DOCUMENT RESUME

ED 099 406 95 TM 004 060

AUTHOR Tatsuoka, Maurice M.
TITLE An Examination of the Statistical Properties of a Multivariate Measure of Strength of Relationship. Final Report.
SPONS AGENCY Office of Education (DHEW), Washington, D.C. Bureau of Research.
BUREAU NO BR-2-E-020-FR
PUB DATE Dec 73
GRANT OEG-5-72-0027(509)
NOTE 58p.

EDRS PRICE MF-$0.75 HC-$3.15 PLUS POSTAGE
DESCRIPTORS *Analysis of Variance; Computer Programs; Matrices; *Sampling; *Statistical Bias; *Tests of Significance
IDENTIFIERS MANOVA; *Multivariate Analysis of Variance

ABSTRACT

A computer-simulated study was made of the sampling distribution of omega squared, a measure of strength of relationship in multivariate analysis of variance which had earlier been proposed by the author. It was found that this measure was highly positively biased when the number of variables is large and the sample size is small. A correction formula for reducing the bias was developed by the method of least squares and was found to yield nearly unbiased corrected values. A simpler, rule-of-thumb correction formula was also presented. Charts of 90 percent confidence intervals for omega squared for the three-variate, five-group case for three sample sizes were drawn, and data that would enable the construction of similar confidence intervals for the five-variate and ten-variate cases (also for five groups) were presented. (Author/RC)

FINAL REPORT

Project No. 2-E-020
Grant No. OEG-5-72-0027(509)

AN EXAMINATION OF THE STATISTICAL PROPERTIES OF A MULTIVARIATE MEASURE OF STRENGTH OF RELATIONSHIP

Maurice M. Tatsuoka
University of Illinois
Urbana, Illinois 61801
December, 1973

The research reported herein was performed pursuant to a grant with the Office of Education, U. S. Department of Health, Education, and Welfare. Contractors undertaking such projects under Government sponsorship are encouraged to express freely their professional judgment in the conduct of the project. Points of view or opinions stated do not, therefore, necessarily represent official Office of Education position or policy.

U. S. DEPARTMENT OF
HEALTH, EDUCATION, and WELFARE
Office of Education, Bureau of Research

TABLE OF CONTENTS

AN EXAMINATION OF THE STATISTICAL PROPERTIES OF A MULTIVARIATE MEASURE OF STRENGTH OF RELATIONSHIP

1. INTRODUCTION

It is well known that statistical significance, by itself, offers no guarantee that the difference, relationship, or effect found in an experiment is of practical significance. If the sample size is extremely large—as it is in most educational research—virtually any miniscule observed effect will be judged "significant" at conventional alpha levels.

In recognition of this fact, the practice is becoming more and more widespread of computing such a statistic as Hays' (1963) $\hat{\omega}^2$ after finding a significant difference by a t-test or a significant effect in an analysis of variance. This statistic offers an estimate of the proportion of variability of the dependent variable that can be attributed to the relevant independent variable, in the population. As such, it is an extremely important and valuable adjunct to significance tests. This is so especially in educational research, where the practical implementation of a research finding is almost always an expensive affair; one therefore wants to have reasonable assurance that the improvement due to an innovation will be of such a magnitude as to warrant the cost, before embarking on a drastic, large-scale change in educational practices.

Another feature of educational research is that it usually involves (or should involve) a multiplicity of dependent variables. Appropriate significance testing procedures for multivariate experimental designs

have long been known, and are increasingly coming to be used by educational researchers. On the other hand, measures of strength of relationship (as against significance of relationship) in multivariate designs have been virtually unexplored. This is a great drawback for educational research, in view of the crucial importance of having such a measure for reasons mentioned above. As far as could be ascertained, the only such measures advocated to date are those proposed by I. L. Smith (1971) and Tatsuoka (1970). (The fact that both of these were proposed in the context of educational and psychological research attests the importance of a measure of strength of relationship for multivariate analysis in educational research.) The first of these, being based on stepdown procedures, "does not yield [results] that are invariant under alternate orderings" (Smith, 1971) of the dependent variables. This seems to be an undesirable property. The second, $\hat{\omega}^2_{mult}$, is a direct extension of Hays' $\hat{\omega}^2$ to multivariate analysis of variance. It was tentatively proposed on an intuitive basis for one-way MANOVA by Tatsuoka (1970) and, independently, by Sachdeva (1972).

The objective of this study was to examine in detail the statistical properties of $\hat{\omega}^2_{mult}$, with a view to supplying a theoretical justification for it—or, if it turned out to be theoretically unsound, to develop an alternative statistic that is sound. The statistic in question was defined (Tatsuoka, 1970, p. 48) as

$$\hat{\omega}^2_{mult} = \frac{|\underset{\sim}{T}| - |\underset{\sim}{W}| - \frac{K-1}{N-k}|\underset{\sim}{W}|}{|\underset{\sim}{T}| + \frac{1}{N-K}|\underset{\sim}{W}|},$$

where $\underset{\sim}{T}$ and $\underset{\sim}{W}$ are the total and within-group SSCP matrices, N is the total sample size and K is the number of treatment groups.

It is recognized, of course, that for the particular sample at hand, the quantity $1 - \Lambda$, to which $\hat{\omega}^2_{mult}$ converges as N tends to infinty (where Λ is Wilks' likelihood-ratio criterion), is a natural multivariate analogue of the correlation ratio, eta-squared, and hence would be a measure of the proportion of generalized variance of the dependent variable vector that is accounted for by the independent variable(s). But just as Hays found it desirable to define his $\hat{\omega}^2$ as a "nearly unbiased" estimate of what the corresponding proportion might be in the population, it is desirable to have a "nearly unbiased" statistic in the multivariate case. This was the motivation for this study.

2. HEURISTIC DERIVATION OF ESTIMATED ω^2

Given K p-variate normal populations $N(\underset{\sim}{\mu}_k, \underset{\sim}{\Sigma})$ with common covariance matrix $\underset{\sim}{\Sigma}$ and possibly different centroids $\underset{\sim}{\mu}_k$ $(k = 1, 2, \ldots, K)$, the proportion of generalized variance of the p variates attributable to differences among centroids may be defined as

$$\omega^2_{mult} = 1 - \frac{|\underset{\sim}{\Sigma}|}{|\underset{\sim}{\Sigma} + (\underset{\sim}{\alpha}\underset{\sim}{\alpha}')/K|} \tag{2.1}$$

where $\underset{\sim}{\alpha} = (\alpha_{jk})$, $[j = 1, 2, \ldots, p;\ k = 1, 2, \ldots, K]$

and $\alpha_{jk} = \mu_{jk} - \mu_j$

is the deviation of the k-th population mean from the general mean for the j-th variate (or the effect parameter of the k-th group on the j-th variate).

It is desired to get an estimate, as nearly unbiased as possible, of ω^2_{mult} based on independent random samples of n observations each from the K populations. Purely from analogy with Hays' (1963) estimated ω^2 in the univariate case, namely

$$\hat{\omega}^2 = \frac{SS_b - (K-1)MS_w}{SS_t + MS_w},$$

Tatsuoka (1970) proposed the following quantity as a possible such estimate:

$$\hat{\omega}^2_{mult} = \frac{|\underset{\sim}{T}| - |\underset{\sim}{W}| - \frac{K-1}{N-K}|\underset{\sim}{W}|}{|\underset{\sim}{T}| + \frac{1}{N-K}|\underset{\sim}{W}|} \tag{2.2}$$

where $\underset{\sim}{W}$ and $\underset{\sim}{T}$ are within-groups and total sums-of-squares-and-cross-products (SSCP) matrices, and N (=Kn) is the total sample size.

Here we present a more rational justification for this statistic, and consider other possible formulations which (as it turns out) are rejected as being less suitable than expression (2.2). For notational convenience, the subscript "mult" will henceforth be omitted from both "ω^2_{mult}" and "$\hat{\omega}^2_{mult}$."

Let us denote, in addition to the two SSCP matrices cited above, the between-groups SSCP matrix by $\underset{\sim}{B}$ (=$\underset{\sim}{T} - \underset{\sim}{W}$). It can then be shown, in exact parallel to the univariate case discussed by Hays, that

$$E(\underset{\sim}{B}) = (K-1)\underset{\sim}{\Sigma} + n(\underset{\sim}{\alpha}\underset{\sim}{\alpha}')$$

and

$$E(\underset{\sim}{W}) = (N-K)\underset{\sim}{\Sigma}.$$

From these relations it follows that

$$E[\underset{\sim}{W}/(N-K)] = \underset{\sim}{\Sigma} \tag{2.3}$$

and

(2.4) $$E[\underset{\sim}{T}/N + \underset{\sim}{W}/N(N - K)] = \underset{\sim}{\Sigma} + (\underset{\sim}{\alpha}\underset{\sim}{\alpha}')/K .$$

Since the matrices on the right-hand side of equations (2.3) and (2.4) are those whose determinants appear in expression (2.1), it seems reasonable to use some determinantal functions of the matrices under the expected-value operators on the left to replace these determinants in getting an estimate of ω^2. Just what determinantal functions are appropriate, ho. r, is not immediately obvious, because $E(\underset{\sim}{P}) = \underset{\sim}{Q}$ does not imply $E(|\underset{\sim}{P}|) = |\underset{\sim}{Q}|$.

Several alternative ways of "slicing the matrix pie" to construct determinants yield the following candidates as estimators of ω^2:

(2.5) $$1 - \frac{|N\underset{\sim}{W}/(N - K)|}{|\underset{\sim}{T} + \underset{\sim}{W}/(N - K)|}$$

(2.6) $$\frac{|\underset{\sim}{B} - (K - 1)\underset{\sim}{W}/(N - K)|}{|\underset{\sim}{T} + \underset{\sim}{W}/(N - K)|}$$

(2.7) $$\frac{|\underset{\sim}{T}| - |\underset{\sim}{W}| - (K - 1)|\underset{\sim}{W}|/(N - K)^p}{|\underset{\sim}{T}| + |\underset{\sim}{W}|/(N - K)^p}$$

(2.8) $$\frac{|\underset{\sim}{T}| - (N - 1)|\underset{\sim}{W}|/(N - K)}{|\underset{\sim}{T}| + |\underset{\sim}{W}|/(N - K)}$$

These four expressions were examined for their convergence properties as $N \to \infty$. A necessary condition for an admissible estimator is that it converge to $1 - \Lambda$, where

$$\Lambda = |W|/|T|$$

is Wilks' (1932) likelihood-ratio criterion. It can be shown that $1 - \Lambda$ itself converges to ω^2 as $N \to \infty$. Also considered was the behavior of each expression as p increases. The results of these investigations

were as follows:

Expression (2.5) converges to $1 - \Lambda$ as $N \to \infty$. However, for fixed N, it can become negative for large values of p. This expression is therefore disqualified.

Expression (2.6) converges to $|\underset{\sim}{I} - \underset{\sim}{T}^{-1}\underset{\sim}{W}|$ instead of $|\underset{\sim}{I}| - |\underset{\sim}{T}^{-1}\underset{\sim}{W}| = 1 - \Lambda$; therefore it is disqualified.

Expression (2.7) converges to $1 - \Lambda$ as $N \to \infty$, but it also so converges as $p \to \infty$ for fixed N, which is ridiculous.

Expression (2.8) converges to $1 - \Lambda$ as $N \to \infty$, and exhibits no "ridiculous" behavior as p increases.

It therefore appears that expression (2.8) is the most plausible candidate, among those listed above, as an estimator of ω^2. This expression is equivalent to that originally proposed by Tatsuoka (1970) and given in equation (2.2) above. It was also proposed, independently, by Sachdeva (1972) in a slightly different form. This statistic, denoted $\hat{\omega}^2$, is the only one examined in detail hereunder. It should be noted that, besides expressions (2.2) and (2.8), two other equivalent expressions for $\hat{\omega}^2$ are:

$$\hat{\omega}^2 = 1 - \frac{N\Lambda}{(N - K) + \Lambda} \tag{2.9}$$

and

$$\hat{\omega}^2 = 1 - \frac{N}{(N - K)\lambda_1\lambda_2\ldots\lambda_p + 1}, \tag{2.10}$$

where the λ_j are the eigenvalues of $\underset{\sim}{W}^{-1}\underset{\sim}{T}$ ($p - K + 1$ of which will be unity when $p > K - 1$). Expression (2.9) most directly shows that $\hat{\omega}^2 \to 1 - \Lambda$ as $N \to \infty$, while (2.10) is probably the most convenient for computational purposes.

3. GENERATING THE POPULATIONS

The proportion of generalized variance of the dependent variates accounted for by membership in the different populations was defined in the previous section by

$$\omega^2 = 1 - \frac{|\underset{\sim}{\Sigma}|}{|\underset{\sim}{\Sigma} + (\underset{\sim}{\alpha}\underset{\sim}{\alpha}')/K|} \tag{3.1}$$

where $\underset{\sim}{\alpha} = (\alpha_{jk}) \qquad [j = 1, 2, \ldots, p;\ k = 1, 2, \ldots, K]$

is the p x K matrix of effect parameters, and $\underset{\sim}{\Sigma}$ is the common covariance matrix of the K p-variate populations.

In order to make ω^2 take on a preassigned value, it is convenient to diagonalize $\underset{\sim}{\Sigma}$ and $(\underset{\sim}{\alpha}\underset{\sim}{\alpha}')$. Of course, in real life, it is inconceivable that these two matrices will be simultaneously diagonalized by the same transformation. Forcing them to do so, therefore, imposes some peculiar constraint on the configuration of the K population centroids. However, the constraint is not such as to reduce the dimensionality of the manifold in which the K centroids lie (which is p or K - 1, whichever is smaller). Hence, the constraint should not result in an artifactual loss of generality of the sampling distributions of centroids of independent random samples from the K populations.

