The University of Manitoba

The Graduate School

Department of Psychology

A Comparison of Univariate and Multivariate Analysis Strategies for Repeated Measures Designs

A Thesis in

Quantitative Psychology

by

Joanne C. Rogan

Submitted in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

April 1978

A COMPARISON OF UNIVARIATE AND MULTIVARIATE ANALYSIS

STRATEGIES FOR REPEATED MEASURES DESIGNS

BY

JOANNE C. ROGAN

A dissertation submitted to the Faculty of Graduate Studies of the University of Manitoba in partial fulfillment of the requirements of the degree of

DOCTOR OF PHILOSOPHY

To Harvey

Table of Contents

Table of Contents (Cont'd)

List of Tables

List of Tables (Cont'd)

Abstract

While it is generally assumed that the within subjects F ratios in repeated measures designs are valid only when the assumption of uniformity exists among the population variances and covariances, recent literature (e.g., Rouanet & Lepine, 1970) demonstrates that these ratios are exactly distributed as F variates under a less restrictive condition referred to as circularity. For the treatments (J) X occasions (K) design where 'treatments' and 'occasions' are the generic terms for the non-repeated and repeated factor, respectively, F_K and F_{JK} follow exact F distributions if $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \sigma^2\underset{\sim}{I}$ $(j = 1,\ldots,J)$ where $\underset{\sim}{C}$ is a matrix of coefficients representing the K-1 orthonormal contrasts among the levels of the repeated factor, $\underset{\sim}{\Sigma}_j$ is the variance-covariance matrix associated with group j, σ^2 is a scalar where $\sigma^2 > 0$ and $\underset{\sim}{I}$ is an identity matrix. The present research examined the Type I error and power rates of several recommended tests of these repeated measures effects, a test for circularity and various sequential testing strategies where the chosen mean equality test is based on the outcome of a preliminary test for circularity when separately and jointly violating the assumptions of (a) circularity, (b) multivariate normality, and (c) multivariate homoscedasticity $(\underset{\sim}{\Sigma}_j = \underset{\sim}{\Sigma},\ j = 1,\ldots,J)$.

Monte Carlo techniques were used to simulate an orthogonal treatments X occasions design with 3 treatment groups, 4 measurement occasions and 13 observations per treatment group. For both effects, the investigated univariate tests were: (a) the conventional F test, (b)

the conservative F test (Greenhouse & Geisser, 1959), (c) the $\hat{e}$-adjusted F test (Collier, Baker, Mandeville, & Hayes, 1967), and (d) the $\tilde{e}$-adjusted F test (Huynh & Feldt, 1976). The multivariate test of the occasions effect was evaluated by Hotelling's (1931) T^2 while Roy's (1953) largest root, Wilk's (1932) likelihood ratio, and the Pillai-Bartlett (Bartlett, 1939; Pillai, 1955) trace criteria were used to assess the multivariate test of the within subjects interaction effect. The circularity hypothesis was tested in a two-stage procedure using Box's (1949) heteroscedasticity and Mauchly's (1940) sphericity criteria.

In general, results indicated that, of the six uniformly adopted procedures, the $\hat{e}$-adjusted F, the $\tilde{e}$-adjusted F and the multivariate test were the most robust statistics. The data did not, however, suggest a uniform rule for choosing among these procedures. For all conditions, the $\hat{e}$-adjusted F test resulted in the fewest Type I errors and the least sensitive test of the effects. The multivariate tests were, on the average, more powerful than either of the sample-adjusted univariate procedures. However, for non-normal populations, when interest was in the occasions main effect, the univariate procedures provided better Type I error control.

Finally, preliminary testing for circularity was not useful in choosing between subsequent analysis strategies. Because the heteroscedasticity and sphericity criteria were highly sensitive to non-normality and to departures from their null hypotheses, the Type I error rates and sensitivity of the various sequential testing strategies were almost identical to those of the corresponding uniformly adopted mean equality test.

INTRODUCTION

Perhaps the most common experimental design in psychological research is the repeated measures design (Edgington, 1974; Lana & Lubin, 1963). Such designs are characterized by repeated observation on a single variable for every experimental unit or subject, with each successive observation being associated with a different level or combination of levels of the design.

Typically, experimenters adopt the repeated measures design as a means of reducing error variability. Here, treatment effects for a given subject are measured relative to the average response made by that subject on all treatments; in essence each subject serves as his own control. In this way, variability due to differences in the average responsiveness of the subject is eliminated from the experimental error (Winer, 1971).

Analysis of Repeated Measures Designs

While the data from repeated measures designs have a distinct univariate appearance in that there is basically only one dependent variable measured on the same subject on more than one occasion, data from such designs can be equally well described by a univariate or multivariate model: the choice of a model depends upon the nature of the assumption(s) the researcher is willing to make concerning the data. According to Cole and Grizzle (1966), when the observations on successive measurements are correlated (as is most often the case in repeated measures designs), the data, while arising from an apparently univariate experiment, are essentially multivariate in nature and should be analyzed as such. Traditionally, however, the univariate mixed model analysis of variance (ANOVA) has been the conventional test statistic for such designs.

Conventional Analysis

Consider, for example, the case where there are J levels of a treatment factor with I subjects exposed to the jth level and each subject measured on one dependent variable for K occasions following treatments (see Table 1 for schematic). Analysis of this treatments X occasions design would typically proceed as a three factor mixed model ANOVA (see Table 2) with treatments and occasions being completely crossed fixed factors and subjects, a random factor nested within treatment groups. Letting X_{ijk} represent the kth measurement on the ith individual in the jth treatment group, a univariate linear model describing this design is

$$X_{ijk} = \mu + \alpha_j + \beta_k + \pi_{i(j)} + \alpha\beta_{jk} + \beta\pi_{ki(j)} + \varepsilon_{ijk} , \qquad (1)$$

$$i = 1,\ldots,I;\ j = 1,\ldots,J;\ k = 1,\ldots,K$$

with side conditions

$$\sum^{J} \alpha_j = \sum^{K} \beta_k = \sum^{J} \alpha\beta_{jk} = \sum^{K} \alpha\beta_{jk} = \sum^{K} \beta\pi_{ki(j)} = 0.0 ,$$

where μ, α_j, and β_k are the overall population mean, jth treatment effect and kth occasion effect, respectively; $\pi_{i(j)}$ is a constant associated with the ith subject in the jth treatment group; $\alpha\beta_{jk}$ is the jkth interaction effect; $\beta\pi_{ki(j)}$ is the interaction of occasion k and subject i within j; and ε_{ijk} is the random error component. The average over the population from which the subjects are drawn may be

Table 1

Schematic for the Treatments (J) X Occasions (K) Repeated Measures Design

		K_1	K_2	, ...,	K_K	
	S_1	$(X_{111}$	X_{112}	, ...,	$X_{11K})$	
	S_2	$(X_{211}$	X_{212}	, ...,	$X_{21K})$	
J_1	⋮			⋮		
	S_I	$(X_{I11}$	X_{I12}	, ...,	$X_{I1K})$	
	S_{I+1}	$(X_{121}$	X_{122}	, ...,	$X_{12K})$	
	S_{I+2}	$(X_{221}$	X_{222}	, ...,	$X_{22K})$	
J_2	⋮			⋮		
	S_{2I}	$(X_{I21}$	X_{I22}	, ...,	$X_{I2K})$	
	⋮			⋮		
	$S_{(J-1)I+1}$	$(X_{1J1}$	X_{1J2}	, ...,	$X_{1JK})$	
	$S_{(J-1)I+2}$	$(X_{2J1}$	X_{2J2}	, ...,	$X_{2JK})$	
J_J	⋮			⋮		
	S_{N}[a]	$(X_{IJ1}$	X_{IJ2}	, ...,	$X_{IJK})$	

[a]N is the total number of subjects (S).

Table 2

Three Factor Mixed Model ANOVA Summary Table for the Treatments (J) X Occasions (K) Repeated Measures Design

Source	df	Sum of Squares[a]	F
Between subjects	JI-1	(6) - (1)	
J	J-1	Q_1 = (3) - (1)	$\frac{[Q_1/(J-1)]}{[Q_2/J(I-1)]}$
S(J)	J(I-1)	Q_2 = (6) - (3)	
Within subjects	JI(K-1)	(2) - (6)	
K	K-1	Q_3 = (4) - (1)	$\frac{[Q_3/(K-1)]}{[Q_5/J(I-1)(K-1)]}$
JK	(J-1)(K-1)	Q_4 = (5) - (3) - (4) + (1)	$\frac{[Q_4/(J-1)(K-1)]}{[Q_5/J(I-1)(K-1)]}$
KS(J)	J(I-1)(K-1)	Q_5 = (2) - (5) - (6) + (3)	
Total	IJK-1	(2) - (1)	

[a] (1) = $X^2_{...}/IJK$; (2) = $\sum\sum\sum^{IJK} X_{ijk}{}^2$; (3) = $\sum^{J} X_{.j.}{}^2/IK$; (4) = $\sum^{K} X_{..k}{}^2/IJ$; (5) = $\sum\sum^{JK} X_{.jk}{}^2/I$;
(6) = $\sum\sum^{IJ} X_{ij.}{}^2/K$

denoted by $E(X_{ijk})$ where

$$E(X_{ijk}) = \mu_{jk} = \mu + \alpha_j + \beta_k + \alpha\beta_{jk} . \tag{2}$$

Expressing the random departure of X_{ijk} from μ_{jk} as m_{ijk},

$$X_{ijk} = \mu_{jk} + m_{ijk} , \tag{3}$$

where

$$m_{ijk} = \pi_{i(j)} + \beta\pi_{ki(j)} + \varepsilon_{ijk} . \tag{4}$$

These random components m_{ijk} are assumed to be normally distributed with

$$E(m_{ijk}) = 0 \tag{5}$$

and

$$\mathrm{Cov}(m_{ijk}, m_{i'j'k'}) = \mathrm{Cov}(X_{ijk}, X_{i'j'k'}) = \begin{cases} \sigma^2, \text{ if } i = i', \\ \quad j = j', k = k' \\ \rho\sigma^2, \text{ if } i = i', \\ \quad j = j', k \neq k' \\ 0, \text{ otherwise,} \end{cases} \tag{6}$$

where σ^2 is the common variance of the observations and ρ is the constant correlation between observations at any two measurement occasions. Expressing these assumptions concerning the K variances and the $\binom{K}{2}$ covariances in population variance-covariance matrices results in an overall K X K matrix of the form

$$\underset{\sim}{\Sigma} = \begin{bmatrix} \sigma^2 & \rho\sigma^2 & \cdots & \rho\sigma^2 \\ \rho\sigma^2 & \sigma^2 & \cdots & \rho\sigma^2 \\ \cdot & \cdot & \cdots & \cdot \\ \rho\sigma^2 & \rho\sigma^2 & \cdots & \sigma^2 \end{bmatrix} = \sigma^2 \begin{bmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \cdot & \cdot & \cdots & \cdot \\ \rho & \rho & \cdots & 1 \end{bmatrix} , \tag{7}$$

with the variance-covariance matrices associated with each level of treatment J, $\underset{\sim}{\Sigma}_j$, being common to all J populations (i.e., $\underset{\sim}{\Sigma}_j = \underset{\sim}{\Sigma}$, $j = 1,\ldots,J$). A matrix of the form in (7) is said to possess the property of compound symmetry or uniformity (Geisser, 1963).[1,2]

For mixed designs in which a different group of randomly assigned subjects is observed under each factor level combination, the random error components and hence the observations can be assumed to be independent and thus the assumption of constant covariance met (Kirk, 1968, p. 139). However, this is not the case for designs involving repeated measurements. Here, the essential feature is that the random error components associated with a particular subject reflect fairly constant characteristics of that subject and must be treated as correlated (Morrison, 1976, p. 141). Moreover, in such designs, it is common to find that two successive or adjacent measurements are more

---

[1] As will be discussed subsequently, the assumptions of uniformity of $\underset{\sim}{\Sigma}$ and equality of the $\underset{\sim}{\Sigma}_j$'s are sufficient but not necessary conditions for exact or valid univariate F ratios in repeated measures designs (Huynh & Feldt, 1970; Rouanet & Lepine, 1970). However, because the majority of the literature to be reviewed makes reference to these assumptions in their traditional form, they are initially stated as such.

[2] Note that the constant covariance requirement applies only to the within subjects effects (i.e., K and JK) and not to the between subjects effect J. The significance of the between subjects factor is assessed using an error term (S(J)) corresponding to the pooled variation among the subjects within the J groups <u>summed over</u> the K measurement occasions. Thus, analysis of this factor is a one way ANOVA and is valid regardless of the assumption concerning the covariances among the measurement occasions. The sole requirement for a valid F_J is that the variation within each of the J treatment populations be homogeneous (Danford, Hughes, & McNee, 1960). In contrast, the significance of the within subjects effects is assessed using an error term (KS(J)) which is not composed of variation collapsed over the measurement occasions but which is the pooled interaction of the occasions factor K with the subjects within the J groups.

highly correlated than two nonadjacent measurements, with the correlation between these measurements decreasing the further apart the measurements are in the series (Danford, Hughes, & McNee, 1960). A variance-covariance matrix based on such serially correlated observations is said to possess the simplex form (Guttman, 1955; Humphreys, 1960).

When the unequal correlations are 'timed-based' (Lana & Lubin, 1963), that is, when correlation heterogeneity is due solely to the temporal order of the repeated measurements (e.g., repeated testing effects, sequence effects), randomization or counterbalancing have been suggested to achieve homogeneous covariances (Box & Müller, 1959; Gaito, 1961). In randomization, the repeated measurements are obtained in a random order determined separately for each subject with the rationale being that, over the entire design, any order effects will be distributed approximately equally over the measurement occasions and consequently will result in relatively constant covariances. A counterbalanced design, on the other hand, is achieved by collecting the repeated measurements in all possible orders using an equal number of subjects within each order. An additional 'order' factor may then be included in the subsequent analysis to detect possible effects due to this counterbalancing.

According to McCall and Appelbaum (1973), both randomization and counterbalancing have major weaknesses. For example, in the randomization strategy, any order effects are incorporated in the error term(s) for the analysis resulting in less sensitive tests of effects, with the sensitivity decreasing with increases in the magnitude of the order effects. Gaito (1961) has discussed the disadvantages of counterbalanced designs which include loss of power, increases in the required number of subjects and additional, seldom-satisfied assumptions concerning the variances and covariances.

When the unequal correlations are 'treatment-bound' (Lana & Lubin, 1963), that is, when correlation heterogeneity is directly the result of the specific nature of the levels of the repeated factor (e.g., a repeated 'test' factor where specific tests will always correlate more highly with others regardless of the order of administration), neither randomization nor counterbalancing have any effect on eliminating covariance heterogeneity. Moreover, in many research situations, such procedures are not applicable, for example, when the repeated assessments must be made in a single order. Common research paradigms in psychology where testing order cannot be manipulated include longitudinal developmental studies assessing various behavioural changes over age or time and classical learning experiments where behaviour is assessed several times during training (i.e., on each trial) or during extinction. It is precisely these types of studies that seldom, if ever, meet the required covariance assumption, rather yielding variance-covariance matrices of the simplex pattern.

Consequences of assumption-violation. The effects of covariance heterogeneity were first postulated by Kogan (1948) who suggested that positive intercorrelations between measurement occasions would result in a liberal F test of the occasions effect which overestimated the significance of the differences between the K correlated means. Investigating the approximate effects of unequal correlations and unequal variances in a simple repeated measures design[3], Box (1954b) found Kogan's (1948) conjecture to be correct. As the magnitude of the first order serial correlation ($\rho = \pm.20, \pm.40$) and/or the degree of variance heterogeneity between the measurement occasions increased, the occasions F test become increasingly liberal with the true probability of a Type I error being greater than the nominal significance level. Similar results were reported by Imhof (1962).

Adjusted Univariate Analyses

'e-adjusted' F test. Using the exact multivariate approach and a formula previously developed (Box, 1954a, p. 297) for the approximation to the distribution of the ratio of two independent quadratic forms, Box (1954b) demonstrated that, under the null hypothesis, the true distribution of the measurement occasions test ratio, F_K, with (K-1) and (K-1)(N-1) degrees of freedom (d.f.) is approximately distributed as ($\overset{\text{approx.}}{\sim}$)

$$F[(K-1)e,\ (K-1)(N-1)e], \tag{8}$$

where

$$e = \frac{K^2(\bar{\sigma}_{kk} - \bar{\sigma}_{..})^2}{(K-1)(\sum\sum_{kk'} \sigma^2_{kk'} - 2\sum_k \bar{\sigma}^2_{k.} + K^2\bar{\sigma}_{..}^2)}, \quad k,k' = 1,\ldots,K \tag{9}$$

[3] A simple repeated measures design is defined as the case when N subjects randomly sampled from an infinite population of subjects are each measured on some characteristic on K occasions. Typically, this design would be analyzed by a two factor mixed model design where subjects is the random factor and occasions, the fixed classification.

and

$\sigma_{kk'}$ is an element of the K X K variance-covariance matrix ($\underset{\sim}{\Sigma}$),

$\bar{\sigma}_{kk}$ is the mean of the diagonal elements of $\underset{\sim}{\Sigma}$,

$\bar{\sigma}_{k.}$ is the mean of the kth row (or column) of $\underset{\sim}{\Sigma}$, and

$\bar{\sigma}_{..}$ is the mean of all elements of $\underset{\sim}{\Sigma}$.

When the K variances and $\binom{K}{2}$ covariances are constant, that is, when $\underset{\sim}{\Sigma}$ is uniform, e equals its upper bound of unity and Box's (1954b) approximation yields exact results. However, as $\underset{\sim}{\Sigma}$ departs from uniformity, the value of e decreases from one and F_K is distributed with a reduced number of d.f..

Geisser and Greenhouse (1958) and Greenhouse and Geisser (1959) extended the work of Box (1954b) to the analysis of the treatments X occasions design (see Tables 1 and 2) showing that, using Box's (1954b) procedure, the F ratios for the tests of the occasions effect and the treatments X occasions interaction could be tested in the same approximate fashion. That is,

$$F_K \overset{\text{approx.}}{\sim} F[(K-1)e,\ J(I-1)(K-1)e] \quad \text{and} \tag{10}$$

$$F_{JK} \overset{\text{approx.}}{\sim} F[(J-1)(K-1)e,\ J(I-1)(K-1)e]\ .^4 \tag{11}$$

Because the test of the treatment effects is valid regardless of the form of $\underset{\sim}{\Sigma}$ (see footnote 2), F_J remains exactly distributed as

$$F[(J-1),\ J(I-1)]\ . \tag{12}$$

Further, these authors demonstrated that for all designs involving a single repeated factor (K), the lower bound of e was equal to $(K-1)^{-1}$.

[4]Note that for such designs involving between subjects factor(s), the parameters in (9) are those of the overall population $\underset{\sim}{\Sigma}$.

Thus,

$$(K-1)^{-1} \leq e \leq 1.00 \ . \tag{13}$$

According to Geisser and Greenhouse (1958), these results are generalizable to higher order repeated measures designs where one of the factors is individuals subdivided into groups or treatments and the remaining classifications are fixed. However, it is important to note that this correction in the d.f. only applies to tests of within subjects main effects or any interaction effects involving within subjects factors; for tests of between subjects effects, the F ratios are exactly distributed as F variates with the usual d.f.. In addition, the value of the lower bound of e is dependent upon the number of repeated factors in the design and the specific F test in question.[5]

<u>Conservative F test</u>. Based on the facts that Box's (1954b) e-adjusted test requires the often laborious computation of e which is based on the elements of a typically unknown population $\underset{\sim}{\Sigma}$ and that use of a sample-estimated e on the approximate F distributions was also unknown, Geisser and Greenhouse (1958) suggested a procedure in which the value of e is always set at its lower limit. Thus, for the treatments X occasions design where the lower limit of e is equal to $(K-1)^{-1}$,

$$F_K \overset{\text{approx.}}{\sim} F[1, J(I-1)] \text{ and} \tag{14}$$

$$F_{JK} \overset{\text{approx.}}{\sim} F[(J-1), J(I-1)] \ .^6 \tag{15}$$

---

[5]For example, for a J X K X L factorial design with two repeated factors, K and L, for the test of factor KL, the lower bound of e is equal to $[(K-1)(L-1)]^{-1}$ while for F_{JK}, e takes on a minimum value of $(K-1)^{-1}$.

[6]For the factorial design discussed in footnote 5, the conservative test of the KL effect is approximately distributed as an F variable with 1 and J(I-1) d.f. while $F_{JK} \overset{\text{approx.}}{\sim} F[(J-1), J(I-1)]$.

However, while using the lower bound of e has the advantages of being computationally expedient and independent of the elements of $\underset{\sim}{\Sigma}$, such a procedure may be overly conservative and consequently lacking in sensitivity: setting e equal to its lower bound assumes maximum nonuniformity (an assumption which is not always valid) and results in a maximum reduction in the d.f. and, consequently, a loss in power. As a possible means of handling this problem, Greenhouse and Geisser (1959) proposed a three step approach to testing significance. The F ratio in question is first tested with the conventional d.f.. If the F ratio is not significant, the analysis stops for the null hypothesis would not be rejected by reducing the d.f.. On the other hand, if the observed F is significant, the conservative test is performed with the d.f. reduced by the value of the lower limit of e. An F ratio larger than this second critical value may be declared significant without further testing. However, if the conservative test is not significant, it is suggested that e be estimated from the sample variance-covariance matrix and the F ratio be tested with d.f. reduced by this estimate.[7]

<u>'$\hat{e}$-adjusted' F test</u>. The effect of using a sample-estimated value of e on the approximate F distributions of F_K and F_{JK} in the treatments X occasions design was investigated by Collier, Baker, Mandeville, and Hayes (1967). By substituting sample values for the parameters in (9), the formula for a sample estimate of e, $\hat{e}$, is

$$\hat{e} = \frac{K^2(\overline{S}_{kk} - \overline{S}_{..})^2}{(K-1)(\sum\sum_{kk'} S^2_{kk'} - 2K\sum_k \overline{S}^2_{k.} + K^2\overline{S}^2_{..}} \qquad (16)$$

[7]Lana and Lubin (1963) recommend a multivariate test when a significant conventional F test is followed by a nonsignificant conservative F test.

where

$s_{kk'}$ is an element of the overall sample variance-covariance matrix, $\underset{\sim}{S}$,

$\bar{s}_{kk}$ is the mean of the diagonal elements of $\underset{\sim}{S}$,

$\bar{s}_{k.}$ is the mean of the kth row (or column) of $\underset{\sim}{S}$, and

$\bar{s}_{..}$ is the mean of all the elements of $\underset{\sim}{S}$.

According to Anderson (1958), if the parent population is multivariate normal, $\hat{e}$ is a maximum likelihood estimator for e.

Results of the sampling study of Collier et al. (1967) indicated that, in general, for varying degrees of $\underset{\sim}{\Sigma}$ nonuniformity ($.45 \leq e \leq 1.00$), the distributions of $\hat{e}$ were negatively skewed for very high values of e and positively skewed for very low values of e. However, for values of e in the middle of the possible range of values, the bias in the estimation of e was negligible. Moreover, for large sample sizes (I = 15), the estimates of very high and very low values of e were less biased. Motivated by these results, Collier et al. (1967) suggested an approximate F test where the d.f. are adjusted by a sample estimate of e, $\hat{e}$. Thus, for the treatments X occasions design,

$$F_K \overset{\text{approx.}}{\sim} F[(K-1)\hat{e},\ J(I-1)(K-1)\hat{e}] \quad \text{and} \tag{17}$$

$$F_{JK} \overset{\text{approx.}}{\sim} F[(J-1)(K-1)\hat{e},\ J(I-1)(K-1)\hat{e}]\ . \tag{18}$$

<u>'$\tilde{e}$-adjusted' F test</u>. Based on the premise that values of e > .75 are fairly common in educational and psychological research, coupled with Collier et al.'s (1967) findings that $\hat{e}$ may be seriously biased when e is near or above .75 especially when sample size is small, Huynh and Feldt (1976) have presented an alternative estimate of e which,

according to their sampling study, is less biased and less dependent on large sample size when $\underset{\sim}{\Sigma}$ deviates only moderately from uniformity.[8] While the optimum statistic would be an unbiased estimate of e, because of the difficulties of deriving the expected value of an intractable ratio[9] (Hajek, 1962), Huynh and Feldt (1976) derived an estimate of e, $\tilde{e}$, which is the ratio of unbiased estimators of the numerator and denominator of the formula for e. Thus, for the treatments X occasions design,

$$\tilde{e} = \frac{I[K^2(\bar{S}_{kk'} - \bar{S}_{..})^2] - 2(\sum\sum_{kk'} S^2_{kk'} - 2K\sum_{k}\bar{S}^2_{k.} + K\bar{S}^2_{..})}{[K-1][(I-1)(\sum\sum_{kk'} S^2_{kk'} - 2K\sum_{k}\bar{S}^2_{k.} + K\bar{S}^2_{..}) - K^2(\bar{S}_{kk'} - \bar{S}_{..})^2]}, \qquad (19)$$

with terms as defined in (16)[10]. Expressing $\tilde{e}$ in terms of $\hat{e}$,

$$\tilde{e} = \frac{N(K-1)\hat{e} - 2}{(K-1)[N-J-(K-1)\hat{e}]} \qquad (20)$$

where N is equal to the total sample size.

