This is to certify that the

thesis entitled

Robustness and Power of Multivariate Tests
for Trends in Repeated Measures Data
Under Variance-Covariance Heterogeneity

presented by

Gabriella Belli

has been accepted towards fulfillment
of the requirements for

Ph.D. degree in Counseling, Educational Psychology, & Special Education (Statistics & Research Design)

A C Porter
Major professor

Date 12/20/83

ROBUSTNESS AND POWER OF MULTIVARIATE TESTS FOR TRENDS IN REPEATED MEASURES DATA UNDER VARIANCE-COVARIANCE HETEROGENEITY

by

Gabriella M. Belli

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Department of Counseling, Educational
Psychology and Special Education

1983

Mu

of homo

groups).

ordered

group d

and (3)

effects

group h

tests o

An argu

for int

but tha

Th

was to

differe

of main

evaluate

two wit

and (2)

ABSTRACT

ROBUSTNESS AND POWER OF MULTIVARIATE TESTS
FOR TRENDS IN REPEATED MEASURES DATA
UNDER VARIANCE-COVARIANCE HETEROGENEITY

By

Gabriella M. Belli

Multivariate statistics are subject to the assumption of homoscedasticity (i.e., equal covariace matrices across groups). In a repeated measures (RM) design with time ordered data, three hypotheses are tested: (1) between-group differences, (2) within-group trends over occasions, and (3) group by occasion interactions. Although the effects of assumption violation on tests of the between-group hypothesis have been investigated, the effects on tests of within-group and interaction hypotheses have not. An argument is presented indicating that multivariate tests for interactions should behave like between-group tests, but that tests for within-group trends should not.

The primary purpose of this Monte Carlo investigation was to determine whether heteroscedasticity has a differential effect on the robustness of multivariate tests of main effects in a RM case. A secondary purpose was to evaluate the robustness and power of multivariate tests of two within-group hypotheses: (1) overall tests of trends, and (2) subsequent tests of trends higher than linear,

unde

samp

Hotel

Wilks

inves

trend

than

(2) W

sligh

(3) D

for w

size,

betwe

group

robus

heter

facto

increa

heter

facto

a dec

rathe

under various combinations of number of groups and equal sample sizes.

The test statistics were: Roy's largest root, R, Hotelling-Lawley trace, T, Pillai-Bartlett trace, V, and Wilks' likelihood ratio, W.

The following are the major conclusions drawn from the investigation. (1) Multivariate tests of within-group trends are considerably more robust to heteroscedasticity than are multivariate tests of between-group differences. (2) Within-group tests of trends higher than linear are slightly more robust than overall tests of trends. (3) Departures of empirical Type I error from nominal alpha for within-group tests increase as heterogeneity, sample size, or alpha increase, but not as dramatically as for between-group tests. (4) Increasing the number of equal groups does not have a consistent detrimental effect on robustness of within-group tests. (5) For low and moderate heterogeneity (i.e., covariance matrices differing by factors of two or four), power of within-group tests increases as total sample size, N, increases. (6) For high heterogeneity (i.e., covariance matrices differing by a factor of nine), power of within-group tests increases with a decrease in the number of discrepant score vectors, rather than with an increase in N.

In a

and comme

wish to t

friend, c

strong mo

teacher ar

programs.

I wou

the many t

constant a

of this re

rest of my

Floden, an

providing

I wou

Dr. Joe L.

to Mr. Jef

were essen

Most

husband, D

and encoura

His patien

ACKNOWLEDGEMENTS

In appreciation for his support, insightful questions and comments, and willingness to let me pursue my ideas, I wish to thank Dr. Andrew C. Porter. He has been a valuable friend, colleague, and chairperson and has provided a strong model for professionalism and excellence as both teacher and researcher throughout my masters' and doctoral programs.

I would also like to thank Dr. Richard F. Houang for the many thought provoking discussions, as well as for his constant availability, during the conceptualization stage of this research. I wish to express my appreciation to the rest of my committee, Drs. James H. Stapleton, Robert E. Floden, and William H. Schmidt, for reviewing my work and providing suggestions for improvement.

I would further like to express my appreciation to Dr. Joe L. Byers, who helped provide the computer time, and to Mr. Jeff Glass, who coded the FORTRAN program. Both were essential components in making this research possible.

Most importantly, an expression of gratitute to my husband, Dr. Robert E. Krapfel, without whose moral support and encouragement this work would not have been completed. His patience and understanding have been immeasureable.

TABLE OF CONTENTS

LIST OF TABLES

Figure

LIST OF FIGURES

CHAPTER I

STATEMENT OF THE PROBLEM

Classical experimental research involves investigating the effect of manipulating one or more independent variables on a single dependent variable. This involves either testing the null hypothesis of equal group means against a general alternative or testing for specific planned comparisons among the group means. The test statistic used is the F-test (or t-test for two groups). Given parametric assumptions, this is the uniformly most powerful test that is invariant with respect to linear transformations (Scheffé, 1959).