Let the diagonalized forms of $\underset{\sim}{\Sigma}$ and $(\underset{\sim}{\alpha}\underset{\sim}{\alpha}')/K$ be

$$\underset{\sim}{\Sigma}^* = \underset{\sim}{V}'\underset{\sim}{\Sigma}\underset{\sim}{V}\ , \tag{3.2}$$

where the columns of $\underset{\sim}{V}$ are the orthonormal eigenvectors of $\underset{\sim}{\Sigma}$, and

$$(\underset{\sim}{\alpha}^*\underset{\sim}{\alpha}^{*\prime})/K = (\underset{\sim}{V}'\underset{\sim}{\alpha})(\underset{\sim}{V}'\underset{\sim}{\alpha})'/K, \tag{3.3}$$

respectively. Denote the diagonal elements of $\underset{\sim}{\Sigma}^*$ by $\underset{\sim}{\sigma}^*_{jj}$ (j = 1, 2, ..., p) and the non-zero diagonal elements of $(\underset{\sim}{\alpha}^*\underset{\sim}{\alpha}^{*\prime})/K$ by $\underset{\sim}{a}^*_{kk}$ [k = 1, 2, ..., r = min(p,K-1)]. The determinantal ratio in expression

(3.1) for ω^2 then reduces, successively, as follows:

$$(3.4) \qquad \frac{|\underset{\sim}{\Sigma}|}{|\underset{\sim}{\Sigma} + (\underset{\sim}{\alpha}\underset{\sim}{\alpha}')/K|} = \frac{|\underset{\sim}{V}\underset{\sim}{\Sigma}^*\underset{\sim}{V}'|}{|\underset{\sim}{V}(\underset{\sim}{\Sigma}^* + (\underset{\sim}{\alpha}^*\underset{\sim}{\alpha}^{*\prime})/K)\underset{\sim}{V}'|}$$

$$= \frac{|\underset{\sim}{\Sigma}^*|}{|\underset{\sim}{\Sigma}^* + (\underset{\sim}{\alpha}^*\underset{\sim}{\alpha}^{*\prime})/K|}$$

$$= \prod_{j=1}^{p} \left(\frac{\sigma^*_{jj}}{\sigma^*_{jj} + a^*_{jj}}\right)$$

$$= \prod_{j=1}^{r} (1 + a^*_{jj}/\sigma^*_{jj})^{-1}$$

since $a^*_{jj} = 0$ for $j > r$.

Thus, in order to assign a specific value, say P, to ω^2, it is necessary only to let a^*_{jj} take values that satisfy the condition

$$\sum_{j=1}^{r} \log(1 + a^*_{jj}/\sigma^*_{jj}) = -\log(1 - P) \ .$$

Since it does no violence to let the a^*_{jj} be proportional to σ^*_{jj}, we may for simplicity let

$$\log(1 + a^*_{jj}/\sigma^*_{jj}) = -[\log(1 - P)]/r \ ,$$

or

$$(3.5) \qquad a^*_{jj} = \begin{cases} \sigma^*_{jj}[(1 - P)^{-1/r} - 1] & \text{for } j = 1, 2, \ldots, r \\ 0 & \text{for } j = r+1, \ldots, p \text{ (when } K-1 < p). \end{cases}$$

The elements σ^*_{jj} of Σ^* are, of course, predetermined once we specify $\underset{\sim}{\Sigma}$. These, together with the assigned value P [= 0.1, (0.2), 0.9, say] of ω^2, determine a^*_{jj} in accordance with equation (3.5), and the generation of the K populations is complete. Varying $\underset{\sim}{\Sigma}$ so that the average intercorrelations among the variates will be low or moderate, we have a means for simulating sets of populations of any type encountered in educational research.

4. SAMPLING PROCEDURES

Now that the populations have been generated so as to have various preassigned values for ω^2, the next step is to simulate sampling from these populations. Since sample covariance matrices and sample centroids are independently distributed when samples are drawn from multivariate normal populations, these two aspects of the sampling may be done quite separately from each other.

For generating simulated sample covariance matrices, the Odell-Feiveson (1966) procedure is well established. A computer program written by Montanelli (1971) was utilized for this purpose. The only modification necessary in the present context is that we need to simulate the sampling of a covariance matrix from each of K populations with identical covariance matrix $\underset{\sim}{\Sigma}$. However, since we ultimately need only the pooled within-groups SSCP matrix $\underset{\sim}{W}$, we may, for expedience, simulate the sample covariance matrix $\underset{\sim}{C}$ for a single sample of nominal size $N - K + 1$ (where $N = n_1 + n_2 + \ldots + n_K$), and then multiply this covariance matrix by $N - K$ to produce the desired matrix $\underset{\sim}{W}$:

(4.1) $$\underset{\sim}{W} = (N - K)\underset{\sim}{C} .$$

Simulating the sampling of centroids so as to get the desired between-groups SSCP matrix was, as far as could be ascertained, a new problem encountered in this study that is not discussed in the literature. It was accomplished as follows.

The diagonal matrix $\underset{\sim}{\alpha}^*\underset{\sim}{\alpha}^{*\prime}$ whose diagonal elements were generated in accordance with equation (3.5) is, by definition, the cross product of the effect-parameter matrix when the variates are expressed in canonical form. Any p x K factor matrix $\underset{\sim}{F}$ of $\underset{\sim}{\alpha}^*\underset{\sim}{\alpha}^{*\prime}$ that is centered by

rows (since $\sum_{k=1}^{K} \alpha^*_{jk} = 0$ by definition) should, therefore, qualify as a matrix of effect parameters. Assuming (as will usually be the case) that $p > K - 1$, such a matrix $\underset{\sim}{F}$ can be generated by taking any $p \times (K-1)$ factor matrix $\underset{\sim}{G}$ of $\underset{\sim}{\alpha}^*\underset{\sim}{\alpha}^{*\prime}$ and postmultiplying it by a $(K-1) \times K$ matrix $\underset{\sim}{H}$ that is centered <u>and</u> orthonormal by rows. The simplest matrix to use as $\underset{\sim}{G}$ is

$$\underset{(p,K-1)}{\underset{\sim}{G}} = \left[\begin{array}{cccc} \sqrt{a^*_{11}} & & & \\ & \sqrt{a^*_{22}} & & 0 \\ & & \ddots & \\ 0 & & & \sqrt{a^*_{K-1,K-1}} \\ \hline & & \underset{\sim}{0}_{p-K+1,K-1} & \end{array}\right]$$

that is, the upper-left $(K-1) \times (K-1)$ segment of $(\underset{\sim}{\alpha}^*\underset{\sim}{\alpha}^{*\prime})^{1/2}$ augmented below by a null matrix of order $(p-K+1) \times (K-1)$. One matrix which satisfies the requirements for H is that whose rows consist of the coefficients of the set of Helmert contrasts, with the set of coefficients for each contrast normalized to unity; that is,

$$\underset{(K-1,K)}{H} = \begin{bmatrix} (K-1)c_1 & -c_1 & -c_1 & \cdots & -c_1 \\ & (K-2)c_2 & -c_2 & \cdots & -c_2 \\ & & (K-3)c_3 & & -c_3 \\ & 0 & \ddots & & \vdots \\ & & & c_{K-1} & -c_{K-1} \end{bmatrix}$$

where the c_k $(k = 1, 2, \ldots, K-1)$ are the row-wise normalizing multipliers $[(K-k)^2 + (K-k)]^{-1/2}$.

Thus, taking $\underset{\sim}{F} = \underset{\sim}{G}\underset{\sim}{H}$ as our effect-parameter matrix, we have an eligible $\underset{\sim}{\alpha}^*$ matrix of the form

$$(4.2) \qquad \underset{\sim}{\alpha}^* = \left[\begin{array}{c} (\underset{\sim}{\alpha}^*\underset{\sim}{\alpha}^{*\prime})^{1/2}_{K-1}\underset{\sim}{H} \\ \hline \underset{\sim}{0}_{p-K+1,K} \end{array}\right]$$

where $(\underset{\sim}{\alpha}^*\underset{\sim}{\alpha}^{*\prime})^{1/2}_{K-1}$ stands for the upper-left $(K-1)x(K-1)$ segment of $(\underset{\sim}{\alpha}^*\underset{\sim}{\alpha}^{*\prime})^{1/2}$. [When $p \leq K - 1$, contrary to the assumption above, $\underset{\sim}{G}$ will be $(\underset{\sim}{\alpha}^*\underset{\sim}{\alpha}^{*\prime})^{1/2}$ itself without the augmenting null matrix, and will be of order $p \times p$. We then take as $\underset{\sim}{H}$ the matrix consisting of the first p rows of the $\underset{\sim}{H}$ displayed above, and our $\underset{\sim}{\alpha}$ will be simply $(\underset{\sim}{\alpha}^*\underset{\sim}{\alpha}^{*\prime})^{1/2}\underset{\sim}{H}$.]

By definition, the general element of $\underset{\sim}{\alpha}$ is

$$\alpha^*_{jk} = \mu^*_{jk} - \mu^*_j ,$$

but we may let each $\mu^*_j = 0$ without loss of generality, since this merely involves a translation of the axes so that the origin coincides with the general centroid. Thus, we may take the (j,k)-element of $\underset{\sim}{\alpha}^*$ as defined by equation (4.2) as the mean μ^*_{jk} of the k-th population on the j-th variate in canonical form. Consequently, we may simulate sampling the k-th sample mean on the j-th canonical variate (independently over both k and j) by taking

$$(4.3) \qquad \bar{X}^*_{jk} = \alpha^*_{jk} + z(\sigma^*_{jj}/n_k)^{1/2} ,$$

where σ^*_{jj} is the variance of the j-th canonical variate (i.e., the j-th eigenvalue of the common population covariance matrix $\underset{\sim}{\Sigma}$);

n_k is the k-th sample size (=N/K in the present context);

and $z \sim N(0, 1)$.

Once $\bar{X}^*_{jk}$ has been determined for each sample (k) for a given canonical variate, the grand mean for that variate is obtained as

$$\bar{X}^*_j = \left(\sum_{k=1}^{K} n_k \bar{X}^*_{jk}\right)/N \ , \tag{4.4}$$

and the j-th row elements of the between-groups SSCP matrix are obtained as

$$(\underset{\sim}{B}^*)_{jj'} = \sum_{k=1}^{K} n_k (\bar{X}^*_{jk} - \bar{X}^*_j)(\bar{X}^*_{j'k} - \bar{X}^*_{j'}) \quad [j' = 1, 2, \ldots, p]. \tag{4.5}$$

Repeating this for j = 1, 2, ..., p, we get the between-groups SSCP matrix $\underset{\sim}{B}^*$ in canonical form.

The matrix $\underset{\sim}{B}^*$ is then "uncanonicalized" by the transformation inverse to that displayed in equation (3.2); namely,

$$\underset{\sim}{B} = \underset{\sim}{V}\underset{\sim}{B}^*\underset{\sim}{V}' \ , \tag{4.6}$$

and we get the SSCP matrix $\underset{\sim}{B}$ in the original dependent-variate space.

For each set of K simulated samples under each combination of sampling conditions described in the next section, the matrices $\underset{\sim}{W}$ and $\underset{\sim}{T}$ needed for computing $\hat{\omega}^2$ from equation (2.10) were determined by getting $\underset{\sim}{W}$ in accordance with (4.1), $\underset{\sim}{B}$ in accordance with (4.6), and finding $\underset{\sim}{T} = \underset{\sim}{W} + \underset{\sim}{B}$. The complete computer program is shown in Appendix A.

5. SAMPLING CONDITIONS

The types of populations from which samples were to be drawn have already been partially indicated in the preceding section. More specifically, it was planned to have two levels of average intercorrelation among the variates in the common covariance matrix: low

(0.10 - 0.30) and moderate (0.40 - 0.60). The sets of populations were also designed to have five levels of ω^2 values: 0.1, 0.3, 0.5, 0.7, and 0.9. Additionally, three levels were used for the number of variates: $p = 3$, 5, and 10. The number of populations in each set was fixed at $K = 5$.

From each of the $2 \times 5 \times 3 = 30$ sets of five populations each, samples of three sizes were to be drawn; namely, $n = 15$, 30, and 60 from each of the five populations in each set, yielding total sample sizes of $N = 75$, 150, and 300 for each set. (Since this is not a study of robustness of a statistical test under violations of assumptions, it was not deemed necessary to vary the sample sizes across populations in each set.)

Thus, there were a total of 90 sampling conditions. Under each of these conditions, 1,000 independent, random samples were to be drawn. It was anticipated that this number of samples would be more than sufficient to achieve adequate approximation to the true sampling distributions of $\hat{\omega}^2$.

6. RESULTS

Although it was originally planned to use sets of populations having two levels of average intercorrelation among the variates, preliminary investigations quickly showed that the magnitude of average correlations had virtually no effect on the sampling distribution of $\hat{\omega}^2$. The means for three pairs of sets of sampling distributions under sampling conditions differing only in the average correlations in the populations were as shown in Table 1. (In retrospect, this lack of dependence of the sampling distributions on the average correlation could have been anticipated: the sampling distribution of the squared

multiple correlation coefficient—of which $1 - \Lambda$ is a generalization for $K > 2$—does not depend on the particular covariance structure which produces a given population ρ^2 value.) It was therefore decided to confine further work to population sets with average correlation in the medium-low range of 0.20 - 0.30. The (common) correlation matrix for each of the three sets of populations used are shown in Appendix B.