According to Huynh and Feldt (1976), for any value of I and K, $\tilde{e} \geq \hat{e}$ with the equality holding when $\hat{e} = (K-1)^{-1}$ and $J = 1$. In addition, unlike $\hat{e}$, values of $\tilde{e}$ can exceed unity. However, because the value of the parameter e cannot exceed 1.0, whenever $\tilde{e}$ exceeds one, the estimate is equated to one.

---

[8] However, when $\underset{\sim}{\Sigma}$ is strongly nonuniform, $\hat{e}$ is a less biased estimator of e than $\tilde{e}$. Moreover, Davidson (1972, p. 447) maintains that strong violations of uniformity do not seem unlikely in practice and are especially likely when the data result from repeated measurements over time and there is sequential dependency between succeeding measurement occasions.

[9] A fraction or ratio whose numerator and denominator involves statistics that are not independent is said to be intractable.

[10] For designs involving between subjects factor(s), the sample values in (16) and (19) are pooled estimates of the parameters of the overall population $\underset{\sim}{\Sigma}$.

Using $\tilde{e}$ as a sample estimate of the population value of e,

$$F_K \overset{\text{approx.}}{\sim} F[(K-1)\tilde{e},\ J(I-1)(K-1)\tilde{e}] \quad \text{and} \tag{21}$$

$$F_{JK} \overset{\text{approx.}}{\sim} F[(J-1)(K-1)\tilde{e},\ J(I-1)(K-1)\tilde{e}] \ . \tag{22}$$

Multivariate Analysis

Equation (1) presents a general univariate linear model for the treatments X occasions design. Assuming the K measurements on the ith individual in the jth treatment to be a randomly sampled observation vector from a K-variate normal distribution with mean vector $\vec{\mu}_j$ and arbitrary covariance matrix $\underset{\sim}{\Sigma}$ which is common to all J populations, that is,

$$\vec{X}'_{ij} = [X_{ij1},\ X_{ij2}, \ldots, X_{ijK}] \sim \text{IN}(\vec{\mu}_j,\ \underset{\sim}{\Sigma}) \ , \tag{23}$$

the general multivariate linear model after parameterization applicable to this design is:

$$\vec{X}'_{ij} = \vec{\mu}' + \vec{\Xi}'_j + \vec{\varepsilon}'_{ij}, \qquad i = 1,\ldots,I;\ j = 1,\ldots,J \tag{24}$$

where $\vec{X}'_{ij}$ is the 1 X K vector of observations across the K levels of the repeated factor for individual i in treatment j; $\vec{\mu}'$ is the 1 X K vector of means of the K measurements; $\vec{\Xi}'_j$ is a 1 X K vector of the effect of the treatment j on the K measurements; and $\vec{\varepsilon}'_{ij}$ is a 1 X K vector of error. Expressing this linear model in terms of population cell means results in the model

$$\vec{X}'_{ij} = \vec{\mu}'_j + \vec{\varepsilon}'_{ij}, \qquad i = 1,\ldots,I;\ j = 1,\ldots,J \tag{25}$$

where $\vec{\mu}'_j$ is a 1 X K vector of the mean of treatment j for the K responses and $\vec{X}'_{ij}$ and $\vec{\varepsilon}'_{ij}$ are as defined previously. In matrix notation,

$$\underset{\sim}{X} = \underset{\sim}{A}\,\underset{\sim}{\xi} + \underset{\sim}{R} \tag{26}$$

where $\underset{\sim}{X}$ is an N X K matrix of observations, $\underset{\sim}{A}$ is an N X J 'design' matrix indicating the incidence of the terms in the model in the data, $\underset{\sim}{\xi}$ is a J X K matrix of unknown population means and $\underset{\sim}{R}$ is an N X K matrix of random error components (Timm, 1976, p. 369).

Both the univariate and multivariate models rest on the assumption that the population random error components and thus the observations follow a normal distribution. However, unlike the univariate model which stipulates a particular form of the population $\underset{\sim}{\Sigma}$ (i.e., uniformity), the multivariate model makes no such specification. Rather, the sole requirement concerning $\underset{\sim}{\Sigma}$ is that it be common to all populations, that is, to each of the J treatment groups.

Given the model in (26), multivariate procedures test hypotheses of the general form

$$H_0\text{:}\ \ \underset{\sim}{B}\,\underset{\sim}{\xi}\,\underset{\sim}{W} = \underset{\sim}{0}\ , \tag{27}$$

where $\underset{\sim}{B}$ is a t X J ($t \leq J-1$) matrix of full row rank consisting of t linearly independent row vectors representing contrasts among elements within given columns of the unknown parameter matrix $\underset{\sim}{\xi}$ and $\underset{\sim}{W}$ is a K X u ($u \leq K-1$) full column rank matrix consisting of u linearly independent column vectors representing contrasts among the elements within different rows of $\underset{\sim}{\xi}$ (Morrison, 1976, p. 176).[11] As Cole and Grizzle (1966) state, a hypothesis of this form is, in fact, the joint statement of $(t \cdot u)$ single d.f. component hypotheses of the form

$$H_0: \quad \vec{b}_j' \underset{\sim}{\xi} \vec{w}_k = 0 , \tag{28}$$

where the vectors $\vec{b}_j'$ $(j = 1,\ldots,t)$ and w_k $(k = 1,\ldots,u)$ compose $\underset{\sim}{B}$ and $\underset{\sim}{W}$, respectively. The choice of these $\vec{b}_j'$ and $\vec{w}_k$ vectors determine the component to be tested.

In the application of multivariate analysis to repeated measures designs (in this case, to the treatments X occasions design), two general types of vectors are used. Consider the columns of the matrix $\underset{\sim}{W}$, that is, the $\vec{w}_k$'s. If $\vec{w}_k$ is a 'summation' vector of the form $\vec{w}_k' = \{1, 1,\ldots,1\}$, the component hypotheses corresponding to the elements of $\underset{\sim}{B} \underset{\sim}{\xi} \vec{w}_k$ involve the sum of the responses for each subject across the K measurement occasions. This summation 'smoothes out' or 'collapses over' variations within the rows of $\underset{\sim}{\xi}$ due to the different measurement occasions represented by the elements of the observation

[11] The rank of a matrix X is the number of linearly independent rows or columns of X (e.g., see Searle, 1966). In general, V row or column vectors of the same order are said to be linearly dependent if non-zero constants $c_1, c_2,\ldots,c_V$ exist such that $c_1\vec{u}_1 + c_2\vec{u}_2 + ,\ldots, c_V\vec{u}_V = 0$. If zeros are the only values of the c's for which this equation is true, then the vectors are said to be linearly independent. Because the number of independent rows in a matrix always equals the number of independent columns and vice-versa, the rank of a rectangular matrix $\underset{\sim}{X}$, where p is the number of rows and q is the number of columns, is equal to or less than the smaller of p and q(min(p,q)). If the rank of the matrix is equal to min(p,q), then the matrix is said to be of full rank; if, on the other hand, the rank of the matrix is less than min(p,q), then the matrix is said to be of less than full rank. Further, if the rank of the matrix is equal to q, the number of columns, and q < p, then the matrix is said to be of full column rank; if the rank of the matrix is equal to p and p < q, the matrix is termed a full row rank matrix.

vector, $\vec{X}'_{ij}$. Accordingly, such a matrix, $\underset{\sim}{W}_1$, consisting of a single vector of this form is typically used to construct hypotheses concerning aspects of the main effect of the between subjects factor, J.

If, on the other hand, $\vec{w}_k$ is a 'contrast' vector whose elements sum to zero, the resulting component hypotheses corresponding to the elements of $\underset{\sim}{B}\, \underset{\sim}{\xi}\, \vec{w}_k$ will involve the associated contrast between the elements of each row of $\underset{\sim}{\xi}$. Unlike hypotheses involving 'summation' matrices of the $\underset{\sim}{W}_1$ type discussed above which collapse over variations within the rows of $\underset{\sim}{\xi}$, these component hypotheses concern differences that distinguish the measurements comprising the observation vector. As a result, matrices composed of these 'contrast' $\vec{w}_k$'s, $\underset{\sim}{W}_2$, are used to test hypotheses about the main effect of the within factor K and interactions between this factor and the between factor J.

Like the columns of $\underset{\sim}{W}$, the t rows of $\underset{\sim}{B}$ are typically chosen to sum the results over the levels of the between factor, thereby ignoring the differences due to treatments, or to calculate differences between results for treatment groups, thereby estimating the effects of any treatment group differences. Thus, if $\vec{b}'_j$ is the summation vector ($\vec{b}'_j = \{1, 1, \ldots 1\}$), the component hypotheses corresponding to the elements of $\vec{b}'_j\, \underset{\sim}{\xi}\, \underset{\sim}{W}$ disregard any differences between the treatment group means. Accordingly, a matrix, $\underset{\sim}{B}_1$, consisting of a single summation vector is used to form hypotheses concerning aspects of the within factor main effect. However, if $\vec{b}'_j$ is a 'contrast' vector, the resulting component hypotheses corresponding to the elements of $\vec{b}'_j\, \underset{\sim}{\xi}\, \underset{\sim}{W}$ will involve the corresponding contrast between the elements of each column of $\underset{\sim}{\xi}$. Matrices consisting of contrast $\vec{b}'_j$'s, $\underset{\sim}{B}_2$, are used in the construction of hypotheses concerning the main effect of the between factor and the interaction of this factor and the within factor.

Through various combinations of these four summation and contrast matrices ($\underset{\sim}{B}_1$, $\underset{\sim}{B}_2$, $\underset{\sim}{W}_1$, $\underset{\sim}{W}_2$), one can construct matrix statements of the primary hypotheses of interest in the treatments X occasions design. For example, if the contrast matrix $\underset{\sim}{B}_2$ is combined with the summation matrix $\underset{\sim}{W}_1$, the resulting joint null hypothesis,

$$H_0: \; \underset{\sim}{B}_2 \, \underset{\sim}{\xi} \, \underset{\sim}{W}_1 = \vec{0} \tag{29}$$

is that of no difference between the levels of the between subjects treatment factor J. Similarly, combining the summation matrix, $\underset{\sim}{B}_1$, with the contrast matrix, $\underset{\sim}{W}_2$, results in the joint null hypothesis,

$$H_0: \; \underset{\sim}{B}_1 \, \underset{\sim}{\xi} \, \underset{\sim}{W}_2 = \vec{0}' \; , \tag{30}$$

of equal occasion effects. Finally, combining the contrast matrices $\underset{\sim}{B}_2$ and $\underset{\sim}{W}_2$ yields the null hypothesis of no treatments X occasions interaction or

$$H_0: \; \underset{\sim}{B}_2 \, \underset{\sim}{\xi} \, \underset{\sim}{W}_2 = \underset{\sim}{0} \; . \tag{31}$$

Full matrix representations of hypotheses (29), (30), and (31) for any values of J and K are presented in Table 3. Table 4 contains the matrix expansion of hypotheses (29) - (31) for a treatments X occasions design with three treatment levels and four measurement occasions.

To test hypotheses of the form given in (27), a host of multivariate test criteria are available. However, unlike the univariate case where the usual F test of a given hypothesis is the uniformly most powerful invariant[12] test criterion against any alternative to the null hypothesis, in the multivariate case, no test criteria possessing the

---

[12] Tests possessing the property of invariance are unaffected by linear transformations, that is, by changes of scale and/or origin (Olson, 1976).

Table 3

Full Matrix Representations of Hypotheses of the Treatments

X Occasions Repeated Measures Design

Treatment Factor J: H_0: $\underset{\sim}{B}_2 \; \underset{\sim}{\xi} \; \underset{\sim}{W}_1 = \vec{0}$.

$$H_0: \; {}_{J-1}\begin{bmatrix} 1 & -1 & 0,\ldots, & 0 \\ 1 & 0 & -1,\ldots, & 0 \\ \cdot & \cdot & \cdot\;\cdot\;\cdot & \cdot \\ 1 & 0 & 0,\ldots, & -1 \end{bmatrix}^{J} \cdot {}_{J}\begin{bmatrix} \mu_{11} & \mu_{12},\ldots, & \mu_{1K} \\ \mu_{21} & \mu_{22},\ldots, & \mu_{2K} \\ \cdot & \cdot\;\cdot\;\cdot & \cdot \\ \mu_{J1} & \mu_{J2},\ldots, & \mu_{JK} \end{bmatrix}^{K} \cdot {}_{K}\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix}^{1} = {}_{J-1}\begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}^{1} .$$

Occasion Factor K: H_0: $\underset{\sim}{B}_1 \; \underset{\sim}{\xi} \; \underset{\sim}{W}_2 = \vec{0}'$.

$$H_0: \; {}_{1}\begin{bmatrix} 1 & 1,\ldots,1 \end{bmatrix}^{J} \cdot {}_{J}\begin{bmatrix} \mu_{11} & \mu_{12},\ldots, & \mu_{1K} \\ \mu_{21} & \mu_{22},\ldots, & \mu_{2K} \\ \cdot & \cdot\;\cdot\;\cdot & \cdot \\ \mu_{J1} & \mu_{J2},\ldots, & \mu_{JK} \end{bmatrix}^{K} \cdot {}_{K}\begin{bmatrix} 1 & 1 & \cdot & 1 \\ -1 & 0 & \cdot & 0 \\ 0 & -1 & \cdot & 0 \\ \vdots & \vdots & \cdot & \vdots \\ 0 & 0 & \cdot & -1 \end{bmatrix}^{K-1} = {}_{1}[0 \quad 0,\ldots,0]^{K-1}$$

Table 3 (Cont'd)

Treatments X Occasions Interaction: H_0: $\underset{\sim}{B}_2 \underset{\sim}{\xi} \underset{\sim}{W}_2 = \underset{\sim}{0}$.

$$H_0: \; {}_{J-1}\begin{bmatrix} 1 & -1 & 0,\ldots,0 \\ 1 & 0 & -1,\ldots,0 \\ \cdot & \cdot & \cdot\;\cdot\;\cdot \\ 1 & 0 & 0,\ldots,-1 \end{bmatrix}^{J} \cdot {}_{J}\begin{bmatrix} \mu_{11} & \mu_{12},\ldots,\mu_{1K} \\ \mu_{21} & \mu_{22},\ldots,\mu_{2K} \\ \cdot & \cdot\;\cdot\;\cdot \\ \mu_{J1} & \mu_{J2},\ldots,\mu_{JK} \end{bmatrix}^{K} \cdot {}_{K}\begin{bmatrix} 1 & 1 & \cdot & 1 \\ -1 & 0 & \cdot & 0 \\ 0 & -1 & \cdot & 0 \\ \vdots & \vdots & \cdot & \vdots \\ 0 & 0 & \cdot & -1 \end{bmatrix}^{K-1} = {}_{J-1}\begin{bmatrix} 0 & 0,\ldots,0 \\ 0 & 0,\ldots,0 \\ \cdot & \cdot\;\cdot\;\cdot \\ 0 & 0,\ldots,0 \end{bmatrix}^{K-1}$$

Table 4

Matrix Expansion of the Hypotheses of the Treatments X Occasions Repeated Measures Design

Treatment Factor J: H_0: $\underset{\sim}{B}_2 \, \underset{\sim}{\xi} \, \underset{\sim}{W}_1 = \vec{0}$.

$$\begin{bmatrix} \mu_{11} - \mu_{21} + \mu_{12} - \mu_{22} + \mu_{13} - \mu_{23} + \mu_{14} - \mu_{24} \\ \mu_{11} - \mu_{31} + \mu_{12} - \mu_{32} + \mu_{13} - \mu_{33} + \mu_{14} - \mu_{34} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Occasion Factor K: H_0: $\underset{\sim}{B}_1 \, \underset{\sim}{\xi} \, \underset{\sim}{W}_2 = \vec{0}'$.

$$\begin{bmatrix} \mu_{11} + \mu_{21} + \mu_{31} - (\mu_{12} + \mu_{22} + \mu_{32}) & \mu_{11} + \mu_{21} + \mu_{31} - (\mu_{13} + \mu_{23} + \mu_{33}) & \mu_{11} + \mu_{21} + \mu_{31} - (\mu_{14} + \mu_{24} + \mu_{34}) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix} .$$

Treatments X Occasions Interaction: $\underset{\sim}{H}_0$: $\underset{\sim}{B}_2 \, \underset{\sim}{\xi} \, \underset{\sim}{W}_2 = \underset{\sim}{0}$.

$$\begin{bmatrix} \mu_{11} - \mu_{21} - (\mu_{12} - \mu_{22}) & \mu_{11} - \mu_{21} - (\mu_{13} - \mu_{23}) & \mu_{11} - \mu_{21} - (\mu_{14} - \mu_{24}) \\ \mu_{11} - \mu_{31} - (\mu_{12} - \mu_{32}) & \mu_{11} - \mu_{31} - (\mu_{13} - \mu_{33}) & \mu_{11} - \mu_{31} - (\mu_{14} - \mu_{34}) \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} .$$

required invariance has the uniformly greatest power to recommend it. Here, different invariant tests are most powerful in different situations, depending on the nature of the alternative hypothesis (Olson, 1976).

All multivariate test criteria are functions of the eigenvalues or characteristic roots, λ_i $(i = 1,\ldots,s)$ of the matrix product $\underset{\sim}{H}\underset{\sim}{E}^{-1}$ of rank s, where $\underset{\sim}{H}$ is the sum of products matrix for the effect in question and $\underset{\sim}{E}$ is the appropriate error sum of products matrix. Expressing $\underset{\sim}{H}$ and $\underset{\sim}{E}$ in matrix notation

$$\underset{\sim}{H} = \underset{\sim}{W}'\underset{\sim}{X}'\underset{\sim}{A}(\underset{\sim}{A}'\underset{\sim}{A})^{-1}\underset{\sim}{B}'[\underset{\sim}{B}(\underset{\sim}{A}'\underset{\sim}{A})^{-1}\underset{\sim}{B}']^{-1}\underset{\sim}{B}(\underset{\sim}{A}'\underset{\sim}{A})^{-1}\underset{\sim}{A}'\underset{\sim}{X}\underset{\sim}{W} \quad \text{and} \tag{32}$$

$$\underset{\sim}{E} = \underset{\sim}{W}'\underset{\sim}{X}'[\underset{\sim}{I} - \underset{\sim}{A}(\underset{\sim}{A}'\underset{\sim}{A})^{-1}\underset{\sim}{A}']\underset{\sim}{X}\underset{\sim}{W} \tag{33}$$

where $\underset{\sim}{I}$ is an N X N identity matrix and $\underset{\sim}{B}$, $\underset{\sim}{W}$, $\underset{\sim}{A}$ and $\underset{\sim}{X}$ are as defined previously.

Of the many proposed test criteria, the four major contenders are Roy's (1953) largest root criterion, θ, the Hotelling-Lawley (Hotelling, 1951; Lawley, 1938) trace criterion, U, Wilk's (1932) likelihood ratio, Λ, and the Pillai-Bartlett (Bartlett, 1931; Pillai, 1955) trace statistic, V. According to these four criteria, the multivariate hypothesis of no difference among the group mean vectors is rejected if

$$\theta = \frac{\lambda_L}{1 + \lambda_L} > \theta_{\alpha(s,m,n)} , \tag{34}$$

$$U = \sum_{i=1}^{s} \lambda_i > U_{\alpha(s,m,n)} , \tag{35}$$

$$\Lambda = \prod_{i=1}^{s} (1 + \lambda_i)^{-1} < \Lambda_{\alpha(u,\nu_H,\nu_E)} , \text{ and} \tag{36}$$

$$V = \sum_{i=1}^{s} \frac{\lambda_i}{1 + \lambda_i} > V_{\alpha(s,m,n)} \ , \tag{37}$$

where λ_L is the largest characteristic root of the matrix product $\underset{\sim}{H}\underset{\sim}{E}^{-1}$, $s = \min[\text{rank}(\underset{\sim}{B}), \text{rank}(\underset{\sim}{W})] = \min(t, u)$, $\nu_H = t$, the hypothesis degrees of freedom, $m = .5(|u - t| - 1)$, $n = .5(\nu_E - u - 1)$, where $\nu_E = N - \text{rank}(\underset{\sim}{A}) = N - J$, the error degrees of freedom, $\theta_{\alpha(s,m,n)}$ is the upper 1-α percentage point of the distribution of θ with parameters s, m, and n, $U_{\alpha(s,m,n)}$ is the upper 1-α percentage point of the distribution of U with the same parameters, $\Lambda_{\alpha(u,\nu_H,\nu_E)}$ is the lower 1-α percentage point of the distribution of Λ with parameters u, ν_H and ν_E, and $V_{\alpha(s,m,n)}$ is the upper 1-α percentage point of the distribution of V with parameters s, m, and n.[13]

Using Monte Carlo methods, Olson (1974) compared the Type I error rates and power of six multivariate tests including Roy's θ, the Hotelling-Lawley U, Wilk's Λ, and the Pillai-Bartlett V, under varying degrees of non-normality or population dispersion matrix heterogeneity for different numbers of dependent variables (K = 2,3,6,10), treatment

---

[13]When the rank of the matrix $\underset{\sim}{H}\underset{\sim}{E}^{-1}$ is equal to one, that is, when $s = \min(t,u) = 1$, all multivariate test criteria necessarily lead to identical conclusions. When u is equal to one (as in (29)), the single root of $\underset{\sim}{H}\underset{\sim}{E}^{-1}$ is equal to the univariate F ratio; when t is equal to one (as in (30)), the single root of $\underset{\sim}{H}\underset{\sim}{E}^{-1}$ is equal to Hotelling's (1931) one sample T^2 statistic

$$T^2 = N(N-J)\vec{x}'\underset{\sim}{W}_2(\underset{\sim}{W}_2'\underset{\sim}{E}\underset{\sim}{W}_2)^{-1}\underset{\sim}{W}_2'\vec{x} \tag{38}$$

where $\vec{x}$ is the K X 1 grand mean vector with kth element, $x_k = \sum^{I}\sum^{J} X_{\cdot\cdot k}/N$ and N, J, $\underset{\sim}{W}_2$ and $\underset{\sim}{E}$ are as defined previously.

groups ($J = 2,3,6,10$), equal group sample sizes ($I = 5,10,50$) and nominal significance levels ($\alpha = .10, .05, .01$). Results indicated that, in general, the MANOVA test criteria were robust to departures from multivariate normality being only slightly more conservative and slightly less powerful when compared to their Type I error and power characteristics under conditions of multivariate normality. However, heterogeneity of covariance matrices proved to be a more serious violation. While Λ, U, and V were found to be robust to mild degrees of dispersion matrix heterogeneity, for moderate to large degrees of heterogeneity, the Type I error rates of all the test criteria became increasingly liberal accompanied by a consistent increase in power. U, Λ and particularly θ were prone to excessively high Type I error rates. However, increases in the Type I error rate of V tended to be much less severe. In addition, the Type I error rates for all criteria, with the exception of V, became more liberal as the number of dependent variables increased and slightly more liberal as the number of groups increased, although the effect on Λ and U were modest. As the group sample size increased, the Type I error rates for all test criteria decreased and approached the nominal significance level. Olson (1974) concluded by rejecting Roy's largest root criterion due to its high Type I error rates, recommending the Pillai-Bartlett trace criterion, V, as the most robust of the MANOVA tests for general protection against departures from normality and homogeneity of population dispersion matrices. While Λ, U, and V are asymptotically equivalent, according to Olson, as a general 'rule of thumb' only when ν_E is at least 10u times larger than ν_H can Λ and U be considered equal to V.

In summary, multivariate procedures can be used to arrive at tests of the same hypotheses as those tested through the conventional univariate approach without any restriction as to the form of $\underset{\sim}{\Sigma}$. Given this demonstrated flexibility of multivariate procedures, the question arises as to why one should not uniformly adopt this procedure for repeated observation experiments. According to Cole and Grizzle (1966) and Rouanet and Lepine (1970), the general use of multivariate analysis cannot be recommended as there are costs to such a procedure in addition to the increased difficulty of calculation (a problem virtually nonexistent given the advent of high speed computers). One liability of the multivariate procedure is its potential lack of sensitivity when compared to univariate techniques. While both procedures are asymptotically equivalent (Danford, Hughes & McNee, 1960), for sample sizes less than infinity, if the univariate assumptions concerning $\underset{\sim}{\Sigma}$ are met, the conventional univariate ANOVA is more powerful than the multivariate test. In general, when the model assumptions are weakened, a loss of power will follow particularly when K is large relative to N (McCall & Appelbaum, 1973).

In addition, because the $\text{rank}(\underset{\sim}{E}) = \min(\nu_E, u)$, to test any hypothesis of the form given in (27) it is necessary that the error degrees of freedom (N-J) be equal to or greater u in order that $\underset{\sim}{E}$ be nonsingular. In the event that $N-J < u$, $\underset{\sim}{E}$ is a singular matrix with no inverse, $\underset{\sim}{E}^{-1}$, which is necessary for computation of the matrix product $\underset{\sim}{H}\underset{\sim}{E}^{-1}$. According to Greenhouse and Geisser (1959), under such conditions, adjusted univariate procedures are also not applicable as their computation is based on the elements of a singular population or sample variance-covariance matrix. Thus, when $N-J < u$, only the conventional and conservative F-tests may be computed.