Generalizing to the multivariate case, where there are two or more dependent variables (say, p), the corresponding null hypothesis is that of no differences among the k group vectors, where each vector consists of the group means on the p dependent measures. The F-test is a univariate test statistic, and several generalizations of it have been proposed for significance testing in the multivariate case. Among those tests that are invariant under linear transformation of the dependent variables, Hotelling's T^2 statistic is the uniformly most powerful for one-sample tests of means and two-sample tests of mean differences (Anderson, 1958).

Four other commonly used test statistics are Roy's largest root, R, Hotelling-Lawley trace, T, Pillai-Bartlett trace, V, and Wilks' likelihood ratio, W. However, for situations where there are multiple dependent variables or more than two groups, no test has emerged that is both invariant with respect to linear transformations and uniformly most powerful.

A specialized case of the multivariate analysis of variance (MANOVA) deals with situations where the same measure is repeatedly taken over the same individuals. The design on the measures, or occasions of testing, may reflect the passage of time, with the same measure taken at equally spaced intervals, or it may represent a factorial structure, with the same measure taken after various treatment interventions. In addition to the usual multivariate hypothesis of group differences, hypotheses about the occasions and, if there are multiple groups, about group by occasion interaction may be tested. The null hypothesis for occasions is that of no differences among the p occasion vectors, where each vector consists of the occasion means for the k groups. When there is only one group or when no group by occasion interaction exists, of even greater interest is the testing of hypotheses about the trend the data follow, assuming equally spaced time points, or about contrasts among the various measures, assuming a factorial design. Tests for these hypotheses are a

tests

test

assu

vecto

(2) r

a co

conc

to vi

null

simi

then

viol

can

the

oper

thre

the

robu

bala

Sand

sign

wher

are all within-group tests as opposed to between-group tests in the usual MANOVA sense.

In both the univariate and the multivariate cases, the test statistics used are based on certain distributional assumptions. These are that the random errors or error vectors for the p measures are: (1) independently and (2) normally or multivariate normally distributed (3) with a common variance or variance-covariance matrix.

Violations of these assumtions may lead to erroneous conclusions. However, if a particular test is insensitive to violation of one or more of the assumptions when the null hypothesis is true (i.e., if it leads to conclusions similar to what would be expected given the assumptions), then the test is said to be "robust" with respect to the violation.

The assumption of independence is critical and no test can favorably withstand its violation. Non-independence of the observartions or of observational vectors due to faulty operationalization of experimental design is a serious threat to nominal alpha levels. In univariate situations, the F-test for fixed effects has been shown to be fairly robust with respect to violation of normality and, for balanced designs, of homogeneity (see Glass, Peckham, and Sanders, 1972). However, severe departure from nominal significance level may occur under heterogeneity conditions when samples are small and unequal (Scheffé, 1959).

Re

situati

violati

1980).

result:

are sim

robust

heteros

varianc

but are

two gro

number

With mo

are equ

moderat

level a

To

violati

within-

situati

(1976) s

to high

larger

likely t

treatmen

between

Regarding between-group differences in multivariate situations, the several tests respond differently to violations of the assumptions (for a review, see Ito, 1980). Generalizing, it may be said that robustness results for fixed effects of at least some of these tests are similar to those in the univariate case. They are robust to non-normality and also fairly robust to heteroscedasticity (i.e., violation of homogeneity of variance and covariance) in balanced, two-group designs, but are not so for unbalance designs. However, even with two groups, the tests become liberal with increases in number of dependent variables or amount of heterogeneity. With more than two groups, tests are robust only if samples are equal and extremely large. If they are unequal, even moderate heterogeneity has large effects on significance level and power (Ito and Schull, 1964).

To date, no studies have considered the robustness to violation of multivariate test assumptions for tests of within-group differences in a repeated measures (RM) situation. Due to the nature of RM studies, Morrison (1976) states that "many experimental conditions which lead to higher mean values may also produce responses with larger variances" (p. 141). Different populations are likely to respond differently to successive measurements or treatment conditions, thereby also causing correlations between measures to differ from group to group. This is particularly true in studies of naturally occurring groups (e.g., a comparison of learning disabled and normal children on learning retention rates over time). Subjects within a classification group may be expected to respond in a similar fashion, but it is unrealistic to expect that scores for the two groups come from the same multivariate normal population. Hence, it is important to determine the validity of the multivariate tests of RM in the presence of heterogeneity conditions.

Just as findings from robustness studies for tests of between group differences have parallels in the univariate and multivariate cases, it may be presumed that similar parallels would hold for tests of within-group differences when homogeneity is violated. However, results from mixed-model RM studies would not apply to multivariate tests since the univariate tests are based on the assumptions of equal variances and equal pairwise correlations across the measures, which are unnecessary for multivariate tests to be valid. The effect on within-group tests when using a covariance matrix that is pooled from heterogeneous population covariance matrices is not known.

