$. Hence the BQUE of σ_a^2 is

$$ {}_B\hat{\sigma}_a^2 = (rs-t^2)^{-1}\left[\sum_{i=1}^{c}(rn_i-t)(1+n_i\rho)^{-2}(y_{i.}^2/n_i) - t\left(\sum_{i=1}^{c}\sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{c} y_{i.}^2/n_i\right)\right], \quad (A19)$$

and its variance is

$$\mathrm{var}({}_B\hat{\sigma}_a^2) = 2r\sigma_e^4/(rs-t^2). \qquad (A20)$$

Figure 1

Relationship Between Variance of ANOVA Estimator and Variance of Approximate BQUE of σ_a^2

$\mu = 0$

$n_i = 3,4,5,6,7$

$p_a = [\mathrm{Var}(\hat{\sigma}_a^2) - \mathrm{Var}(\tilde{\sigma}_a^2)]\ /\mathrm{Var}(\tilde{\sigma}_a^2)$

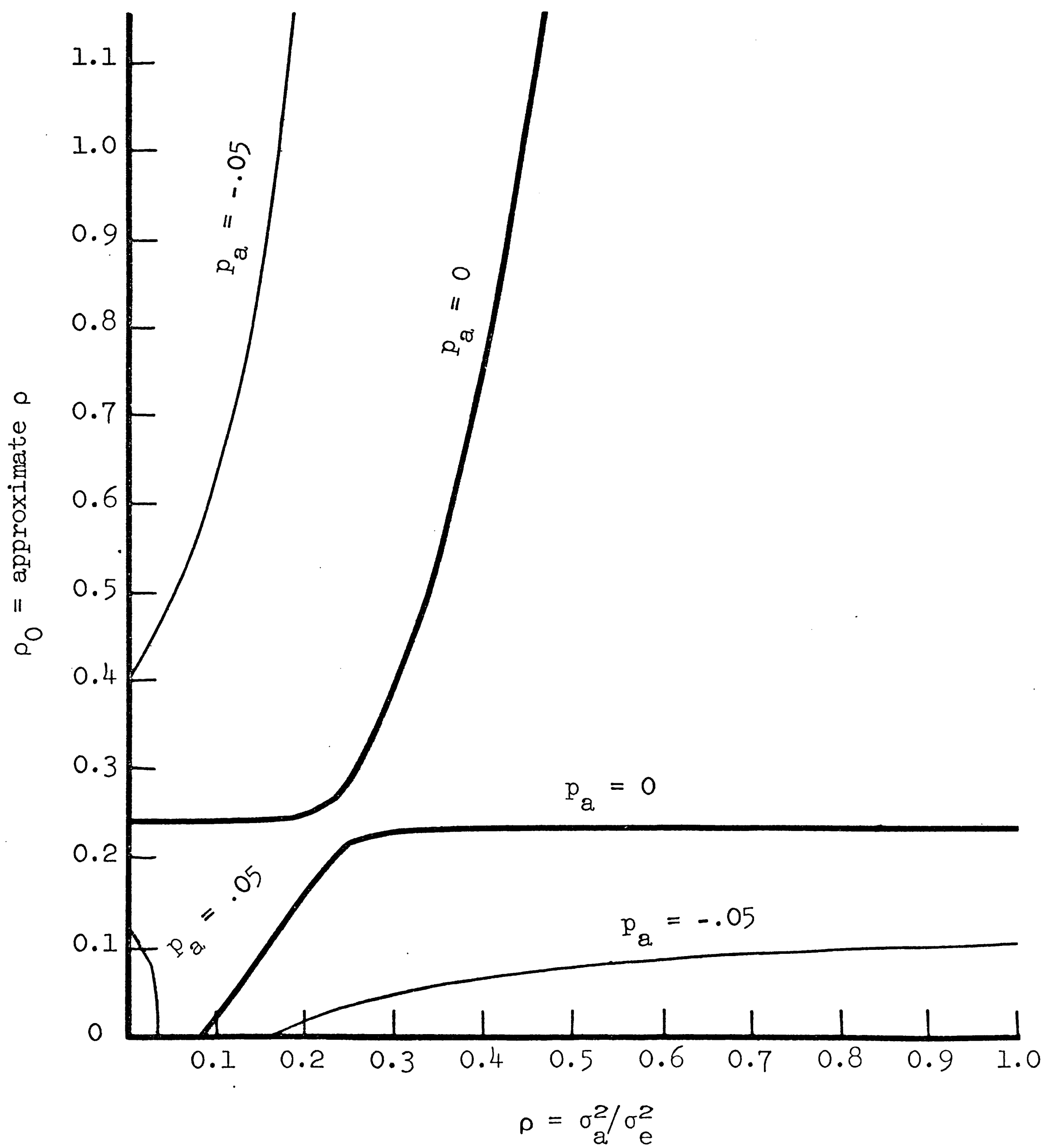

Figure 1 a

Relationship Between Variance of ANOVA Estimator and Variance of Approximate BQUE of σ_a^2