Comparisons of Analysis Strategies

Using Monte Carlo methods, Collier et al. (1967) investigated the effects of varying degrees of $\underset{\sim}{\Sigma}$ nonuniformity ($.45 \leq e \leq 1.00$) on the empirical Type I error rates of the conventional, conservative, e-adjusted and $\hat{e}$-adjusted tests of the within subjects hypotheses in the treatments X occasions design (J = 3, K = 4). For varying nominal significance levels (α = .10, .05, .025, .01) and treatment group or block sizes (I = 5,15), results indicated that for all 15 conditions of nonuniformity investigated, the empirical rates of Type I error for the conventional F tests were equal to or greater than their nominal values with the magnitude of this discrepancy increasing with decreases in the value of e. For α = .10, the empirical estimates attained a maximum value of .15, while for α = .05, .025, and .01, the maximum rates of Type I error were .11, .07, and .06, respectively. In contrast, the empirical estimates of the conservative tests were markedly less than their associated nominal values except when e was close to its lower bound indicating extremely nonuniform $\underset{\sim}{\Sigma}$'s. For tests with d.f. reduced by e, the Type I error rates were very close to the nominal values. Similarly, the obtained sizes for the $\hat{e}$-adjusted tests agreed well with their nominal significance levels, especially when e $\leq$ .74. For those conditions characterized by highly homogeneous $\underset{\sim}{\Sigma}$'s ($e \geq .90$), the empirical estimates were consistently lower than the nominal size. Based on these results, Collier et al. (1967) recommended that when the assumptions of the univariate model are suspect, tests of the within subjects hypotheses should be conducted by referrring the usual univariate F ratios to the F distribution with d.f. reduced by $\hat{e}$.

The relative robustness of the $\hat{e}$-adjusted and $\tilde{e}$-adjusted F tests of the within subjects hypotheses in a simple repeated measures design (K = 5) to varying degrees of $\underset{\sim}{\Sigma}$ nonuniformity ($.36 \leq e \leq 1.00$) was examined by Huynh and Feldt (1976). For varying nominal significance levels (α = .10, .05, .025, .01), results indicated that for relatively small values of e ($e < .75$) or when the number of subjects was large (N = 15-20), the $\hat{e}$-adjusted test provided better control of Type I errors. However, for large values of e ($e > .75$), tests with d.f. adjusted by $\tilde{e}$ maintained the empirical estimates closer to the nominal values and, moreover, were less dependent on the size of the sample, performing well even when the sample size was only twice the number of levels of K. Based on the belief that the condition of $e > .75$ best typifies educational and psychological research, Huynh and Feldt (1976) recommended the use of $\tilde{e}$-adjusted tests in the presence of suspected $\underset{\sim}{\Sigma}$ nonuniformity.

Noe (Note 1) obtained empirical rates of Type I error for four nominal significance levels (α = .10, .05, .025, .01) for the conventional, conservative, and multivariate tests (one sample Hotelling's (1931) T^2 and Rao's (1952) F approximation to Λ) in the treatments X occasions design (J = 3, K = 4, I = 15) when $\underset{\sim}{\Sigma}$ nonuniformity ($.45 \leq e \leq 1.00$) and equality of the $\underset{\sim}{\Sigma}_j$'s (j = 1,...,J) was violated jointly or separately. As expected, the multivariate tests of the within subjects effects and the conventional test of the between subjects effect were unaffected by $\underset{\sim}{\Sigma}$ nonuniformity, consistently controlling the rate of Type I error at or slightly below the nominal significance levels. However, for a given value of e, the conventional tests of both within subjects effects approximately equally overestimated the nominal significance levels with the degree of liberalness increasing with decreases in the value of e. While the Type I error rates of the conservative tests of these effects also increased as the value of e decreased, even for the smallest values of e, the estimates remained below the nominal significance levels, especially for the interaction test.

Violation of the assumption of equality of the $\underset{\sim}{\Sigma}_j$'s for all levels of the between factor J in the presence of $\underset{\sim}{\Sigma}$ uniformity led to biased rates of Type I error for all the tests investigated with the effect being largest for tests of interaction effects and decreasing for the conventional test of treatment effects and further for tests of the occasions effect. When treatment group sizes were equal, the Type I error rates of all three conventional tests and the multivariate tests of occasions and interaction effects were slightly liberal. However, for unequal group sizes, this bias was more severe, with its direction depending on the pairing of the unequal sample sizes and the unequal $\underset{\sim}{\Sigma}_j$'s. When the unequal sample sizes were directly paired with unequal $\underset{\sim}{\Sigma}_j$'s, that is, when a fewer number of observations were sampled from a population with the smaller $\underset{\sim}{\Sigma}_j$, all conventional and multivariate tests were conservative. For indirect pairings of unequal sample sizes and unequal $\underset{\sim}{\Sigma}_j$'s, that is, when a few number of observations were sampled from a population with the larger $\underset{\sim}{\Sigma}_j$, all conventional and multivariate tests proved to be liberal with empirical rates of Type I error exceeding the nominal value.[14] Regardless of the distribution of the sample sizes, the conservative tests of the within subjects hypotheses maintained their conservative nature.

When the assumptions of uniformity and common population $\underset{\sim}{\Sigma}_j$'s were simultaneously violated, as the value of e decreased, the bias in the Type I error rates of the conventional and conservative tests of the within subjects effects became increasingly larger than those present when only the assumption of common population $\underset{\sim}{\Sigma}_j$'s was violated. However, because the conventional test of the between subjects factor and the multivariate tests of the within subjects factors are not dependent on the form of $\underset{\sim}{\Sigma}$, simultaneous violation of both assumptions produced results identical to those where only the assumption of common population $\underset{\sim}{\Sigma}_j$'s was violated. Noe (Note 1) thus concluded that while there is little difference in the Type I error protection afforded by the conventional and multivariate tests of the within subjects effects when only the assumption of common $\underset{\sim}{\Sigma}_j$'s is violated, when both assumptions are violated, the multivariate tests are clearly superior.

The power of the univariate conservative F test and a multivariate technique (Hotelling's (1931) T^2) was compared by Davidson (1972) for a simple repeated measures design (K = 3,6). The relative power of the two tests was found to be dependent on (a) the total sample size, (b)

---

[14] Similar results were reported by Ito and Schull (1964) investigating the robustness of Hotelling's (1951) U to unequal $\underset{\sim}{\Sigma}_j$'s in the presence of equal sample sizes and Ito (1969) examining the robustness of Hotelling's (1951) U and Wilk's (1932) Λ to direct and indirect pairings of unequal sample sizes and unequal $\underset{\sim}{\Sigma}_j$'s.

the degree to which $\underset{\sim}{\Sigma}$ was nonuniform, and (c) when $\underset{\sim}{\Sigma}$ was nonuniform, the nature of the alternative hypothesis. When $\underset{\sim}{\Sigma}$ was uniform, the conservative test proved to be less sensitive than the multivariate test with the power of the multivariate test increasing with increases in the total sample size. When $\underset{\sim}{\Sigma}$ was substantially nonuniform and when the levels of the repeated factor could be divided into two sets with the data being perfectly correlated within each set and less well correlated between sets, the conservative test was somewhat more powerful than the multivariate test. However, under similar conditions of nonuniformity, only the multivariate test was sensitive to small, reliable differences between highly correlated measurement occasions when other less correlated conditions were present; under such conditions, the conservative test was relatively powerless. Because an often-stated advantage of the repeated measures design is its ability to detect small but reliable differences (Myers, 1966), Davidson (1972) recommended that, when the form of $\underset{\sim}{\Sigma}$ is not known in advance, multivariate tests are the method of choice.

Unfortunately, the majority of Davidson's (1972) conclusions concerning the relative power of univariate and multivariate tests are based on the conservative adjustment of the univariate test. While Davidson (1972) did describe a form of the univariate test based on a sample estimate of e and acknowledged that it is a less conservative and apparently more successful procedure for estimating the appropriate correction factor, he nevertheless derived his power tables using the most conservative value of e. Given the extremely conservative nature of such tests, it would seem that the power of univariate tests with d.f. adjusted by a sample value of e would be larger than that of the conservative test and would compare more favorably to the multivariate test (Wilson, 1975).

Using a simple repeated measures design, Stoloff (1970) compared the effect of varying degrees of $\underset{\sim}{\Sigma}$ nonuniformity ($.54 \leq e \leq .99$) on the Type I error rates and power of the conventional, conservative, e-adjusted, $\hat{e}$-adjusted and multivariate (T^2) tests for varying nominal significance levels ($\alpha = .05, .01$), number of measurement occasions ($K = 3,5$) and total sample size (for $K = 3$, $N = 4,6,10,15$; for $K = 5$, $N = 6,7,10,15$). With respect to Type I errors, the obtained results were similar to those reported by Collier, et al. (1967) and Noe (Note 1). Thus, as the degree of $\underset{\sim}{\Sigma}$ nonuniformity increased, the conventional F test became increasingly liberal while the empirical estimates of the conservative test remained consistently less than the nominal values only providing a relatively unbiased test when e approached its lower bound of $(K-1)^{-1}$. Both the e-adjusted and $\hat{e}$-adjusted tests consistently maintained the rate of Type I error close to the levels of significance with the $\hat{e}$-adjusted procedure showing a slight negative bias under conditions of minimal $\underset{\sim}{\Sigma}$ nonuniformity. The empirical estimates of the multivariate tests were found to be similar to those associated with the $\hat{e}$-adjusted tests with rates of Type I error close to the nominal sizes under conditions of $\underset{\sim}{\Sigma}$ nonuniformity. For all tests of the null hypothesis, increasing the number of measurement occasions led to slight increases in the rate of Type I error while increasing the total sample size resulted in negligible effects for all tests except the conventional F test which showed a decrease in bias.

With respect to sensitivity, when $e = 1.00$, the conventional and e-adjusted tests were generally found to be most powerful, followed by the $\hat{e}$-adjusted, the conservative, and the multivariate test, in decreasing order.[15] When $\underset{\sim}{\Sigma}$ was nonuniform, the power of the approximate tests using the F distribution (i.e., conservative, e-adjusted and $\hat{e}$-adjusted tests) increased with decreases in the value of e and approached that of the conventional F test. Thus, the various correction procedures effectively maintained relatively high levels of power even though a curtailment in the d.f. resulted when $\underset{\sim}{\Sigma}$ was not uniform.

Mendoza, Nicewander and Toothaker (1974) extended the work of Stoloff (1970) to the treatments X occasions design ($J = 3$, $K = 4$, $I = 9$) comparing the sensitivity and Type I error rates of the conventional, conservative, e-adjusted, $\hat{e}$-adjusted and multivariate tests (T^2 and Roy's (1953) θ) for varying degrees of $\underset{\sim}{\Sigma}$ nonuniformity ($.54 \leq e \leq 1.00$) and/or varying population shapes (normal, chi-squared (χ^2) with 3 d.f.). Results indicated that the $\hat{e}$-adjusted and multivariate tests of the within subjects effects were the most robust tests when subjected to violation of the assumption of $\underset{\sim}{\Sigma}$ uniformity and/or normality. While the conventional tests had comparable power to these tests, they also had liberal Type I error rates. Similarly, although the conservative tests consistently maintained the rate of Type I error below the nominal

[15] Recall that Davidson (1972) found the power of the multivariate test to be superior to that of the conservative test when Σ was uniform. This discrepancy in results is most likely due to the different sample sizes used in the different investigations. For all types of mean differences investigated only when N was greater than approximately K + 6 did Davidson report the multivariate test to be uniformly more powerful than the conservative test. However, in the majority of conditions investigated by Stoloff (1970) N was less than K + 6. As Davidson (1972) points out, the power of the multivariate test drops rapidly as N nears K.

size, these tests were less powerful than either the $\hat{e}$-adjusted or multivariate tests.

No uniform rule could be stated for choosing between the $\hat{e}$-adjusted tests and their multivariate counterparts. For both small and larger population mean differences, while the multivariate test of the occasion effect was, on the average, more powerful, it also made more Type I errors, especially when sampling from the skewed distribution. The multivariate test of the interaction effect was found to be a function of not only the size of the population mean differences but also of $\underset{\sim}{\Sigma}$. Moreover, given certain $\underset{\sim}{\Sigma}$'s, the power of the multivariate test decreased with increases in the population mean differences, resulting in a less sensitive test of effects than the corresponding $\hat{e}$-adjusted test. While this result is counter-intuitive, Mendoza, et al. (1974) demonstrated that because the power of the multivariate interaction test is a function of the variance-covariance matrix, their empirical power data is consonant with theoretical expectation.

Necessary and Sufficient Conditions for Valid F Ratios in Repeated Measures Designs

As noted previously, the traditionally stated assumption of uniformity and equality of the $\underset{\sim}{\Sigma}_j$'s are sufficient but not necessary conditions for valid within subjects F ratios in repeated measures designs. Rather, these ratios are exactly distributed as F variates under more general conditions, referred to as circularity, of which uniformity and equality of the $\underset{\sim}{\Sigma}_j$'s represent a specific case (Mendoza, Toothaker & Crain, 1976; Rouanet & Lepine, 1970).

The following definitions are pertinent to the presentation of the circularity assumption.

A contrast between levels of the repeated factor(s) is a vector $\vec{c}$ whose t components sum to zero, where t is the total number of repeated factor(s) levels in the design. The standard deviation or norm of a contrast $\vec{c}$ is equal to $(\vec{c}'\vec{c})^{\frac{1}{2}}$; a standardized contrast is a contrast of norm 1.

A t X p matrix is referred to as a contrast matrix, $\underset{\sim}{C}$, if its p column vectors are contrasts. If the p columns are pairwise orthogonal, $\underset{\sim}{C}$ is said to be p-orthogonal and p-orthonormal if, in addition, each column of $\underset{\sim}{C}$ is standardized. Thus, a p-orthonormal contrast matrix, $\underset{\sim}{C}$, with order t X p is of the general form

$$\underset{\sim}{C} = \begin{bmatrix} 1/c_1 & 1/c_2 & \cdots & 1/c_p \\ -1/c_1 & 1/c_2 & \cdots & 1/c_p \\ 0 & -2/c_2 & \cdots & 1/c_p \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & -p/c_p \end{bmatrix}, \tag{39}$$

where $c_i = (i(i+1))^{\frac{1}{2}}$, i = 1,...,p (Mendoza, Toothaker & Crain, Note 2). Note that a matrix of this form satisfies the equality

$$\underset{\sim}{C}'\underset{\sim}{C} = \underset{\sim}{I}_p \tag{40}$$

where $\underset{\sim}{I}_p$ is a p X p identity matrix.

A comparison among repeated factor(s) levels associated with a t X p contrast matrix is defined as the vector space spanned by the p contrasts of this matrix where p is the degrees of freedom of the comparison. Thus, the contrast matrix is said to represent the associated comparison. Indeed, every comparison with p degrees of freedom can be represented by a contrast matrix of rank p and more specifically by a p-orthonormal contrast matrix.

According to Rouanet and Lepine (1970), there are three general types of comparisons. A comparison with one d.f. and thus represented by a single contrast is referred to as a unidimensional comparison. Partial comparisons, on the other hand, are those comparisons with p d.f. where $p < t-1$. Finally, if $p = t-1$, the comparison is called overall. There is only one overall comparison since all contrast matrices of rank t-1 span the same vector space. Further, every comparison whether unidimensional or partial is a subcomparison of the overall comparison.

Given these definitions, consider the simple repeated measures design where the N observation vectors, $\vec{X}_i$, are a random sample from a K-dimensional normal distribution with mean vector $\vec{\mu}$ and variance-covariance matrix $\underset{\sim}{\Sigma}$. According to Rouanet and Lepine (1970), the F ratio

$$F_K = \frac{SS_K/(K-1)}{SS_{NK}/(K-1)(N-1)} \tag{41}$$

will be valid if and only if the following three conditions hold:

(a) SS_K is distributed as a χ^2 variable with (K-1) degrees of freedom (central under the null hypothesis and noncentral under the alternative);

(b) SS_{NK} is distributed as a central χ^2 with (K-1) (N-1) degrees of freedom; and

(c) SS_K and SS_{NK} are independently distributed.

Condition (c) is an immediate consequence of $\vec{X}_i' \sim IN(\vec{\mu}, \underset{\sim}{\Sigma})$. However, in order for conditions (a) and (b) to be satisfied, some constraints must be placed on the elements of $\underset{\sim}{\Sigma}$. Specifically, conditions (a) and (b) are true if and only if

$$\underset{\sim}{C}'\underset{\sim}{\Sigma}\underset{\sim}{C} = \sigma^2 \underset{\sim}{I}_{(K-1)} \text{ ,} \tag{42}$$

where C is a K X (K-1) orthonormal contrast matrix of the form given in (39) representing the comparison of interest, in this case, the K main effect, and σ^2 is a scalar number where $\sigma^2 > 0$. Thus, F_K has an exact F distribution if and only if the K-1 standardized contrasts between the elements of $\underset{\sim}{\Sigma}$ as defined by the coefficients of $\underset{\sim}{C}$ are independent and equally variable. A variance-covariance matrix satisfying this condition is said to satisfy the circularity assumption for the particular comparison represented by $\underset{\sim}{C}$. For example, consider the following population variance-covariance matrix associated with a simple repeated measures design (K = 3).

$$\underset{\sim}{\Sigma} = \begin{bmatrix} 1.0 & 0.5 & 1.5 \\ 0.5 & 3.0 & 2.5 \\ 1.5 & 2.5 & 5.0 \end{bmatrix} .$$

Clearly, this matrix is not uniform. However, the test of the within subjects main effect represented by the orthonormal contrast matrix

$$\underset{\sim}{C} = \begin{bmatrix} .707 & .408 \\ -.707 & .408 \\ 0 & -.816 \end{bmatrix}$$

is valid as $\underset{\sim}{C}'\underset{\sim}{\Sigma}\underset{\sim}{C} = (1.5)\underset{\sim}{I}$.

For the treatments X occasions design, the within subjects F ratios (see Table 2) are distributed exactly as F variates if and only if

(a) SS_K is distributed as a χ^2 variable with (K-1) degrees of freedom (central under the null hypothesis);

(b) SS_{JK} is distributed as a central χ^2 with (J-1)(K-1) degrees of freedom;

(c) $SS_{KS(J)}$ is distributed as a central χ^2 with J(I-1)(K-1) degrees of freedom; and

(d) SS_K and $SS_{KS(J)}$ and SS_{JK} and $SS_{KS(J)}$ are independently distributed.

Again, while condition (d) is immediately satisfied by $\vec{X}'_{ij} \sim IN(\vec{\mu}_j, \underset{\sim}{\Sigma})$, additional constraints must be placed on the elements of the $\underset{\sim}{\Sigma}_j$'s (j = 1,...,J) in order to satisfy conditions (a) - (c). Specifically, conditions (a) - (c) are true if and only if

$$\underset{\sim}{C}'\underset{\sim}{\Sigma}_1\underset{\sim}{C} = \underset{\sim}{C}'\underset{\sim}{\Sigma}_2\underset{\sim}{C} = ,\ldots, = \underset{\sim}{C}'\underset{\sim}{\Sigma}_J\underset{\sim}{C} = \sigma^2\underset{\sim}{I}_{(K-1)} \tag{43}$$

or

$$\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \sigma^2\underset{\sim}{I}_{(K-1)} \text{ , } j = 1,\ldots,J \text{ ,} \tag{44}$$

where $\underset{\sim}{C}$ is the contrast matrix representing the overall comparison among the K occasions and $\underset{\sim}{I}_{(K-1)}$ and σ^2 are as defined previously. Thus, F_K and F_{JK} have exact F distributions if and only if for each level of the between subjects factor J, the K-1 standardized contrasts between the elements of $\underset{\sim}{\Sigma}_j$ as defined by the coefficients of $\underset{\sim}{C}$ are independent and equally variable. Further, this variability must be constant for all levels of J. Variance-covariance matrices satisfying this condition are said to satisfy the circularity property for the specific comparison represented by $\underset{\sim}{C}$. Note that for this design, there is a single circularity assumption necessary for the validity of both within subjects F ratios. This situation occurs because, while F_J and F_{JK} have different degrees of freedom (see Table 2), both involve the same number of degrees of freedom (i.e., (K-1)) among the levels of the repeated factor.

Mendoza, et al. (1976) subdivide the concept of circularity into overall and local circularity. According to Mendoza, et al. (1976), when the contrast matrix in equations (42) - (44) represents the overall comparison of the total number of repeated measures, the assumption of overall circularity is satisfied.[16] Local circularity, on the other hand, is satisfied if $\underset{\sim}{C}$ is a contrast matrix representing any comparison of interest other than the overall comparison. If the assumption of overall circularity is satisfied, then local circularity is necessarily satisfied for any subcomparison. The reverse, however, is not the case. Thus, it is possible for variance-covariance matrices to satisfy the local circularity property for a given comparison while, at the same time, fail to satisfy the circularity property for a different comparison or for the overall comparison. This fact is not apparent from an examination of the treatments X occasions design as the circularity assumption pertaining to both within subjects F ratios involves the overall contrast matrix and hence the overall circularity assumption. However, consider

---

[16] As Mendoza, et al. (Note 2) demonstrate, the concept of overall circularity is equivalent to Huynh and Feldt's (1970) Type H matrix. According to Huynh and Feldt (1970), in order for within subjects F ratios in a treatments X occasions design to be valid, the $\underset{\sim}{\Sigma}_j$'s ($j = 1,\ldots,j$) must be Type H matrices where all possible differences between the levels of the repeated factor are equally variable.

a J X K X L factorial design where there are repeated observations on factors K and L. In order for the overall circularity assumption to be satisfied

$$\underset{\sim}{C}'\underset{\sim j}{\Sigma}\underset{\sim}{C} = \sigma^2 \underset{\sim}{I}_{(KL-1)} \quad , \; j = 1,\ldots,J \tag{45}$$

where $\underset{\sim}{C}$ is a KL X (KL-1) orthonormal contrast matrix representing the overall comparison among the KL repeated measures and $\underset{\sim j}{\Sigma}$ is the KL X KL variance-covariance matrix associated with group j. However, the necessary and sufficient condition for F_K to be exactly distributed as an F variate is

$$\underset{\sim}{C}'\underset{\sim j}{\Sigma}\underset{\sim}{C} = \sigma^2 \underset{\sim}{I}_{(K-1)} \quad , \; j = 1,\ldots,J \tag{46}$$

where $\underset{\sim}{C}$ is a KL X (K-1) orthonormal contrast matrix representing the comparison among the levels of the repeated factor K. Thus, it is possible for a given design to contain both valid and invalid F ratios.

Table 5 presents a partial ANOVA summary table for the J X K X L design and the necessary and sufficient conditions for the validity of each within subjects F ratio. Note that for each of the three sets of within subject effects tested with a common error term, there is a single common circularity assumption necessary to yield valid F ratios while for effects tested with different error terms, the validity of the F ratios rest on different circularity assumptions.

Relationship of overall circularity and uniformity. According to Rouanet and Lepine (1970), the property of uniformity implies the property of overall circularity. Therefore, if $\underset{\sim}{\Sigma}$ is uniform and σ_{kk} is the common value of the variances and $\sigma_{kk'}, k \neq k'$, the common value of the covariances, the overall circularity assumption is satisfied, with its scalar equalling $\sigma_{kk} - \sigma_{kk'}$. The converse, however, is not true:

Table 5

Partial ANOVA Summary Table for a J X K X L Factorial Design with Two Repeated Factors (K and L)

Source	d.f.	Sums of Squares[a]	F	NSC[b]
Within Subjects	IJ(KL-1)			
K	K-1	$Q_1 = (4)-(1)$	$\frac{[Q_1/(K-1)]}{[Q_3/J(I-1)(K-1)]}$	
JK	(J-1)(K-1)	$Q_2 = (6)-(3)-(4)+(1)$	$\frac{[Q_2/(J-1)(K-1)]}{[Q_3/J(I-1)(K-1)]}$	$\underset{\sim}{C}'\Sigma_j\underset{\sim}{C} = \sigma^2\underset{\sim}{I}_{(K-1)}, j=1,\ldots,J$
KS(J)	J(I-1)(K-1)	$Q_3 = (11)-(6)-(10)+(3)$		
L	L-1	$Q_4 = (5)-(1)$	$\frac{[Q_4/(L-1)]}{[Q_6/J(I-1)(L-1)]}$	
JL	(J-1)(L-1)	$Q_5 = (7)-(3)-(5)+(1)$	$\frac{[Q_5/(J-1)(L-1)]}{[Q_6/J(I-1)(L-1)]}$	$\underset{\sim}{C}'\Sigma_j\underset{\sim}{C} = \sigma^2\underset{\sim}{I}_{(L-1)}, j=1,\ldots,J$
LS(J)	J(I-1)(L-1)	$Q_6 = (12)-(7)-(10)+(3)$		
KL	(K-1)(L-1)	$Q_7 = (8)-(4)-(5)+(1)$	$\frac{[Q_7/(K-1)(L-1)]}{[Q_9/J(I-1)(K-1)(L-1)]}$	
JKL	(J-1)(K-1)(L-1)	$Q_8 = (9)-(6)-(7)-(8)+(3)+(4)+(5)-(1)$	$\frac{[Q_8/(J-1)(K-1)(L-1)]}{[Q_9/J(I-1)(K-1)(L-1)]}$	$\underset{\sim}{C}'\Sigma_j\underset{\sim}{C} = \sigma^2\underset{\sim}{I}_{(KL-1)}, j=1,\ldots,J$
KLS(J)	J(I-1)(K-1)(L-1)	$Q_9 = (2)-(9)-(11)-(12)+(6)+(7)+(10)-(3)$		

[a] $(1) = X_{....}^2/IJKL$; $(2) = \sum\sum\sum\sum^{IJKL} X_{ijkl}^2$; $(3) = \sum^{J} X_{.j..}^2/IKL$; $(4) = \sum^{K} X_{..k.}^2/IJL$; $(5) \sum^{L} X_{...l}^2/IJK$;

$(6) = \sum\sum^{JK} X_{.jk.}^2/IL$; $(7) = \sum\sum^{JL} X_{.j.l}^2/IK$; $(8) = \sum\sum^{KL} X_{..kl}^2/IJ$; $(9) = \sum\sum\sum^{JKL} X_{.jkl}^2/I$; $(10) = \sum\sum^{IJ} X_{ij..}^2/KL$;

$(11) = \sum\sum\sum^{IJK} X_{ijk.}^2/L$; $(12) = \sum\sum\sum^{IJL} X_{ij.l}^2/K$

[b] NSC = necessary and sufficient condition for the validity of F .

overall circularity does not imply uniformity. The uniformity assumption amounts to placing $\frac{1}{2}[(p+1)(p+2)] - 2$ linear constraints among the variances and covariances of $\underset{\sim}{\Sigma}$ as compared to $\frac{1}{2}[p(p+1)] - 1$ for the overall circularity property.[17] Thus, an F test of a comparison, in this case, the overall comparison can be valid even when the uniformity assumption is not met.