The robustness of a parametric test is idiosyncratic rather than general with respect to any violation and changes in one parameter may produce different levels of departures from nominal significance level. Tests of within-group differences are based on transformations of the dependent variables and the assumptions are made on the transformed scores. It will be shown in Chapter II that multivariate tests of between-group and within-group differences are based on sums of squares and cross products (SSCP) matrices that are different in both form and size, and that the relationship between the eigenvalues needed for calculating the test statistics for the two tests is not obvious. Hence, it is not possible to predict the behavior of one type of test from that of the other. Since the current robustness results from studies of multivariate between-group tests may not apply directly to within-group tests, separate investigations need to be made.

Furthermore, subtests of particular trends for RM data make use of subcomponents of the appropriate SSCP matrices for hypothesis and error. Since it is known that between-group tests become more robust with lower dimensionality of variables, it is expected that tests of successively higher order trends should show greater robustness than tests of lower order trends.

The present research was an investigation of the robustness and power of multivariate within-group tests for a repeated measures design with the same measure taken over a series of equally spaced time points. Non-normality does not seem to cause serious problems under any circumstances thus far investigated, whereas heterogeneity may be a serious problem in certain cases. Therefore, given that he

th

vi

ha

st

to

gr

co

ex

di

of

com

tes

lev

amo

Als

tru

com

pre

mul

hyp

lit

and

dis

heterogeneity is typically a violation of greater concern, the focus of this study was limited to the effect that violation of the assumption of a common covariance matrix has on the sampling distributions of four multivariate test statistics.

The purpose of the first part of the investigation was to determine whether tests of between-group and within-group hypotheses differ in their reactions to heterogeneous covariance matrices across groups. The second part was to examine whether covariance matrix heterogeneity produces differential effects on within-group tests when the number of groups or of subjects within groups are varied. Third, comparisons were made between overall tests of trends and tests of non-linearity. In all cases, actual significance levels obtained under a true null hypothesis and a given amount of heterogeneity were compared to nominal values. Also, actual powers for within-group tests obtained under a true alternative and a given amount of heterogeneity were compared to expected nominal powers if no violation was present.

The following chapters will present the general multivariate and repeated measures models, along with their hypotheses and test statistics, a review of the robustness literature, the method used for investigating robustness and power of multivariate within-group tests, results and discussion of results.

CHAPTER II

MULTIVARIATE ANALYSIS OF VARIANCE

In this chapter, the mathematical models for the general multivariate analysis of variance (MANOVA) and for the multivariate generalization to repeated measures (RM) are described. These are followed by a description of the hypothesis testing procedures through the separation of the total source of variation into component parts, the tests of significance used in multivariate analyses, and the assumptions on which they are based. The final section deals with a comparison of the sums of squares and cross products (SSCP) matrices used to test between-group and within-group differences.

General Multivariate Linear Model

Assuming there are n_j $(j = 1,\ldots,k)$ independent observations in each of k groups, the ith observation in the jth group is a px1 vector consisting of a constant term $\underline{\mu}$, a group effect $\underline{\alpha}_j$, and a random error component $\underline{\varepsilon}_{ij}$

$$y_{ij} = \underline{\mu} + \underline{\alpha}_j + \underline{\varepsilon}_{ij}.$$

The y_{ij} and the $\underline{\varepsilon}_{ij}$ are distributed in the population of subjects as $N(\underline{\mu}, \Sigma)$ and $N(\underline{0}, \Sigma)$, respectively, where Σ is any pxp symmetric positive definite matrix.

The null hypothesis tested in MANOVA is that the px1 mean vectors of all groups are equal,

$$H_0: \underline{\mu}_1 = \underline{\mu}_2 = \ldots = \underline{\mu}_k.$$

By letting $\underline{\mu}_j = \underline{\mu} + \underline{\alpha}_j$, this hypothesis is equivalent to testing that all the $\underline{\alpha}_j = 0$ (i.e., that all the treatment or group effects are equal) (Bock, 1975).

The general MANOVA model for k group means may be expressed in matrix terms as

$$Y. = A\Xi + E.$$

where:

$Y.$ = a kxp data matrix of k group means on p measures

A = a kxm known design matrix

Ξ = an mxp matrix of unknown paramenters

$E.$ = a kxp matrix of random errors

The error matrix E. is distributed $N(\underline{0}, D^{-1} \bullet \Sigma)$ where $D = diag(n_1, n_2, \ldots, n_k)$.