$\mu = 0$

$n_i = 3,4,5,6,7$

$p_a = [\mathrm{Var}(\hat{\sigma}_a^2) - \mathrm{Var}(\tilde{\sigma}_a^2)] / \mathrm{Var}(\tilde{\sigma}_a^2)$

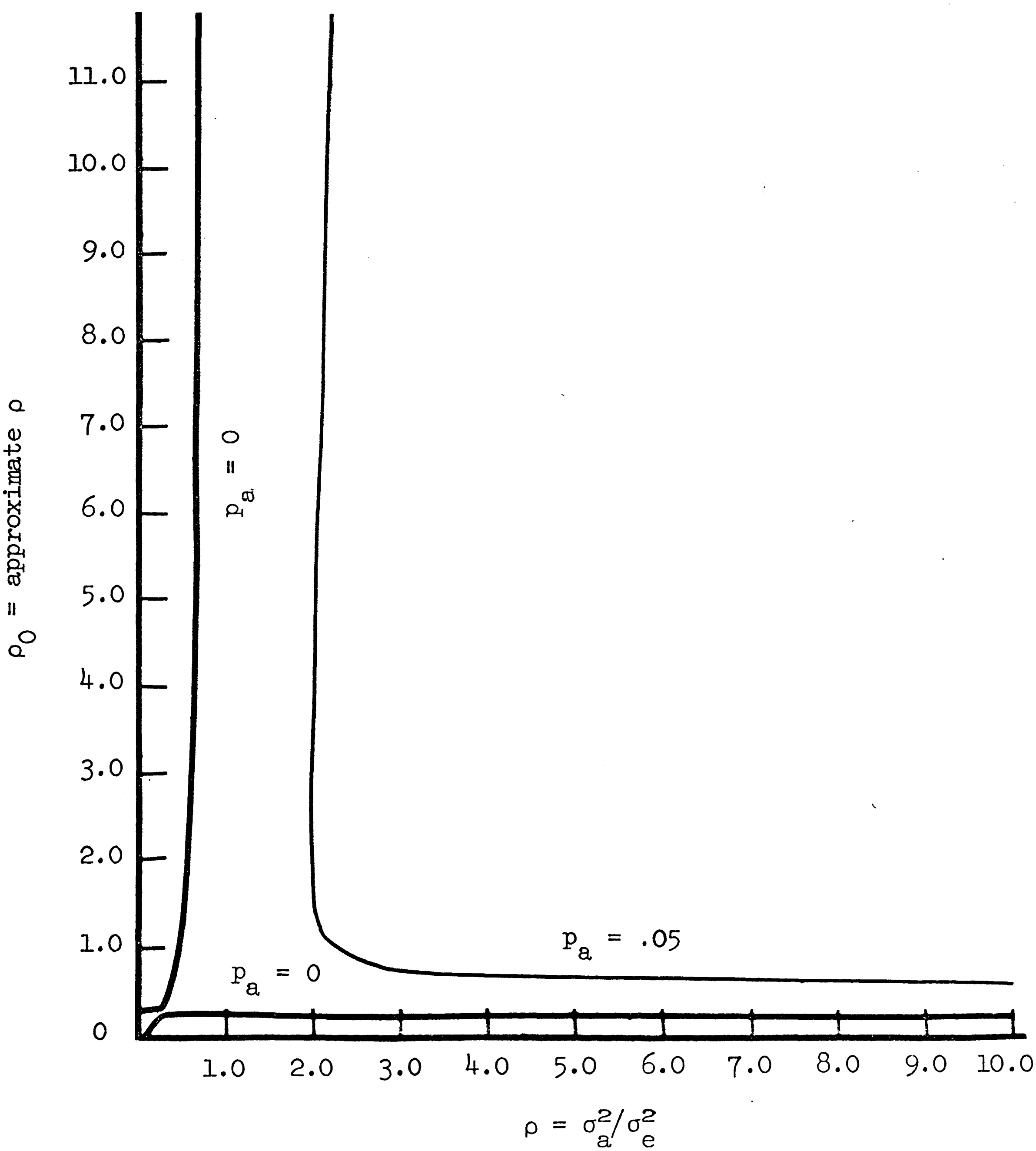

Figure 2

Relationship Between Variance of ANOVA Estimator and Variance of Approximate BQUE of σ_a^2

$\mu = 0$

$n_i = 1,1,1,11,11$

$p_a = [Var(\hat{\sigma}_a^2) - Var(\tilde{\sigma}_a^2)] / Var(\tilde{\sigma}_a^2)$