In addition, Rouanet and Lepine (1970) state that when the overall circularity is satisfied, e can be shown to be equal to unity with its value decreasing with increases in the degree of overall noncircularity of $\underset{\sim}{\Sigma}$. Thus, the correction factor e is an index of departure from the overall circularity assumption. This fact is most important in reinterpreting the reported investigations in light of the less restrictive assumption of circularity. Because these studies used e as an index of the degree of $\underset{\sim}{\Sigma}$ nonuniformity, their obtained results are relevant to an investigation of the effect of violating overall circularity on the various data-analytic strategies for repeated measures designs and, as before, yield no uniform rule as to a choice among the alternative test procedures when overall circularity is not satisfied.

Tests for Circularity

Traditionally, satisfaction of the assumptions of repeated measures designs has been judged by evaluating departures from the uniformity assumption, that is, departures from homogeneity of variance

[17] Note that for single degree of freedom tests, the number of linear constraints placed on $\underset{\sim}{\Sigma}$ by the overall circularity assumption is equal to zero and the assumption is therefore trivially satisfied.

and homogeneity of covariance in the sample covariance matrix (e.g., Box, 1949, 1950). However, uniformity constitutes a sufficient but not a necessary condition for valid F ratios in repeated measures designs. Rather, the necessary and sufficient condition is the satisfaction of the circularity assumption. According to Huynh and Feldt (1970), a statistic which bears more precisely on this assumption is Mauchly's (1940) sphericity criterion, W, which tests the hypothesis of independence and homoscedasticity of multivariate normal random variables. Thus, for the simple repeated measures design with K repeated measurements, W tests the hypothesis that

$$H_0: \ \underset{\sim}{C}'\underset{\sim}{\Sigma}\underset{\sim}{C} = \sigma^2\underset{\sim}{I}. \tag{47}$$

Letting $\underset{\sim}{S}$ denote the unbiased sample variance-covariance matrix, then the value of W is given by

$$W = |\underset{\sim}{C}'\underset{\sim}{S}\underset{\sim}{C}|/[\mathrm{tr}(\underset{\sim}{C}'\underset{\sim}{S}\underset{\sim}{C})/p]^p , \tag{48}$$

where 'tr' is the trace operator and p equals the rank of $\underset{\sim}{C}$, in this case, (K-1). Under the null hypothesis, assuming multivariate normality, if I is large and $I \geq K$, the statistic

$$X_1 = -S_1 \ln W \tag{49}$$

where $S_1 = I-1$ and ln represents the natural logarithm is approximately distributed as a central χ^2 variable with d.f. equal to $[p(p+1)/2] - 1$ and is exact when $p = 2$. When I is small, a better approximation is obtained by multiplying S_1 by the correction term, d, where

$$d = 1 - [(2p^2 + p + 2)/6pS_1] . \tag{50}$$

According to Rouanet and Lepine (1970), because this correction always tends to be conservative and may not be small, its use should generally be recommended as a routine procedure. Thus, to test the circularity hypothesis in (47), the test statistic is given by

$$X_1 = -S_1 d \ln W . \tag{51}$$

For designs involving between subjects factor(s), the circularity assumption is tested in two stages (Mendoza, et al., Note 2). For example, for the treatments X occasions design, to test the circularity hypothesis

$$H_0\text{:} \quad \underset{\sim}{C}'\underset{\sim}{\Sigma}_1\underset{\sim}{C} = \underset{\sim}{C}'\underset{\sim}{\Sigma}_2\underset{\sim}{C} = ,\ldots,= \underset{\sim}{C}'\underset{\sim}{\Sigma}_J\underset{\sim}{C} = \sigma^2\underset{\sim}{I} , \tag{52}$$

the hypothesis

$$H_{01}\text{:} \quad \underset{\sim}{C}'\underset{\sim}{\Sigma}_1\underset{\sim}{C} = \underset{\sim}{C}'\underset{\sim}{\Sigma}_2\underset{\sim}{C} = ,\ldots, = \underset{\sim}{C}'\underset{\sim}{\Sigma}_J\underset{\sim}{C} = \underset{\sim}{\Sigma} \tag{53}$$

is first tested. If H_{01} is not rejected, the hypothesis

$$H_{02}\text{:} \quad \underset{\sim}{C}'\underset{\sim}{\Sigma}\underset{\sim}{C} = \sigma^2\underset{\sim}{I} \tag{54}$$

is then tested. If both H_{01} and H_{02} are not rejected, then H_0 is not rejected and the circularity assumption is assumed satisfied. However, if either H_{01} or H_{02} are rejected, then H_0 is rejected indicating a violation of the assumption of circularity.

To test the first requirement (H_{01}) that the variance-covariance matrix of the p standardized contrasts is constant for each level of the between subject factor J, Box's (1949) multivariate analogue to Bartlett's (1937) homogeneity of variance test can be employed (Huynh & Feldt, 1970). If $\underset{\sim}{S}_j$ is the unbiased sample variance-covariance matrix corresponding to group j and $\underset{\sim}{S}$ is the pooled sample variance-covariance matrix, then the value of Box's (1949) heteroscedasticity criterion, M, is given by

$$M = (N-J) \ln |\underset{\sim}{C}'\underset{\sim}{S}\underset{\sim}{C}| - (I-1) \sum_{j=1}^{J} \ln |\underset{\sim}{C}'\underset{\sim}{S}_j\underset{\sim}{C}| . \tag{55}$$

Under the null hypothesis, assuming multivariate normality, and when $I \geq K$, the statistic

$$X = M/b \, , \tag{56}$$

is approximately distributed as an F variable with ν_1 and ν_2 degrees of freedom where

$$\nu_1 = .5[(J-1)p(p+1)] \, , \tag{57}$$

$$\nu_2 = (\nu_1 + 2)/(\theta_2 - \theta_1^2) \, , \text{ and} \tag{58}$$

$$b = \nu_1/[1 - \theta_1 - (\nu_1/\nu_2)] \text{ with} \tag{59}$$

$$\theta_1 = \frac{(2p^2 + 3p - 1)(J+1)}{6(p+1)JS_1} \text{ and} \tag{60}$$

$$\theta_2 = [(p-1)(p+2)(J^2 + J + 1)]/6J^2S_1^2 \, .^{18} \tag{61}$$

If the Box (1949) test does not lead to rejection of H_{01}, Mauchly's (1940) sphericity criterion, as applied to the pooled variance-covariance matrix, may be used to test the second requirement given in (54). Thus, the value of the test statistic, X_2, is given by

$$X_2 = -S_2 d \ln W \tag{62}$$

where $S_2 = N-J$,

$$d = 1 - [(2p^2 + p + 2)/6pS_2] \, , \tag{63}$$

W is defined previously and X_2 is a central χ^2 variable with d.f. = $[p(p+1)/2] - 1$.

[18]The F approximation is presented here rather than the χ^2 approximation as, according to Box (1950), when $p \leq 6$ and $J \leq 6$, the χ^2 approximation is reasonable only when $I > 20$. However, when I is small and/or J or p are larger than six, a more precise approximation is obtained by using the F distribution.

Given these methods for testing the circularity assumption corresponding to a given comparison, Rouanet and Lepine (1970) have suggested that these tests be used as preliminary tests with their outcomes forming a base for a decision among the various potential data-analytic strategies. Thus, if tests for circularity suggest that $\underset{\sim}{\Sigma}$ does satisfy the circularity assumption for a given comparison, the F ratio of the conventional ANOVA can be assumed valid and provides the most powerful test of the null hypothesis. However, if the circularity hypothesis is rejected, the conventional F ratio is not valid and one should adopt an adjusted univariate technique or a multivariate technique to analyze the data.

However, according to Mendoza, et al. (Note 2) and Rouanet and Lepine (1970), there are several problems with such a strategy. First, when one is working in a situation where $I < K$, circularity tests are not applicable as their calculation is based on the elements of a singular sample variance-covariance matrix. Secondly, and of most importance, is the fact that very little is known about these tests. While Box's (1949) heteroscedasticity criterion has been shown to be highly sensitive to departures from multivariate normality (Hopkins & Clay, 1963; Korin, 1972; Olson, 1974), the robustness of Mauchly's (1940) sphericity criterion to such as assumption violation is not known. Moreover, the usefulness of circularity tests in choosing between subsequent analysis strategies has not been investigated. As Rouanet and Lepine (1970) have noted, preliminary tests may be too powerful detecting assumption departures of little consequence to the 'would-be' subsequent analysis or not powerful enough failing to detect violations of serious consequences. In fact, Mendoza, et al. (Note 2) suggest that if one suspects that the circularity assumption is not met, a technique such as the three-stage approach described by Greenhouse and Geisser (1959) might be used in lieu of a test for circularity to decide upon an analysis strategy.

The purpose of the proposed research, therefore, was to investigate the operating characteristics of tests for circularity, repeated measures mean equality tests and various sequential testing procedures where the chosen mean equality test is based on the outcome of the preliminary test for circularity. For all individual tests and testing sequences, empirical estimates of Type I error and power were determined when violating the assumptions of normality, circularity, and multivariate homoscedasticity ($\underset{\sim}{\Sigma}_j = \underset{\sim}{\Sigma}$, $j = 1,\ldots,J$). The four sequential testing strategies examined were: (a) using the conservative F test following a significant test for circularity or using the conventional F test following a nonsignificant test for circularity, (b) using the $\hat{e}$-adjusted F test following a significant test for circularity or using the conventional F test following a nonsignificant test for circularity, (c) using the $\tilde{e}$-adjusted F test following a significant test for circularity or using the conventional F test following a nonsignificant test for circularity, and (d) using a multivariate test following a significant test for circularity or using the conventional test following a nonsignificant test for circularity. The empirical estimates obtained from the sequential strategies were compared to the estimates obtained when each of the proposed test procedures was uniformly adopted. Finally, following the recommendation of Mendoza, et al. (Note 2), the empirical estimates of the uniformity adopted test procedures and the sequential testing strategies were compared to the Type I error rates and power of the three stage approach due to Greenhouse and Geisser (1959).

METHOD

Design

The experimental design simulated in the present research was the three factor mixed model design schematized in Table 1 and represented by the univariate linear model given in (1) or the multivariate linear model given in (25). To afford comparability with previous investigations (e.g., Collier, et al., 1967; Mendoza, et al., 1974; Noe, Note 1), the number of levels of the between subjects or treatment factor J was set equal to three and the number of levels of the within subjects or occasion factor K was set equal to four.

Data Generation

In order to obtain the empirical results, a computer program was written to generate a pseudorandom vector $\vec{X}'_{ij} = [X_{ij1}, X_{ij2}, \ldots, X_{ijK}]$ approximating a K-variate normal distribution with mean vector $\vec{\mu}'_j = [\mu_1, \mu_2, \ldots, \mu_K]$ and variance-covariance matrix $\underset{\sim}{\Sigma}_j$. According to Schuer and Stoller (1966), generation of such a vector first requires generation of a pseudorandom vector $\vec{Z}'_{ij} = [Z_{ij1}, Z_{ij2}, \ldots, Z_{ijK}]$ where each Z_{ijk} is approximately independently and normally distributed with mean zero and variance one (i.e., Z_{ijk} is a pseudorandom unit normal deviate). Upon determination of $\vec{Z}'_{ij}$, $\vec{X}'_{ij}$ can then be obtained from the relationship

$$\vec{X}_{ij} = \vec{\mu}_j + \underset{\sim}{L}\vec{Z}_{ij} , \tag{64}$$

where L is a lower triangular matrix satisfying the equality $\underset{\sim}{\Sigma}_j = \underset{\sim}{L}\underset{\sim}{L}'$.

This triangular decomposition of $\underset{\sim}{\Sigma}_j$ is known by various names such as the Cholesky factorization and the square root method (Harman, 1967).

Pseudorandom unit normal deviates were generated by the procedure due to Marsaglia, MacLaren and Bray (1964). $\vec{X}_{ij}$ was then obtained from the relationship stated in (64).

In order to assess the effects of non-normality, a pseudorandom vector $\vec{Y}'_{ij} = [Y_{ij1}, Y_{ij2}, \ldots, Y_{ijK}]$ approximating a K-variate skewed distribution with mean vector $\vec{\mu}'_j = [\mu_1, \mu_2, \ldots, \mu_K]$ and variance-covariance matrix $\underset{\sim}{\Sigma}_j$ was also generated according to a procedure similar to that outlined above. Again, to allow comparability with prior research (Mendoza, et al., 1974), the skewed distribution was derived from a χ^2 distribution with 3 d.f. and hence moments $\mu'_1 = 3$, $\mu_2 = \sigma^2 = 6$, $\mu_3 = 24$, $\mu_4 = 252$, a skewness measure $\gamma_1 = 1.63$ and a kurtosis measure $\gamma_2 = 4$. Using the relationship

$$\chi^2_{(n)} = \sum_{i=1}^{n} Z_i^2 , \tag{65}$$

the pseudorandom χ^2 variates were generated by summing the squares of three pseudorandom unit normal deviates. By subtracting three from each χ^2 variate and then multiplying each by $1/\sigma$ where $\sigma^2 = 6$, each χ^2 variate was standardized to mean zero and variance one. Finally, using the same procedure as was used for the normal distribution,

$$\vec{Y}_{ij} = \vec{\mu}_j + \underset{\sim}{L}\vec{Z}_{ij} , \tag{66}$$

where $\vec{Z}_{ij}$ is a 1 X K vector of standardized pseudorandom skewed variates and $\vec{\mu}_j$ and $\underset{\sim}{L}$ are as defined previously.

Table 6 presents a schematic of the six general conditions simulated for both the normal and skewed distributions in order to assess the

Table 6

Schematic of Simulated Conditions

- $\mathbf{\Sigma}_j = \mathbf{\Sigma},\ j = 1,\ldots,J$
  - $\mathbf{C}'\mathbf{\Sigma}_j\mathbf{C} = \mathbf{\Sigma},\ j = 1,\ldots,J$
    - $\mathbf{C}'\mathbf{\Sigma}\mathbf{C} = \sigma^2\mathbf{I}$ (1)
    - $\mathbf{C}'\mathbf{\Sigma}\mathbf{C} \neq \sigma^2\mathbf{I}$ (2)
  - Not Possible
- $\mathbf{\Sigma}_j \neq \mathbf{\Sigma},\ j = 1,\ldots,J$
  - $\mathbf{C}'\mathbf{\Sigma}_j\mathbf{C} = \mathbf{\Sigma},\ j = 1,\ldots,J$
    - $\mathbf{C}'\mathbf{\Sigma}\mathbf{C} = \sigma^2\mathbf{I}$ (3)
    - $\mathbf{C}'\mathbf{\Sigma}\mathbf{C} \neq \sigma^2\mathbf{I}$ (4)
  - $\mathbf{C}'\mathbf{\Sigma}_j\mathbf{C} \neq \mathbf{\Sigma},\ j = 1,\ldots,J$
    - $\mathbf{C}'\mathbf{\Sigma}\mathbf{C} = \sigma^2\mathbf{I}$ (5)
    - $\mathbf{C}'\mathbf{\Sigma}\mathbf{C} \neq \sigma^2\mathbf{I}$ (6)

single and combined effects of violation of the assumptions of multivariate homoscedasticity and circularity. Note that the conditions presented in Table 6 do not represent a completely crossed design as it is impossible for the assumption of $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$, $j = 1,\ldots,J$ to be violated when the assumption of multivariate homoscedasticity is satisfied.

A modified form of Box's (1949) heteroscedasticity criterion was used to quantify the degree of departure from the assumption of multivariate homoscedasticity. According to Box (1949), when sample sizes are equal, the probability that a set of J K X K sample variance-covariance matrices, $\underset{\sim}{S}_1, \underset{\sim}{S}_2,\ldots,\underset{\sim}{S}_J$, originate from a common population variance-covariance matrix can be obtained from the expression

$$M_1 = (N-J)\ \ln|\underset{\sim}{S}| - (I-1)\sum_{j=1}^{J} \ln|\underset{\sim}{S}_j| \quad \text{or}$$
$$M_1 = (I-1)(J \cdot \ln|\underset{\sim}{S}| - \sum_{j=1}^{J} \ln|\underset{\sim}{S}_j|)\ , \tag{67}$$

with $\underset{\sim}{S}_j$ and $\underset{\sim}{S}$ as defined previously. For a given sample size, when the $\underset{\sim}{S}_j$'s originate from a common population matrix the value of M_1 is zero; however, as the value of M_1 increases from zero, the probability that the set of $\underset{\sim}{S}_j$'s shares a common origin decreases. Similarly, if population matrices are used in (67), the magnitude of the resulting quantity will reflect the degree of population variance-covariance matrix inequality. Moreover, since parameter values are unaffected by sample size, the quantity (I-1) can be disregarded and (67) can be redefined as

$$M_1 = J \cdot \ln|\underset{\sim}{\Sigma}| - \sum_{j=1}^{J} \ln|\underset{\sim}{\Sigma}_j| \ . \tag{68}$$

An additional modification to (68) provided a means of quantifying the assumption that $\mathbf{C}'\mathbf{\Sigma}_j\mathbf{C} = \mathbf{\Sigma}$, $j = 1,\ldots,J$. That is, by substituting $|\mathbf{C}'\mathbf{\Sigma}\mathbf{C}|$ for $|\mathbf{\Sigma}|$ and $|\mathbf{C}'\mathbf{\Sigma}_j\mathbf{C}|$ for $|\mathbf{\Sigma}_j|$, one arrives at the expression

$$M_2 = J \cdot \ln |\mathbf{C}'\mathbf{\Sigma}\mathbf{C}| - \sum_{j=1}^{J} |\mathbf{C}'\mathbf{\Sigma}_j\mathbf{C}| \quad . \tag{69}$$

As with (69), when $\mathbf{C}'\mathbf{\Sigma}_j\mathbf{C} = \mathbf{\Sigma}$ for all j, the value of M_2 is zero; however, as the $\mathbf{C}'\mathbf{\Sigma}_j\mathbf{C}$'s depart from equality, the value of M_2 increases.

Finally, the degree of violation of the assumption that $\mathbf{C}'\mathbf{\Sigma}\mathbf{C} = \sigma^2\mathbf{I}$ was quantified by Box's (1954a) correction factor, e, given in (9). When $\mathbf{C}'\mathbf{\Sigma}\mathbf{C} = \sigma^2\mathbf{I}$, the value of e equals unity; when $\mathbf{C}'\mathbf{\Sigma}\mathbf{C} \neq \sigma^2\mathbf{I}$, e is less than one, depending on the degree of assumption violation.

Table 7 presents a summary of the fifteen sets of variance-covariance matrices and their associated M_1, M_2, and e values generated to depict the assumption violation conditions specified in Table 6. The exact elements of these matrices are presented in the Appendix. All matrices were chosen so that for each of the 15 matrix sets, the overall matrix, $\mathbf{\Sigma}$, (a) had an average variance, $\bar{\sigma}_{kk}$, equal to 10 and an average covariance, $\bar{\sigma}_{kk'}$, $(k \neq k')$, equal to 5 and (b) was characterized by diagonal elements whose magnitude decreased in successive diagonals from the main diagonal, a property of the simplex form common to psychological research investigations involving repeated measurements over age and/or time. In addition, following the recommendation of Huynh and Feldt (1970), for each matrix set, the values of the three individual group matrices and those of the overall matrix were chosen such that the resulting matrices were positive definite, that is, had K strictly positive eigenvalues (Tatsuoka, 1971).

Table 7

Summary of the Matrix Sets Associated With the Simulated Conditions

Condition (1):	$\boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}$, $j=1,\ldots,J$;	$\boldsymbol{C}'\boldsymbol{\Sigma}_j\boldsymbol{C} = \boldsymbol{\Sigma}$, $j=1,\ldots,J$;	$\boldsymbol{C}'\boldsymbol{\Sigma}\boldsymbol{C} = \sigma^2\boldsymbol{I}$
Matrix Set 1:	$M_1 = 0.0$	$M_2 = 0.0$	$e = 1.00$
Condition (2):	$\boldsymbol{\Sigma}_j = \boldsymbol{\Sigma}$, $j=1,\ldots,J$;	$\boldsymbol{C}'\boldsymbol{\Sigma}_j\boldsymbol{C} = \boldsymbol{\Sigma}$, $j=1,\ldots,J$;	$\boldsymbol{C}'\boldsymbol{\Sigma}\boldsymbol{C} \neq \sigma^2\boldsymbol{I}$
Matrix Set 2:	$M_1 = 0.0$	$M_2 = 0.0$	$e = .96$
Matrix Set 3:	$M_1 = 0.0$	$M_2 = 0.0$	$e = .75$
Matrix Set 4:	$M_1 = 0.0$	$M_2 = 0.0$	$e = .57$
Matrix Set 5:	$M_1 = 0.0$	$M_2 = 0.0$	$e = .48$
Condition (3):	$\boldsymbol{\Sigma}_j \neq \boldsymbol{\Sigma}$, $j=1,\ldots,J$;	$\boldsymbol{C}'\boldsymbol{\Sigma}_j\boldsymbol{C} = \boldsymbol{\Sigma}$, $j=1,\ldots,J$;	$\boldsymbol{C}'\boldsymbol{\Sigma}\boldsymbol{C} = \sigma^2\boldsymbol{I}$
Matrix Set 6:	$M_1 = 2.20$	$M_2 = 0.0$	$e = 1.00$
Condition (4):	$\boldsymbol{\Sigma}_j \neq \boldsymbol{\Sigma}$, $j=1,\ldots,J$;	$\boldsymbol{C}'\boldsymbol{\Sigma}_j\boldsymbol{C} = \boldsymbol{\Sigma}$, $j=1,\ldots,J$;	$\boldsymbol{C}'\boldsymbol{\Sigma}\boldsymbol{C} \neq \sigma^2\boldsymbol{I}$
Matrix Set 7:	$M_1 = 1.64$	$M_2 = 0.0$	$e = .96$
Matrix Set 8:	$M_1 = 1.27$	$M_2 = 0.0$	$e = .75$
Matrix Set 9:	$M_1 = 1.54$	$M_2 = 0.0$	$e = .57$
Matrix Set 10:	$M_1 = 2.03$	$M_2 = 0.0$	$e = .48$
Condition (5):	$\boldsymbol{\Sigma}_j \neq \boldsymbol{\Sigma}$, $j=1,\ldots,J$;	$\boldsymbol{C}'\boldsymbol{\Sigma}_j\boldsymbol{C} \neq \boldsymbol{\Sigma}$, $j=1,\ldots,J$;	$\boldsymbol{C}'\boldsymbol{\Sigma}\boldsymbol{C} = \sigma^2\boldsymbol{I}$
Matrix Set 11:	$M_1 = 1.83$	$M_2 = 1.37$	$e = 1.00$
Condition (6):	$\boldsymbol{\Sigma}_j \neq \boldsymbol{\Sigma}$, $j=1,\ldots,J$;	$\boldsymbol{C}'\boldsymbol{\Sigma}_j\boldsymbol{C} \neq \boldsymbol{\Sigma}$, $j=1,\ldots,J$;	$\boldsymbol{C}'\boldsymbol{\Sigma}\boldsymbol{C} \neq \sigma^2\boldsymbol{I}$
Matrix Set 12:	$M_1 = 1.83$	$M_2 = 1.37$	$e = .96$
Matrix Set 13:	$M_1 = 1.83$	$M_2 = 1.37$	$e = .75$
Matrix Set 14:	$M_1 = 1.83$	$M_2 = 1.37$	$e = .57$
Matrix Set 15:	$M_1 = 1.83$	$M_2 = 1.37$	$e = .48$

As seen in Table 7, matrix sets resulting in five values of the parameter e were generated in order to investigate the assumption that $\tilde{C}'\tilde{\Sigma}\tilde{C} = \sigma^2\tilde{I}$. When $\tilde{C}'\tilde{\Sigma}\tilde{C} = \sigma^2\tilde{I}$, the chosen matrix sets yielded e values of unity; however, when $\tilde{C}'\tilde{\Sigma}\tilde{C} \neq \sigma^2\tilde{I}$, the values of e for the matrix sets were less than one, decreasing with increases in the degree of assumption violation. Because e can only range between .333 to 1.00 for the present design, values of e equal to .96, .75, .57, and .48 were chosen to represent extremely small, small, medium, and large degrees of assumption violation, respectively.