Since A typically is not of full rank, A^{-1} does not exist, and therefore solving for the unknown parameters is not possible. One solution is to reparameterize the model, which may be done by factoring A into the product of two matrices, K and L,

$$A = KL$$

where L is an lxm contrast matrix that describes a set of l linear combinations of the paramenters in Ξ and K is the corresponding kxl column basis for the design matrix A. Then,

$$E(Y.) = A\Xi \quad = KL\Xi \quad = K\Theta$$

where Θ is an lxp matrix of new paramenters describing the resulting linear combinations that reflect the research interest regarding differences among the groups (Bock, 1975, pp. 239-240).

Multivariate Generalization for Repeated Measures

Multivariate analysis of variance of repeated measures (MANOVA of RM) is a variation of MANOVA that includes a test for the occasions or repeated measures. What distinguishes these data from general multivariate data is that in RM the multiple dependent scores are assumed to be in the same metric (i.e., having the same origin and unit), whereas in general the scores are qualitatively distinct (i.e., having different origin and unit).

The underlying model for the ith observation in the jth group is a pxl vector that contains a component for occasions $\underline{\tau}$, for groups $\underline{\theta}_j$, and for random subject error $\underline{\varepsilon}_{ij}$,

$$\underline{y}_{ij} = \underline{\tau} + \underline{\theta}_j + \underline{\varepsilon}_{ij}.$$

As before, the $\underline{\varepsilon}_{ij}$ are distributed $N(\underline{0}, \Sigma)$. But, unlike the general MANOVA model, where the common term $\underline{\mu}$ does not provide any additional information, the common term in this model, $\underline{\tau}$, represents a pxl vector of constants and general means for the p occasions. The second term, $\underline{\theta}_j$, is a pxl vector of effects for the jth group that incorporates both group and group by occasion interaction effects. The model allows for a design on the occasions and a design on the subjects (Bock, 1975).

In the one-sample case or, assuming no interactions, in the k-sample case, the objective is to characterize the occasion vector $\underline{\tau}$. The appropriate characterization depends on the structure of the repeated measures dimension. If the measures correspond to points along a continuum, a polynomial representation is used,

$$\underline{\tau} = X\,\underline{\beta}$$

where X is a regression model matrix and $\underline{\beta}$ is a vector of unknown regression coefficients. If the measures correspond to a factorial classification, then a treatment contrasts and interaction representation is used,

$$\underline{\tau} = A\,\underline{\xi}$$

where A is a design matrix for the occasions and $\underline{\xi}$ is a vector of unknown occasion effects. In the former case, X is of full rank while, in the latter case, A is not and the model may be reparamenterized a second time. While this reparamenterization follows the same pattern as before, with A = KL, A is now the design matrix for the occasions and not for the groups.

Under the usual MANOVA model, the general occasion effect $\underline{\tau}$ is not estimable and hypotheses on it are not testable in the presence of group effects. Bock (1963) and Potthoff and Roy (1964) have suggested a variation of MANOVA that involves transforming the dependent variables to within-subject differences. A new set of measured variables is formed as linear combinations of the original measures,

$$y_{ij}^* = P'y_{ij}$$

where P is a matrix representation of the design over the measures. In terms of the previous discussion of the characterization of $\underline{y}$, P is either: (1) the regression model matrix X, if the measures are taken at ordered time points or (2) the orthonormalization of K, where K is the basis for the reparameterization of A, the design matrix for the occasions.

Assuming a full rank model for group means, the transformation in matrix terms consists of postmultiplying the components of the MANOVA model by a known matrix P, which may be any pxq matrix. Preferably, P should be an orthogonal matrix and this is now assumed. Then,

$$Y.P = K \Theta P + E.P$$

or equivalently,

$$Y.^* = K \Theta P + E.^*$$

where:

$Y.^*$= a kxp matrix of transformed scores

K = a kxl basis matrix for transformations on groups

Θ = an lxp matrix of parameters

P = a pxp basis matrix for transformations on occasions

$E.^*$= a kxp matrix of transformed errors

Analysis now proceeds as usual with the transformed scores in $Y.^*$ replacing the original dependent measures. The fact that the standard procedures apply can be seen since the transformation

$$Y. = Y.^{*}P^{-1}$$

reduces the RM model to a standard MANOVA model (Timm, 1980, p. 76). Furthermore, if P is orthogonal (i.e., $P'P = I$), so that $P^{-1} = P'$, each vector of scores may be transformed using P', as was shown previously. When P is either non-singular or has rank p, the transformation has nice properties with respect to the distributional assumptions. Given that the y_{ij} are independent and distributed $N(\underline{\mu},\Sigma)$, then the y_{ij}^{*} are also independent and are distributed $N(P'\underline{\mu}, P'\Sigma P)$ (Bock, 1975, p. 140).

Three basic hypotheses are of interest with k-sample RM data. These deal with comparisons among the mean curves or profiles of the groups, and may be phrased in terms of the following questions: (1) Are the curves or profiles of the k groups parallel? (2) If parallel, are they also coincident? and (3) If coincident, are they also constant? (Bock, 1975). The first question is asking about the presence of any group by occasion interactions. The second relates to group differences and the third to occasion differences.