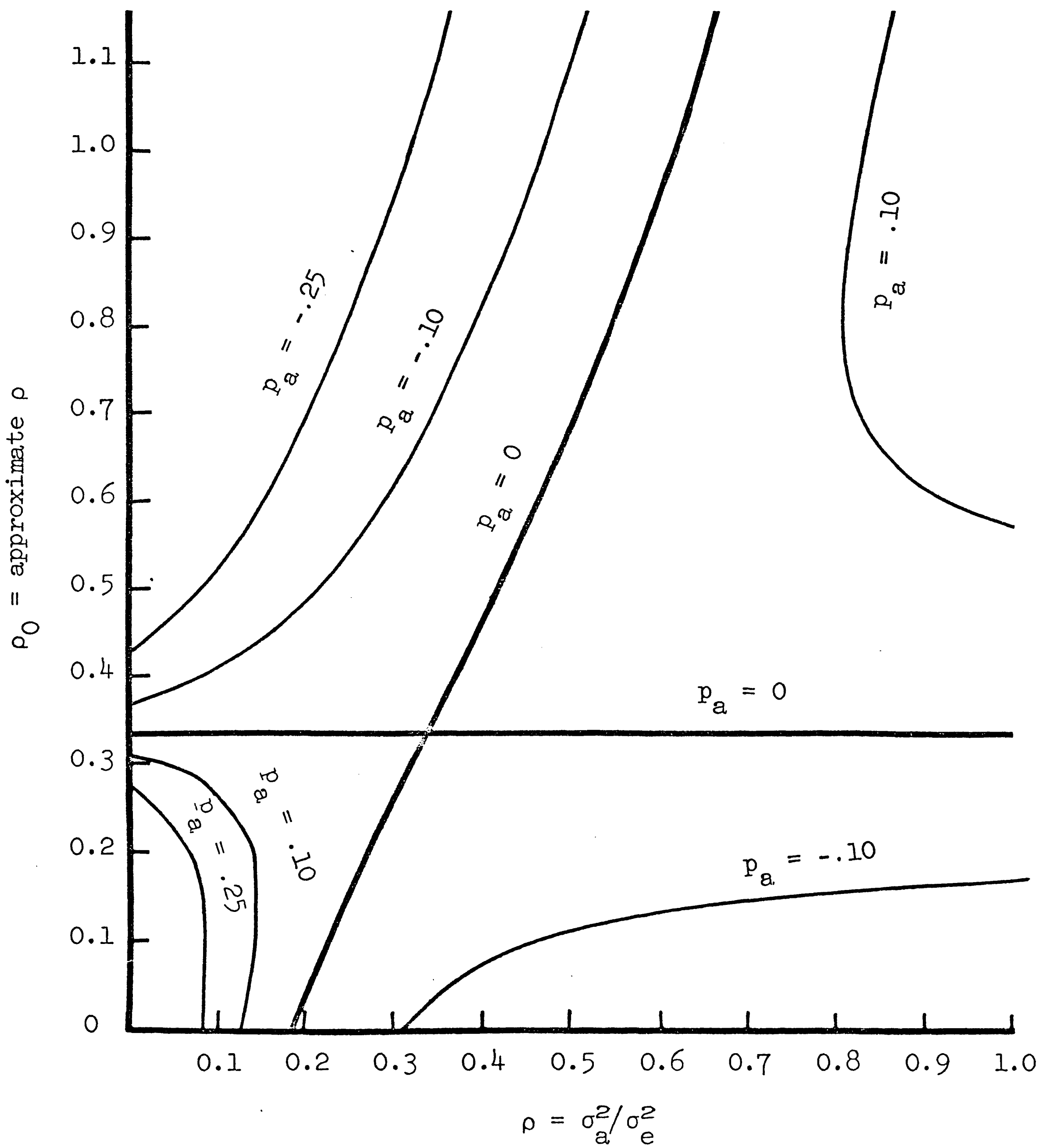

Figure 2a

Relationship Between Variance of ANOVA Estimator and
Variance of Approximate BQUE of σ_a^2

$\mu = 0$

$n_i = 1,1,1,11,11$

$p_a = [\mathrm{Var}(\hat{\sigma}_a^2) - \mathrm{Var}(\tilde{\sigma}_a^2)] / \mathrm{Var}(\hat{\sigma}_a^2)$