To investigate the assumption that $\tilde{C}'\tilde{\Sigma}_j\tilde{C} = \tilde{\Sigma}$, j = 1,...,J, matrix sets resulting in two values of the parameter M_2 were generated. When $\tilde{C}'\tilde{\Sigma}_j\tilde{C} = \tilde{\Sigma}$, j = 1,...,J, the generated matrix sets resulted in M_2 values equal to zero; when $\tilde{C}'\tilde{\Sigma}_j\tilde{C} \neq \tilde{\Sigma}$, j = 1,...,J, the associated M_2 value of the matrix sets equalled 1.37. Because, unlike the parameter e, the limits of M_2 are not known, this value of M_2 was arrived at by forming individual group matrices, $\tilde{\Sigma}_j$'s, in the ratio of 1:3:4 (a situation thought to be not uncommon in psychological research) and calculating the common associated M_2 value.

The restrictions imposed upon the matrices by the particulars of these assumption violation conditions (i.e., the desired e and M_2 values) combined with the matrix constraints previously enumerated (positive definite matrices with overall matrices displaying characteristics of the simplex form and where $\bar{\sigma}_{kk} = 10$ and $\bar{\sigma}_{kk'} = 5$) made generation of matrix sets possessing a constant nonzero value of M_1 impossible. As seen in Table 7, when the assumption that $\tilde{C}'\tilde{\Sigma}_j\tilde{C} = \tilde{\Sigma}$, j = 1,...,J, and the assumption of multivariate homoscedasticity were met, all matrix sets possessed a common M_1 value equal to zero.

Similarly, when $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} \neq \underset{\sim}{\Sigma}$, $j = 1,\ldots,J$, and the assumption of multivariate homoscedasticity was violated, all matrix sets had M_1 values equal to 1.83. However, such was not the case when $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$, $j = 1,\ldots,J$, and multivariate heteroscedasticity was present. For these conditions ((3) and (4)), matrix sets were chosen which possessed M_1 values as similar to one another as possible and which resulted in a range of M_1 values bracketing the value of M_1 in conditions (5) and (6) and the constant value of M_2. For conditions of multivariate heteroscedasticity, matrix sets were found such that the value of M_1 only ranged from 1.27 to 2.20 (see Table 7).

For both normal and skewed distributions and for all 15 matrix sets specified in Table 7, the probability of a Type I error was generated when the J population mean vectors were null vectors, that is, when $\vec{\mu}_j' = [0, 0,\ldots,0]$, $j = 1,\ldots,J$. Table 8 contains the values of the K-variate population mean vectors under which the empirical power estimates were generated. The values of these mean vectors were chosen to yield within subjects effects (i.e., K and JK) of medium magnitudes as defined by Cohen (1969).

According to Cohen (1969), for equal group sizes, the size of a given effect can be indexed by the square root of the sum of the squared effects of a given factor divided by that factors' d.f. + 1 in ratio to the common within population standard deviation. Thus, for the within subjects occasions effect K, the effect size index, denoted by f, is equal to

$$f_K = \frac{\left[\sum_{k=1}^{K} \beta_k^2/(K-1)+1\right]^{\frac{1}{2}}}{\sigma}, \tag{70}$$

Table 8

Values of the Population Mean Vectors for the Non-Null Distribution

	$\vec{\mu}_1'$	.20	1.15	2.20	3.00	1.6375
$\vec{\mu}_j'$	$\vec{\mu}_2'$	.40	1.00	1.50	2.25	1.2875
	$\vec{\mu}_3'$	.90	.85	.80	.75	.825
		.50	1.00	1.50	2.00	1.25

where β_k is the kth occasion effect and σ is the common within population standard deviation. Similarly, for the JK interaction effect,

$$f_{JK} = \frac{\left[\sum_{j=1}^{J} \sum_{k=1}^{K} \alpha\beta_{jk}^2 / [(J-1)(K-1)] + 1\right]^{\frac{1}{2}}}{\sigma} \tag{71}$$

where $\alpha\beta_{jk}$ is the jkth interaction effect and σ is as previously defined. As Cohen (1969) states, σ refers to the square root of the variance being estimated by the denominator mean square of the F test(s) of the effects in question. Therefore, for the present design, because both within subjects effects are tested over $MS_{KS(J)}$ (see Table 2), $\sigma = [\sigma^2(1-\rho)]^{\frac{1}{2}} = (\sigma^2 - \rho\sigma^2)^{\frac{1}{2}}$, where σ^2 and $\rho\sigma^2$ are the common population variance and covariance of the K repeated measures, respectively (Kirk, 1968, p. 256). Substituting an average variance value of 10 and an average covariance value of 5 and population mean vector values given in Table 8 into (70) and (71) resulted in values of f_K and f_{JK} equal to .25 and .27, respectively. According to Cohen (1969), values of f approximating .25 are indicative of medium-sized treatment effects in behavioral science research.

For all conditions investigated, the total number of observations in the simulated design was set equal to 39, with 13 observations being sampled from each of the J treatment populations. The value of I = 13 was chosen to yield a univariate F test of the hypothesis of no occasions effects with an a priori power of .75 to detect an occasion effect with means such as those given in Table 8 (f_K = .25). While Cohen (1969) proposes that the behavioral science researcher should adopt a power value of .80, in the present investigation a power value of .75 was chosen to minimize the possible occurrence of 'ceiling' effects under conditions of assumption violation, where the empirical power estimates of all test criteria attain their maximum value of 100% and thus preclude a comparison of the test procedures.

Statistics

For both population shapes and for all matrix sets specified in Table 7, empirical estimates of the probability of a Type I error and power were determined for the following test statistics and combinations of test statistics:

A. Test for Circularity

B. Tests for Mean Equality

(a) the conventional F test[19]

(b) the conservative F test

(c) the $\hat{e}$-adjusted F test

(d) the $\tilde{e}$-adjusted F test

(e) a multivariate test

C. Tests for Mean Equality following the Test for Circularity

(a) the conservative F test following a significant test for circularity and the conventional F test following a nonsignificant test for circularity

(b) the $\hat{e}$-adjusted F test following a significant test for circularity and the conventional F test following a nonsignificant test for circularity

[19]Following the recommendation of Gabriel (Note 3), precipitated by the finding of Collier, et al. (1967) that the empirical Type I error rates of the conventional F test rarely exceed twice the value of the nominal significance level even for extreme Σ nonuniformity, the conventional F test was not only tested at the nominal significance level set for the other test statistics, but at a significance level of one-half this size.

(c) the $\tilde{e}$-adjusted F test following a significant test for circularity and the conventional F test following a nonsignificant test for circularity

(d) a multivariate test following a significant test for circularity and the conventional F test following a nonsignificant test for circularity

D. Greenhouse and Geisser (1959) Three Stage Approach

(a) using the $\hat{e}$-adjusted F test as the final step

(b) using the $\tilde{e}$-adjusted F test as the final step

(c) using a multivariate test as the final step

Because little is known concerning the operating characteristics of the component tests (i.e., Box's (1949) heteroscedasticity criterion and Mauchly's (1940) sphericity criterion) of the test for circularity (Part A), Type I error rates and power characteristics of both these component tests were determined, in addition to those of the overall two stage test for circularity.

All test statistics enumerated in Parts B through D above were computed for both the within subjects effects in the simulated treatments X occasions design. In all cases, the multivariate test of the hypothesis of no occasions effect was tested through use of Hotelling's (1931) one sample T^2 statistic given in footnote 13. Evaluation of the hypothesis of no treatments X occasions interaction was based on three multivariate criteria. Following the recommendation of Olson (1976), the within subjects interaction hypothesis was tested by the Pillai-Bartlett (Bartlett, 1939; Pillai, 1955) trace criterion given in (37). Wilk's (1932) likelihood ratio criterion specified in (36) and Roy's

(1953) largest root criterion specified in (34) were chosen as the second and third test statistics because of their frequent use in currently popular packaged programs (e.g., Finn, 1976).

For each of the 30 conditions to be investigated (15 matrix sets X two population shapes), a total of 2,000 sets of data were generated. For the individual test statistics specified in Parts A and B above, if the observed value of the test statistic exceeded its critical value when the J population mean vectors were null, a Type I error was counted for the test statistic. On the other hand, if the observed value of the test statistic exceeded its critical value when the vectors of means were non-null, an incidence of power was recorded.

For the sequential strategies enumerated in Part C, empirical estimates of Type I error and power were determined by the following counting procedure. Because a sequential testing procedure leads the experimenter to perform only one mean equality test on the data from a given experiment, that being the test recommended by the test for circularity, for each test of mean equality, the number of times it was chosen by the test for circularity was initially counted. Further, of this number, for each mean equality test, the number of times it was significant was recorded. With this information, the empirical estimates of Type I error and power for a particular sequential strategy were determined. For example, consider the first sequential strategy outlined in Part C where a nonsignificant test for circularity is followed by the conventional F test and a significant circularity test is followed by the conservative F test. For a certain subset of the 2,000 data sets, the circularity test was significant and, of this subset, the number of significant conservative F tests was recorded. Similarly, for the complimentary subset of the 2,000 data sets where the circularity test was nonsignificant, the number of significant conventional F tests in this subset was recorded. By summing the number of times the conventional F test was significant when the circularity test was non-significant and the number of times the conservative F test was significant when the test for circularity was also significant and dividing this sum by the total number of generated data sets, the empirical Type I error (when the population mean vectors were null) and power (when the population mean vectors were non-null) rates for the sequential testing strategy were determined.

Type I error and power rates for the Greenhouse and Geisser (1959) three stage approaches of Part D were obtained by counting the number of times the 'stopping point' (see p. 12) in the sequence judged the effect in question to be significant. For example, consider the three stage approach using the $\hat{e}$-adjusted F test as the final step. If, for a given data set, the conventional F test proved nonsignificant, the testing sequence stopped and no incidence of Type I error or power was recorded. However, data sets resulting in significant conventional and conservative F tests or in a significant conventional F test followed by a nonsignificant conservative F test and a significant $\hat{e}$-adjusted F test were counted as incidences of Type I error or power.

For all test statistics and combinations of test statistics enumerated in Parts A through D, the nominal level of significance, α, was set equal to .05. In addition, for all individual test statistics or sequential strategies involving multivariate tests, the upper $100(1-\alpha)$ percentage points of Hotelling's (1931) T^2 distribution were obtained from the relationship

$$F = \frac{N - J - K + 2}{(N-J)(K-1)} T^2 \sim F_{K-1,N-J-K+2} \tag{71}$$

while the upper 100(1-α) percentage points of the Pillai-Bartlett (Bartlett, 1931; Pillai, 1955) trace criterion, V, the lower 100(1-α) percentage points of Wilk's (1932) likelihood ratio criterion, Λ, and the upper 100(1-α) percentage points of Roy's (1953) largest root criterion, θ, were obtained from tables generated by Pillai (1960), Wall (1968) and Heck (1960), respectively.

RESULTS AND DISCUSSION

Effects of Assumption Violations

Tests for Circularity

Table 9 presents the empirical Type I error and power rates for the component tests (i.e., Box's (1949) heteroscedasticity criterion and Mauchly's (1940) sphericity criterion) and the overall test for circularity when sampling from normal and non-normal populations. To distinguish between the two estimates, the rates of Type I error are enclosed in parentheses.

As seen from Table 9, violation of the assumption of multivariate normality results in inflated rates of Type I error for the component and overall circularity tests. The Type I error estimates of Box's (1949) heteroscedasticity criterion are approximately six times larger than those obtained when sampling from normal populations while the estimates for Mauchly's (1940) sphericity criterion and the overall test for circularity are three to four times as large as the normal distribution values.

When sampling from non-normal distributions, the empirical power rates of Box's (1949) test of the hypothesis H_0: $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$, $j = 1,\ldots,J$ are approximately 10 percentage points larger than the rates obtained from normal populations. For Mauchly's (1940) test of the hypothesis H_0: $\underset{\sim}{C}'\underset{\sim}{\Sigma}\underset{\sim}{C} = \sigma^2\underset{\sim}{I}$ and the overall test of the hypothesis H_0: $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \sigma^2\underset{\sim}{I}$, $j = 1,\ldots,J$, the increased sensitivity afforded by conditions of non-normality varies as a function of the degree of departure from the tests' respective H_0's. For extremely small departures ($\bar{e} = .96$), the power

Table 9

Empirical Type I Error and Power Rates for the Test for Circularity

Condition			Component Tests: BOX Distribution		Component Tests: MAUCHLY Distribution		Overall Test Distribution	
M_1	M_2	e	N(0, 1)	$\chi^2_{(3)}$	N(0, 1)	$\chi^2_{(3)}$	N(0, 1)	$\chi^2_{(3)}$
0.00	0.00	1.00	(.051)	(.300)	(.046)	(.198)	(.095)	(.433)
0.00	0.00	.96	(.048)	(.294)	17.10	33.82	21.20	52.95
0.00	0.00	.75	(.051)	(.274)	96.04	94.87	96.44	95.92
0.00	0.00	.57	(.054)	(.268)	100.00	99.97	100.00	99.97
0.00	0.00	.48	(.052)	(.279)	100.00	100.00	100.00	100.00
2.20	0.00	1.00	(.053)	(.290)	(.051)	(.203)	(.102)	(.423)
1.64	0.00	.96	(.050)	(.301)	15.37	32.32	19.82	52.12
1.27	0.00	.75	(.058)	(.290)	95.99	94.79	96.19	96.17
1.54	0.00	.57	(.054)	(.278)	100.00	99.92	100.00	99.92
2.03	0.00	.48	(.048)	(.287)	100.00	100.00	100.00	100.00
1.83	1.37	1.00	70.10	80.72	(.094)	(.276)	72.74	85.04
1.83	1.37	.96	68.82	80.34	23.22	41.72	76.14	87.37
1.83	1.37	.75	70.17	79.84	95.72	94.54	98.94	98.74
1.83	1.37	.57	71.44	81.32	100.00	99.92	100.00	99.94
1.83	1.37	.48	70.44	80.50	100.00	99.97	100.00	100.00

Note. Values are based on 4,000 simulations.

estimates for Mauchly's (1940) test are 17 to 19 percentage points greater than those obtained in the presence of multivariate normality. However, as the value of e decreases, the difference between the empirical estimates becomes negligible: for $e \leq .75$, the sensitivity of the sphericity criterion ranges from 95% to 100% for either population form.

When $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$ for all j ($M_2 = 0.0$) and the departure from $\underset{\sim}{C}'\underset{\sim}{\Sigma}\underset{\sim}{C} = \sigma^2\underset{\sim}{I}$ is extremely small ($e = .96$), the empirical power rates of the overall test for circularity obtained under conditions of non-normality are approximately 30 percentage points larger than the normal distribution values. However, when $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} \neq \underset{\sim}{\Sigma}$ for all j ($M_2 \neq 0.0$), this discrepancy decreases with the power estimates from non-normal distributions now being approximately 12 percentage points larger. Further, like the test for sphericity, as the degree of departure from $\underset{\sim}{C}'\underset{\sim}{\Sigma}\underset{\sim}{C} = \sigma^2\underset{\sim}{I}$ increases, the effect of non-normality on the empirical power estimates becomes minimal: for $e \leq .75$, the power of the overall circularity test ranges from 96% to 100% for both population shapes.

An unexpected finding is the apparent lack of robustness of Mauchly's (1940) sphericity criterion when $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} \neq \underset{\sim}{\Sigma}$ for all j. Under such conditions, the sphericity criterion yields inflated Type I error and power rates for both population shapes.

Tests for Mean Equality

Empirical Type I error rates: Within subjects main effect. The empirical Type I error rates for the tests of the occasions main effect when sampling from normal and non-normal populations are presented in Tables 10 and 11, respectively.

Table 10

Empirical Type I Error Rates for Tests of the Occasions Effect for Normal Populations

Condition			Test[a]					
M_1	M_2	e	$F_{conv.}$	$F_{cons.}$	$F_{\hat{e}}$	$F_{\tilde{e}}$	T^2	$F_{\alpha/2}$
0.00	0.00	1.00	.046	.006	.040	.045	.054	.025
0.00	0.00	.96	.050	.010	.042	.048	.046	.026
0.00	0.00	.75	.058	.016	.046	.052	.051	.040
0.00	0.00	.57	.074	.028	.044	.046	.044	.046
0.00	0.00	.48	.086	.036	.052	.056	.050	.061
2.20	0.00	1.00	.044	.008	.039	.044	.045	.026
1.64	0.00	.96	.051	.008	.044	.050	.046	.024
1.27	0.00	.75	.058	.010	.044	.050	.048	.034
1.54	0.00	.57	.070	.028	.046	.048	.044	.046
2.03	0.00	.48	.098	.039	.056	.060	.046	.066
1.83	1.37	1.00	.053	.012	.046	.052	.057	.028
1.83	1.37	.96	.051	.006	.044	.051	.055	.027
1.83	1.37	.75	.076	.022	.056	.064	.066	.046
1.83	1.37	.57	.082	.028	.053	.055	.055	.054
1.83	1.37	.48	.084	.038	.046	.050	.049	.054

[a] $F_{conv.}$ - conventional F test; $F_{cons.}$ - conservative F test; $F_{\hat{e}}$ - $\hat{e}$-adjusted F test; $F_{\tilde{e}}$ - $\tilde{e}$-adjusted F test; T^2 - Hotelling's (1931) T^2; $F_{\alpha/2}$ - conventional F test at $\alpha/2$ significance level.

Table 11

Empirical Type I Error Rates for Tests of the Occasions Effect for Chi-Square Populations

Condition			Test[a]					
M_1	M_2	e	$F_{conv.}$	$F_{cons.}$	$F_{\hat{e}}$	$F_{\tilde{e}}$	T^2	$F_{\alpha/2}$
0.00	0.00	1.00	.060	.010	.048	.056	.078	.030
0.00	0.00	.96	.052	.010	.042	.050	.065	.028
0.00	0.00	.75	.064	.018	.044	.052	.056	.039
0.00	0.00	.57	.092	.038	.060	.064	.064	.062
0.00	0.00	.48	.089	.038	.055	.058	.065	.062
2.20	0.00	1.00	.060	.010	.050	.056	.070	.031
1.64	0.00	.96	.056	.009	.050	.052	.063	.032
1.27	0.00	.75	.061	.018	.046	.054	.054	.037
1.54	0.00	.57	.076	.028	.048	.054	.065	.050
2.03	0.00	.48	.112	.050	.072	.075	.067	.080
1.83	1.37	1.00	.049	.010	.040	.046	.063	.024
1.83	1.37	.96	.052	.011	.042	.048	.062	.026
1.83	1.37	.75	.072	.016	.050	.058	.065	.048
1.83	1.37	.57	.092	.038	.066	.070	.066	.069
1.83	1.37	.48	.090	.034	.052	.056	.063	.064

[a]See Table 10.

As seen from the tables, violation of the assumption of multivariate normality results in an increase in the Type I error estimates for all test procedures, with this increase being slightly larger for the multivariate test (T^2). Further, the magnitude of this increase varies as a function of the degree of departure from the assumption that $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$ for all j. When this assumption is satisfied ($M_2 = 0.0$), the empirical estimates obtained when sampling from non-normal populations are typically one and two percentage points larger than their normal distribution counterparts for the univariate tests and the multivariate test, respectively. When $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} \neq \underset{\sim}{\Sigma}$ for all j ($M_2 \neq 0.0$), the effect of non-normality decreases: for the univariate procedures, the difference between the Type I error rates obtained from normal and non-normal populations is typically less than one percentage point, with the estimates from the non-normal distributions being slightly smaller for moderate to large values of e (e = 1.00, .96, .75); for T^2, the empirical rates obtained when sampling from non-normal populations are typically one percentage point larger for all values of e.

In general, violation of the assumption of multivariate homoscedasticity ($\underset{\sim}{\Sigma}_j = \underset{\sim}{\Sigma}$, j = 1,...,J) has a negligible effect on the Type I error estimates for all test procedures. For both population shapes, the rates obtained under conditions of multivariate homoscedasticity ($M_1 = 0.0$) and multivariate heteroscedasticity ($M_1 \neq 0.0$) are typically within one-half of a percentage point of one another. When sampling is from non-normal populations and the value of e is small (e = .57, .48), this difference is slightly larger for the univariate procedures; here, the estimates obtained when $\underset{\sim}{\Sigma}_j \neq \underset{\sim}{\Sigma}$ for all j are approximately one and one-half percentage points smaller and one and one-half percentage points larger than those obtained when $\underset{\sim}{\Sigma}_j = \underset{\sim}{\Sigma}$, j = 1,...,J for e = .57 and e = .48, respectively.

The effect of violating the assumption that $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$ for all j varies as a function of the population shape and/or the value of e. When the population is normal, for all values of e, the empirical rates of T^2 obtained under conditions where $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} \neq \underset{\sim}{\Sigma}$ for all j ($M_2 \neq 0.0$) are approximately one to one and one-half percentage points larger than those obtained when $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$ for all j ($M_2 = 0.0$). When e $\geq$ .57, the presence of $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} \neq \underset{\sim}{\Sigma}$ for all j has a similar effect on the estimates for the univariate procedures. However, when e = .48, the probability of a Type I error for the univariate procedures is approximately one percentage point less when $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} \neq \underset{\sim}{\Sigma}$ for all j.

For non-normal populations, violation of the assumption that $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$ for all j generally has an even smaller effect on the Type I error rates of all test procedures than that noted under conditions of multivariate normality. Only when e = .57 and e = .48 is the effect of $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} \neq \underset{\sim}{\Sigma}$ for all j larger with the estimates for the univariate tests being one and one-half to two percentage points greater than and one and one-half to two percentage points less than those obtained when $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$ for all j, respectively.

Violation of the assumption that $\underset{\sim}{C}'\underset{\sim}{\Sigma}\underset{\sim}{C} = \sigma^2\underset{\sim}{I}$ (e = 1.00) has a negligible effect on the rates of Type I error for the multivariate test. However, as the value of e decreases from one, the estimates for all univariate procedures increase with this increase varying as a function of the specific test procedure. The conventional ($F_{conv.}$ and $F_{\alpha/2}$) and conservative F tests are most sensitive to departures from $\underset{\sim}{C}'\underset{\sim}{\Sigma}\underset{\sim}{C} = \sigma^2\underset{\sim}{I}$ with the rates obtained under the largest degree of assumption violation (e = .48) being from two to four times larger than those obtained when $\underset{\sim}{C}'\underset{\sim}{\Sigma}\underset{\sim}{C} = \sigma^2\underset{\sim}{I}$. For the $\hat{e}$- and $\tilde{e}$-adjusted F tests, the increase in the Type

I error rates is slight with the estimates typically remaining within one and one-half percentage points of one another regardless of the degree of assumption violation.

Empirical Type I error rates: Within subjects interaction effect. Tables 12 and 13 contain the Type I error rates for the tests of the treatments X occasions interaction effect when sampling from normal and non-normal populations, respectively.

While the presence of non-normality results in slight increases in the empirical estimates for the tests of the occasions main effect, the effect of non-normality on the tests of the treatments X occasions interaction is to slightly decrease the probability of a Type I error. Like the tests of the main effect, this decrease varies as a function of the degree of departure from the assumption that $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$ for all j. When this assumption is met, for all test procedures, the rates obtained when sampling from non-normal populations are approximately one percentage point smaller than those obtained when sampling from normal populations. When $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} \neq \underset{\sim}{\Sigma}$ for all j, the effect of non-normality decreases with the empirical estimates obtained from both population forms typically being within one-half of a percentage point of one another.

In general, the effect of multivariate heteroscedasticity is similar to that noted for the tests of the within subjects main effect. For both population shapes, the rates obtained when $\underset{\sim}{\Sigma}_j = \underset{\sim}{\Sigma}$ for all j and those obtained under conditions where $\underset{\sim}{\Sigma}_j \neq \underset{\sim}{\Sigma}$ for all j are typically within one-half of a percentage point of one another. When sampling is from normal populations, this difference in the Type I error rates is slightly larger for the multivariate tests with the estimates obtained under conditions of multivariate heteroscedasticity

Table 12

Empirical Type I Error Rates for Tests of the Treatments X Occasions Effect for Normal Populations

Condition			Test[a,b]							
M_1	M_2	e	$F_{conv.}$	$F_{cons.}$	$F_{\hat{e}}$	$F_{\tilde{e}}$	θ	V	Λ	$F_{\alpha/2}$
0.00	0.00	1.00	.054	.008	.048	.053	.050	.044	.044	.027
0.00	0.00	.96	.052	.009	.045	.050	.042	.039	.039	.025
0.00	0.00	.75	.069	.016	.044	.053	.049	.049	.050	.041
0.00	0.00	.57	.092	.024	.053	.059	.044	.050	.048	.060
0.00	0.00	.48	.103	.028	.056	.062	.043	.046	.044	.073
2.20	0.00	1.00	.048	.006	.044	.048	.048	.048	.048	.026
1.64	0.00	.96	.057	.008	.050	.055	.054	.053	.050	.028
1.27	0.00	.75	.074	.011	.052	.060	.048	.049	.048	.040
1.54	0.00	.57	.081	.020	.045	.052	.038	.038	.036	.052
2.03	0.00	.48	.110	.041	.059	.066	.064	.060	.062	.080
1.83	1.37	1.00	.058	.007	.054	.058	.066	.057	.054	.036
1.83	1.37	.96	.056	.011	.048	.056	.062	.054	.055	.034
1.83	1.37	.75	.075	.018	.052	.064	.070	.056	.058	.048
1.83	1.37	.57	.090	.023	.050	.058	.056	.050	.050	.060
1.83	1.37	.48	.110	.041	.058	.064	.070	.062	.064	.076

[a]See Table 10.