Table 1. Means of sampling distributions of $\hat{\omega}^2$ for three pairs of sets of populations, the sets in each pair differing only in the average correlation ($\bar{\rho}$) in the populations

		ω^2				
		0.9	0.7	0.5	0.3	0.1
p=3,N=75	$\bar{\rho}$ = .16	.9113	.7296	.5482	.3696	.1970
	$\bar{\rho}$ = .28	.9112	.7297	.5479	.3705	.1970
p=5,N=75	$\bar{\rho}$ = .24	.9192	.7551	.5956	.4382	.2777
	$\bar{\rho}$ = .56	.9187	.7596	.6040	.4482	.2857
p=5,N=1000	$\bar{\rho}$ = .24	.9017	.7047	.5071	.3109	.1143
	$\bar{\rho}$ = .56	.9000	.7041	.5067	.3140	.1141

It also became apparent as computations proceeded that $\hat{\omega}^2$ was extremely positively biased, especially for population sets with low ω^2 values, when the ratio N/p (of total sample size to number of variates) was any lower than 40 or so. The original plan of using samples of sizes 15, 30, and 60 from each population (i.e., total sample size N = 75, 150, 300) was therefore abandoned in favor of one in which the N/p ratio had values 50, 100, and 200. This change was deemed

appropriate since it appeared that, in order to get anything resembling a realistic estimate of ω^2, the N/p ratio had to be of these orders of magnitude. In retrospect, however, the decision may not have been a wise one, for reasons described below.

Due to the exorbitant cost of computations for the 10-variate cases, the number of samples drawn was reduced from 1,000 to either 200, 300, or 500.

With the foregoing modifications of the original plans for sampling conditions, the means of the generated empirical sampling distributions of $\hat{\omega}^2$ under the various conditions were as shown in Table 2.

Table 2. Means of sampling distributions of $\hat{\omega}^2$ for various numbers of variates (p) and sample sizes (N), for five sets of populations with ω^2 values as indicated. (Each set consists of K=5 populations.)

		ω^2				
p	N	0.9	0.7	0.5	0.3	0.1
3	75	.9112	.7297	.5479	.3705	.1970
3	150	.9059	.7129	.5262	.3398	.1461
3	300	.9030	.7073	.5115	.3181	.1249
3	600	.9016	.7038	.5058	.3101	.1127
5	75	.9192	.7551	.5956	.4382	.2777
5	250	.9056	.7159	.5287	.3425	.1539
5	500	.9033	.7077	.5156	.3214	.1283
5	1000	.9017	.7047	.5071	.3109	.1143
10	75	.9437	.8205	.7104	.6087	.4716
10	500	.9071	.7222	.5331	.3452	.1650
10	1000	.9038	.7116	.5167	.3233	.1313
10	2000	.9016	.7052	.5088	.3124	.1164

Inspection of Table 2 shows that, even for an N/p ratio as large as 200, the positive bias of $\hat{\omega}^2$ is considerable. At this point it was suspected that, despite the reasoning in Section 3, the peculiar method of constructing the populations with preassigned ω^2 values might have led to anomalous sampling distributions of $\hat{\omega}^2$. Therefore, an alternative (and more expensive) method, described in Appendix C, that did not depend on the simultaneous diagonalization of $\underset{\sim}{\Sigma}$ and $\underset{\sim}{\alpha}\underset{\sim}{\alpha}'$ was employed to generate a set of five populations with $p = 5$ and $\omega^2 = 0.1$. One thousand samples of size 15 were drawn from each of the five populations thus generated, and the distribution of $\hat{\omega}^2$ was constructed and compared with that under the comparable sampling condition based on the simultaneous-diagonalization procedure. (See Appendix C.) The mean of this sampling distribution was 0.2859, or 0.0082 larger than the mean (0.2777) in the corresponding cell of Table 2. It was therefore concluded that the positive bias of $\hat{\omega}^2$ was not due to the nature of the populations generated by the simultaneous diagonalization method. Evidently, our attempt to develop a "quasi-unbiased" estimator of ω^2 was unsuccessful.

Correcting the Bias in $\hat{\omega}^2$

Various attempts to develop an alternative statistic that would estimate ω^2 unbiasedly (or nearly so) were made but proved to be of no avail. It was therefore decided to try to develop a formula for correcting the bias in $\hat{\omega}^2$.

Close scrutiny of Table 2 revealed that, within each row (i.e., for fixed p and N), the amount of bias, $\hat{\omega}^2 - \omega^2$, seemed very nearly

a linear function of $1 - \hat{\omega}^2$. This impression was put to a test as follows. For each row in Table 2, a straight line

$$(\hat{\omega}^2 - \omega^2)' = m(1 - \hat{\omega}^2) \tag{6.1}$$

was fitted by the least-squares method. (That the intercept in equation (6.1) must be 0 follows from the fact that $\hat{\omega}^2$ can theoretically show no bias when $\omega^2 = 1$.) From this equation, "corrected" values of $\hat{\omega}^2$ were computed as

$$\hat{\omega}^2_{corr} = \hat{\omega}^2 - m(1 - \hat{\omega}^2), \tag{6.2}$$

and these were correlated with the true ω^2 values within each row. The results were as shown in Table 3.

Table 3. Means of sampling distributions of $\hat{\omega}^2_{corr}$ [corrected by equation (6.2)], the proportionality constant m, and the correlation between ω^2 and $\hat{\omega}^2_{corr}$, for various p and N

		ω^2						
p	N	0.9	0.7	0.5	0.3	0.1	m	r
3	75	.9010	.6985	.4958	.2979	.1044	.1153	1.0000
3	150	.9007	.6969	.4999	.3031	.0986	.0556	1.0000
3	300	.9004	.6994	.4983	.2997	.1013	.0270	1.0000
3	600	.9002	.6997	.4989	.3005	.1003	.0139	1.0000
5	75	.8995	.6955	.4972	.3015	.1020	.2433	.9909
5	250	.8996	.6980	.4990	.3010	.1005	.0631	1.0000
5	500	.9002	.6984	.5002	.2998	.1006	.0318	1.0000
5	1000	.9001	.7000	.4993	.3000	.1003	.0158	1.0000
10	75	.9027	.6896	.4992	.3234	.0863	.7291	.9989
10	500	.9002	.7016	.4984	.2966	.1030	.0742	1.0000
10	1000	.9004	.7013	.4995	.2992	.1004	.0356	1.0000
10	2000	.8998	.6998	.4998	.2999	.1003	.0182	1.0000

It seemed indisputable, from the results exhibited in Table 3, that a correction term linear in $1 - \hat{\omega}^2$ would suffice for each p and N (for fixed K) in order to get a nearly unbiased estimator of ω^2. The problem was to determine the functional dependence of the proportionality coefficient m on p and N, <u>and</u> on the hitherto unvaried K (the number of populations).

At this point, the earlier decision to use different sample sizes for different numbers of variates (in order to control the N/p ratio) proved detrimental. We were left with insufficient data points adequately to conjecture the relation of m with p for fixed N, and that of m with N for fixed p. Nevertheless, there were enough grounds for surmising that m was approximately inversely proportional to N and roughly directly proportional to p, as a scanning of Table 3 shows.

The foregoing observations, coupled with knowledge that the quantities

$$N - 1 - (p + K)/2 \quad \text{and} \quad p(K - 1)$$

often appear in p-variate, K-sample problems, led to the tentative conjecture that the m in equations (6.1) and (6.2) might be expressible, approximately, as

$$m(K, p, N) = c \frac{p(K - 1)}{N - 1 - (p + K)/2}$$

where c is a proportionality constant to be determined.

In order to permit greater flexibility in the curve-fitting venture, however, it was decided to use a more general form as the conjectured relation between m and its three arguments:

$$m(K, p, N) = cM^a Q^b \ , \tag{6.3}$$

with

$$M = N - 1 - (p + K)/2 \tag{6.4}$$

or some similar quantity dependent largely on N, and

$$Q = p(K - 1) \tag{6.5}$$

or some other quantity dependent symmetrically on p and K - 1. The symmetry with respect to p and K - 1 was conjectured from the fact that a p-variate, K-sample problem can be recast as a canonical-correlation problem with p variates in one set and K-1 in the other.

Combining equations (6.1) and (6.3), we arrive at the least-squares problem

$$(\hat{\omega}^2 - \omega^2)' = cM^aQ^b(1 - \hat{\omega}^2) \ , \tag{6.6}$$

where the constants c, a, and b are to be determined on a least-squares basis, and M and Q are as conjectured in equations (6.4) and (6.5) or similar expressions in N, p, and K. Taking logarithms of both sides of equation (6.6), the quantity to be minimized is

$$E = \Sigma[\log\underline{c} + \underline{a}\log M + \underline{b}\log Q + \log(1 - \hat{\omega}^2)/(\hat{\omega}^2 - \omega^2)]^2 \ , \tag{6.7}$$

where the summation is taken over all data points.

The normal equation for determining the optimal values of c, a, and b is

$$\begin{bmatrix} N_D & \Sigma\log M & \Sigma\log Q \\ \Sigma\log M & \Sigma(\log M)^2 & \Sigma(\log M)(\log Q) \\ \Sigma\log Q & \Sigma(\log Q)(\log M) & \Sigma(\log Q)^2 \end{bmatrix} \begin{bmatrix} \log\underline{c} \\ a \\ b \end{bmatrix} = \begin{bmatrix} -\Sigma\log W \\ -\Sigma(\log M)(\log W) \\ -\Sigma(\log Q)(\log W) \end{bmatrix} \tag{6.8}$$

where N_D is the number of data points, and $W = (1 - \hat{\omega}^2)/(\hat{\omega}^2 - \omega^2)$.

Equation (6.8) enables us to solve for least-squares optimal values of a, b, and c under various plausible conjectures for M and Q

besides those stated in equations (6.4) and (6.5). We would then favor choices of M and Q that lead to the smallest value of E as defined in equation (6.7)--subject, of course, to review upon cross-validating on data points generated by sampling conditions not included in the optimizing process.

To carry out the calculations, it was now necessary to vary K instead of fixing it at 5 as was done up to this point. It was decided this time to fix p at 4 and use K = 3, 7, and 10 in combination with N = 75 and 500. To keep down computer costs, the number of samples drawn was reduced to 200 for these runs. The means of the resulting sampling distributions were as shown in Table 4.

Table 4. Means of sampling distributions of $\hat{\omega}^2$ for two sample sizes (N) for sets of populations with ω^2 values as indicated. The sets consist of varying numbers (K) of populations. The number of variates is p = 4 throughout.

		ω^2				
N	K	0.9	0.7	0.5	0.3	0.1
75	3	.9071	.7189	.5304	.3513	.1704
75	7	.9234	.7720	.6058	.4666	.3054
75	10	.9344	.7976	.6543	.5293	.3922
500	3	.9008	.7004	.5046	.3062	.1087
500	7	.9034	.7086	.5135	.3274	.1315
500	10	.9051	.7152	.5235	.3343	.1499

The data in the 12 rows of Table 2 and those in the six rows of Table 4 (90 data points in all) were fed into equation (6.8) for

determining the least-square estimates of the parameters in equation (6.6), with M and Q as initially conjectured in equations (6.4) and (6.5). Solution of equation (6.8) with these inputs yielded

$$\log \underline{c} = -0.4341 \text{ (or } c = .3680), \quad a = -1.0677, \quad b = 1.3631.$$

Thus, the correction formula for $\hat{\omega}^2$ based on the data of Tables 2 and 4 and the conjectures of equations (6.4) and (6.5) is as follows:

$$(6.9) \quad \hat{\omega}^2_{corr} = \hat{\omega}^2 - .368[N-1-(p+K)/2]^{-1.0677}[p(K-1)]^{1.3631}(1-\hat{\omega}^2).$$

The 90 corrected $\hat{\omega}^2$ values resulting from use of this formula were as shown in Table 5. The root mean squared error for these data points was 0.0108, and the correlation between $\hat{\omega}^2$ and ω^2 was 0.9993.

Table 5. Means of sampling distributions of $\hat{\omega}^2$ for 22 combinations of p (number of variables), K (number of populations) and N (total sample size), corrected in accordance with equation (6.9).

			ω^2				
P	K	N	0.9	0.7	0.5	0.3	0.1
3	5	75	.9008	.6982	.4952	.2971	.1033
3	5	150	.9009	.6975	.5008	.3044	.1003
3	5	300	.9006	.7000	.4992	.3010	.1029
3	5	600	.9004	.7003	.4999	.3019	.1022
5	5	75	.9000	.6969	.4995	.3047	.1060
5	5	250	.8998	.6984	.4996	.3019	.1017
5	5	500	.9005	.6992	.5015	.3017	.1030
5	5	1000	.9003	.7006	.5003	.3014	.1021
10	5	75	.9079	.7064	.5262	.3599	.1356
10	5	500	.9001	.7013	.4980	.2960	.1023
10	5	1000	.9004	.7014	.4995	.2993	.1044
10	5	2000	.8999	.7002	.5005	.3008	.1015
4	3	75	.9009	.7002	.4991	.3081	.1151
4	3	500	.9000	.6979	.5005	.3004	.1013
4	7	75	.8999	.7020	.4847	.3028	.0921
4	7	500	.8998	.6977	.4954	.3023	.0991
4	10	75	.8986	.6870	.4654	.2721	.0601
4	10	500	.8989	.6967	.4925	.2910	.0946

Various other conjectures for M and Q, such as $M = N-(p+K)$, $M = N$, $Q = p+K-1$, $Q = p^2+(K-1)^2$ were tried out, each combination giving rise to a table essentially similar to Table 5. These are not included here because there is little point in presenting a lengthy series of similar tables. The three most promising combinations besides that consisting of the expressions given in equations (6.4) and (6.5) led to the following estimated values for the parameters c, a, and b:

$M = N-1-(p+K)/2$, $Q = p^2+(K-1)^2$:	$c = .2801$, $a = -1.0692$, $b = 1.1343$
$M = N$, $Q = p(K-1)$:	$c = .4358$, $a = -1.1048$, $b = 1.3899$
$M = N$, $Q = p^2+(K-1)^2$:	$c = .3041$, $a = -1.1066$, $b = 1.1579$

In order to test the effectiveness of the corrections using these combinations of parameter values on a new set of data, the means of sampling distributions of $\hat{\omega}^2$ for four new combinations of p, K, and N were computed. The resulting 20 data points were as shown in block (i) of Table 6.