Subhypotheses to assess the effect of the treatment structure or the trend over the occasions may also be tested. Assuming a polynomial representation for the RM dimension, this involves partitioning the sources of variation for occasion and for group by occasion into constant, linear, quadratic, etc. terms. Then a hypothesized trend may be tested by a multivariate test that all higher order trends are zero. The interpretation for these tests on occasions is straightforward and relates information about the type of trend the RM follow over time. However, a q-degree trend among the interactions implies that "any contrast among the groups can presumably be described as a polynomial of this degree. For example, a degree-2 interaction would imply that differences between groups, in addition to a possible linear trend, are accelerating or decelerating with respect to occasions" (Bock, 1975, p. 474).

<u>Hypothesis Testing</u>

The multivariate hypothesis testing stage involves partitioning the sums of squares and cross products (SSCP) matrix for total variation into a constant, a between-groups, and a within-groups part. The MANOVA table for the general multivariate analysis is given in Table 2-1 (adapted from Bock, 1975).

The SSCP matrices for RM may be calculated directly by substituting Y^* for Y in Table 2-1. The same results may be obtained by transforming the MANOVA SSCP matrices as shown in Table 2-2 (Bock, 1975).

Table 2-1

Multivariate Analysis of Variance (k-sample case)

Source of Variation	df	SSCP (pxp)* Equal n's	General
Constant (occasion effect)	1	$Q_c = (n/k)Y.'\underline{1}\underline{1}'Y.$	$(1/N)Y.'D\underline{1}\underline{1}'DY.$
Between groups (group effect)	k-1	$Q_b = nY.'Y. - Q_c$	$Y.'DY. - Q_c$
Within groups error	N-k	$Q_w = Y'Y - nY.'Y$	$Y'Y - Y.'DY.$
Total	N	$Q_t = Y'Y$	$Y'Y$

* where $D = \text{diag}(n_1,\ldots,n_k)$ and $\underline{1}$ = a unit vector.

Table 2-2

Multivariate Analysis of Variance for Repeated Measures

Source of Variation	df	SSCP (pxp)
Constant	1	$Q_c^* = P'Q_cP$
Between groups	k-1	$Q_b^* = P'Q_bP$
Within groups error	N-k	$Q_w^* = P'Q_wP$
Total	N	$Q_t^* = P'Q_tP$

The multivariate test statistics are functions of the appropriate SSCP for hypothesis and error (say, H and E, respectively). The MANOVA hypothesis of equal group means may be tested by setting $H = Q_b$ and $E = Q_w$. For RM, the matrices in Table 2-2 may be partitioned in the following manner:

$$Q_c^* = \left[\begin{array}{c|c} c & \\ \hline & C \end{array}\right] \qquad Q_b^* = \left[\begin{array}{c|c} b & \\ \hline & B \end{array}\right] \qquad Q_w^* = \left[\begin{array}{c|c} w & \\ \hline & W \end{array}\right]$$

Assuming a polynomial decomposition, the scalars c, b, and w represent the sums of squares for constant, group effect, and error terms that would be used in a univariate analysis. The (p-1)x(p-1) matrices C, B, and W are the SSCP for occasion effects, group by occasion effects, and subject within group by occasion error. The diagonal elements of these submatrices are the univariate sums of squares for the respective linear, quadratic, etc. trends. Table 2-3 shows how these matrices are used for the three omnibus tests in a RM situation.

With no group by occasion interaction, the full matrices Q_b^* and Q_w^* are the H and E matrices for group effect and corresponding error for a multivariate test of group differences. When P is orthogonal, a test using these transformed matrices gives the identical results as with Q_b and Q_w, because test statistics based on either determinants or trace functions remain invariant under orthogonal transformation (Anderson, 1958, p. 277).

Table 2-3

SSCP Matrices for RM Tests

Hypothesis	H	E	Dimension
Parallelism (interaction)	B	W	(p-1)x(p-1)
Coincidence (group effect)	Q_b^*	Q_w^*	pxp
Constancy (occasion effect)	C	W	(p-1)x(p-1)

The submatrices C, B, and W may be partitioned further to provide tests for particular trends. To test for any q<p degree trend in the data, H and E are submatrices corresponding to the lower right (p-q-1)x(p-q-1) corners of the appropriate matrices (Bock, 1975, p. 480). The required submatrices would be of rank p-q-1 and may be represented by another transformation, R, such that

$$H_q^* = R'H^*R \qquad \text{and} \qquad E_q^* = R'E^*R$$

where,

$$R = \begin{bmatrix} 0 \\ \hline I \end{bmatrix} \begin{matrix} (q+1) \text{ rows} \\ (p-q-1) \text{ rows} \end{matrix}$$

$$(p-q-1) \text{ columns}$$

For example, let p = 5 and a linear trend (q = 1) be hypothesized. Then R', with rank p-q-1 = 3, is

$$R' = \left[\begin{array}{cc|ccc} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{array}\right]$$

yielding the lower right 3x3 corners of the 5x5 SSCP matrices to test if trends higher than linear are zero.