[b]θ - Roy's (1953) largest root criterion; V - Pillai-Bartlett (Bartlett, 1937; Pillai, 1955) trace criterion; Λ - Wilk's (1932) likelihood ratio criterion.

Table 13

Empirical Type I Error Rates for Tests of the Treatments X Occasions Effect for Chi-Square Populations

Condition			Test[a]							
M_1	M_2	e	$F_{conv.}$	$F_{cons.}$	$F_{\hat{e}}$	$F_{\tilde{e}}$	θ	V	Λ	$F_{\alpha/2}$
0.00	0.00	1.00	.048	.007	.038	.047	.044	.046	.040	.022
0.00	0.00	.96	.048	.004	.039	.046	.046	.051	.046	.025
0.00	0.00	.75	.078	.015	.048	.056	.042	.047	.045	.042
0.00	0.00	.57	.088	.021	.046	.052	.038	.043	.038	.052
0.00	0.00	.48	.086	.029	.044	.049	.034	.042	.038	.057
2.20	0.00	1.00	.048	.004	.038	.046	.040	.036	.036	.022
1.64	0.00	.96	.058	.006	.044	.054	.045	.049	.049	.030
1.27	0.00	.75	.064	.015	.048	.053	.042	.046	.041	.042
1.54	0.00	.57	.086	.018	.045	.051	.044	.046	.044	.052
2.03	0.00	.48	.100	.024	.044	.050	.044	.044	.043	.066
1.83	1.37	1.00	.050	.006	.039	.046	.054	.046	.048	.024
1.83	1.37	.96	.068	.010	.056	.064	.062	.060	.058	.040
1.83	1.37	.75	.072	.018	.048	.056	.063	.058	.056	.044
1.83	1.37	.57	.090	.028	.060	.064	.054	.050	.050	.062
1.83	1.37	.48	.106	.038	.056	.062	.063	.058	.058	.075

[a]See Table 12.

remaining within one to one and one-half percentage points of the estimates obtained when $\underset{\sim}{\Sigma}_j = \underset{\sim}{\Sigma}$ for all j.

For all test procedures, violating the assumption that $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$ for all j results in an increase in the probability of a Type I error with this increase being greatest for the largest root criterion, θ. When sampling is from normal populations, the empirical estimates obtained when $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$ for all j are within one percentage point of one another for all tests but θ. The effect of $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} \neq \underset{\sim}{\Sigma}$ for all j on these test procedures is slightly larger for non-normal populations; here, the rates obtained when $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} \neq \underset{\sim}{\Sigma}$ for all j are approximately one to one and one-half percentage points larger than those obtained when $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$ for all j. For both population shapes, the Type I error estimates for the largest root criterion are one and one-half to two percentage points greater when $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} \neq \underset{\sim}{\Sigma}$ for all j than when $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$ for all j.

Finally, violation of the assumption that $\underset{\sim}{C}'\underset{\sim}{\Sigma}\underset{\sim}{C} = \sigma^2\underset{\sim}{I}$ (e = 1.00) yields effects similar to those noted for the tests of the within subjects main effect. Thus, as the value of e decreases from one, the empirical estimates of the univariate procedures increase with this increase being largest for the conventional ($F_{conv.}$ and $F_{\alpha/2}$) and conservative F tests and smallest for the $\hat{e}$- and $\tilde{e}$-adjusted F tests. Similarly, departures from the assumption that $\underset{\sim}{C}'\underset{\sim}{\Sigma}\underset{\sim}{C} = \sigma^2\underset{\sim}{I}$ has a negligible effect on all three multivariate test criteria.

<u>Empirical power rates: Within subjects main effect</u>. The empirical power rates for the tests of the occasions main effect when sampling from normal and non-normal populations are presented in Tables 14 and 15, respectively.

Table 14

Empirical Power Rates for Tests of the Occasions Effect for Normal Populations

Condition			Test[a]					
M_1	M_2	e	$F_{conv.}$[b]	$F_{cons.}$	$F_{\hat{e}}$	$F_{\tilde{e}}$	T^2	$F_{\alpha/2}$
0.00	0.00	1.00	72.84	45.80	71.44	72.84	67.60	60.75
0.00	0.00	.96	71.24	45.95	69.30	70.74	66.00	62.00
0.00	0.00	.75	66.50	44.50	61.45	63.00	60.05	58.05
0.00	0.00	.57	62.65	42.00	52.00	53.75	65.40	53.70
0.00	0.00	.48	58.40	41.35	47.15	48.35	56.00	51.05
2.20	0.00	1.00	72.40	47.25	71.14	72.20	69.70	63.10
1.64	0.00	.96	71.14	46.85	69.34	70.74	66.20	62.15
1.27	0.00	.75	64.90	42.60	59.20	61.20	58.50	56.15
1.54	0.00	.57	63.20	42.35	52.90	54.85	66.85	54.65
2.03	0.00	.48	61.15	43.75	50.35	51.70	58.35	54.35
1.83	1.37	1.00	72.20	46.55	70.74	71.84	69.64	62.15
1.83	1.37	.96	72.44	45.85	70.04	72.14	68.24	62.50
1.83	1.37	.75	65.60	45.65	60.45	61.95	60.85	57.60
1.83	1.37	.57	62.45	42.75	53.40	55.05	66.40	54.65
1.83	1.37	.48	59.50	43.15	48.35	49.85	56.35	52.25

[a] See Table 10.

[b] Theoretical power rate approximately equals 75.00.

Table 15

Empirical Power Rates for Tests of the Occasions Effect for Chi-Square Populations

Condition			Test[a]					
M_1	M_2	e	$F_{conv.}$	$F_{cons.}$	$F_{\hat{e}}$	$F_{\tilde{e}}$	T^2	$F_{\alpha/2}$
0.00	0.00	1.00	72.04	49.40	70.20	72.80	70.00	62.90
0.00	0.00	.96	70.64	48.25	68.24	70.00	67.34	62.05
0.00	0.00	.75	65.45	47.20	60.70	62.00	61.80	58.50
0.00	0.00	.57	62.15	46.95	55.45	56.45	71.25	56.55
0.00	0.00	.48	60.25	45.75	50.55	52.35	59.55	54.45
2.20	0.00	1.00	70.90	48.25	68.80	70.80	69.30	61.50
1.64	0.00	.96	71.20	48.05	68.14	70.50	68.64	61.40
1.27	0.00	.75	66.35	46.75	60.70	62.40	61.60	58.35
1.54	0.00	.57	62.95	46.10	54.90	56.25	71.00	55.75
2.03	0.00	.48	60.35	45.55	51.10	52.10	61.95	54.40
1.83	1.37	1.00	72.94	50.20	71.14	72.60	71.24	64.25
1.83	1.37	.96	72.30	49.55	68.80	71.34	70.20	62.50
1.83	1.37	.75	68.00	47.50	62.60	64.25	64.40	59.95
1.83	1.37	.57	62.75	44.95	53.25	54.70	71.20	54.90
1.83	1.37	.48	59.30	44.45	49.90	50.80	59.00	52.80

[a] See Table 10.

As seen from the tables, violation of the assumption of multivariate normality results in slight increases in the sensitivity of all tests procedures. When sampling is from non-normal populations, the power rates of the conservative F test and the multivariate test are approximately two to three percentage points larger than those rates obtained under conditions of normality. For the remaining test procedures, the effect of non-normality is minimal with the empirical estimates remaining within one to two percentage points of one another.

The presence of multivariate heteroscedasticity has a negligible effect on the sensitivity of all tests. In general, the power rates obtained when $\underset{\sim}{\Sigma}_j \neq \underset{\sim}{\Sigma}$ ($M_1 = 0.0$) for all j are one and one-half percentage points greater than and one and one-half percentage points less than those obtained when $\underset{\sim}{\Sigma}_j = \underset{\sim}{\Sigma}$ ($M_1 = 0.0$) for all j for normal and non-normal populations, respectively.

Similarly, for all criteria, the effect of violating the assumption that $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$ for all j is minimal. When the population shape is normal, the estimates remain within one percentage point of one another; for non-normal populations, the rates rarely differ by more than two percentage points.

Finally, as the degree of departure from the assumption that $\underset{\sim}{C}'\underset{\sim}{\Sigma}\underset{\sim}{C} = \sigma^2\underset{\sim}{I}$ (e = 1.00) increases, the power of all test procedures decreases. The $\hat{e}$- and $\tilde{e}$-adjusted F tests are most sensitive to departures from the assumption that $\underset{\sim}{C}'\underset{\sim}{\Sigma}\underset{\sim}{C} = \sigma^2\underset{\sim}{I}$, followed by the conventional F test and the multivariate test. Least affected are the conventional ($F_{\alpha/2}$) and conservative F tests where the difference in the power rates when e = 1.00 and $\underset{-}{e}$ = .48 is approximately four and nine percentage points, respectively.

Empirical power rates: Within subjects interaction effect. Tables 16 and 17 contain the empirical power rates for the tests of the treatments X occasions interaction when sampling is from normal and non-normal populations, respectively.

Violation of the assumption of multivariate normality results in an increase in the sensitivity of all procedures with this increase being largest for the multivariate test criteria. Further, the magnitude of this increase varies as a function of the degree of departure from the assumption that $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$ for all j. When this assumption is satisfied ($M_2 = 0.0$), the empirical estimates obtained under conditions of non-normality are approximately one percentage point and three percentage points greater than those obtained under conditions of normality for the univariate and multivariate procedures, respectively. When $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} \neq \underset{\sim}{\Sigma}$ for all j, the effect of non-normality decreases with the power rates for all tests typically remaining within one percentage point of one another.

In general, the presence of multivariate heteroscedasticity has a negligible effect on the sensitivity of all tests. For normal populations, all estimates remain within one percentage point of one another; for non-normal populations, the effect of multivariate heteroscedasticity is slightly larger for the multivariate criteria with the power rates typically differing by less than two percentage points.

For all procedures, the effect of violating the assumption that $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$ for all j varies as a function of the distribution form. When sampling is from normal populations, the presence of $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$ for all j has little effect on the power of the univariate tests and only slightly increases the sensitivity of the multivariate tests. When the

Table 16

Empirical Power Rates for Tests of the Treatments X Occasions Effect for Normal Populations

Condition			Test[a]							
M_1	M_2	e	$F_{conv.}$[b]	$F_{cons.}$	$F_{\hat{e}}$	$F_{\tilde{e}}$	θ	V	Λ	$F_{\alpha/2}$
0.00	0.00	1.00	41.60	12.75	39.45	41.45	36.90	36.00	36.35	29.90
0.00	0.00	.96	42.40	15.00	39.60	41.85	35.95	34.35	34.55	31.75
0.00	0.00	.75	41.10	18.50	34.05	36.50	29.05	30.25	30.10	31.55
0.00	0.00	.57	39.30	19.95	28.75	30.40	32.85	33.25	33.45	31.40
0.00	0.00	.48	39.10	22.05	27.70	28.60	27.95	28.00	28.15	32.05
2.20	0.00	1.00	42.20	15.25	40.15	42.00	38.05	36.95	37.20	31.80
1.64	0.00	.96	40.95	14.80	38.45	40.50	35.10	33.90	34.05	30.25
1.27	0.00	.75	40.85	17.40	33.85	36.50	30.75	30.30	30.70	31.55
1.54	0.00	.57	40.00	19.95	29.75	31.70	34.55	33.25	33.50	32.40
2.03	0.00	.48	40.50	22.10	27.75	28.95	29.55	30.25	29.95	33.05
1.83	1.37	1.00	41.50	15.40	38.85	41.30	40.60	38.05	37.75	30.65
1.83	1.37	.96	41.75	16.25	38.85	41.15	38.55	36.30	36.70	31.95
1.83	1.37	.75	40.30	18.55	33.75	35.95	32.90	30.65	31.35	31.40
1.83	1.37	.57	39.25	18.90	29.35	30.80	35.25	34.25	33.90	31.95
1.83	1.37	.48	39.05	21.65	26.50	28.10	31.10	30.85	31.10	31.35

[a]See Tables 10 and 12.

[b]Theoretical power rate approximately equals 45.00.

Table 17

Empirical Power Rates for Tests of the Treatments X Occasions Effect for Chi-Square Populations

Condition			Test[a]							
M_1	M_2	e	$F_{conv.}$	$F_{cons.}$	$F_{\hat{e}}$	$F_{\tilde{e}}$	θ	V	Λ	$F_{\alpha/2}$
0.00	0.00	1.00	42.75	16.30	40.60	42.35	41.00	39.65	39.65	32.35
0.00	0.00	.96	43.70	15.90	39.30	42.80	39.10	38.35	38.50	33.25
0.00	0.00	.75	42.25	18.85	35.45	37.40	31.65	30.20	30.35	33.45
0.00	0.00	.57	40.85	19.95	29.35	31.35	33.90	34.00	34.15	31.95
0.00	0.00	.48	41.00	21.60	28.40	29.65	30.05	31.60	31.40	33.85
2.20	0.00	1.00	44.05	15.65	41.25	43.70	41.25	39.70	39.80	32.60
1.64	0.00	.96	41.65	15.75	37.40	40.55	37.90	35.80	36.30	30.85
1.27	0.00	.75	44.30	19.90	36.70	39.25	33.10	31.80	31.70	34.50
1.54	0.00	.57	42.50	20.95	30.60	32.85	38.40	37.10	37.25	33.55
2.03	0.00	.48	42.20	22.45	29.00	30.90	33.50	33.50	32.90	33.30
1.83	1.37	1.00	41.15	15.55	38.35	40.55	39.55	37.30	37.80	31.15
1.83	1.37	.96	41.50	15.65	36.65	39.60	37.70	34.85	35.50	30.40
1.83	1.37	.75	38.65	17.35	31.90	34.25	32.00	30.80	30.95	30.95
1.83	1.37	.57	40.50	18.95	28.80	30.75	38.40	37.15	37.35	32.00
1.83	1.37	.48	38.45	19.10	25.20	26.60	32.95	31.60	31.45	30.75

[a]See Tables 10 and 12.

population form is non-normal, the power of all tests decreases; for the univariate and multivariate criteria, the estimates obtained when $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} \neq \underset{\sim}{\Sigma}$ for all j are approximately two to three percentage points and one to two percentage points less than those obtained when $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$ for all j, respectively.

For both population shapes, the effect of violating the assumption that $\underset{\sim}{C}'\underset{\sim}{\Sigma}\underset{\sim}{C} = \sigma^2\underset{\sim}{I}$ (e = 1.00) varies as a function of the test criterion. With the exception of the conservative and conventional ($F_{\alpha/2}$) F tests, as the value of e decreases, the power of the tests decreases. For the conservative and conventional ($F_{\alpha/2}$) tests, decreases in the value of e results in increases in the empirical power rates. The $\hat{e}$- and $\tilde{e}$-adjusted procedures are most sensitive to departures from $\underset{\sim}{C}'\underset{\sim}{\Sigma}\underset{\sim}{C} = \sigma^2\underset{\sim}{I}$ followed by the conservative and multivariate tests and, finally, the conventional ($F_{conv.}$ and $F_{\alpha/2}$) F tests.

Sequential Testing Strategies

The empirical results for the sequential testing strategies are presented in Tables 18 through 23. For all tables, asterisked values denote conditions where the estimates of a given sequential procedure and those of the corresponding uniformly adopted mean equality test are equal.

Empirical Type I error rates. Table 18 contains the empirical Type I error rates for the sequential tests of the occasions main effect when sampling from normal and non-normal populations. The Type I error estimates of the sequential tests of the treatments X occasions interaction for normal and non-normal distributions are presented in Tables 19 and 20, respectively.

Table 18

Empirical Type I Error Rates for Sequential Tests of the Occasions Effect

Condition			Test[a]							
			$SQ_{cons.}$		$SQ_{\hat{e}}$		$SQ_{\tilde{e}}$		SQ_{T^2}	
			Distribution		Distribution		Distribution		Distribution	
M_1	M_2	e	N(0, 1)	$\chi^2_{(3)}$	N(0, 1)	$\chi^2_{(3)}$	N(0, 1)	$\chi^2_{(3)}$	N(0, 1)	$\chi^2_{(3)}$
0.00	0.00	1.00	.044	.035	.046	.051	.046	.056*	.052	.073
0.00	0.00	.96	.041	.025	.046	.044	.048*	.050*	.051	.067
0.00	0.00	.75	.018	.020	.046*	.044*	.052*	.052*	.052	.056*
0.00	0.00	.57	.028*	.038*	.044*	.060*	.046*	.064*	.044*	.064*
0.00	0.00	.48	.036*	.038*	.052*	.055*	.056*	.058*	.050*	.065*
2.20	0.00	1.00	.040	.034	.044	.052	.044*	.056*	.046	.072
1.64	0.00	.96	.044	.026	.050	.050	.050*	.052*	.051	.062
1.27	0.00	.75	.012	.023	.044*	.047	.050*	.054*	.048*	.056
1.54	0.00	.57	.028*	.028*	.046*	.048*	.048*	.054*	.044*	.066
2.03	0.00	.48	.039*	.050*	.056*	.072*	.060*	.075*	.046*	.067*
1.83	1.37	1.00	.023	.014	.048	.041	.052*	.046*	.056	.064
1.83	1.37	.96	.014	.017	.045	.042*	.051*	.048*	.056	.060
1.83	1.37	.75	.023	.016*	.056*	.050*	.064*	.058*	.066*	.066
1.83	1.37	.57	.028*	.038*	.053*	.066*	.055*	.070*	.055*	.066*
1.83	1.37	.48	.038*	.034*	.046*	.052*	.050*	.056*	.049*	.063*

[a] $SQ_{cons.}$, $SQ_{\hat{e}}$, $SQ_{\tilde{e}}$, and SQ_{T^2} - sequential tests where nonsignificant circularity tests are followed by the conventional F test and significant circularity tests are followed by the the conservative, $\hat{e}$-adjusted F, $\tilde{e}$-adjusted F and T^2 test, respectively.

Table 19

Empirical Type I Error Rates for Sequential Tests of the Treatments X Occasions Effect for Normal Populations

Condition			Test[a,b]					
M_1	M_2	e	$SQ_{cons.}$	$SQ_{\hat{e}}$	$SQ_{\tilde{e}}$	SQ_{θ}	SQ_V	SQ_{Λ}
0.00	0.00	1.00	.048	.053	.054	.058	.056	.056
0.00	0.00	.96	.041	.048	.050*	.053	.052	.052
0.00	0.00	.75	.018	.045	.053*	.050	.050	.052
0.00	0.00	.57	.024*	.053*	.059*	.044*	.050*	.048*
0.00	0.00	.48	.028*	.056*	.062*	.043*	.046*	.044*
2.20	0.00	1.00	.046	.048	.048*	.053	.051	.051
1.64	0.00	.96	.050	.055	.055*	.062	.060	.061
1.27	0.00	.75	.016	.052*	.060*	.052	.054	.052
1.54	0.00	.57	.020*	.045*	.052*	.038*	.038*	.036*
2.03	0.00	.48	.041*	.059*	.066*	.064*	.060*	.062*
1.83	1.37	1.00	.022	.055	.058*	.062	.058	.056
1.83	1.37	.96	.024	.050	.056*	.064	.056	.058
1.83	1.37	.75	.018*	.052*	.064*	.070*	.056*	.058*
1.83	1.37	.57	.023*	.050*	.058*	.056*	.050*	.050*
1.83	1.37	.48	.041*	.058*	.064*	.070*	.062*	.064*

[a] See Table 18.

[b] SQ_{θ}, SQ_V, and SQ_{Λ} - sequential tests where nonsignificant circularity tests are followed by conventional F tests and significant circularity tests are followed by θ, V, and Λ, respectively.

Table 20

Empirical Type I Error Rates for Sequential Tests of the Treatments X Occasions Effect for Chi-Square Populations

Condition			Test[a]					
M_1	M_2	e	$\underset{\sim}{SQ}_{cons.}$	$\underset{\sim}{SQ}_{\hat{e}}$	$\underset{\sim}{SQ}_{\tilde{e}}$	$\underset{\sim}{SQ}_{\theta}$	$\underset{\sim}{SQ}_{V}$	$\underset{\sim}{SQ}_{\Lambda}$
0.00	0.00	1.00	.026	.042	.048	.059	.052	.048
0.00	0.00	.96	.020	.042	.046*	.050	.051	.048
0.00	0.00	.75	.018	.050	.056*	.044	.048	.046
0.00	0.00	.57	.021*	.046*	.052*	.038*	.043*	.038*
0.00	0.00	.48	.029*	.044*	.049*	.034*	.042*	.038*
2.20	0.00	1.00	.022	.042	.048	.042	.042	.040
1.64	0.00	.96	.024	.048	.054*	.052	.054	.054
1.27	0.00	.75	.019	.048*	.053*	.046	.048	.044
1.54	0.00	.57	.018*	.045*	.051*	.044*	.046*	.044*
2.03	0.00	.48	.024*	.044*	.050*	.044*	.044*	.043*
1.83	1.37	1.00	.012	.040	.046*	.056	.049	.051
1.83	1.37	.96	.020	.058	.064*	.066	.062	.061
1.83	1.37	.75	.019	.048*	.056*	.064	.060	.057
1.83	1.37	.57	.028*	.060*	.064*	.054*	.050*	.050*
1.83	1.37	.48	.038*	.056*	.062*	.063*	.058*	.058*

[a] See Table 19.

For both effects and for all sequential procedures excluding the strategy where significant circularity tests are followed by conservative F tests ($SQ_{cons.}$), joint and/or separate violation of any of the investigated assumptions produces the same effects as those reported for the corresponding uniformly adopted tests for mean equality. When $e \leq .75$, the estimates of the $SQ_{cons.}$ test of both within subjects effects are affected by assumption violation(s) in a manner similar to that of the uniformly adopted conservative F test; however, when $e > .75$, the sequential and uniformly adopted tests respond differently to certain assumption violation(s).

For example, violation of the assumption of multivariate normality results in a decrease in the empirical estimates with the magnitude of this decrease varying as a function of the degree of departure from the assumption that $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$ for all j. When this assumption is satisfied, the estimates obtained from non-normal populations are approximately two percentage points less than those obtained from normal populations; when $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} \neq \underset{\sim}{\Sigma}$ for all j, the difference in the estimates decreases with the rates typically remaining within one to one and one-half percentage points of one another.

In a similar fashion, the effect of violating the assumption that $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$ for all j is to decrease the Type I error estimates with the magnitude of this decrease varying as a function of population shape. For normal and non-normal populations respectively, the estimates obtained when $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} \neq \underset{\sim}{\Sigma}$ for all j are two to three and one to two percentage points less than those obtained when $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$ for all j.

Finally, while the Type I error rates of the uniformly adopted conservative test consistently increase as the degree of departure from the assumption that $\underset{\sim}{C}'\underset{\sim}{\Sigma}\underset{\sim}{C} = \sigma^2\underset{\sim}{I}$ increases, when $e > .75$, as the value of e decreases, the empirical estimates of the $SQ_{cons.}$ test decrease.

Empirical power rates. Table 21 contains the empirical power rates for the sequential tests of the occasions main effect when sampling from normal and non-normal populations. The power estimates for the sequential tests of the treatments X occasions interaction for normal and non-normal populations are presented in Tables 22 and 23, respectively.

Like the Type I error results, only for the $SQ_{cons.}$ strategy and only when the departure from $\underset{\sim}{C}'\underset{\sim}{\Sigma}\underset{\sim}{C} = \sigma^2\underset{\sim}{I}$ is null or extremely small ($e = 1.00$, $.96$) do the effects of certain assumption violations differ from those effects reported for the corresponding uniformly adopted mean equality test. For all other conditions, the power of all sequential procedures is affected by the various assumption violations in a manner similar to that reported for the corresponding test for mean equality.

When $e = 1.00$ or $e = .96$, the effect of non-normality is to decrease the sensitivity of the $SQ_{cons.}$ strategy with this decrease varying as a function of the degree of departure from the assumption that $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$ for all j. For both effects, when this assumption is satisfied, the estimates obtained from non-normal populations are approximately seven to 10 percentage points less than their normal distribution counterparts. When $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} \neq \underset{\sim}{\Sigma}$ for all j, the difference in the rates decreases with the estimates obtained from non-normal populations being approximately equal to and four percentage points less than those obtained from normal populations for the main and interaction effects, respectively.