Blocks (ii) - (v) of Table 6 show the corrected $\hat{\omega}^2$ values using the four combinations of M and Q with the values for the parameters c, a, and b as cited above. Also shown, at the bottom of each of these blocks, is the root-mean-square error (RMSE) for that set of corrected $\hat{\omega}^2$ values. Examination of these RMSE values and a scanning of the table indicate that the corrections seem quite adequate. All the $\hat{\omega}^2_{corr}$ values are correct at least to one digit, and about two-thirds are correct to two or more digits.

In terms of the RMSE values, the two combinations using $Q = p(K-1)$ yield somewhat better corrections than do those using $Q = p^2+(K-1)^2$.

Table 6. Means of sampling distributions of $\hat{\omega}^2$ for four combinations of p (number of variables), K (number of populations), and N (total sample size): (i) as observed; and (ii) - (v) corrected for bias by subtracting out the following correction terms:

(ii) $.3680[N-1-(p+K)/2]^{-1.0677}[p(K-1)]^{1.3631}(1 - \hat{\omega}^2)$

(iii) $.2801[N-1-(p+K)/2]^{-1.0692}[p^2+(K-1)^2]^{1.1343}(1 - \hat{\omega}^2)$

(iv) $.4358\ N^{-1.1048}[p(K-1)]^{1.3899}(1 - \hat{\omega}^2)$

(v) $.3041\ N^{-1.1066}[p^2+(K-1)^2]^{1.1579}(1 - \hat{\omega}^2)$

Also shown are the root mean squared error, MSE, for each set of corrected $\hat{\omega}^2$ values.

				ω^2				
	p	K	N	0.9	0.7	0.5	0.3	0.1
(i)	4	6	150	.9118	.7288	.5489	.3628	.1909
	4	5	300	.9048	.7128	.5164	.3245	.1355
	7	4	300	.9050	.7194	.5291	.3403	.1522
	7	8	400	.9092	.7309	.5502	.3673	.1900
(ii)	4	6	150	.9022	.6994	.5003	.2938	.1032
	4	5	300	.9013	.7021	.4984	.2993	.1033
	7	4	300	.8999	.7042	.5036	.3046	.1063
	7	8	400	.8977	.6969	.4934	.2874	.0877
					(RMSE = .00518)			
(iii)	4	6	150	.9036	.7035	.5069	.3035	.1156
	4	5	300	.9017	.7034	.5006	.3024	.1072
	7	4	300	.8989	.7013	.4987	.2978	.0976
	7	8	400	.9014	.7077	.5116	.3130	.1204
					(RMSE = .00774)			
(iv)	4	6	150	.9021	.6988	.4990	.2924	.1015
	4	5	300	.9012	.7020	.4982	.2990	.1029
	7	4	300	.8998	.7040	.5032	.3040	.1056
	7	8	400	.8974	.6959	.4917	.2851	.0847
					(RMSE = .00594)			
(v)	4	6	150	.9034	.7031	.5062	.3024	.1143
	4	5	300	.9017	.7033	.5004	.3022	.1070
	7	4	300	.8988	.7010	.4982	.2969	.0965
	7	8	400	.9012	.7073	.5107	.3118	.1189
					(RMSE = .00718			

In particular, the M and Q originally conjectured and stated in equations (6.4) and (6.5) give the smallest RMSE. Pitted against this advantage, however, is the fact that the values of b for the two combinations involving $Q = p^2+(K-1)^2$ are closer to unity than they are for the other two. (Note that all four values of a are quite close to -1.) Observing further that the value of c for the last combination is close to 1/3, we are led to an alternative, simpler formula that should be sufficient for providing a rough, "rule-of-thumb" correction; namely

$$\hat{\omega}^2_{corr} = \hat{\omega}^2 - \frac{p^2 + (K-1)^2}{3N}(1 - \hat{\omega}^2). \tag{6.10}$$

Using this rule-of-thumb correction for the 22 (p, K, N)-combinations (110 data points in all) that were considered in the foregoing, the corrected $\hat{\omega}^2$ values showed the following distribution of number of significant digits in agreement with the true ω^2 values:

20 agreed to 3 digits,
55 agreed to 2 digits,
33 agreed to 1 digit,
and 2 agreed to 0 digit.

(The two showing 0-digit accuracy were larger by 0.1 than the true ω^2 values, when rounded to one decimal place.)

It therefore seems safe to conclude that, at least within the limits of the p (number of variates), K (number of populations) and N (total sample size) values that were examined, the simple correction formula presented in equation (6.10) will suffice to reduce the bias in $\hat{\omega}^2$ to less than 0.05. The constraints are that $p(K-1) \leq 49$ and $75 \leq N \leq 2000$. It remains to be seen how the correction formula will work outside these limits.

Confidence Intervals for ω^2

Another approach to estimating the population ω^2 from even a badly biased statistic is to construct charts for getting confidence intervals. For this purpose it is unnecessary first to eliminate (or even reduce) the bias in $\hat{\omega}^2$--provided we are willing to adopt confidence intervals that will often not even include the observed statistic value, especially when p and/or K is large.

Arbitrarily deciding to construct symmetric 90% confidence intervals (symmetric in the sense that 5% is excluded at each tail end, but not symmetric about the observed $\hat{\omega}^2$), the 5th and 95th centile points of the sampling distributions for selected combinations of p and N, with K fixed at 5, were computed. The selection was based essentially on those combinations represented in Table 2, excluding the largest N for each p. (Inspection of the sampling distributions for the largest-N cases showed that the confidence intervals would reduce practically to point estimates within the accuracy of graphing for these cases.) Sampling distributions represented in Tables 4 and 6 were not considered because they were based on only 200 samples, which would make the C_5 and C_{95} values unreliable. Additionally, the sampling distributions for N = 150 with p = 5 were computed anew, since the jump from N = 75 to N = 250 seemed large in this case.

The 5th and 95th centiles, C_5 and C_{95}, of the selected sampling distributions were as shown in Table 7. The values for the null distributions (with $\omega^2 = 0$) were computed backwards from Rao's (1952) approximate equation relating Λ to an F distribution under the null hypothesis. This relation states that

Table 7. 5th and 95th centiles of sampling distributions of $\hat{\omega}^2$ for 3, 5, and 10 variables, each with three selected sample sizes and six true ω^2 values.

	p = 3					
	N = 75		N = 150		N = 300	
ω^2	C_5	C_{95}	C_5	C_{95}	C_5	C_{95}
0	.0198	.2082	.0092	.1068	.0043	.0550
.1	.0725	.3333	.0675	.2425	.0713	.1828
.3	.2275	.5256	.2380	.4505	.2418	.3945
.5	.4083	.6796	.4288	.6225	.4410	.5828
.7	.6208	.8268	.6411	.7854	.6560	.7568
.9	.8728	.9466	.8767	.9348	.8804	.9225

	p = 5					
	N = 75		N = 150		N = 250	
ω^2	C_5	C_{95}	C_5	C_{95}	C_5	C_{95}
0	.0981	.3189	.0478	.1681	.0282	.1042
.1	.1513	.4125	.1131	.2925	.0934	.2225
.3	.2915	.5808	.2645	.4850	.2621	.4280
.5	.4650	.7164	.4575	.6425	.4550	.6041
.7	.6569	.8411	.6615	.8025	.6586	.7708
.9	.8798	.9547	.8808	.9406	.8838	.9299

	p = 10					
	N = 75		N = 500		N = 1000	
ω^2	C_5	C_{95}	C_5	C_{95}	C_5	C_{95}
0	.2963	.5305	.0451	.0989	.0225	.0506
.1	.3475	.6075	.1245	.2163	.1011	.1669
.3	.4550	.7009	.2925	.3975	.2875	.3650
.5	.5958	.8083	.4850	.5900	.4818	.5583
.7	.7425	.8963	.6775	.7669	.6852	.7420
.9	.9100	.9694	.8939	.9260	.8940	.9188

$$(6.11) \qquad \frac{1 - \Lambda^{1/s}}{\Lambda^{1/s}} \; \frac{ms - p(K-1)/2 + 1}{p(K-1)} \sim F_{\nu_1, \nu_2} ,$$

where

$$\nu_1 = p(K-1), \; \nu_2 = ms - p(K-1)/2 + 1 ,$$

and

$$m = N - 1 - (p+K)/2$$

$$s = \sqrt{\frac{p^2(K-1)^2 - 4}{p^2 + (K-1)^2 - 5}} .$$

From relation (6.11), it follows that

$$\Lambda^{-1/s} - 1 \sim hF ,$$

where $\quad h = \dfrac{p(K-1)}{ms - p(K-1)/2 + 1}$.

Consequently, the $100\alpha\%$ point of the distribution of Λ when $\omega^2 = 0$ may be calculated as

$$\Lambda_\alpha = \text{antilog}[-s\log(1+h\, F_{\nu 1,\, \nu 2;\, 1-\alpha})].$$

Upon substituting this value in equation (2.9), namely

$$\hat{\omega}^2 = 1 - \frac{N\Lambda}{(N-K) + \Lambda} ,$$

we get the $100(1-\alpha)\%$ point of the null distribution of $\hat{\omega}^2$. Using $\alpha = .95$ and $.05$, therefore, gives us the C_5 and C_{95} values, respectively, of the sampling distribution of $\hat{\omega}^2$ when $\omega^2 = 0$.

Figure 1 shows the graphs of the upper and lower limits of the symmetric 90% confidence interval for $N = 75$, 150, and 300 in the three-variate case. Similar graphs may be constructed for $p = 5$ and $p = 10$ from the data given in Table 7.

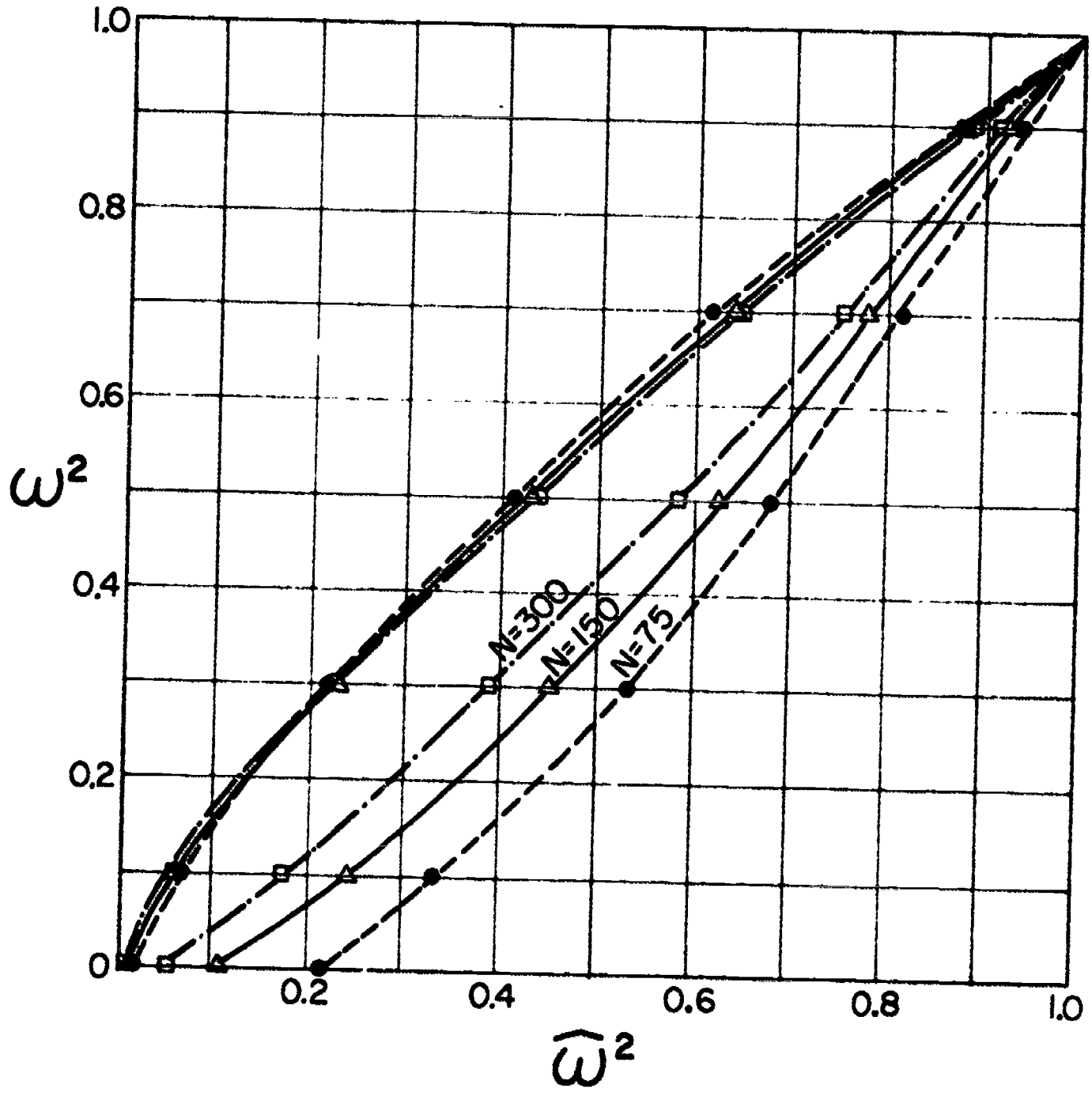

Figure 1. 90% confidence limits for ω^2 for three sample sizes, with three variates and five populations.