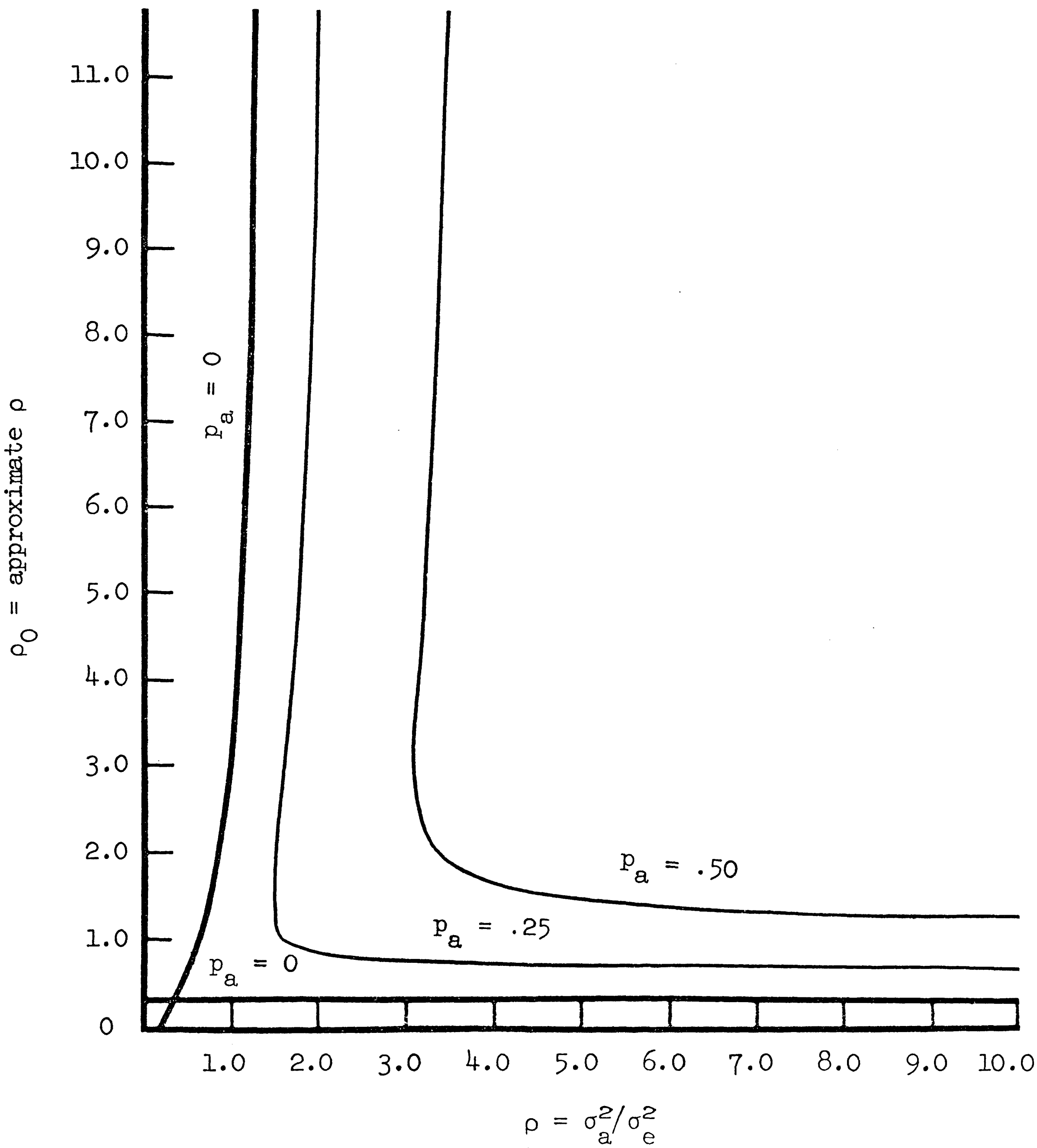

Figure 3

Relationship Between Variance of ANOVA Estimator and Variance of Approximate BQUE of σ_a^2

$\mu = 0$

$n_i = 2,2,2,22,22$

$p_a = [\mathrm{Var}(\hat{\sigma}_a^2) - \mathrm{Var}(\tilde{\sigma}_a^2)] / \mathrm{Var}(\tilde{\sigma}_a^2)$