Table 21

Empirical Power Rates for Sequential Tests of the Occasions Effect

Condition			Test[a]							
			$SQ_{cons.}$		$\hat{SQ}_e$		$\tilde{SQ}_e$		SQ_{T^2}	
			Distribution		Distribution		Distribution		Distribution	
M_1	M_2	e	N(0, 1)	$\chi^2_{(3)}$	N(0, 1)	$\chi^2_{(3)}$	N(0, 1)	$\chi^2_{(3)}$	N(0, 1)	$\chi^2_{(3)}$
0.00	0.00	1.00	70.65	62.70	72.70	71.35	72.85	71.90	72.50	72.00
0.00	0.00	.96	66.00	58.30	70.50	68.75	70.75	70.00*	70.75	69.75
0.00	0.00	.75	45.40	47.75	61.55	60.70*	63.00*	62.00*	60.55	61.90
0.00	0.00	.57	42.00*	46.95*	52.00*	55.45*	53.75*	56.45*	65.40*	71.25*
0.00	0.00	.48	41.35*	45.75*	47.15*	50.55*	48.35*	52.35*	56.00*	59.55*
2.20	0.00	1.00	70.15	60.05	72.30	69.70	72.35	70.80*	72.15	70.75
1.64	0.00	.96	65.80	58.25	70.55	69.15	70.75	70.50*	70.90	71.05
1.27	0.00	.75	43.40	47.35	59.30	60.85	61.25	62.40*	58.80	62.05
1.54	0.00	.57	42.35*	46.10*	52.90*	54.90*	54.85*	56.25*	66.85*	71.00*
2.03	0.00	.48	43.75*	45.55*	50.35*	51.10*	51.70*	52.10*	58.35*	61.95*
1.83	1.37	1.00	53.20	52.80	70.90	71.30	71.85	72.60*	70.35	71.45
1.83	1.37	.96	51.25	51.70	70.45	69.00	72.20	71.34*	69.05	71.00
1.83	1.37	.75	45.75	47.65	60.45*	62.60*	61.95*	64.25*	60.90	64.35
1.83	1.37	.57	42.75*	44.95*	53.40*	53.25*	55.05*	54.70*	66.40*	71.20*
1.83	1.37	.48	43.15*	44.45*	48.35*	49.90*	49.85*	50.80*	56.35*	59.00*

[a]See Table 18.

Table 22

Empirical Power Rates for Sequential Tests of the Treatments X Occasions Effect for Normal Populations

Condition			Test[a]					
M_1	M_2	e	$SQ_{cons.}$	$SQ_{\hat{e}}$	$SQ_{\tilde{e}}$	SQ_{θ}	SQ_V	SQ_{Λ}
0.00	0.00	1.00	38.70	41.30	41.50	41.65	41.65	41.70
0.00	0.00	.96	36.75	41.50	41.95	42.30	41.80	41.85
0.00	0.00	.75	19.35	34.10	36.50*	29.80	31.05	30.80
0.00	0.00	.57	19.95*	28.75*	30.40*	32.85*	33.25*	33.45*
0.00	0.00	.48	22.05*	27.70*	28.60*	27.95*	28.00*	28.15*
2.20	0.00	1.00	39.50	41.95	42.05	42.20	42.15	42.15
1.64	0.00	.96	35.30	40.05	40.55	40.90	40.45	40.70
1.27	0.00	.75	18.40	33.95	36.50*	31.15	30.95	31.30
1.54	0.00	.57	19.95*	29.75*	31.70*	34.55*	33.25*	33.50*
2.03	0.00	.48	22.10*	27.75*	28.95*	29.55*	30.25*	29.95*
1.83	1.37	1.00	23.45	39.50	41.30*	41.45	39.45	39.20
1.83	1.37	.96	22.65	39.45	41.15*	39.20	37.75	37.95
1.83	1.37	.75	18.85	33.80	35.95*	32.95	30.90	31.50
1.83	1.37	.57	18.90*	29.35*	30.80*	35.25*	34.25*	33.90*
1.83	1.37	.48	21.65*	26.50*	28.10*	31.10*	30.85*	31.10*

[a]See Table 19.

Table 23

Empirical Power Rates for Sequential Tests of the Treatments X Occasions Effect for Chi-Square Populations

Condition			Test[a]					
M_1	M_2	e	$SQ_{cons.}$	$SQ_{\hat{e}}$	$SQ_{\tilde{e}}$	SQ_{θ}	SQ_V	SQ_{Λ}
0.00	0.00	1.00	31.90	41.55	42.40	43.40	42.70	42.65
0.00	0.00	.96	30.05	41.10	42.95	42.60	42.35	42.30
0.00	0.00	.75	19.55	35.50	37.40*	32.15	30.75	30.80
0.00	0.00	.57	19.95*	29.35*	31.35*	33.90*	34.00*	34.15*
0.00	0.00	.48	21.60*	28.40*	29.65*	30.05*	31.60*	31.40*
2.20	0.00	1.00	32.35	42.45	43.70*	43.30	43.60	43.35
1.64	0.00	.96	28.65	38.95	40.65	41.35	40.10	40.50
1.27	0.00	.75	20.90	36.70*	39.25*	33.50	32.45	32.30
1.54	0.00	.57	20.95*	30.60*	32.85*	38.40*	37.10*	37.25*
2.03	0.00	.48	22.45*	29.00*	30.90*	33.50*	33.50*	32.90*
1.83	1.37	1.00	18.55	38.45	40.55*	39.85	37.85	38.40
1.83	1.37	.96	18.60	36.90	39.60*	37.85	35.95	36.45
1.83	1.37	.75	17.75	31.95	34.25*	32.15	30.95	31.10
1.83	1.37	.57	18.95*	28.80*	30.75*	38.40*	37.15*	37.35*
1.83	1.37	.48	19.10*	25.20*	26.60*	32.95*	31.60*	31.45*

[a]See Table 19.

Similarly, violating the assumption that $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$ for all j decreases the sensitivity of the $SQ_{cons.}$ test with the magnitude of this decrease varying as a function of the distribution form. For normal populations, the rates obtained when $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} \neq \underset{\sim}{\Sigma}$ for all j are approximately 15 percentage points less than those obtained when $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$ for all j. For non-normal populations, the magnitude of this decrease is less with the estimates obtained when $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} \neq \underset{\sim}{\Sigma}$ for all j being approximately seven and 12 percentage points smaller for the main and interaction effects, respectively.

Lastly, while small departures from the assumption that $\underset{\sim}{C}'\underset{\sim}{\Sigma}\underset{\sim}{C} = \sigma^2\underset{\sim}{I}$ increase the power of the uniformly adopted conservative F test of the interaction effect, the sensitivity of the $SQ_{cons.}$ test of this effect decreases as the departure from $\underset{\sim}{C}'\underset{\sim}{\Sigma}\underset{\sim}{C} = \sigma^2\underset{\sim}{I}$ increases.

Greenhouse and Geisser (1959) Three Stage Approaches

For all conditions investigated, when an adjusted univariate procedure (i.e., an $\hat{e}$- or $\tilde{e}$-adjusted F test) is calculated at the third stage of this approach, the Type I error and power rates for the tests of both within subjects effects are identical to the rates for the corresponding uniformly adopted adjusted univariate tests. Only when a multivariate test is used at the final step do the empirical estimates for the three stage approach and the corresponding uniformly adopted test differ. The Type I error and power rates for these approaches are presented in Tables 24 and 25, respectively.

In general, for both within subjects effects, joint and/or separate violation of any of the investigated assumptions affects the Type I error and power estimates of the 'multivariate' three stage approaches in the same manner as that reported for the corresponding uniformly

Table 24

Empirical Type I Error Rates for the Greenhouse and Geisser (1959) Three Stage Approach

Condition			Test[a]							
			GG_{T^2}		GG_{θ}		GG_V		GG_{Λ}	
			Distribution		Distribution		Distribution		Distribution	
M_1	M_2	e	N(0, 1)	$\chi^2_{(3)}$	N(0, 1)	$\chi^2_{(3)}$	N(0, 1)	$\chi^2_{(3)}$	N(0, 1)	$\chi^2_{(3)}$
0.00	0.00	1.00	.034	.046	.034	.024	.036	.031	.036	.026
0.00	0.00	.96	.036	.034	.026	.025	.030	.030	.030	.028
0.00	0.00	.75	.029	.035	.029	.027	.034	.032	.034	.031
0.00	0.00	.57	.034	.048	.036	.028	.038	.032	.038	.030
0.00	0.00	.48	.044	.046	.038	.036	.040	.037	.040	.036
2.20	0.00	1.00	.034	.044	.032	.022	.036	.028	.036	.026
1.64	0.00	.96	.032	.042	.034	.026	.038	.032	.036	.032
1.27	0.00	.75	.029	.034	.026	.028	.032	.032	.030	.030
1.54	0.00	.57	.036	.036	.029	.030	.028	.034	.028	.034
2.03	0.00	.48	.046	.056	.052	.033	.055	.035	.054	.034
1.83	1.37	1.00	.042	.033	.042	.030	.042	.028	.042	.028
1.83	1.37	.96	.036	.036	.034	.044	.036	.046	.036	.044
1.83	1.37	.75	.041	.036	.043	.036	.038	.037	.039	.036
1.83	1.37	.57	.041	.050	.036	.043	.038	.040	.038	.042
1.83	1.37	.48	.044	.044	.050	.046	.055	.046	.054	.046

[a] GG_{T^2}, GG_{θ}, GG_V, and GG_{Λ} - Greenhouse and Geisser (1959) procedure with T^2, θ, V, and Λ as third step, respectively.

Table 25

Empirical Power Rates for the Greenhouse and Geisser (1959) Three Stage Approach

Test[a]

Condition			GG_{T^2} Distribution		GG_{θ} Distribution		GG_{V} Distribution		GG_{Λ} Distribution	
M_1	M_2	e	N(0, 1)	$\chi^2_{(3)}$	N(0, 1)	$\chi^2_{(3)}$	N(0, 1)	$\chi^2_{(3)}$	N(0, 1)	$\chi^2_{(3)}$
0.00	0.00	1.00	65.75	66.85	32.65	33.95	33.00	34.90	33.20	34.70
0.00	0.00	.96	64.35	64.40	32.00	33.75	32.35	33.95	32.25	34.00
0.00	0.00	.75	56.90	57.60	26.35	28.25	27.75	28.35	27.55	28.35
0.00	0.00	.57	55.85	57.30	27.10	27.75	28.20	28.65	28.00	28.45
0.00	0.00	.48	50.15	53.25	26.90	28.00	27.40	29.30	27.30	29.00
2.20	0.00	1.00	66.95	65.80	33.15	35.30	34.40	35.50	34.60	35.65
1.64	0.00	.96	64.20	64.95	31.20	31.90	31.10	31.30	31.40	31.65
1.27	0.00	.75	54.90	56.40	27.05	29.65	27.20	30.25	27.35	29.95
1.54	0.00	.57	56.80	57.60	28.25	30.10	29.15	30.80	28.60	30.75
2.03	0.00	.48	52.90	53.95	28.55	29.60	29.15	30.65	29.00	30.20
1.83	1.37	1.00	67.15	67.60	34.25	32.75	34.40	32.75	33.95	32.80
1.83	1.37	.96	65.40	66.85	33.15	31.50	32.70	30.65	32.80	30.95
1.83	1.37	.75	56.80	59.85	28.65	26.85	28.30	27.05	28.65	27.05
1.83	1.37	.57	55.80	56.95	27.85	29.25	28.30	29.15	28.15	29.25
1.83	1.37	.48	51.45	51.90	27.50	26.40	27.45	26.55	27.60	26.50

[a]See Table 24.

adopted tests. The effect of non-normality on the three stage test of the occasions main effect (GG_{T^2}) is slightly less than that noted for the uniformly adopted test, T^2. Similarly, departures from the assumption that $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$ for all j have less effect on the three stage tests of the treatments X occasions interaction (GG_θ, GG_V, and GG_Λ) than on the uniformly adopted multivariate tests (θ, V, and Λ).

Comparison of the Analysis Strategies

Tests for Mean Equality

As seen from Tables 10 through 17, while the conventional F test is generally the most powerful test of either within subjects hypothesis, it fails to control the rate of Type I error for all values of the parameter e, becoming increasingly liberal as e decreases from one. In contrast, the conservative F test consistently maintains the Type I error rate at or below the nominal significance level. However, when compared to the other tests for mean equality, it is overly conservative and lacking in sensitivity.

Over all conditions, both the $\hat{e}$- and $\tilde{e}$-adjusted F tests ($F_{\hat{e}}$ and $F_{\tilde{e}}$) consistently maintain the rate of Type I error at or near the nominal value and have power rates comparable to the conventional F test. For large values of e (e = 1.00, .96), the $\hat{e}$-adjusted F test proves to be slightly conservative with the $\tilde{e}$-adjusted F test providing better Type I error control. However, as the value of e decreases, the $\tilde{e}$-adjusted F test becomes slightly liberal and the $\hat{e}$-adjusted F test affords better Type I error protection. With respect to sensitivity, for $e \geq .75$ and $e < .75$, the power rates of the conventional and $\hat{e}$- and $\tilde{e}$-adjusted F tests never differ by more than five percentage points and ten percentage points, respectively. In all cases, $F_{\tilde{e}}$ is slightly more sensitive than $F_{\hat{e}}$.

Like the conventional F test, the procedure of testing the usual F ratio at an $\alpha/2$ level of significance fails to control the rate of Type I error for all values of e, resulting in a liberal test statistic when $e < .75$. Moreover, while the empirical estimates are typically less than the nominal value for $e \geq .75$, under such conditions, the $F_{\alpha/2}$ test is more conservative and less sensitive than other mean equality tests (i.e., $F_{\hat{e}}$ and $F_{\tilde{e}}$).

When the population form is normal, the multivariate test of the within subjects main effect (T^2) generally affords Type I error protection comparable to $F_{\hat{e}}$ and $F_{\tilde{e}}$ and, when $e < .75$, yields a more powerful test statistic. However, when the population form is non-normal, while T^2 is again more powerful for $e < .75$, its rates of Type I error are consistently liberal, ranging from .056 to .078.

When $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$ for all j, all three multivariate tests of the within subjects interaction effect result in Type I error rates comparable to $F_{\hat{e}}$ and $F_{\tilde{e}}$ regardless of the population shape. However, when $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} \neq \underset{\sim}{\Sigma}$ for all j, only the trace and likelihood ratio criteria maintain this Type I error control; under such conditions, the largest root criterion proves to be liberal with empirical estimates ranging from .054 to .070. Like T^2, when $e \geq .75$ and when $e < .75$, the sensitivity of the three multivariate criteria is, on the average, five percentage points less than and five percentage points greater than $F_{\hat{e}}$ and $F_{\tilde{e}}$, respectively.

Sequential Testing Strategies

For both within subjects effects, the relationships between the empirical Type I error and power rates of the sequential testing strategies are the same as those noted between the empirical estimates of the uniformly adopted tests for mean equality (see Tables 18 through 23). Comparing each sequential procedure to the corresponding mean equality test (e.g., $SQ_{cons.}$ compared to $F_{cons.}$), one finds that when the procedures are not equal, the rates of Type I error and power for the sequential tests are generally larger. In these cases, the sequential strategies involving only univariate procedures (i.e., $SQ_{cons.}$, $SQ_{\hat{e}}$, and $SQ_{\tilde{e}}$) are not as conservative as the corresponding tests for mean equality and therefore offer better Type I error control. However, for the sequential procedures involving a multivariate test (SQ_{T^2}, SQ_{θ}, SQ_{V}, and SQ_{Λ}), the increased rates of Type I error result in liberal testing sequences.

With respect to sensitivity, only the $SQ_{cons.}$ test shows a marked increase in power when compared to the corresponding mean equality test. For both within subjects effects, the sensitivity of the $SQ_{cons.}$ strategy is approximately 20 to 25 percentage points and 10 to 15 percentage points greater than that of the conservative F test for normal and non-normal populations, respectively. For all other comparisons, the difference in the empirical power estimates is generally less than five percentage points.

Greenhouse and Geisser (1959) Three Stage Approaches

For both within subjects effects, the Type I error rates of the approaches using a multivariate test at the final stage are consistently less than the rates of the approaches using an $\hat{e}$- or $\tilde{e}$-adjusted F test at the third step. For all conditions investigated, these 'multivariate' three stage approaches consistently maintain the rate of Type I error at or, more frequently, below the nominal significance level with empirical estimates ranging from .022 to .055 (see Table 24). The relationship between the power rates of the 'univariate' and 'multivariate' three stage approaches is the same as that noted between the empirical estimates of the uniformly adopted tests for mean equality: when $e \geq .75$, the 'univariate' three stage approaches are consistently more powerful than the 'multivariate' three stage approaches; when $e < .75$, the approaches using a multivariate test provide the greater sensitivity (see Table 25).

Finally, on comparing the 'multivariate' three stage approaches to the corresponding uniformly adopted mean equality tests, one notes that, over all conditions, the Type I error and power rates of all approaches are consistently less than those of the corresponding uniformly adopted procedures.

Concluding Remarks

In general, the data from this investigation indicate that preliminary testing for circularity is not useful in choosing between subsequent analysis strategies. Because both the heteroscedasticity and sphericity criteria are highly sensitive to departures from multivariate normality and from their respective null hypotheses (a finding consonant with the results of Olson (1974)), the Type I error and power rates of the various sequential testing strategies differ little from those of the corresponding uniformly adopted mean equality tests.

Similarly, assessing the within subjects effects with a Greenhouse and Geisser (1959) three stage approach in lieu of a uniformly adopted test for mean equality cannot be recommended. Adopting an adjusted univariate procedure ($\hat{e}$- or $\tilde{e}$-adjusted F test) at the third step in this approach results in rates of Type I error and power identical to those of the uniformly adopted adjusted univariate procedures. When a multivariate procedure is used, the resulting test is more conservative and less powerful than the corresponding uniformly adopted procedure.

Of the six uniformly adopted procedures for testing mean equality, the $\hat{e}$-adjusted F, $\tilde{e}$-adjusted F and multivariate tests are the most robust statistics when subjected to conditions of assumption violation. For the within subjects main effect, the effect of non-normality is to slightly increase the Type I error and power rates of these procedures, with the increase in the Type I error estimates being largest for the multivariate statistic (T^2). For these tests of the within subjects interaction, departures from normality result in slight decreases and slight increases in the rates of Type I error and power, respectively.

For both effects, the presence of multivariate heteroscedasticity has a negligible effect on the operating characteristics of the adjusted univariate and multivariate criteria.[20] Violation of the assumption that $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$ for all j results in slight increases in the Type I error and power rates of these procedures with the increase in the Type I error estimates being largest for Roy's (1953) test of the interaction effect. Departures from the assumption that $\underset{\sim}{C}'\underset{\sim}{\Sigma}\underset{\sim}{C} = \sigma^2\underset{\sim}{I}$ (e = 1.00) have no effect on the Type I error rates of the multivariate procedures and cause only slight increases in the probability of a Type I error for the $\hat{e}$- and $\tilde{e}$-adjusted F tests. However, as the value of e decreases from one, the sensitivity of the adjusted univariate and multivariate procedures decreases with this decrease being largest for the adjusted univariate procedures.

The data from this investigation do not suggest a clear-cut rule for choosing between the $\hat{e}$- and $\tilde{e}$-adjusted univariate F tests and their

---

[20] While the multivariate model describing the treatments X occasions design assumes that the observation vectors, $\vec{X}'_{ij}$, are independently and normally distributed with mean vector $\vec{\mu}_j$ and arbitrary covariance matrix $\underset{\sim}{\Sigma}$ which is common to all J populations, exact multivariate tests of the within subjects hypotheses associated with this design can be achieved when the assumption of multivariate homoscedasticity is not satisfied (Timm, Note 4). Because the multivariate approach arrives at tests of these hypotheses by transforming the K original measurements into K-1 contrasts among the K measurements, the necessary condition for valid multivariate tests is that the variance-covariance matrix of the K-1 contrasts among the levels of the repeated factor be common to all J populations, i.e., $\underset{\sim}{C}'\underset{\sim}{\Sigma}_j\underset{\sim}{C} = \underset{\sim}{\Sigma}$, $j = 1,\ldots,J$. Note, however, that this assumption is specific to multivariate tests of the within subjects effects; for other hypotheses in this design (see Timm, 1975), the validity of the multivariate tests rest on different assumption(s) concerning the covariance matrices. However, if $\underset{\sim}{\Sigma}_j = \underset{\sim}{\Sigma}$ for all j, multivariate procedures may be used to arrive at exact tests of all hypotheses of interest. For this reason, Timm (Note 4) recommends that the assumption concerning the equality of the dispersion matrices be stated in terms of the total number of measurement occasions.

multivariate counterparts. In general, if the value of the parameter e is greater than or equal to .75, the $\tilde{e}$-adjusted procedures provides the most robust and most powerful test of both within subjects effects, followed by the $\hat{e}$-adjusted F test and finally the multivariate tests. If $e < .75$, the multivariate procedures are consistently the most powerful tests of either within subjects hypothesis followed by the $\tilde{e}$- and $\hat{e}$-adjusted F tests. However, if the population is non-normal in form, the multivariate test of the within subjects main effect, while most powerful, is liberal and better Type I error control is achieved by adopting either of the adjusted univariate procedures.

Reference Notes

1. Noe, M. J. A monte carlo study of several test procedures in the repeated measures design. Paper presented at the meeting of the American Education Research Association, San Francisco, April, 1976.

2. Mendoza, J. L., Toothaker, L. E., & Crain, B. R. Necessary and sufficient conditions for F ratios in repeated measures designs. Unpublished manuscript, 1976. (Available from the Department of Psychology, University of Georgia, Athens, Georgia, 30602.)

3. Gabriel, R. M. Personal communication, September 23, 1977.

4. Timm, N. H. Personal communication, January 20, 1978.

References

Anderson, T. W. _An introduction to multivariate statistical analysis._ New York: Wiley & Sons, Inc., 1958.

Bartlett, M. S. Properties of sufficiency and statistical tests. _Proceedings of the Royal Society of London_, 1937, _160_, 268-282.

Bartlett, M. S. A note on tests of significance in multivariate analysis. _Proceedings of the Cambridge Philosophical Society_, 1939, _35_, 180-185.

Box, G. E. P. A general distribution theory for a class of likelihood criteria. _Biometrika_, 1949, _36_, 317-346.

Box, G. E. P. Problems in the analysis of growth and wear curves. _Biometrics_, 1950, _6_, 362-389.

Box, G. E. P. Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. _Annals of Mathematical Statistics_, 1954, _25_, 290-302. (a)

Box, G. E. P. Some theorems on quadratic forms applied in the study of analysis of variance problems, II. Effects of inequality of variance and correlation between errors in the two-way classification. _Annals of Mathematical Statistics_, 1954, _25_, 484-498. (b)

Box, G. E. P., & Müller, M. E. Randomization and least squares estimates. _Journal of the American Statistical Association_, 1959, _54_, 489.

Cohen, J. _Statistical power analysis for the behavioral sciences._ New York: Academic Press, 1969.

Cole, J. W. L., & Grizzle, J. E. Applications of multivariate analysis to repeated measurements experiments. _Biometrics_, 1966, _22_, 810-828.

Collier, R. O., Jr., Baker, F. B., Mandeville, G. K., & Hayes, T. F. Estimates of test size for several test procedures on conventional variance ratios in the repeated measures design. _Psychometrika_, 1967, _32_, 339-353.

Danford, M. B., Hughes, H. M., & McNee, R. C. On the analysis of repeated-measurements experiments. _Biometrics_, 1960, _16_, 547-565.

Davidson, M. L. Univariate versus multivariate tests in repeated-measures experiments. _Psychological Bulletin_, 1972, _77_(6), 446-452.

Edgington, E. S. A new tabulation of statistical procedures used in APA journals. _American Psychologist_, 1974, _29_, 25-26.

Finn, J. D. _Multivariance: Univariate and multivariate analysis of variance, covariance, and regression. A Fortran IV program._ Chicago: International Educational Services, 1976.

Gaito, J. Repeated measurements designs and counterbalancing. _Psychological Bulletin_, 1961, _58_, 46-54.

Geisser, S. Multivariate analysis of variance for a special covariance case. _Journal of the American Statistical Association_, 1963, _58_, 600-669.

Geisser, S., & Greenhouse, S. W. An extension of Box's results on the use of the F distribution in multivariate analysis. _Annals of Mathematical Statistics_, 1958, _29_, 885-891.

Greenhouse, S. W., & Geisser, S. On methods in the analysis of profile data. Psychometrika, 1959, 24, 95-112.

Guttman, L. A. A generalized simplex for factor analysis. Psychometrika, 1955, 20, 173-192.

Hajek, J. Inequalities for the generalized student's distribution and their applications. Selected Translations in Mathematical Statistics and Probabilities, Institute of Mathematical Statistics, 1962, 2, 63-74.

Harman, H. H. Modern factor analysis. (2nd ed.). Chicago: University of Chicago Press, 1967.

Heck, D. L. Charts on some upper percentage points of the distribution of the largest characteristic root. Annals of Mathematical Statistics, 1960, 31, 625-642.

Hopkins, J. W., & Clay, P. P. F. Some empirical distributions of bivariate T^2 and homoscedasticity criterion M under unequal variance and leptokurtosis. Journal of the American Statistical Association, 1963, 58, 1048-1053.