7. SUMMARY AND CONCLUDING REMARKS

The sampling distribution of a measure of strength of relationship in multivariate analysis of variance (MANOVA), proposed by Tatsuoka (1970), was examined by a computer simulation technique developed especially for this purpose. The measure, denoted $\hat{\omega}^2$ and defined in equation (2.10) as

$$\hat{\omega}^2 = 1 - \frac{N}{(N-K)\lambda_1\lambda_2 \ldots \lambda_p + 1},$$

where the λ_i are the eigenvalues of $\underset{\sim}{W}^{-1}\underset{\sim}{T}$ ($\underset{\sim}{W}$ being the within-groups, and $\underset{\sim}{T}$ the total SSCP matrix) was found to be highly positively biased--especially when the population value ω^2 is small.

Although the amount of bias steadily decreases with increasing N (indicating that $\hat{\omega}^2$ is a consistent estimator of ω^2), it does not become negligible until the N/p ratio exceeds 50 for $\omega^2 \geq .50$, and exceeds 100 for $\omega^2 \leq .30$. This means that a study involving p = 10 variables must use at least 1,000 subjects before any realistic estimate of the population ω^2 can be obtained.

Since such large samples are not ordinarily used in typical studies (with the exception of large-scale statewide or nationwide studies), it becomes important to have a means for eliminating, or at least reducing the bias in $\hat{\omega}^2$.

Various attempts to develop, theoretically, an alternative statistic with little or no bias proved futile. An empirical approach was therefore adopted. Careful study of graphical plots of the amount of bias revealed that a linear correction of the form

$$\hat{\omega}^2_{corr} = \hat{\omega}^2 - m(1 - \hat{\omega}^2)$$

would suffice for any fixed number of variables (p), number of populations or levels in the MANOVA factor (K), and total sample size (N). Furthermore, the proportionality constant m appeared to be approximately inversely proportional to N and (very roughly) directly proportional to p for fixed K.

The above observations, combined with knowledge that p and K-1 should affect the amount of bias in the same way, led to the conjecture that a correction of the form

$$\hat{\omega}^2_{corr} = \hat{\omega}^2 - cM^aQ^b(1 - \hat{\omega}^2)$$

should be adequate. Here M is an expression largely dependent on N, Q an expression symmetrical in p and K-1, and c (> 0), a (< 0), and b (> 0) are parameters to be estimated.

Several different expressions for M and Q were tried out, and the parameters c, a, and b were estimated by the least-squares method, based on 18 different combinations of p, K, and N, with five ω^2 values (0.1, 0.3, 0.5, 0.7, and 0.9), for a total of 90 data points.

Cross-validation on 20 data points not used in the process of selecting optimal expressions for M and Q and estimating c, a, and b led to the choice of the following correction formula:

$$\hat{\omega}^2_{corr} = \hat{\omega}^2 - .3041\ N^{-1.1066}[p^2 + (K-1)^2]^{1.1579}(1 - \hat{\omega}^2).$$

This was simplified to a rule-of-thumb correction formula

$$\hat{\omega}^2_{corr} = \hat{\omega}^2 - \frac{p^2 + (K-1)^2}{3N}(1 - \hat{\omega}^2).$$

The simplified formula, when tested against all 110 data points (90 derivation points and 20 cross-validation points), yielded very

satisfactory results: agreement with ω^2 to at least one significant digit in all but two cases, and agreement to two or three significant digits in slightly less than two-thirds of the cases. This formula was therefore deemed to be adequate at least when $p(K-1) \leq 49$ and $75 \leq N \leq 2000$.

At this point, the question naturally arises, if a further correction is needed on $\hat{\omega}^2$ which was designed to be a correction of sorts for $1 - \Lambda$ (where Λ is Wilks' likelihood-ratio criterion), why should one bother with a statistic more complicated than $1 - \Lambda$? The answer is, obviously, that one need not bother. This conclusion was implicit in Huberty's (1972, p. 525) statement that "It is clear that numerically the difference between η^2_W, η^2_H, and $\hat{\omega}^2_{mult}$ is practically nil," made in a study comparing four multivariate indices of strength of association, including $1 - \Lambda$ (η^2_W in his notation) and $\hat{\omega}^2$. However, the further conclusion that anyone of these "may be employed as indices of discriminatory power of a set of variables" has now been shown to miss the mark. To put it bluntly the present study, in combination with Huberty's findings that they all yield nearly equal numerical values, shows that they are all equally poor instead of equally good.

The natural thing to do would seem to be to develop a correction formula to be applied to either of the purely sample-descriptive indices η^2_W (i.e., $1 - \Lambda$) or η^2_H (which is based on the trace of $\underset{\sim}{W}^{-1}\underset{\sim}{B}$). In view, again, of Huberty's findings, it is likely that the rule-of-thumb correction formula given in equation (6.10) and cited above will suffice for all practical purposes. In particular, the correction for $1 - \Lambda$ would be

$$(1 - \Lambda)_{corr} = (1 - \Lambda) - \frac{p^2 + (K-1)^2}{3N} \Lambda .$$

It may seem odd that virtually nothing is known about the noncentral distribution of so widely used and long established a statistic as Wilks' Λ, but this appears to be the case. Gupta (1971) derived the distribution in the special case when $\underset{\sim}{\alpha}\underset{\sim}{\alpha}'$ is of unit rank, and asserted that, for the general case, the distribution "has not been expressed in a numerically feasible form" (p. 1259). As far as could be ascertained, nothing has appeared in the literature since 1971 to negate this assertion.

Besides the correction-formula method, another means was presented for estimating ω^2 from $\hat{\omega}^2$; namely the use of confidence intervals. Charts were given in Figure 1 for $p = 3$ with $N = 75$, 105, and 300, and data for constructing similar charts for $p = 5$ and 10 were presented in Table 7.

REFERENCES

Gupta, A. K. Noncentral distribution of Wilks' statistic in MANOVA. Annals of Mathematical Statistics, 1971, 42, 1254-1261.

Hays, W. L. Statistics for Psychologists. New York: Holt, Rinehart & Winston, 1963.

Huberty, C. J. Multivariate indices of strength of association. Multivariate Behavioral Research, 1972, 7 523-526.

Montanelli, R. G. An investigation of the goodness of fit of the maximum likelihood estimation procedure in factor analysis. Report No. 451, University of Illinois at Urbana-Champaign, 1971.

Odell, P. L., & Feiveson, A. H. A numerical procedure to generate a sample covariance matrix. Journal of the American Statistical Association, 1966, 61, 199-203.

Sachdeva, D. Multivariate analog of Hays' ω^2. (Paper presented at the Annual Meeting of the American Educational Research Association, April, 1972.)

Smith, I. L. Strength of relationship in multivariate analysis of variance. (Paper presented at the Annual Meeting of the American Educational Research Association, February, 1971.)

Tatsuoka, M. M. Discriminant Analysis: The Study of Group Differences. Champaign, Ill.: Institute for Personality and Ability Testing, 1970.

Wilks, S. S. Certain generalizations in the analysis of variance. Biometrika, 1932, 24, 471-494.