Tests of Significance

The multivariate tests of significance are derived under the assumptions of multivariate normality and independence between pairs of subjects. The observed data vectors are independent random samples from a population in which any linear combination of variables in the observed vector is normally distributed (Harris, 1975, p. 231). In terms of the error components, the distributional assumptions are that the errors for the p measures for each subject are independently distributed and follow a p-variate normal distribution with expectation zero and a common pxp covariance matrix, Σ. Whenever more than one group is involved, it is assumed that the sampled data for all groups come from populations that have identical covariance matrices (Harris, 1975, p. 231).

Numerous criteria are available to test multivariate hypotheses. However, they are all functions of the non-zero characteristic roots, or eigenvalues λ_i, of HE^{-1}, where H and E are SSCP matrices due to hypothesis and error, respectively. These roots may be obtained by solving the determinental equation

$$|H - \lambda E| = 0.$$

For this equation to have real-valued solutions, it is necessary for E to be positive definite (i.e., that the quadratic form $\mathbf{x}'E\mathbf{x} > 0$ for all $\mathbf{x} = 0$) (Anderson, 1958, p. 337). This will usually be the case if the number of dependent variables (p) is less than the degrees of freedom for error (df_e).

Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_s > 0$ where $s = \min(df_h, u)$ with df_h = degrees of freedom for hypothesis and u = the number of variables after any transformation. Then, four commonly used multivariate test criteria are defined in Table 2-4 (Timm, 1975). These are exact tests, with known central and noncentral distributions. When $s = 1$ (i.e., if $p = 1$ or $k = 2$), they are equivalent and may be represented as an exact F distribution. There also are F approximations for the multivariate tests (see e.g., Tatsuoka, 1971).

The only parameters necessary to define the distribution of the statistics under valid assumptions and true null hypothesis are number of variates, degrees of freedom for hypothesis, and degrees of freedom for error (Ito, 1962). Additionally, noncentrality parameters are needed under true alternatives. Based on these parameters, Timm (1975) provides tables for the upper percentile points of R, T, and V and for the lower percentile points of W. The null hypothesis is rejected at significance level α if the obtained value of W is less than the 100α-centile of the null distribution. For the other tests, the null is rejected if the obtained value of a statistic exceeds the $100(1-\alpha)$-centile of the corresponding distribution.

Table 2-4

Multivariate Test Statistics

Roy's largest root	$R = \frac{\lambda_1}{1+\lambda_1}$	
Hotelling-Lawley trace	$T = \sum_{i=1}^{s} \lambda_i$	$= tr(HE^{-1})$
Pillai-Bartlett trace	$V = \sum_{i=1}^{s} \frac{\lambda_i}{1+\lambda_i}$	$= tr(H(H+E)^{-1})$
Wilks' likelihood ratio	$W = \prod_{i=1}^{s} \frac{1}{1+\lambda_i}$	$= \lvert E\rvert \cdot (\lvert H+E\rvert)^{-1}$

Theoretical Comparison of Tests

Although the multivariate test statistics for tests of between-group and within-group differences are identical, they operate on different SSCP matrices. The question of interest, then, is whether these matrices, whose expected values are functions of the common covariance matrix, are equally subject to violations of homoscedasticity. The following discussion outlines the relationship between the matrices used for the two tests.

Multivariate test criteria are functions of the eigenvalues of HE^{-1}, where H and E are the SSCP matrices for hypothesis and error, respectively. For a within-group test HE^{-1} is the lower (p-1)x(p-1) submatrix of the appropriately transformed

$$Q_c Q_w^{-1} \quad (1)$$

and for a between-group test, it is the pxp matrix

$$Q_b Q_w^{-1} \quad (2)$$

where Q_c, Q_b, and Q_w are defined in Table 2-1.

From the robustness literature, which is reviewed in Chapter III, we have general conclusions about the effects of particular types of homogeneity violations in the population covariance matrices when (2) is used to test for group differences. These results are based on distributions of the p eigenvalues of (2).

Tests for group by occasion interactions are based on the eigenvalues of the order-(p-1) submatrices of (2).

Since the same SSCP matrices are used for both interaction and between-group tests, the lower dimensionality in the portion of those matrices used for interaction tests should tend to make them slightly more robust than the between-group tests.