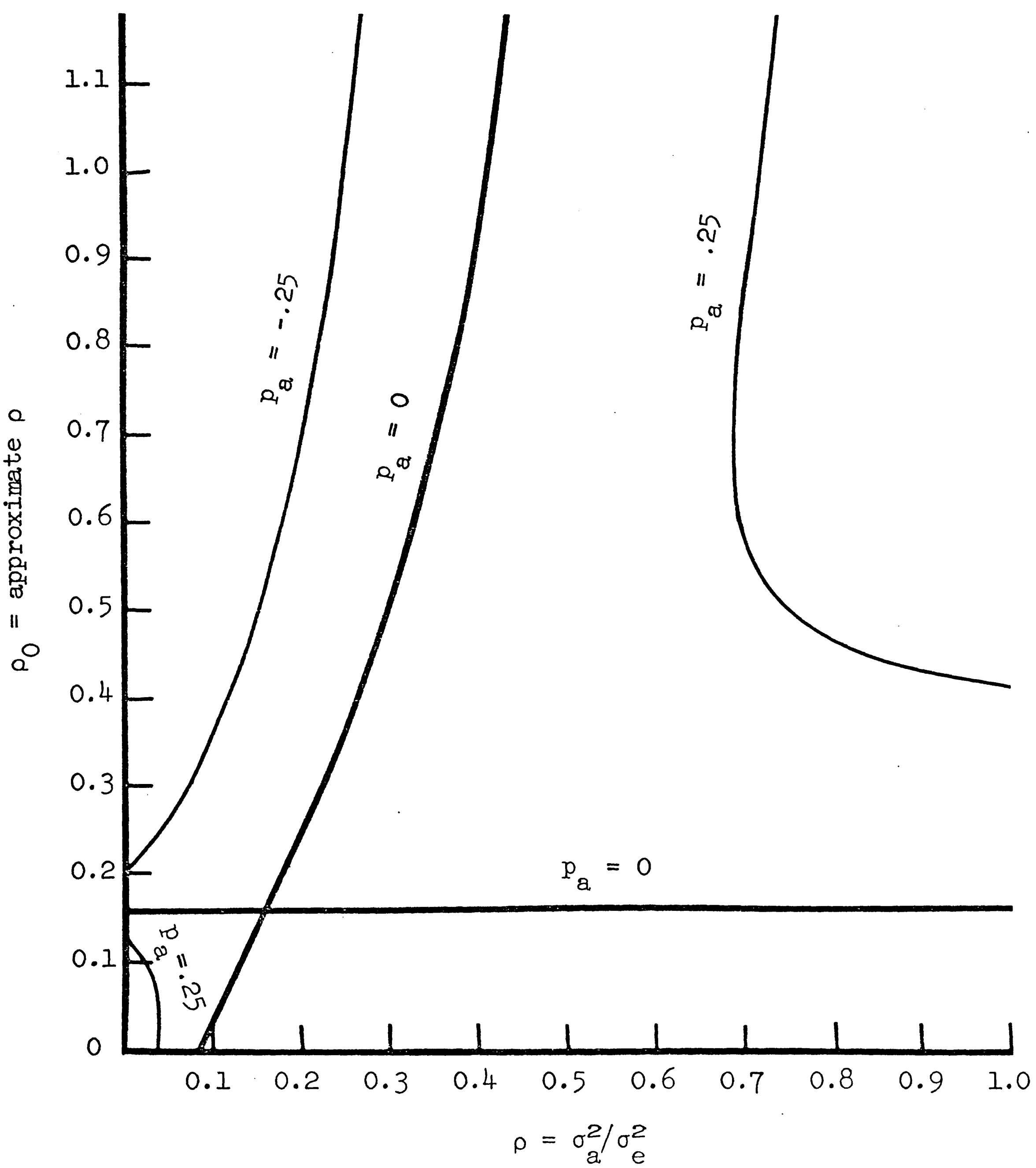

Figure 4

Relationship Between Variance of ANOVA Estimator and Variance of Approximate BQUE of σ_a^2

$\mu = 0$

$n_i = 1,1,1,1,1,4,4,4,4,4$

$p_a = [\mathrm{Var}(\hat{\sigma}^2) - \mathrm{Var}(\tilde{\sigma}^2)] / \mathrm{Var}(\tilde{\sigma}^2)$