APPENDIX A: Computer Program and Sample Output

```
      IMPLICIT REAL*8(A-H,O-Z)
      COMMON KFRQ(200,5),PMEAN(5),OMEAN(5),EMEAN(5),NMTXS
     1,ICOUNT,OMETR(30,30),AMDA(30),TEMPE(900),EMC(5,10,10)
     1,INDX,KIM,SDP(30),ISKIP
      LOGICAL   IPAR(8)
      INTEGER*4 IDUM(30,2),IDET
      INTEGER FMT(20),BLANK(20)/20*'    '/,ID(20)
      DIMENSION SIGMA(30,30),T(30,30),SDS(30),IQ(5)
      EQUIVALENCE(NROW,IP)
C***** INPUT PARAMETERS
C ID 1 CARD OF IDENTIFICATION.FIRST 4 COLUMNS APPEAR IN
C PUNCHED OUTPUT
C FMT INPUT FORMAT FOR POPULATION CORRELATIONS.
C IPAR LOGICAL FLAGS FOR CONTROLLING PRINTED OUTPUT
C ISKIP FLAG FOR CONTROLLING PRINTED OUTPUT IN SUBROUTINE
C KGR NUMBER OF GROUPS
C IX INTEGER STARTING POINT FOR RANDOM NUMBER GENERATOR
C NMTXS NUMBER OF SAMPLE CORRELATION MATRICES TO BE GENERATED
C N SAMPLE SIZE
C NROW NUMBER OF ROWS IN INPUT MATRIX
      READ(5,1001) ID,FMT,IPAR,IX,NMTXS,N,NROW,KGR,IQ,ISKIP
      READ(5,FMT) (SDP(I),I=1,NROW)
      ISKIP=ISKIP+1
      NINT=200
      WRITE(6,1004) ID,IX
      DO 777 I=1,5
      PMEAN(I)=0.
      OMEAN(I)=0.
      EMEAN(I)=0.
      DO 778 J=1,NINT
        KFRQ(J,I)=0
778     CONTINUE
777     CONTINUE
      WRITE(6,1032) NROW,KGR,N,NMTXS
      CALL RN3INZ(IX)
C READ IN SIGMA,CALCULATE OMEGA TRANSPOSE,UPPER TRIANGULAR
      N=N-KGR+1
        DO 4739 I=1,NROW
4739      READ(5,FMT)(SIGMA(I,J), J=1,NROW)
      S=0.
      DO 250 I=1,NROW
      DO 250 J=1,I
250     S=S+SIGMA(I,J)
      X=(NROW-1)*NROW/2.
      S=S-NROW
      S=S/X
      WRITE(6,2001) S
      I=11-2*J
      WRITE(6,1033) (J,J=1,5),(J,J=1,5),(J,J=1,5)
      DO 2000 I=1,NROW
      DO 1500 J=I,NROW
 1500 SIGMA(I,J)=SIGMA(J,I)
 2000 SIGMA(I,I)=1.
      IF(IPAR(1)) GO TO 911
```

```
      WRITE(6,1005)
      DO 901 L=1,NROW
  901 WRITE(6,1011) L,(SIGMA(L,M),M=1,NROW)
  911 CONTINUE
      DO 10 I=1,NROW
   10 OMETR(1,I)=SIGMA(1,I)
      DO 200 I=2,NROW
      ILESS1=I-1
      IPLUS1=I+1
      SUM=0.0
      DO 110 J=1,ILESS1
  110 SUM=SUM+OMETR(J,I)*OMETR(J,I)
       OMETR(I,I)=DSQRT(SIGMA(I,I)-SUM)
      IF(I.EQ.NROW) GO TO 200
      TEMP=1.0/OMETR(I,I)
      DO 130 J=IPLUS1,NROW
      SUM=0.0
      DO 120 K=1,ILESS1
  120 SUM=SUM+OMETR(K,I)*OMETR(K,J)
  130 OMETR(I,J)=TEMP*(SIGMA(I,J)-SUM)
  200 CONTINUE
      IF(IPAR(2)) GO TO 912
      WRITE(6,1006)
      DO 902 L=1,NROW
  902 WRITE(6,1011) L,(OMETR(M,L),M=1,L)
  912 CONTINUE
        DO 2002 I=1,NROW
          DO 2002 J=1,NROW
 2002      T(I,J)=SIGMA(I,J)*SDP(I)*SDP(J)
        KIM=1
         CALL SBETW(T,N,NROW,IX,KGR,IQ,IP)
        KIM=2
C
C LOOP FOR ALL SAMPLE CORRELATION MATRICES
C
  205 NCODE=N/100
      DO 900 ICOUNT=1,NMTXS
C CALCULATE T,LOWER TRIANGULAR
      DO 300 I=1,NROW
      NDF=N-I
      DF=NDF
      CALL NORMAL(X)
      X2=X*X
      X3=X*X2
      H60=.3080441E-03*X3-.1589653E-03*X2-.9243112E-03*X
     1+.1885979E-03
      H =(60.0/DF)*H60
      TEMP=(2.0/(9.0*DF))
       T(I,I)=DF*(1.0-TEMP+(X-H)*DSQRT(TEMP))**3
       T(I,I)=DSQRT(T(I,I))
      IF(I.EQ.1) GO TO 300
      ILESS1=I-1
      DO 210 J=1,ILESS1
      CALL NORMAL(X)
```

```
  210 T(I,J)=X
  300 CONTINUE
      IF(IPAR(3)) GO TO 913
      WRITE(6,1007)
      DO 903 L=1,NROW
  903 WRITE(6,1011) L,(T(L,M),M=1,L)
  913 CONTINUE
C CALCULATE OMEGA*T,AND STORE IN SIGMA,AS A LOWER TRIANGULAR
      DO 350 I=1,NROW
      DO 350 J=1,I
      SIGMA(I,J)=0.0
      DO 350 K=J,I
  350 SIGMA(I,J)=SIGMA(I,J)+OMETR(K,I)*T(K,J)
      IF(IPAR(4)) GO TO 914
      WRITE(6,1008)
      DO 904 L=1,NROW
  904 WRITE(6,1011) L,(SIGMA(L,M),M=1,L)
  914 CONTINUE
C FORM A MATRIX,IN T(LOWER TRIANGLE,ONLY)
      DO 370 I=1,NROW
      DO 370 J=1,I
      T(I,J)=0.0
      DO 370 K=1,J
  370 T(I,J)=T(I,J)+SIGMA(I,K)*SIGMA(J,K)
      IF(IPAR(5)) GO TO 915
      WRITE(6,1009)
      DO 905 L=1,NROW
  905 WRITE(6,1011) L,(T(L,M),M=1,L)
  915 CONTINUE
      DO 600 I=1,NROW
      DO 600    J=1,NROW
600    T(I,J)=T(I,J)*SDP(I)*SDP(J)
      CALL SBETW(T,N,NROW,IX,KGR,IQ,IP)
      IF(IPAR(8)) GO TO 900
C GET STANDARD DEVIATIONS, STORE IN FMT,AND CALCULATE R,IN T
      DO 380 I=1,NROW
       SDS(I)=DSQRT(T(I,I))
      DO 380 J=1,I
      T(I,J)=T(I,J)/(SDS(I)*SDS(J))
  380 T(J,I)=T(I,J)
      IF(IPAR(6)) GO TO 916
      WRITE(6,1010) N
C     DO 906 L=1,NROW
C PRINT FIRST ROW OF EACH CORRELATION MATRIX
      L=1
  906 WRITE(6,1011) L,(    T(L,M),M=1,NROW)
  916 CONTINUE
      IF(IPAR(7)) GO TO 917
      DO 907 L=1,NROW
      M1=1
      DO 907 K=1,4
      M2=M1+9
      WRITE(7,1020)ID(1),L,ICOUNT,NCODE,K,(T(L,M),M=M1,M2)
  907 M1=M2+1
```

```
      WRITE(7,1021) BLANK
  917 CONTINUE
  900 CONTINUE
  950 CALL RN3NDZ(IX)
      WRITE(6,1999) IX
      CALL EXIT
 1001 FORMAT(20A4/20A4/8L1,2X,I10,13I5)
 1004 FORMAT('1',20A4,',INTEGER STARTING VALUE=',I10)
 1005 FORMAT('0INPUT CORRELATION MATRIX'/)
 1006 FORMAT('1SQUARE ROOT FACTORS OF CORRELATIONK'/)
 1007 FORMAT('1T MATRIX'/)
 1008 FORMAT('1OMEGA*T '/)
 1009 FORMAT(' A MATRIX'/)
C1010 FORMAT('1SAMPLE CORRELATION MATRIX',I6/)
 1010 FORMAT('-SAMPLE CORRELATION MATRIX',I6/)
 1011 FORMAT(I3,10F13.6/(3X,10F13.6))
 1020 FORMAT(A4,2I2,2I1,10F7.5)
 1021 FORMAT(20A4)
1032  FORMAT('  NO. OF DEPENDENT VARIABLES = ',I4/
     1'  NO. OF GROUPS                    = ',I4/
     2'  SAMPLE SIZE                      = ',I4/
     3'  NO. OF SAMPLES                   = ',I4)
1033  FORMAT(/19X,' 1-LAMBDA VALUES',21X
     1, ' OMEGASQUARE VALUES',23X,'      ERRORS'//6X,I6
     2,14I8)
2001  FORMAT('  AVERAGE CORRELATION          = ',F5.4)
 1999 FORMAT('0INTEGER STOPPING POINT=',I10)
      END
      SUBROUTINE SBETW(T,N,NROW,IX,K,IG,IP)
      IMPLICIT REAL*8(A-H,O-Z)
      COMMON KFRQ(200,5),PMEAN(5),OMEAN(5),EMEAN(5),NMTXS
     1,ICOUNT,OMETR(30,30),AMDA(30),TEMPE(900),EMC(5,10,10)
     1,INDX,KIM,SDP(30),ISKIP
       INTEGER*4 IDUM(30,2),INDIC(10),IDET
      DIMENSION SIGMA(30,30),T(30,30),OMEGA(30,30),DUM1(465)
     1,ID(20),FMT(20),BET(30,30),IQ(5),PRT(5),EROR(5),SRT(5)
     2,VEC(900),WVAL(30),TVAL(30)
      NINT=200
C CONVERT CORRELATION MATRIX TO COVARIANCE MATRIX,STORE
C IN SIGMA
      DO 10 L=1,NROW
      DO 10 J=1,NROW
10      T(L,J)=T(J,L)
      DO 11 L=1,NROW
      DO 11 J=1,NROW
11      SIGMA(L,J)=T(L,J)
      N=N+K-1
        SMESQ=0.
C
C CALCULATE EIGENVALUES AND VECTORS OF POPULATION
C COVARIANCE MATRIX, ONLY ON FIRST PASS. OTHERWISE SKIP
C TO 771
C
      GO TO (770,771),KIM
```

```
C
C**** INPUT PARAMETERS
C K  NUMBER OF GROUPS
C IP NUMBER OG VARIABLES
C NROW ORDER OF COVARIANCE MATRIX
C N SAMPLE SIZE
C IQ=10*THE GIVEN VALUE OF OMEGA SQ.
C SIGMA CONTAINS VARIANCE-COVARIANCE MATRIX
C
770     CALL EIGENZ(T,TEMPE,AMDA,DUM1,NROW,30,0)
        IPAT=30-NROW
         DO 1785  I=1,IPAT
        INPT=NROW+I
        AMDA(INPT)=0.
1785        CONTINUE
C
C TEMPE CONTAINS EIGEN-VECTORS IN DESCENDING ORDER OF ROOTS
C AMDA(30) HAVE EIGENVALUES IN DESCENDING ORDER
C
C T HAS EIGENVALUES IN THE DIAGONAL POSITION
C
        IF(ISKIP-2) 751,750,751
750     WRITE(6,1022)
        DO 601        L=1,NROW
        IPAT=1+ NROW*(L-1)
        IR= NROW*L
601        WRITE(6,1011) L,AMDA(L),(TEMPE(M),M=IPAT,IR)
         L=0
          WRITE(6,1011) L,(SDP(I),I=1,NROW)
C
C GENERATE MATRIX ALPHA-STAR,STORE IN T
751        DO1614   INPT=1,5
        DO 914  IPAT=1,30
        DO 914  IST=1,30
        T(IPAT,IST)=0.
914        CONTINUE
        Q=IQ(INPT)
        Q=Q/10.
        IF(IP-K+1) 1892,1891, 1891
1891        S=K-1
        GO TO 1914
1892        S=IP
1914     Q=-DLOG(Q)/S
        Q=DEXP(Q)-1.
        Q=DSQRT(Q)
         I=K-1
        DO 604 L=1,I
        AT=(K-L+1)*(K-L)
       T(L,L)=DSQRT( AMDA(L)/AT) *(K-L)*Q
        M=K-1
        DO 604 J=L,M
        LL=J+1
       T(L,LL)=-1*DSQRT( AMDA(L)/AT)
         T(L,LL)=T(L,LL)*Q
```

```
604         T(LL,L)=0.
        S=K
        DO 516 L=1,IP
        DO 516      J=1,K
516        EMC(INPT,L,J)=T(L,J)*DSQRT(S)
        GO TO (1601,1600), ISKIP
1600    WRITE(6,1615) INPT
        DO 1616    I=1,NROW
1616        WRITE(6,1011)I, (EMC(INPT,I,J),J=1,K)
1601        CONTINUE
C
C RESTORE ALPHA-STAR TO ORIGINAL VARIATE-SPACE ALPHA
C
        DO 630 L=1,NROW
        DO 631   J=1,K
        X=0.
        DO 632   M=1,NROW
        IPAT=L+NROW*(M-1)
632      X=TEMPE(IPAT)*T(M,J)+X
        BET(J,L)=X
631        OMEGA(L,J)=X
630        CONTINUE
C
C GET ALPHA*ALPHA-PRIME
C
        DO 633 L=1,NROW
        DO 634 J=1,NROW
        X=0.
        DO 635 M=1,K
635        X=X+OMEGA(L,M)*BET(M,J)
634        DUM1(J)=X
        DO 636   J=1,NROW
636        T(L,J)=DUM1(J)
633        CONTINUE
        DO 609 L=1,NROW
        DO 609    M=1,NROW
C
C BET HAS ALPHA*ALPHA-PRIME,ORDER IS NROW BY NROW
C  T HAS ALPHA, ORDER IS NROW BY K
C
        BET(L,M)=T(L,M)
        T(L,M)=OMEGA(L,M)
609        CONTINUE
C
C COMPUTE LAMBDA
C
        DO 612 L=1,NROW
        DO 612 J=1,NROW
           OMEGA(L,J)=BET(L,J) +SIGMA(L,J)
612         OMEGA(L,J)=OMEGA(L,J)
        CALL EIGENZ(OMEGA,VEC,TVAL,DUM1,NROW,30,0)
        S=1.
        DO 230   L=1,NROW
230        S=S*AMDA(L)/TVAL(L)
```

```
      WILKS=S
      WRITE(6,1031) INPT,WILKS
1614      CONTINUE
      N=N-K+1
C
C RETURN TO MAIN TO GET SAMPLE CORRELATION MATRIX
C
         RETURN
C
C LOOP FOR A SAMPLE BETWEEN GROUP MATRIX
C
771   SAM=N
      SAM=SAM/K
C
C  GET THE DETERMINANT OF W AND STORE IN SIDG
C
      IGO=1
       Y=DABS(SIGMA(1,1))
730      INDX=2
723      IF(Y-10**INDX) 724,724,725
725      INDX=INDX+1
      GO TO 723
724      INDX=INDX-1
731      DO 727 L=1,NROW
      DO 727  J=1,NROW
727      SIGMA(L,J)=SIGMA(L,J)/10**INDX
C
C COMPUTE EIGENVALUES OF WITHIN GROUPS SSCP MATRIX W
C
      CALL EIGENZ(SIGMA,VEC,WVAL,DUM1,NROW,30,0)
      GO TO (200,201),ISKIP
201      DO 202 L=1,NROW
      IPAT=1+NROW*(L-1)
      IR=NROW*L
202      WRITE(6,1011) L,WVAL(L),(VEC(M),M=IPAT,IR)
200      CONTINUE
C
C****GENERATE BETWEEN GROUPS SSCP MATRIX B
C
      DO 614   INPT=1,5
      DO 701  I=1,IP
      DO 700 J=1,K
      CALL NORMAL(X)
        S=DSQRT(AMDA(I)/SAM)
      STAX=X*S+EMC(INPT,I,J)
      T(I,J)=STAX
700      CONTINUE
701      CONTINUE
      GO TO (203,204),ISKIP
204      DO 205  L=1,NROW
205      WRITE(6,1011) L,(T(L,J),J=1,K)
203      CONTINUE
C
C GET VECTOR OF GRAND MEANS, STORE IN OMEGA
```

```
C
      DO 703 I=1,IP
      S=0.
      DO 704 J=1,K
704      S=S+T(I,J)
      DO 705 J=1,K
705      OMEGA(I,J)=S/K
703      CONTINUE
C
C  MAKE THE BETWEEN GROUP MATRIX
C
      DO 706 I=1,IP
      GO TO (207,206),ISKIP
206      WRITE(6,1011) I,(OMEGA(I,M),M=1,K)
207      CONTINUE
C
C GET DEVIATIONS OF GROUP MEANS FROM GRAND MEAN
C
      DO 707 J=1,K
707      T(I,J)=T(I,J)-OMEGA(I,J)
706      CONTINUE
      DO 709 I=1,IP
      GO TO (208,209),ISKIP
209      WRITE(6,1011) I,(T(I,M),M=1,K)
208      CONTINUE
      DO 709 J=1,K
709      OMEGA(J,I)=T(I,J)
      DO 711 I=1,IP
      DO 713 M=1,IP
      S=0.
      DO 712 J=1,K
712      S=S+T(I,J)*OMEGA(J,M)
      DUM1(M)=S*SAM
713      CONTINUE
      DO 515  M=1,NROW
515      BET(I,M)=DUM1(M)
      GO TO (711,211),ISKIP
211     WRITE(6,1011) I,(BET(I,M),M=1,NROW)
711      CONTINUE
C
C RESTORE B-STAR TO ORIGINAL VARIATE-SPACE B,STORE IN BET
C
      IGO=1
740      DO 736  L=1,NROW
      DO 737     J=1,NROW
      X=0.
      DO 738   M=1,NROW
      IPAT=L+NROW*(M-1)
738      X=TEMPE(IPAT)*BET(M,J)+X
737      OMEGA(L,J)=X
      GO TO (736,213),ISKIP
213      WRITE(6,1011) L,(OMEGA(L,M),M=1,NROW)
736      CONTINUE
      IGO=IGO+1
```

```
      IF(IGO-2) 742,742,741
742   DO 739 L=1,NROW
      DO 739  J=1,NROW
739      BET(J,L)=OMEGA(L,J)
      GO TO 740
741   DO 720  L=1,NROW
      DO 720 J=1,NROW
      BET(L,J)=OMEGA(L,J)/10**INDX
         OMEGA(L,J)=BET(L,J)+SIGMA(L,J)
720      CONTINUE
C
C  CALCULATE WILKS LAMBDA
C BET = BETWEEN GROUP MATRIX
C  OMEGA HAS TOTAL
C  GET THE DETERMINANT OF W+B AND STORE IT IN X
C
      CALL EIGENZ(OMEGA,VEC,TVAL,DUM1,NROW,30,0)
      S=1.
      GO TO (214,215),ISKIP
215      L=0
      WRITE(6,1011) L,(TVAL(L),L=1,IP)
214      DO 216  L=1,IP
216      S=S*TVAL(L)/NVAL(L)
        EWILK=1./S
       S=S*(N-K)+1.
      Y=N
      S=1.-Y/S
      OMESQ=S
      IF(ISKIP-2) 754,755,754
755      DO 756 L=1,NROW
756      WRITE(6,1011) L,(SIGMA(L,J),J=1,NROW)
      DO 757    L=1,NROW
757      WRITE(6,1011) L,(OMEGA(L,J),J=1,NROW)
      DO 758  L=1,NROW
758      WRITE(6,1011) L,(BET(L,J),J=1,NROW)
      WRITE(6,1032) INDX,      NROW,K,N,INPT
      WRITE(6,99) OMESQ
      WRITE(6,99) EWILK
754      CONTINUE
      PMEAN(INPT)=PMEAN(INPT)+1.-EWILK
      PRT(INPT)=1.-EWILK
C
C  OMEGA SQUARE
C
       Y=K-1
       YV=N-K
       SRT(INPT)=OMESQ
      OMEAN(INPT)=OMEAN(INPT)+OMESQ
       EROR(INPT)=OMESQ+IQ(INPT)/10. -1.
      EMEAN(INPT)=EROR(INPT)+EMEAN(INPT)
      S=OMESQ+0.0005
      IS=10**3*S
     LIM=NINT*5
     IF(IS-LIM) 748,749,749
```

```
749       IS=NINT
        GOTO 3750
748     IS=IS/5+1
3750     KFRQ(IS,INPT)=KFRQ(IS,INPT)+1
614         CONTINUE
            WRITE(6,1031) ICOUNT,PRT,SRT,EROR
         IF(ICOUNT-NMTXS) 744,745,745
745         DO 746 J=1,5
         PMEAN(J)=PMEAN(J)/NMTXS
         OMEAN(J)=OMEAN(J)/NMTXS
746         EMEAN(J)=EMEAN(J)/NMTXS
         I=11-2*J
        WRITE(6,1034) (J,J=1,5),(J,J=1,5),(J,J=1,5)
         WRITE(6,1031) NMTXS,PMEAN,OMEAN,EMEAN
         DO 250  I=1,5
250             IQ(I)=10-IQ(I)
          WRITE(6,1033) IQ
         DO 747  J=1,NINT
747         WRITE(6,1032) J,(KFRQ(J,M),M=1,5)
744         N=N-K+1
         RETURN
99         FORMAT(2X,F15.4)
1011         FORMAT(I3,10F13.6/(3X,10F13.6) )
1022         FORMAT(' EIGEN VALUES AND EIGEN VECTORS'/)
1031         FORMAT(2X,I4,15F8.4)
1032         FORMAT( 6I10)
1033  FORMAT(//5X,'   DISTRIBUTION OF OMEGA SQUARE'/10X,
     1 5I10/)
1034  FORMAT(/   19X,'MEAN OF 1-LAMBDA',21X,
     1 'MEAN OF OMEGA SQUARE',23X, 'MEAN OF ERRORS'//6X,I6
     2 ,14I8)
1615        FORMAT(5X,'1 ALPHA-STAR MATRIX FOR POP.WILKS=',I5)
          END
         SUBROUTINE NORMAL(/Y/)
         INTEGER K/1/
         REAL*8 WA,WB
         IF(K.EQ.2) GO TO 3
         WA=RAN3Z(0)
         WB=RAN3Z(0)
         WA=DSQRT(-2.*DLOG(WA))
         WB=WB*6.28318531
         Y=WA*DCOS(WB)
         K=2
         RETURN
       3 Y=WA*DSIN(WB)
         K=1
         RETURN
         END
```