Tests of occasion differences with RM data are based on the p-1 eigenvalues of the order-(p-1) submatrix of (1). By substitution,

$$\begin{aligned} Q_c Q_w^{-1} &= (Y.'DY. - Q_b) Q_w^{-1} \\ &= (Y.'DY.) Q_w^{-1} - Q_b Q_w^{-1}. \end{aligned}$$

Even though a relationship exists between matrices (1) and (2), knowledge about the distributions of eigenvalues of (2) does not provide direct information about the distribution of eigenvalues of (1). Since within-group tests are actually based on a submatrix of (1), it would further be necessary to establish the relationship between the p eigenvalues of the full matrix (1) and the (p-1) eigenvalues of the submatrix used for these tests in order to fully specify the relationship between the matrices for the two types of tests.

Each subsequent within-group test of successively higher order trends is based on submatrices of (1), which decrease in dimension. Therefore, each test of a higher order trend should result in slight increases in robustness over the previous within-group test.

It is not obvious whether heterogeneity in the population covariance matrices would differentially effect the robustness of between-group and within-group tests and the mathematics needed to demonstrate the necessary relationships are intractable. Therefore, an empirical study was conducted to determine if the distributions for any of the four multivariate test statistics presented earlier are comparable for testing the two types of hypotheses. In this way, it could be determined if the tests respond similarly to the same violation to homogeneity. A further comparison of the robustness between within-group tests for any trend across time and the subsequent tests for trends higher than linear was also conducted. The study involved the simulation of a large number of experiments so that the actual significance levels could be compared to nominal levels with minimal standard error.

The second part of the study was an investigation of the effects on robustness and power of within-group tests when sample size and number of groups are varied.

CHAPTER III

REVIEW OF THE LITERATURE

Consequences of assumption violations have been thoroughly investigated for univariate test statistics from both the large sample and small sample points of view. Only recently have similar studies been undertaken for multivariate test statistics. While some of this work has been theoretical, involving large sample theory and asymptotic approximations, most of it has been empirical. Since the mathematics involved in a theoretical study of multivariate statistics are quite complex, Ito and Schull (1964) remarked that "the small sample treatment of the problem ... is very difficult if not impossible" (p. 72).

Researchers in the multivariate area have focused on a one-way fixed effects classification for the independent factor and have considered tests of between-group differences on multiple dependent measures. Robustness studies of within-group tests have dealt only with violations of the univariate mixed-model assumptions of equal variances and covariances across the repeated measures (RM). Typically, comparisons have been made between the usual F-test and the F adjusted by a correction factor (e.g., Collier, Baker, Mandeville, and Hayes, 1967) or between univariate and multivariate analyses (e.g., Scheifley, 1974). However, in all cases with more than one group, groups were assumed to have a common covariance matrix.

The following review will briefly summarize the fixed and mixed model univariate results and then present the multivariate results in greater detail. As a preliminary, an overview of a common strategy used to model heterogeneity will be presented.

Strategies for Investigating Robustness to Heterogeneity

Variance heterogeneity in univariate studies is easily portrayed by a ratio of population variances. For multivariate problems, modeling is more complicated since there are many ways of introducing heterogeneity in population covariance matrices. Two single-valued multivariate analogs to a variance are the trace and determinant of the covariance matrix. The trace represents the total variation and the determinant represents a generalized variance (Tatsuoka, 1971). Ratios of covariance matrix determinants parallel the univariate case, forming a convenient index of multivariate heterogeneity.

A typical tactic used in empirical studies of robustness of multivariate test statistics against violation of the assumption of homoscedasticity is to reduce the problem to canonical form. This procedure, which was used in all but two of the multivariate studies reviewed, produces diagonal covariance matrices, thereby reducing the number of parameters that need to be considered by $p(p-1)/2$.

The procedure is based on theorems for matrix transformations (see Tatsuoka, 1971, pp. 125-129). It consists of applying a linear transformation, say C (where C is orthogonal, i.e., $C'C = I$ and $|C| = 1$), to the matrix of observations X, thus producing a new set of uncorrelated variables $Y = XC$. The matrix C represents a rigid (or angle-preserving) rotation from the original variates to the principal axes and consists of columns of eigenvectors of the original covariance matrix Σ. Using the same transformation matrix, Σ is transformed into a diagonal matrix $C'\Sigma C = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_s)$, with the variances of the canonical or transformed variates (eigenvalues) as diagonal elements. This is called "diagonalizing the matrix" (Tatsuoka, 1971, p. 128). The trace and determinant of the original covariance matrix are equal to the trace and determinant of the transformed matrix. A multivariate analysis on the canonical variates produces the same results as those obtained with the original ones since the MANOVA test criteria are invariant under any linear transformation (Anderson, 1958, p. 277).

The operationalization of this procedure in MANOVA robustness studies of heterogeneity relies on the fact that two population covariance matrices, say V_1 and V_2, may be linearly transformed to the identity matrix, I, and to a diagonal matrix, D, whose diagonal elements are the eigenvalues of $V_2V_1^{-1}$ (Holloway and Dunn, 1967). D is called the diagonal matrix of latent roots.