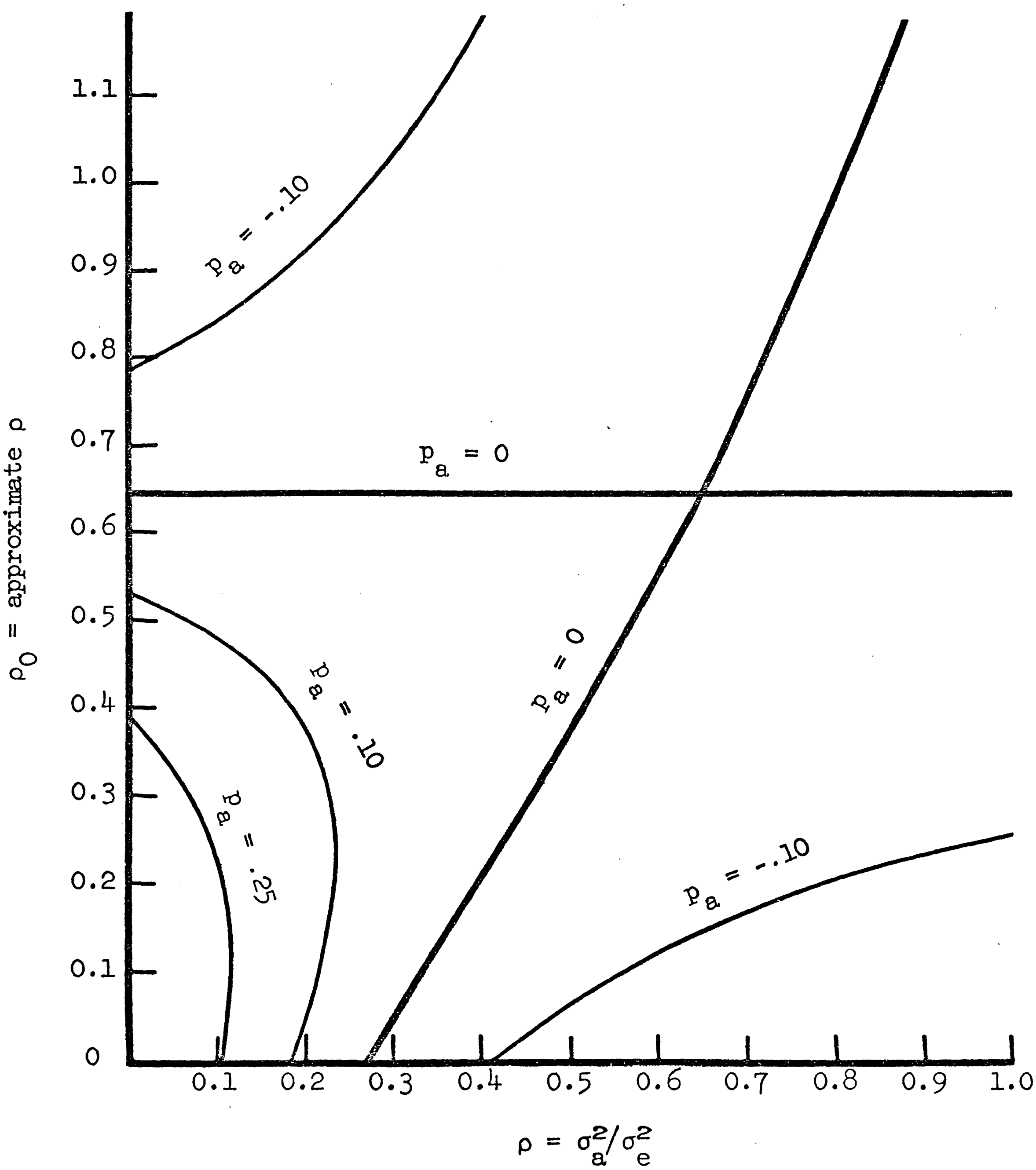

Figure 5

Relationship Between Variance of ANOVA Estimator and Variance of Approximate BQUE of σ_a^2

$\mu = 0$

$$n_i = \begin{array}{l} 1,1,\ 1,\ 1,\ 1,\ 1,\ 1,\ 1,\ 1,\ 1, \\ 2,2,\ 2,\ 2,\ 2,\ 2,\ 2,\ 2,\ 2,\ 2, \\ 3,3,\ 3,\ 3,\ 3,\ 3,\ 3,\ 3,\ 3,\ 3, \\ 4,4,\ 4,\ 4,\ 4,\ 5,\ 5,\ 5,\ 5,\ 5, \\ 6,8,10,12,14,35,40,45,50,55 \end{array}$$

$$p_a = [\mathrm{Var}(\hat{\sigma}_a^2) - \mathrm{Var}(\tilde{\sigma}_a^2)] / \mathrm{Var}(\tilde{\sigma}_a^2)$$