NO. OF DEPENDENT VARIABLES = 5
NO. OF GROUPS = 5
SAMPLE SIZE = 75
NO. OF SAMPLES = 40
AVERAGE CORRELATION = .1569

.INTEGER STARTING VALUE=1111204578

1-LAMBDA VALUES

	1	2	3	4	5
1	0.9462	0.7438	0.6830	0.5354	0.2261
2	0.8871	0.7431	0.6637	0.4090	0.1796
3	0.9538	0.7177	0.6114	0.4102	0.1391
4	0.9314	0.7385	0.3911	0.4327	0.3015
5	0.9032	0.7466	0.6882	0.3298	0.1935
6	0.9029	0.6459	0.4322	0.5622	0.2428
7	0.9160	0.7164	0.6778	0.4798	0.2251
8	0.9267	0.7462	0.6066	0.4658	0.1479
9	0.9287	0.8319	0.5723	0.5571	0.2444
10	0.9025	0.6833	0.4921	0.2871	0.1502
11	0.9046	0.7936	0.6608	0.5901	0.1588
12	0.9713	0.7148	0.4773	0.4121	0.2076
13	0.9092	0.6392	0.5591	0.3332	0.1885
14	0.9224	0.7216	0.5922	0.3679	0.3273
15	0.8818	0.7464	0.5629	0.2552	0.2093
16	0.8769	0.7404	0.5690	0.3390	0.2193
17	0.9067	0.6830	0.5638	0.4650	0.3061
18	0.9386	0.6425	0.3673	0.2913	0.1464
19	0.9161	0.7430	0.6060	0.5242	0.2595
20	0.9475	0.6923	0.5136	0.4269	0.3011
21	0.9279	0.7616	0.5800	0.4088	0.1457
22	0.9362	0.7586	0.7521	0.4321	0.2160
23	0.9205	0.8028	0.6872	0.3436	0.1830
24	0.9096	0.7475	0.6225	0.3762	0.2282
25	0.9334	0.7741	0.5629	0.3566	0.2315
26	0.9202	0.6884	0.4887	0.5066	0.1439
27	0.9445	0.7247	0.6443	0.4558	0.1974
28	0.9187	0.7359	0.6135	0.3108	0.3429
29	0.8860	0.6833	0.6710	0.5377	0.2484
30	0.9235	0.7887	0.6398	0.4292	0.3539
31	0.9170	0.6703	0.6305	0.4566	0.2593
32	0.9099	0.7776	0.5955	0.3002	0.1815
33	0.9304	0.7454	0.6157	0.5220	0.2613
34	0.9109	0.7173	0.5805	0.5137	0.2674
35	0.9375	0.7490	0.4179	0.3746	0.2095
36	0.9319	0.8039	0.6511	0.3784	0.3730
37	0.9412	0.7454	0.5607	0.4467	0.3110
38	0.9112	0.7434	0.5584	0.4052	0.2801
39	0.8547	0.6467	0.5269	0.4368	0.2424
40	0.9145	0.7588	0.6094	0.4234	0.2305

OMEGASQUARE VALUES

	1	2	3	4	5
1	0.9424	0.7265	0.6619	0.5055	0.1799
2	0.8792	0.7258	0.6414	0.3721	0.1311
3	0.9506	0.6988	0.5859	0.3733	0.0889
4	0.9265	0.7209	0.3533	0.3971	0.2591
5	0.8964	0.7295	0.6674	0.2887	0.1457
6	0.8961	0.6225	0.3965	0.5338	0.1974
7	0.9101	0.6974	0.6564	0.4468	0.1788
8	0.9216	0.7291	0.5808	0.4319	0.0981
9	0.9237	0.8203	0.5445	0.5284	0.1991
10	0.8957	0.6622	0.4598	0.2438	0.1004
11	0.8980	0.7795	0.6383	0.5634	0.1095
12	0.9693	0.6957	0.4441	0.3753	0.1605
13	0.9029	0.6154	0.5306	0.2923	0.1405
14	0.9170	0.7029	0.5656	0.3288	0.2861
15	0.8735	0.7293	0.5346	0.2104	0.1623
16	0.8683	0.7229	0.5410	0.2984	0.1728
17	0.9002	0.6619	0.5355	0.4312	0.2638
18	0.9343	0.6189	0.3282	0.2482	0.0964
19	0.9102	0.7257	0.5802	0.4937	0.2149
20	0.9438	0.6718	0.4824	0.3910	0.2586
21	0.9228	0.7454	0.5527	0.3719	0.0958
22	0.9317	0.7422	0.7353	0.3964	0.1693
23	0.9150	0.7894	0.6664	0.3033	0.1348
24	0.9033	0.7305	0.5977	0.3375	0.1821
25	0.9287	0.7588	0.5346	0.3169	0.1855
26	0.9146	0.6676	0.4562	0.4790	0.0939
27	0.9406	0.7062	0.6208	0.4215	0.1498
28	0.9130	0.7181	0.5882	0.2688	0.3025
29	0.8802	0.6622	0.6492	0.5060	0.2032
30	0.9181	0.7743	0.6160	0.3933	0.3140
31	0.9112	0.6484	0.6062	0.4223	0.2147
32	0.9036	0.7625	0.5691	0.2577	0.1332
33	0.9255	0.7282	0.5905	0.4913	0.2168
34	0.9047	0.6983	0.5532	0.4826	0.2232
35	0.9331	0.7320	0.3814	0.3359	0.1625
36	0.9271	0.7905	0.6280	0.3398	0.3341
37	0.937	0.7282	0.5323	0.4139	0.2690
38	0.9050	0.7260	0.5298	0.3681	0.2365
39	0.8446	0.6254	0.4965	0.4014	0.1970
40	0.9085	0.7424	0.5838	0.3872	0.1845

ERRORS

	1	2	3	4	5
1	0.0424	0.0265	0.1619	0.2055	0.0799
2	-0.0208	0.0258	0.1414	0.0721	0.0311
3	0.0506	-0.0012	0.0859	0.0733	-0.0111
4	0.0265	0.0209	-0.1467	0.0971	0.1591
5	-0.0036	0.0295	0.1674	-0.0113	0.0457
6	-0.0039	-0.0775	-0.1035	0.2338	0.0974
7	0.0101	-0.0026	0.1564	0.1468	0.0788
8	0.0216	0.0291	0.0808	0.1319	-0.0019
9	0.0237	0.1203	0.0445	0.2284	0.0991
10	-0.0043	-0.0378	-0.0402	-0.0562	0.0004
11	-0.0020	0.0795	0.1383	0.2634	0.0095
12	0.0693	-0.0043	-0.0559	0.0753	0.0605
13	0.0029	-0.0846	0.0306	-0.0077	0.0405
14	0.0170	0.0029	0.0656	0.0288	0.1861
15	-0.0265	0.0293	0.0346	-0.0896	0.0623
16	-0.0317	0.0229	0.0410	-0.0016	0.0728
17	0.0002	-0.0381	0.0355	0.1312	0.1638
18	0.0343	-0.0811	-0.1718	-0.0518	-0.0036
19	0.0102	0.0257	0.0802	0.1937	0.1149
20	0.0438	-0.0282	-0.0176	0.0910	0.1586
21	0.0228	0.0454	0.0527	0.0719	-0.0042
22	0.0317	0.0422	0.2353	0.0964	0.0693
23	0.0150	0.0694	0.1664	0.0033	0.0348
24	0.0033	0.0305	0.0977	0.0375	0.0821
25	0.0287	0.0588	0.0346	0.0169	0.0859
26	0.0146	-0.0324	-0.0438	0.1750	-0.0061
27	0.0406	0.0062	0.1208	0.1215	0.0498
28	0.0130	0.0161	0.0882	-0.0312	0.2025
29	-0.0198	-0.0378	0.1492	0.2080	0.1032
30	0.0181	0.0743	0.1160	0.0933	0.2140
31	0.0112	-0.0516	0.1062	0.1223	0.1147
32	0.0036	0.0625	0.0691	-0.0423	0.0332
33	0.0255	0.0282	0.0905	0.1913	0.1168
34	0.0047	-0.0017	0.0532	0.1826	0.1232
35	0.0331	0.0320	-0.1186	0.0359	0.0625
36	0.0271	0.0905	0.1280	0.0398	0.2341
37	0.0371	0.0282	0.0323	0.1139	0.1690
38	0.0050	0.0260	0.0298	0.0681	0.1365
39	-0.0554	-0.0766	-0.0035	0.1014	0.0970
40	0.0065	0.0424	0.0838	0.0872	0.0845

MEAN OF 1-LAMBDA

	1	2	3	4	5
40	0.9189	0.7313	0.5825	0.4223	0.2320

MEAN OF OMEGA SQUARE

	1	2	3	4	5
40	0.9152	0.7133	0.5554	0.3862	0.1862

MEAN OF ERRORS

	1	2	3	4	5
40	0.0152	0.0133	0.0554	0.0862	0.0862

DISTRIBUTION OF OMEGA SQUARE

	9	7	5	3	1
1	0	0	0	0	0
2	0	0	0	0	0
3	0	0	0	0	0
4	0	0	0	0	0
5	0	0	0	0	0
6	0	0	0	0	0
7	0	0	0	0	0
8	0	0	0	0	0
9	0	0	0	0	0
10	0	0	0	0	0
11	0	0	0	0	0
12	0	0	0	0	0
13	0	0	0	0	0
14	0	0	0	0	0
15	0	0	0	0	0
16	0	0	0	0	0
17	0	0	0	0	0
18	0	0	0	0	1
19	0	0	0	0	1
20	0	0	0	0	3
21	0	0	0	0	1
22	0	0	0	0	1
23	0	0	0	0	0
24	0	0	0	0	0
25	0	0	0	0	0
26	0	0	0	0	2
27	0	0	0	0	1
28	0	0	0	0	1
29	0	0	0	0	1
30	0	0	0	0	0
31	0	0	0	0	3
32	0	0	0	0	1
33	0	0	0	0	1
34	0	0	0	0	1
35	0	0	0	0	2
36	0	0	0	0	2
37	0	0	0	0	0
38	0	0	0	0	3
39	0	0	0	0	1
40	0	0	0	0	0
41	0	0	0	1	0
42	0	0	0	0	3
43	0	0	0	0	1
44	0	0	0	0	0
45	0	0	0	0	0
46	0	0	0	1	0
47	0	0	0	1	0
48	0	0	0	0	2
49	0	0	0	1	1
50	0	0	0	1	0
51	0	0	0	0	0
52	0	0	0	0	0
53	0	0	0	0	0
54	0	0	0	1	1
55	0	0	0	0	1
56	0	0	0	0	1
57	0	0	0	0	0
58	0	0	0	0	1