Hotelling, H. The generalization of Student's ratio. Annals of Mathematical Statistics, 1931, 2, 360-378.

Hotelling, H. A generalized T test and measure of multivariate dispersion. In J. Newyman (Ed.), Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability. Berkeley: University of California Press, 1951.

Humphreys, L. G. Investigation of the simplex. Psychometrika, 1960, 25, 313-323.

Huynh, H. S., & Feldt, L. Conditions under which mean square ratios in repeated measurements designs have exact F-distributions. Journal of the American Statistical Association, 1970, 65, 1582-1585.

Huynh, H., & Feldt, L. S. Estimation of the Box correction for degrees of freedom from sample data in randomized block and split-plot designs. Journal of Educational Statistics, 1976, 1(1), 69-82.

Imhof, J. P. Testing the hypothesis of no fixed main-effects in Scheffé's mixed model. Annals of Mathematical Statistics, 1962, 33, 1085-1095.

Ito, K. On the effect of heteroscedasticity and nonnormality upon some multivariate test procedures. In P.R. Krishnaiah (Ed.), Multivariate analysis (Vol. 2). New York: Academic Press, 1969.

Ito, K., & Schull, W. J. On the robustness of the T^2 test in multivariate analysis of variance when variance-covariance matrices are not equal. Biometrika, 1964, 51, 71-82.

Kirk, R. E. Experimental design: Procedures for the behavioral sciences. Belmont, Calif.: Brooks-Cole, 1968.

Kogan, L. S. Analysis of variance - repeated measurements. Psychological Bulletin, 1948, 45, 131-143.

Korin, B. P. Some comments on the homoscedasticity criterion M and the multivariate analysis of variance tests T^2, W and R. Biometrika, 1972, 59, 215-216.

Lana, R. E., & Lubin, A. The effect of correlation on the repeated measures design. Educational and Psychological Measurement, 1963, 23, 729-739.

Lawley, D. N. A generalization of Fisher's z test. Biometrika, 1938, 30, 180-187.

Marsaglia, G., MacLaren, M. D., & Bray, T. A. A fast procedure for generating normal random variables. Communication of the ACM, 1964, 7, 4-10.

Mauchly, J. W. Significance test for sphericity of a normal n-variate distribution. The Annals of Mathematical Statistics, 1940, 29, 204-209.

McCall, R. B., & Appelbaum, M. I. Bias in the analysis of repeated measures designs: Some alternative approaches. Child Development, 1973, 44, 401-415.

Mendoza, J. L., Toothaker, L. E., & Nicewander, W. A. A monte carlo comparison of the univariate and multivariate methods for the groups by trials repeated measures design. Multivariate Behavioral Research, 1974, 9, 165-178.

Mendoza, J. L., Toothaker, L. E., & Crain, B. R. Necessary and sufficient conditions for F ratios in the L X J X K factorial design with two repeated factors. Journal of the American Statistical Association, 1976, 71(356), 992-993.

Morrison, D. F. Multivariate statistical methods. (2nd ed.). New York: McGraw-Hill, 1976.

Myers, J. L. Fundamentals of experimental design. Boston, Mass.: Allyn & Bacon, 1966.

Olson, C. L. Comparative robustness of six tests in multivariate analysis of variance. Journal of the American Statistical Association, 1974, 69, 894-908.

Olson, C. L. On choosing a test statistic in multivariate analysis of variance. Psychological Bulletin, 1976, 83(4), 579-586.

Pillai, K. C. S. Some new test criteria in multivariate analysis. Annals of Mathematical Statistics, 1955, 26, 117-121.

Pillai, K. C. S. Statistical tables for tests of multivariate hypotheses. Manila: University of the Philippines, Statistical Center, 1960.

Rao, C. R. Advanced statistical methods in biometric research. New York: Wiley & Sons, Inc., 1952.

Rouanet, H., & Lepine, D. Comparison between treatments in a repeated-measures design: ANOVA and multivariate methods. The British Journal of Mathematical and Statistical Psychology, 1970, 23, 147-163.

Roy, S. N. On a heuristic method of test construction and its use in multivariate analysis. Annals of Mathematical Statistics, 1953, 24, 220-238.

Schuer, E. M., & Stoller, D. S. On the generation of normal random vectors. Technometrics, 1966, 4, 279-280.

Searle, S. R. Matrix algebra for the biological sciences. New York: Wiley & Sons, Inc., 1966.

Stoloff, P. H. Correcting for heterogeneity of covariances for repeated measures designs of the analysis of variance. Educational and Psychological Measurement, 1970, 30, 909-924.

Tatsuoka, M. M. Multivariate analysis: Techniques for educational and psychological research. New York: Wiley & Sons, Inc., 1971.

Timm, N. H. Multivariate analysis with applications in education and psychology. Monterey, Calif.: Brooks-Cole, 1975.

Wall, F. J. The generalized variance ratio or U-statistic. Albuquerque, New Mexico: The Dikewood Corporation, 1968.

Wilks, S. S. Certain generalizations in the analysis of variance. _Biometrika_, 1932, _24_, 471-494.

Wilson, R. S. Analysis of developmental data: Comparison among alternative methods. _Developmental Psychology_, 1975, _11_(6), 676-680.

Winer, B. J. _Statistical principles in experimental design_. (2nd ed.). New York: McGraw-Hill, 1971.

Appendix

Table A

Matrix Sets and Associated Eigenvalues (λ_k)

<u>Matrix Set 1</u>: $M_1 = 0$; $M_2 = 0$; $e = 1.00$

$$\underset{\sim}{\Sigma}_1 = \begin{bmatrix} 16 & 9 & 7 & 5 \\ & 12 & 5 & 3 \\ & & 8 & 1 \\ & & & 4 \end{bmatrix}$$

λ_k: 28.42, 5.0, 5.0, 1.58

$$\underset{\sim}{\Sigma}_2 = \begin{bmatrix} 16 & 9 & 7 & 5 \\ & 12 & 5 & 3 \\ & & 8 & 1 \\ & & & 4 \end{bmatrix}$$

λ_k: 28.42, 5.0, 5.0, 1.58

$$\underset{\sim}{\Sigma}_3 = \begin{bmatrix} 16 & 9 & 7 & 5 \\ & 12 & 5 & 3 \\ & & 8 & 1 \\ & & & 4 \end{bmatrix}$$

λ_k: 28.42, 5.0, 5.0, 1.58

$$\underset{\sim}{\Sigma} = \begin{bmatrix} 16 & 9 & 7 & 5 \\ & 12 & 5 & 3 \\ & & 8 & 1 \\ & & & 4 \end{bmatrix}$$

λ_k: 28.42, 5.0, 5.0, 1.58

<u>Matrix Set 2</u>: $M_1 = 0$; $M_2 = 0$; $e = .96$

$$\underset{\sim}{\Sigma}_1 = \begin{bmatrix} 16 & 9 & 7 & 5 \\ & 12 & 4 & 3 \\ & & 8 & 2 \\ & & & 4 \end{bmatrix}$$

λ_k: 28.18, 5.82, 3.90, 2.07

$$\underset{\sim}{\Sigma}_2 = \begin{bmatrix} 16 & 9 & 7 & 5 \\ & 12 & 4 & 3 \\ & & 8 & 2 \\ & & & 4 \end{bmatrix}$$

λ_k: 28.18, 5.82, 3.90, 2.07

$$\underset{\sim}{\Sigma}_3 = \begin{bmatrix} 16 & 9 & 7 & 5 \\ & 12 & 4 & 3 \\ & & 8 & 2 \\ & & & 4 \end{bmatrix}$$

λ_k: 28.18, 5.82, 3.90, 2.07

$$\Sigma = \begin{bmatrix} 16 & 9 & 7 & 5 \\ & 12 & 4 & 3 \\ & & 8 & 2 \\ & & & 4 \end{bmatrix}$$

λ_k: 28.18, 5.82, 3.90, 2.07

Table A (Cont'd)

Matrix Set 3: $M_1 = 0$; $M_2 = 0$; $e = .75$

$$\underset{\sim}{\Sigma}_1 = \begin{bmatrix} 16 & 9 & 5 & 3 \\ & 12 & 5 & 4 \\ & & 8 & 4 \\ & & & 4 \end{bmatrix}$$

λ_k: 27.54, 6.86, 4.26, 1.36

$$\underset{\sim}{\Sigma}_2 = \begin{bmatrix} 16 & 9 & 5 & 3 \\ & 12 & 5 & 4 \\ & & 8 & 4 \\ & & & 4 \end{bmatrix}$$

λ_k: 27.54, 6.86, 4.26, 1.36

$$\underset{\sim}{\Sigma}_3 = \begin{bmatrix} 16 & 9 & 5 & 3 \\ & 12 & 5 & 4 \\ & & 8 & 4 \\ & & & 4 \end{bmatrix}$$

λ_k: 27.54, 6.86, 4.26, 1.36

$$\underset{\sim}{\Sigma} = \begin{bmatrix} 16 & 9 & 5 & 3 \\ & 12 & 5 & 4 \\ & & 8 & 4 \\ & & & 4 \end{bmatrix}$$

λ_k: 27.54, 6.86, 4.26, 1.36

Matrix Set 4: $M_1 = 0$; $M_2 = 0$; $e = .57$

$$\underset{\sim}{\Sigma}_1 = \begin{bmatrix} 20 & 9 & 5 & 3 \\ & 10 & 5 & 4 \\ & & 6 & 4 \\ & & & 4 \end{bmatrix}$$

λ_k: 28.80, 7.63, 2.76, 0.82

$$\underset{\sim}{\Sigma}_2 = \begin{bmatrix} 20 & 9 & 5 & 3 \\ & 10 & 5 & 4 \\ & & 6 & 4 \\ & & & 4 \end{bmatrix}$$

λ_k: 28.80, 7.63, 2.76, 0.82

$$\underset{\sim}{\Sigma}_3 = \begin{bmatrix} 20 & 9 & 5 & 3 \\ & 10 & 5 & 4 \\ & & 6 & 4 \\ & & & 4 \end{bmatrix}$$

λ_k: 28.80, 7.63, 2.76, 0.82

$$\underset{\sim}{\Sigma} = \begin{bmatrix} 20 & 9 & 5 & 3 \\ & 10 & 5 & 4 \\ & & 6 & 4 \\ & & & 4 \end{bmatrix}$$

λ_k: 28.80, 7.63, 2.76, 0.82

Table A (Cont'd)

Matrix Set 5: $M_1 = 0$; $M_2 = 0$; $e = .48$

$$\underset{\sim}{\Sigma}_1 = \begin{bmatrix} 21 & 10 & 4 & 3 \\ & 9 & 5 & 4 \\ & & 5 & 4 \\ & & & 5 \end{bmatrix} \qquad \underset{\sim}{\Sigma}_2 = \begin{bmatrix} 21 & 10 & 4 & 3 \\ & 9 & 5 & 4 \\ & & 5 & 4 \\ & & & 5 \end{bmatrix}$$

λ_k: 29.45, 8.06, 1.70, 0.78 $\qquad$ λ_k: 29.45, 8.06, 1.70, 0.78

$$\underset{\sim}{\Sigma}_3 = \begin{bmatrix} 21 & 10 & 4 & 3 \\ & 9 & 5 & 4 \\ & & 5 & 4 \\ & & & 5 \end{bmatrix} \qquad \underset{\sim}{\Sigma} = \begin{bmatrix} 21 & 10 & 4 & 3 \\ & 9 & 5 & 4 \\ & & 5 & 4 \\ & & & 5 \end{bmatrix}$$

λ_k: 29.45, 8.06, 1.70, 0.78 $\qquad$ λ_k: 29.45, 8.06, 1.70, 0.78

Matrix Set 6: $M_1 = 2.20$; $M_2 = 0$; $e = 1.00$

$$\underset{\sim}{\Sigma}_1 = \begin{bmatrix} 8 & 2 & 1 & 0 \\ & 6 & 0 & -1 \\ & & 4 & -2 \\ & & & 2 \end{bmatrix} \qquad \underset{\sim}{\Sigma}_2 = \begin{bmatrix} 16 & 9 & 7 & 5 \\ & 12 & 5 & 3 \\ & & 8 & 1 \\ & & & 4 \end{bmatrix}$$

λ_k: 9.47, 5.0, 5.0, 0.53 $\qquad$ λ_k: 28.42, 5.0, 5.0, 1.58

$$\underset{\sim}{\Sigma}_3 = \begin{bmatrix} 24 & 16 & 13 & 10 \\ & 18 & 10 & 7 \\ & & 12 & 4 \\ & & & 6 \end{bmatrix} \qquad \underset{\sim}{\Sigma} = \begin{bmatrix} 16 & 9 & 7 & 5 \\ & 12 & 5 & 3 \\ & & 8 & 1 \\ & & & 4 \end{bmatrix}$$

λ_k: 49.08, 5.0, 5.0, 0.92 $\qquad$ λ_k: 28.42, 5.0, 5.0, 1.58

Table A (Cont'd)

<u>Matrix Set 7:</u> $M_1 = 1.64$; $M_2 = 0$; $e = .96$

$$\underset{\sim}{\Sigma}_1 = \begin{bmatrix} 8 & 2 & 1 & 0 \\ & 6 & -1 & -1 \\ & & 4 & -1 \\ & & & 2 \end{bmatrix} \qquad \underset{\sim}{\Sigma}_2 = \begin{bmatrix} 16 & 9 & 7 & 5 \\ & 12 & 4 & 3 \\ & & 8 & 2 \\ & & & 4 \end{bmatrix}$$

λ_k: 9.30, 5.80, 3.82, 1.07 $\qquad$ λ_k: 28.18, 5.82, 3.90, 2.07

$$\underset{\sim}{\Sigma}_3 = \begin{bmatrix} 24 & 16 & 13 & 10 \\ & 18 & 9 & 7 \\ & & 12 & 5 \\ & & & 6 \end{bmatrix} \qquad \underset{\sim}{\Sigma} = \begin{bmatrix} 16 & 9 & 7 & 5 \\ & 12 & 4 & 3 \\ & & 8 & 2 \\ & & & 4 \end{bmatrix}$$

λ_k: 48.88, 5.83, 3.90, 1.39 $\qquad$ λ_k: 28.18, 5.82, 3.90, 2.07

<u>Matrix Set 8:</u> $M_1 = 1.27$; $M_2 = 0$; $e = .75$

$$\underset{\sim}{\Sigma}_1 = \begin{bmatrix} 8 & 2 & -1 & -2 \\ & 6 & 0 & 0 \\ & & 4 & 1 \\ & & & 2 \end{bmatrix} \qquad \underset{\sim}{\Sigma}_2 = \begin{bmatrix} 16 & 9 & 5 & 3 \\ & 12 & 5 & 4 \\ & & 8 & 4 \\ & & & 4 \end{bmatrix}$$

λ_k: 9.83, 5.36, 3.64, 1.17 $\qquad$ λ_k: 27.51, 6.86, 4.26, 1.36

$$\underset{\sim}{\Sigma}_3 = \begin{bmatrix} 24 & 16 & 11 & 8 \\ & 18 & 10 & 8 \\ & & 12 & 7 \\ & & & 6 \end{bmatrix} \qquad \underset{\sim}{\Sigma} = \begin{bmatrix} 16 & 9 & 5 & 3 \\ & 12 & 5 & 4 \\ & & 8 & 4 \\ & & & 4 \end{bmatrix}$$

λ_k: 48.26, 6.30, 4.29, 1.16 $\qquad$ λ_k: 27.51, 6.86, 4.26, 1.36

Table A (Cont'd)

<u>Matrix Set 9</u>: $M_1 = 1.54$; $M_2 = 0$; $e = .57$

$$\underset{\sim}{\Sigma}_1 = \begin{bmatrix} 10 & 1.5 & -1.5 & -3 \\ & 5 & 1 & 0.5 \\ & & 3 & 1.5 \\ & & & 2 \end{bmatrix} \qquad \underset{\sim}{\Sigma}_2 = \begin{bmatrix} 20 & 9 & 5 & 3 \\ & 10 & 5 & 4 \\ & & 6 & 4 \\ & & & 4 \end{bmatrix}$$

λ_k: 11.59, 5.73, 2.15, 0.53 $\qquad$ λ_k: 28.80, 7.63, 2.76, 0.82

$$\underset{\sim}{\Sigma}_3 = \begin{bmatrix} 30 & 16.5 & 11.5 & 9 \\ & 15 & 9 & 7.5 \\ & & 9 & 6.5 \\ & & & 6 \end{bmatrix} \qquad \underset{\sim}{\Sigma} = \begin{bmatrix} 20 & 9 & 5 & 3 \\ & 10 & 5 & 4 \\ & & 6 & 4 \\ & & & 4 \end{bmatrix}$$

λ_k: 50.25, 6.19, 2.76, 0.80 $\qquad$ λ_k: 28.80, 7.63, 2.76, 0.82

<u>Matrix Set 10</u>: $M_1 = 2.03$; $M_2 = 0$; $e = .48$

$$\underset{\sim}{\Sigma}_1 = \begin{bmatrix} 10.5 & 2.5 & -2.5 & -3.5 \\ & 4.5 & 1.5 & 0.5 \\ & & 2.5 & 1.5 \\ & & & 2.5 \end{bmatrix} \qquad \underset{\sim}{\Sigma}_2 = \begin{bmatrix} 21 & 10 & 4 & 3 \\ & 9 & 5 & 4 \\ & & 5 & 4 \\ & & & 5 \end{bmatrix}$$

λ_k: 12.98, 5.72, 0.88, 0.43 $\qquad$ λ_k: 29.45, 8.06, 1.70, 0.78

$$\underset{\sim}{\Sigma}_3 = \begin{bmatrix} 31.5 & 17.5 & 10.5 & 9.5 \\ & 13.5 & 8.5 & 7.5 \\ & & 7.5 & 6.5 \\ & & & 7.5 \end{bmatrix} \qquad \underset{\sim}{\Sigma} = \begin{bmatrix} 21 & 10 & 4 & 3 \\ & 9 & 5 & 4 \\ & & 5 & 4 \\ & & & 5 \end{bmatrix}$$

λ_k: 50.96, 6.44, 1.82, 0.78 $\qquad$ λ_k: 29.45, 8.06, 1.70, 0.78

Table A (Cont'd)

<u>Matrix Set 11</u>: $M_1 = 1.83$; $M_2 = 1.37$; $e = 1.00$

$$\underset{\sim}{\Sigma}_1 = \begin{bmatrix} 6 & 3.375 & 2.625 & 1.875 \\ & 4.5 & 1.875 & 1.125 \\ & & 3 & 0.375 \\ & & & 1.5 \end{bmatrix} \qquad \underset{\sim}{\Sigma}_2 = \begin{bmatrix} 18 & 10.125 & 7.875 & 5.625 \\ & 13.5 & 5.625 & 3.375 \\ & & 9 & 1.125 \\ & & & 4.5 \end{bmatrix}$$

λ_k: 10.66, 1.88, 1.88, 0.59 $\qquad$ λ_k: 31.97, 5.62, 5.62, 1.78

$$\underset{\sim}{\Sigma}_3 = \begin{bmatrix} 24 & 13.5 & 10.5 & 7.5 \\ & 18 & 7.5 & 4.5 \\ & & 12 & 1.5 \\ & & & 6 \end{bmatrix} \qquad \underset{\sim}{\Sigma} = \begin{bmatrix} 16 & 9 & 7 & 5 \\ & 12 & 5 & 3 \\ & & 8 & 1 \\ & & & 4 \end{bmatrix}$$

λ_k: 42.62, 7.5, 7.5, 2.38 $\qquad$ λ_k: 28.42, 5.0, 5.0, 1.58

<u>Matrix Set 12</u>: $M_1 = 1.83$; $M_2 = 1.37$; $e = .96$

$$\underset{\sim}{\Sigma}_1 = \begin{bmatrix} 6 & 3.375 & 2.625 & 1.875 \\ & 4.5 & 1.5 & 1.125 \\ & & 3 & 0.75 \\ & & & 1.5 \end{bmatrix} \qquad \underset{\sim}{\Sigma}_2 = \begin{bmatrix} 18 & 10.125 & 7.875 & 5.625 \\ & 13.5 & 4.5 & 3.375 \\ & & 9 & 2.25 \\ & & & 4.5 \end{bmatrix}$$

λ_k: 10.57, 2.20, 1.46, 0.78 $\qquad$ λ_k: 31.71, 6.58, 4.38, 2.33

$$\underset{\sim}{\Sigma}_3 = \begin{bmatrix} 24 & 13.5 & 10.5 & 7.5 \\ & 18 & 6 & 4.5 \\ & & 12 & 3 \\ & & & 6 \end{bmatrix} \qquad \underset{\sim}{\Sigma} = \begin{bmatrix} 16 & 9 & 7 & 5 \\ & 12 & 4 & 3 \\ & & 8 & 2 \\ & & & 4 \end{bmatrix}$$

λ_k: 42.28, 8.78, 5.84, 3.10 $\qquad$ λ_k: 28.18, 5.85, 3.90, 2.07

Table A (Cont'd)

Matrix Set 13: $M_1 = 1.83$; $M_2 = 1.37$; $e = .75$

$$\underset{\sim}{\Sigma}_1 = \begin{bmatrix} 6 & 3.375 & 1.875 & 1.125 \\ & 4.5 & 1.875 & 1.5 \\ & & 3 & 1.5 \\ & & & 1.5 \end{bmatrix} \qquad \underset{\sim}{\Sigma}_2 = \begin{bmatrix} 18 & 10.125 & 5.625 & 3.375 \\ & 13.5 & 5.625 & 4.5 \\ & & 9 & 4.5 \\ & & & 4.5 \end{bmatrix}$$

λ_k: 10.32, 2.57, 1.60, 0.51 $\qquad$ λ_k: 30.95, 7.72, 4.80, 1.53

$$\underset{\sim}{\Sigma}_3 = \begin{bmatrix} 24 & 13.5 & 7.5 & 4.5 \\ & 18 & 7.5 & 6 \\ & & 12 & 6 \\ & & & 6 \end{bmatrix} \qquad \underset{\sim}{\Sigma} = \begin{bmatrix} 16 & 9 & 5 & 3 \\ & 12 & 5 & 4 \\ & & 8 & 4 \\ & & & 4 \end{bmatrix}$$

λ_k: 41.27, 10.29, 6.40, 2.04 $\qquad$ λ_k: 27.51, 6.86, 4.27, 1.36

Matrix Set 14: $M_1 = 1.83$; $M_2 = 1.37$; $e = .57$

$$\underset{\sim}{\Sigma}_1 = \begin{bmatrix} 7.5 & 3.375 & 1.875 & 1.125 \\ & 3.75 & 1.875 & 1.5 \\ & & 2.25 & 1.5 \\ & & & 1.5 \end{bmatrix} \qquad \underset{\sim}{\Sigma}_2 = \begin{bmatrix} 22.5 & 10.125 & 5.625 & 3.375 \\ & 11.25 & 5.625 & 4.5 \\ & & 6.75 & 4.5 \\ & & & 4.5 \end{bmatrix}$$

λ_k: 10.80, 2.86, 1.03, 0.30 $\qquad$ λ_k: 32.40, 8.58, 3.10, 0.92

$$\underset{\sim}{\Sigma}_3 = \begin{bmatrix} 30 & 13.5 & 7.5 & 4.5 \\ & 15 & 7.5 & 6 \\ & & 9 & 6 \\ & & & 6 \end{bmatrix} \qquad \underset{\sim}{\Sigma} = \begin{bmatrix} 20 & 9 & 5 & 3 \\ & 10 & 5 & 4 \\ & & 6 & 4 \\ & & & 4 \end{bmatrix}$$

λ_k: 43.20, 11.44, 4.13, 1.22 $\qquad$ λ_k: 28.80, 7.63, 2.76, 0.81

Table A (Cont'd)

<u>Matrix Set 15:</u> $M_1 = 1.83$; $M_2 = 1.37$; $e = .48$

$$\underset{\sim}{\Sigma}_1 = \begin{bmatrix} 7.875 & 3.75 & 1.5 & 1.125 \\ & 3.375 & 1.875 & 1.5 \\ & & 1.875 & 1.5 \\ & & & 1.875 \end{bmatrix}$$

λ_k: 11.04, 3.02, 0.64, 0.29

$$\underset{\sim}{\Sigma}_2 = \begin{bmatrix} 23.625 & 11.25 & 4.5 & 3.375 \\ & 10.125 & 5.625 & 4.5 \\ & & 5.625 & 4.5 \\ & & & 5.625 \end{bmatrix}$$

λ_k: 33.14, 9.07, 1.92, 0.88

$$\underset{\sim}{\Sigma}_3 = \begin{bmatrix} 31.5 & 15 & 6 & 4.5 \\ & 13.5 & 7.5 & 6 \\ & & 7.5 & 6 \\ & & & 7.5 \end{bmatrix}$$

λ_k: 44.18, 12.09, 2.56, 1.17

$$\underset{\sim}{\Sigma} = \begin{bmatrix} 21 & 10 & 4 & 3 \\ & 9 & 5 & 4 \\ & & 5 & 4 \\ & & & 5 \end{bmatrix}$$

λ_k: 29.45, 8.06, 1.70, 0.78

<u>Note</u>. Values of the eigenvalues have been rounded to two decimal places.