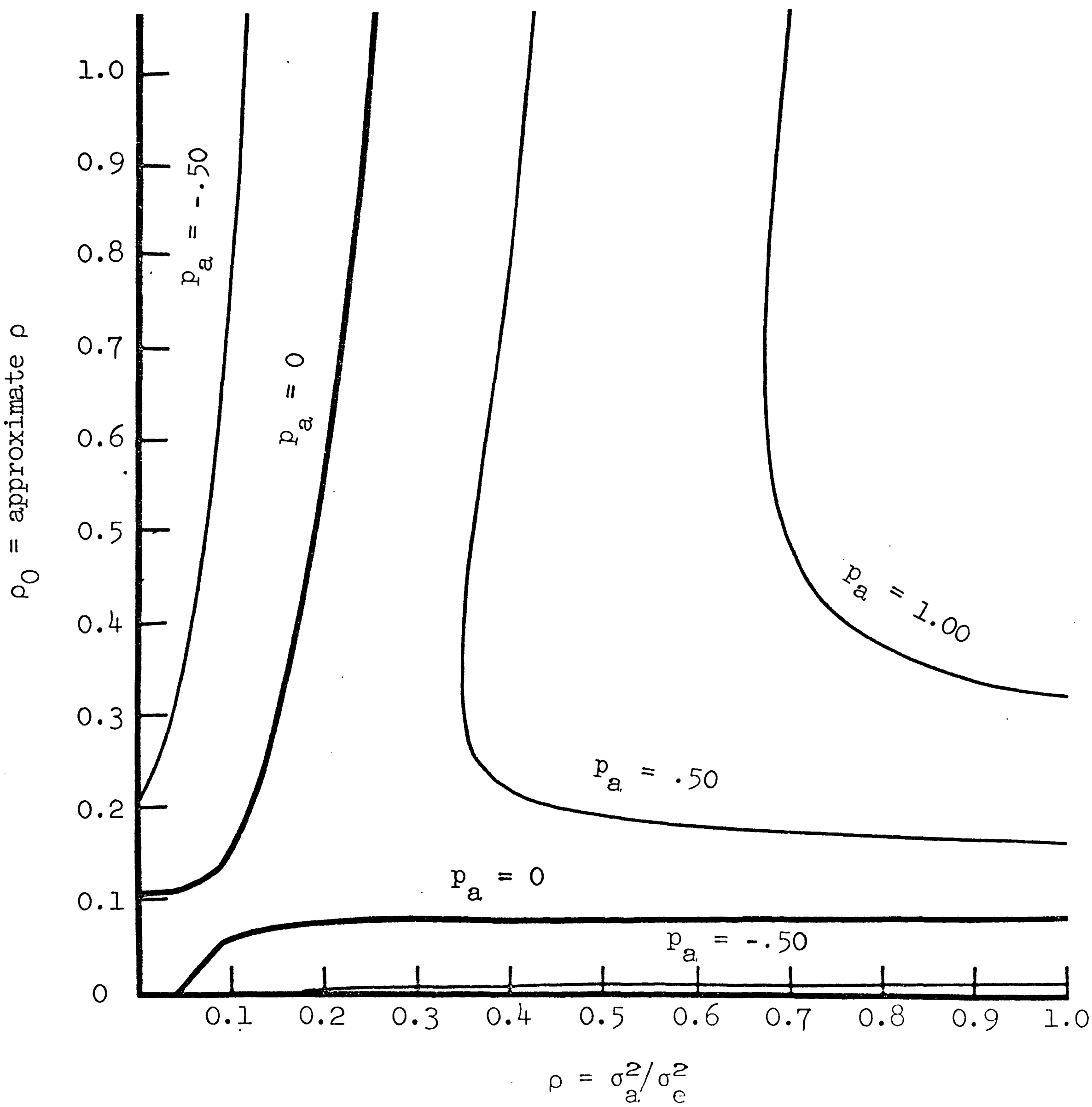

Figure 6

Relationship Between Variance of ANOVA Estimator and Variance of Suggested Estimator of σ_a^2

$\mu \neq 0$

$n_i = 3,4,5,6,7$

$p_a = [\mathrm{Var}(\hat{\sigma}_a^2) - \mathrm{Var}(\tilde{\tilde{\sigma}}_a^2)] / \mathrm{Var}(\tilde{\tilde{\sigma}}_a^2)$

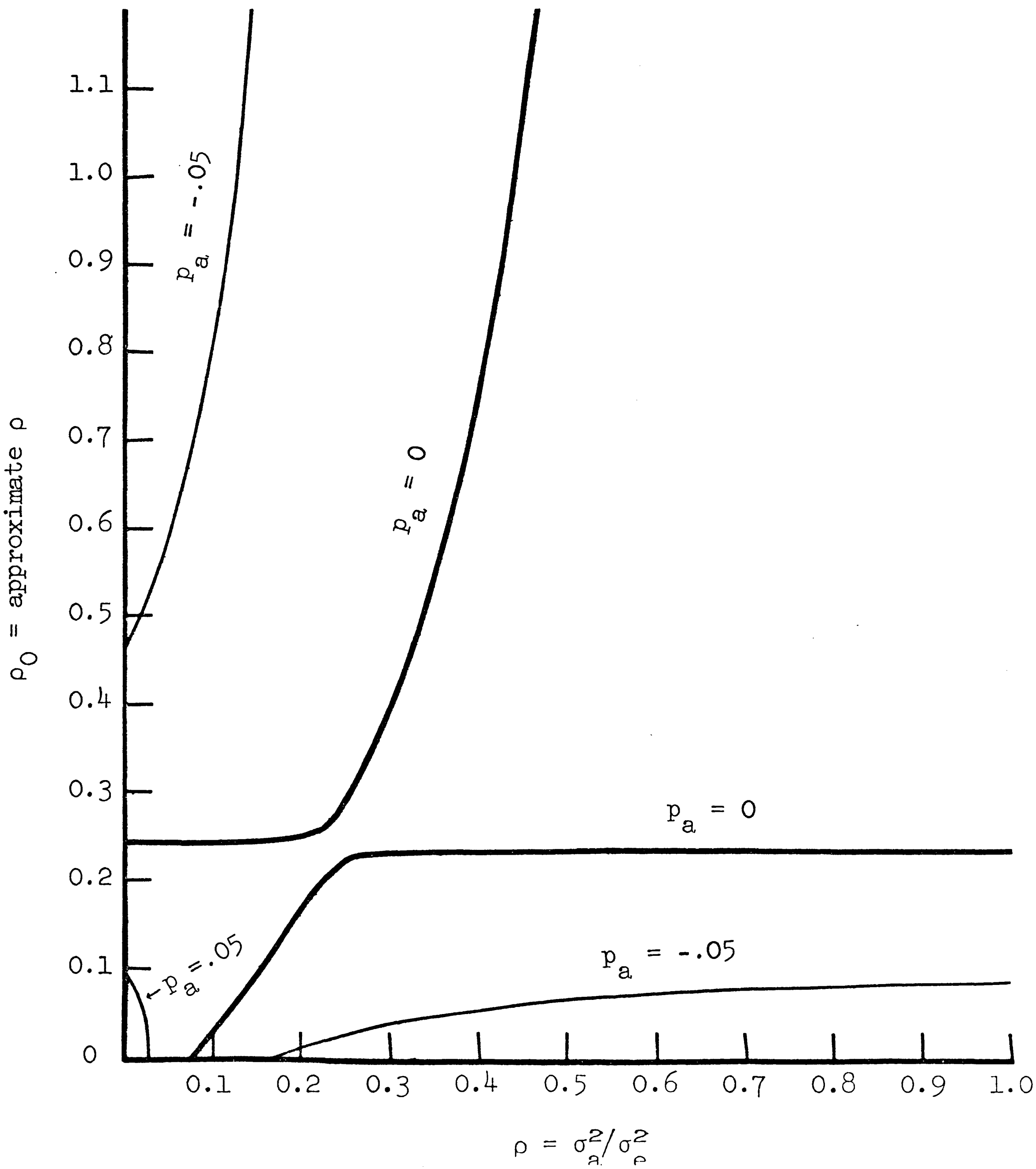

Figure 7

Relationship Between Variance of ANOVA Estimator and Variance of Suggested Estimator of σ_a^2