59	0	0	0	1	0
60	0	0	0	1	0
61	0	0	0	1	1
62	0	0	0	0	0
63	0	0	0	0	1
64	0	0	0	1	0
65	0	0	0	0	0
66	0	0	1	1	0
67	0	0	0	0	1
68	0	0	0	2	0
69	0	0	0	1	0
70	0	0	0	0	0
71	0	0	1	0	0
72	0	0	0	0	0
73	0	0	0	0	0
74	0	0	0	1	0
75	0	0	0	3	0
76	0	0	0	1	0
77	0	0	1	0	0
78	0	0	0	1	0
79	0	0	0	2	0
80	0	0	1	2	0
81	0	0	0	1	0
82	0	0	0	0	0
83	0	0	0	1	0
84	0	0	0	0	0
85	0	0	0	2	0
86	0	0	0	0	0
87	0	0	0	2	0
88	0	0	0	0	0
89	0	0	1	0	0
90	0	0	0	1	0
91	0	0	0	0	0
92	0	0	1	0	0
93	0	0	1	0	0
94	0	0	0	0	0
95	0	0	0	0	0
96	0	0	0	1	0
97	0	0	1	1	0
98	0	0	0	0	0
99	0	0	0	2	0
100	0	0	1	0	0
101	0	0	0	0	0
102	0	0	0	2	0
103	0	0	0	0	0
104	0	0	0	0	0
105	0	0	0	0	0
106	0	0	0	1	0
107	0	0	3	1	0
108	0	0	3	0	0
109	0	0	1	0	0
110	0	0	1	0	0
111	0	0	2	0	0
112	0	0	0	1	0
113	0	0	0	0	0
114	0	0	2	0	0
115	0	0	0	0	0
116	0	0	0	0	0
117	0	0	3	0	0
118	0	0	2	0	0
119	0	0	1	0	0

120	0	0	1	0	0
121	0	0	0	0	0
122	0	0	1	0	0
123	0	0	0	0	0
124	0	2	1	0	0
125	0	2	1	0	0
126	0	0	1	0	0
127	0	0	0	0	0
128	0	0	1	0	0
129	0	0	1	0	0
130	0	1	1	0	0
131	0	0	0	0	0
132	0	0	1	0	0
133	0	3	1	0	0
134	0	1	2	0	0
135	0	1	0	0	0
136	0	0	0	0	0
137	0	0	0	0	0
138	0	0	0	0	0
139	0	4	0	0	0
140	0	1	0	0	0
141	0	1	0	0	0
142	0	0	0	0	0
143	0	1	0	0	0
144	0	2	0	0	0
145	0	9	0	0	0
146	0	2	0	0	0
147	0	0	1	0	0
148	0	2	0	0	0
149	0	1	0	0	0
150	0	0	0	0	0
151	0	1	0	0	0
152	0	1	0	0	0
153	0	0	0	0	0
154	0	1	0	0	0
155	0	0	0	0	0
156	0	1	0	0	0
157	0	1	0	0	0
158	0	1	0	0	0
159	0	0	0	0	0
160	0	0	0	0	0
161	0	0	0	0	0
162	0	0	0	0	0
163	0	1	0	0	0
164	0	0	0	0	0
165	0	0	0	0	0
166	0	0	0	0	0
167	0	0	0	0	0
168	0	0	0	0	0
169	1	0	0	0	0
170	0	0	0	0	0
171	0	0	0	0	0
172	1	0	0	0	0
173	1	0	0	0	0
174	1	0	0	0	0
175	1	0	0	0	0
176	0	0	0	0	0
177	0	0	0	0	0
178	4	0	0	0	0

181	4	0	0	0	0
182	3	0	0	0	0
183	4	0	0	0	0
184	4	0	0	0	0
185	3	0	0	0	0
186	4	0	0	0	0
187	3	0	0	0	0
188	1	0	0	0	0
189	3	0	0	0	0
190	0	0	0	0	0
191	1	0	0	0	0
192	0	0	0	0	0
193	0	0	0	0	0
194	1	0	0	0	0
195	0	0	0	0	0
196	0	0	0	0	0
197	0	0	0	0	0
198	0	0	0	0	0
199	0	0	0	0	0
200	0	0	0	0	0

INTEGER STOPPING POINT=1359084993

<u>Notes</u>: (a) The heading in each column is $10\omega^2$.

(b) Each class interval extends from $5(m-1) \times 10^{-3}$ to $5m \times 10^{-3} - 10^{-8}$ where m is the class-interval number shown in the first column. Thus, for example, the last non-empty interval for $\omega^2 = 0.9$, with $m = 194$, is [.9650, .9699+].

APPENDIX B

The Population Correlation Matrices

For the 10-variate cases, the correlation matrix common to the sets of five populations was as shown below, rounded to four decimal places.

For the five- and three-variate cases, the upper left-hand 5 x 5 and 3 x 3 segments, respectively, of this matrix was used.

In each case, the correlation matrix was pre- and post-multiplied by an arbitrary diagonal matrix to generate the common covariance matrix.

1	2	3	4	5	6	7	8	9	10
1.0000									
.1875	1.0000								
.0833	.2000	1.0000							
.2500	.2500	.1667	1.0000						
.1875	.3125	.5000	.3000	1.0000					
.2917	.0833	.3333	.1667	.4167	1.0000				
.4250	.5000	.3333	.6000	.6125	.3333	1.0000			
.2250	.1250	.2917	.2250	.4000	.6000	.2000	1.0000		
.3750	.2250	.2500	.2000	.3000	.4000	.3000	.1250	1.0000	
.1000	.4000	.2667	.1200	.5000	.2667	.0900	.2000	.1000	1.0000

APPENDIX C

"Orthodox" Sampling Procedure

In view of the unexpectedly high positive bias in $\hat{\omega}^2$, it was deemed advisable to make sure that this was not a result of the method, described in Section 3, that was used for generating the sets of populations so as to have preassigned ω^2 values. That method involved simultaneous diagonalization of the common covariance matrix $\underset{\sim}{\Sigma}$ and the cross product $\underset{\sim}{\alpha}\underset{\sim}{\alpha}'$ of the effect-parameter matrix.

Accordingly, the sampling distribution of $\hat{\omega}^2$ when $\omega^2 = 0.1$ was constructed for the case when $p = K = 5$ and $N = 75$, with the populations generated by an alternative method more true to real life.

A conveniently available data deck containing, among other things, scores on five standardized achievement tests for some 260 ninth-grade students was used as one of the five populations; this is referred to as P_o. (Since sampling was to be done with replacement, this modest population size was deemed sufficient.) Each of the remaining four populations, $P_1, \ldots, P_4$, was conceptually (but not physically) generated as follows. Every student's score on any given test was increased by a small constant amount, the amount varying from test to test (as well as from population to population). This was to insure that the five population covariance matrices would be exactly equal.

The additive constants were determined in the following manner: For the j-th fictitious population P_j ($j = 1, 2, 3, 4$), the amounts by which everyone's scores on $X_1, X_2, \ldots, X_5$ were to be incremented were set at $c\delta_{j1}, c\delta_{j2}, \ldots, c\delta_{j5}$, respectively, where $(\delta_{j1}, \delta_{j2}, \ldots, \delta_{j5})$ is a separate random permutation of $(.1, .2, \ldots, .5)$ for each j, and

c is a constant to be determined so that $\omega^2 = 0.1$.

For each $\hat{\omega}^2$ value to be computed, a random sample S_{jj} should be drawn from each population P_j ($j = 0, 1, 2, 3, 4$), and the within- and between-groups matrices $\underset{\sim}{W}$ and $\underset{\sim}{B}$ computed. In actuality, however, five samples $S_{oo}, S_{o1}, \ldots, S_{o4}$ may be drawn from P_o, and subsequent computational adjustments made where necessary. For computing $\underset{\sim}{W}$, no adjustments are needed; for $\underset{\sim}{B}$ the required adjustments are as described below.

Let the centroids of the five samples S_{oj} be

$$\underset{\sim}{\bar{X}}'_{oj} = [\bar{X}_{oj1}, \bar{X}_{oj2}, \ldots, \bar{X}_{oj5}]. \quad (j = 0, 1, \ldots, 4)$$

Then the centroids that would have been observed if the samples S_{jj} ($j = 0, 1, \ldots, 4$) had been drawn are: $\underset{\sim}{\bar{X}}'_{oo}$ (observed) for $j = 0$, and

$$\underset{\sim}{\bar{X}}'_{jj} = \bar{X}'_{oj} + c[\delta_{j1}, \delta_{j2}, \ldots, \delta_{j5}] \quad \text{for } j = 1, 2, 3, 4 .$$

The vector of grand means for the total sample comprising Soo, S11, ..., S44 is

$$\underset{\sim}{\bar{X}}' = \underset{\sim}{\bar{X}}'_o + c[\delta_1, \delta_2, \ldots, \delta_5] ,$$

where

$$\underset{\sim}{\bar{X}}'_o = (\sum_{j=o}^{4} \underset{\sim}{\bar{X}}'_{oj})/5$$

and

$$\bar{\delta}_p = (\sum_{j=1}^{4} \delta_{jp})/5 \qquad [p = 1, 2, \ldots, 5]$$

Thus, the deviations of the group centroids from the grand centroid are

$$\underset{\sim}{\bar{X}}'_{jj} - \underset{\sim}{\bar{X}}' = (\underset{\sim}{\bar{X}}'_{oj} - \underset{\sim}{\bar{X}}'_o) - c[\delta_{j1} - \bar{\delta}_1, \delta_{j2} - \bar{\delta}_2, \ldots, \delta_{j5} - \bar{\delta}_5] ,$$

where the $\bar{X}'_{oj} - \bar{X}'_o$ are found from the samples actually drawn (all from P_o),

and the adjustment term is computable once the δ_{jp} and c are determined.

Now $c(\delta_{jp} - \bar{\delta}_p)$ is precisely the (j, p) element of the transpose $\underset{\sim}{\alpha}'$ of the effect-parameter matrix. Thus, the population ω^2 defined in equation (2.1) is here expressible as

$$\omega^2 = 1 - \frac{|\underset{\sim}{\Sigma}|}{|\underset{\sim}{\Sigma} + c^2(\underset{\sim}{\Delta}\underset{\sim}{\Delta}')/5|}$$

where $\underset{\sim}{\Sigma}$ is the covariance matrix of P_o (and, equivalently of $P_1, \ldots, P_4$) and

$$\underset{\sim}{\Delta}' = \begin{bmatrix} -\bar{\delta}_1 & -\bar{\delta}_2, & \ldots, & -\bar{\delta}_5 \\ \delta_{11} - \bar{\delta}_1, & \delta_{12} - \bar{\delta}_2, & \ldots, & \delta_{15} - \bar{\delta}_5 \\ \vdots & & & \\ \delta_{41} - \bar{\delta}_1, & \delta_{42} - \bar{\delta}_2, & \ldots, & \delta_{45} - \bar{\delta}_5 \end{bmatrix}.$$

since $\underset{\sim}{\Delta}'$ is determined once we select the four random permutations $(\delta_{j1}, \delta_{j2}, \ldots, \delta_{j5})$ of (.1, .2, ..., .5), and $\underset{\sim}{\Sigma}$ is given by the original data, ω^2 is a function solely of c. (It is clearly a monotonely decreasing function of $|c|$.) By a trial-and-error process, the value of c making $\omega^2 = 0.1$ to four decimal places was determined.

The sampling distribution of 1,000 values of $\hat{\omega}^2$ computed in the foregoing manner, grouped in class intervals of size 0.03 each, was as shown in the row labelled f_A below:

f_A	3	16	23	62	74	120	140	141	131	98	88	44	30	19	8	2	1
f_B	9	14	28	79	82	121	145	123	127	101	82	51	19	11	7	1	0

(The class intervals are, from left to right, .0600-.0899, .0900-.1199, ..., .5400-.5699.) In row f_B above is shown the sampling distribution of 1,000 $\hat{\omega}^2$ values generated by the sampling procedure of Section 3.

A visual comparison of the two distributions shows that they are quite similar. A chi-square test for the significance of the difference between the two distributions (with the last three class intervals collapsed into one) yielded $\chi^2 = 13.31$, $\underline{df} = 14$ ($p \simeq .50$).

Thus, it may safely be concluded that the two distributions differ only by sampling error. The apprehension that the high positive bias of $\hat{\omega}^2$, especially for $\omega^2 = 0.1$ with large p and small N, might have resulted from the peculiar manner in which the populations were generated in this study may therefore be cast aside.