$\mu \neq 0$

$n_i = 1,1,1,11,11$

$p_a = [\mathrm{Var}(\hat{\sigma}_a^2) - \mathrm{Var}(\tilde{\tilde{\sigma}}_a^2)] / \mathrm{Var}(\tilde{\tilde{\sigma}}_a^2)$

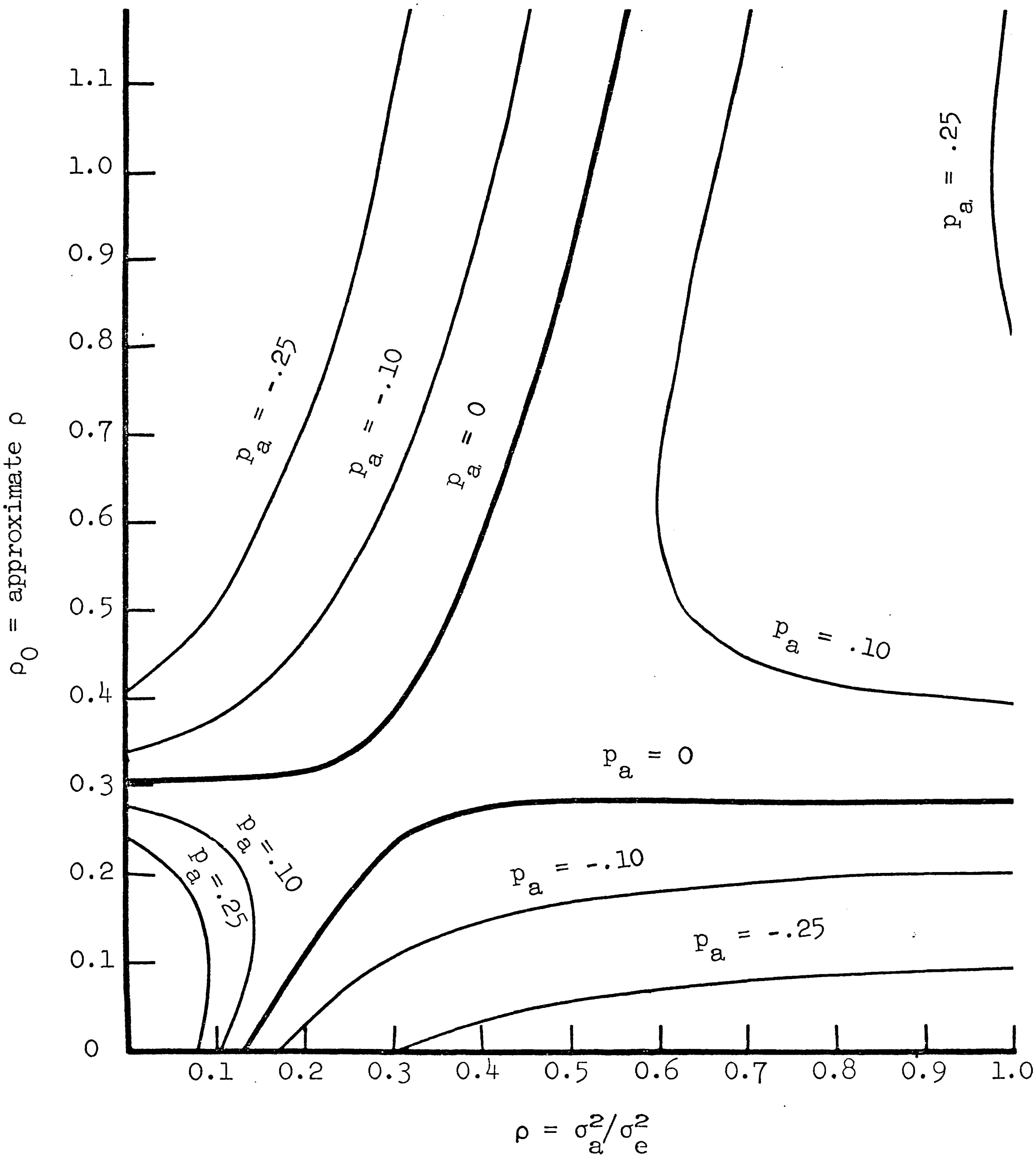

Figure 8

Relationship Between Variance of ANOVA Estimator and Variance of Suggested Estimator of σ_a^2

$\mu \neq 0$

n_i = 1,1, 1, 1, 1, 1, 1, 1, 1, 1,
2,2, 2, 2, 2, 2, 2, 2, 2, 2,
3,3, 3, 3, 3, 3, 3, 3, 3, 3,
4,4, 4, 4, 4, 5, 5, 5, 5, 5,
6,8,10,12,14,35,40,45,50,55

$p_a = [\mathrm{Var}(\hat{\sigma}_a^2) - \mathrm{Var}(\tilde{\tilde{\sigma}}_a^2)] / \mathrm{Var}(\tilde{\tilde{\sigma}}_a^2)$