MANOVA test criteria for a given test based on any mixture of $N(\underline{0},V_1)$ and $N(\underline{0},V_2)$ are equivalent to a mixture of $N(\underline{0},I)$ and $N(\underline{0},D)$ (Olson, 1973). To model situations with non-zero mean vectors, a mixture of $N(\underline{\mu}_1,I)$ and $N(\underline{\mu}_2,D)$ may be used to represent the canonical forms of $N(C'^{-1}\underline{\mu}_1,V_1)$ and $N(C'^{-1}\underline{\mu}_2,V_2)$. This applies to both the central case, with equal population means, and the noncentral case, with unequal population means.

Heterogeneity is typically introduced either equally in all of the canonical dimensions, with $D = dI$, or in only one dimension, with $D = diag(d,1,...,1)$. Variations on this theme allow for heterogeneity to vary across canonical dimensions, with some $d_i = 1$ while other $d_i = d$. In this way, a researcher need only vary values of d to simulate a variety of heterogeneous conditions. For more than two groups, either one or more groups are sampled from a population with covariance matrix D and the rest from a population with covariance matrix I. An alternative is to sample groups from populations with covariance matrices I, D, and multiples of D.

Consequences of Non-independence and Non-normality

Violation of the independence assumption is quite serious. For analysis of variance (ANOVA), positive correlations among the errors yield a liberal test (i.e., too many significant results) and negative correlations yield a conservative test (i.e., too few significant results). This is true for both equal and unequal sample sizes and the discrepancy between nominal and actual levels of significance increases as the absolute value of the correlations increase (see Scheffé, 1959).

For a two-group matched-subjects design in univariate situations, use of the correlated or dependent t-test is an appropriate technique to handle the problem. For correlated observations that arise from a RM situation, the problem is identical to that of a mixed-model analysis and two avenues are open. One is to use the correction factors of Box (1954) or Greenhouse and Geisser (1959). These adjust the degrees of freedom for the F-test and the latter produces conservative results. The other method is to use exact multivariate tests, which do not make the ANOVA assumption of independence of errors across measures taken on the same subject. However, independence of errors between subjects must still be maintained.

Glass, Peckham, and Sanders (1972) provided a thorough review of the univariate literature for fixed-effects designs. General conclusions were that violation of the normality assumption does not present a problem for either the t-test or the F-test in an analysis of mean differences. For both equal and unequal sample sizes, discrepancies between actual and nominal significance levels are slight and, with equal n's, the F-test proves to be robust even in the extreme case of dichotomous data. However, non-normality does effect inferences about variances, such as in tests of random-effects or equality of variances (Scheffé, 1959).

Considering six multivariate tests and using equal n's, Olson (1973, 1974) found that departures from normality in the direction of positive kurtosis (occasional extreme observations) had only minor conservative effects on Type I error rates. From the asymptotic expressions for central and non-central distributions of Hotelling's T^2 and T_o^2 (a generalized T^2 for more than two groups), which were obtained by Ito (1969), approximate values for actual significance and power may be found. In a recent review, Ito (1980) mathematically demonstrated that, for sufficiently large sample sizes, non-normality did not appreciably effect either the significance level or the power of these test statistics. The question of what sample size is to be considered "sufficiently large" was left open, since this is difficult to demonstrate theoretically. Ito (1980) further stated that, from Monte Carlo studies, the T^2 test in the two-sample case has been found to be particularly robust against non-normality for tests about means. However, as in the univariate case, non-normality has serious consequences for tests of equality of covariance matrices.

Consequences of Heterogeneity

Studies of both univariate and multivariate cases indicate that violation of the homogeneity assumption may cause serious discrepancy between actual and nominal significance levels and that this is typically a more serious problem than non-normality. Since this violation is more serious, as well as being the focus of the present research, greater attention will be given to studies of robustness in the face of heterogeneity. Consequences in both the univariate and multivariate cases will be reviewed.

Fixed-model ANOVA

Extensive work has been done to examine the consequences of departures from homogeneity of variance for univariate test procedures (for reviews, see Scheffé, 1959, Chapter 10 and Glass, Peckham, and Sanders, 1972). In the univariate two-sample case, inequality of population variances has little effect on either significance level or power of the t-test if sample sizes are about equal. However, if sample sizes are markedly disparate, large deviations from the nominal error rate occur for both large and small sample cases. The test is conservative if the larger group has the larger variance and is liberal if the larger group has the smaller variance (Scheffé, 1959).

For more than two groups, heterogeneity does have a slight effect on the Type I error rate of the F-test even when groups are of about equal size, in which case the test is liberal (Scheffé, 1959). However, general conclusions from both theoretical and empirical work have been that the ANOVA F-test is robust to heterogeneity of variance. A major exception is in the case of small and unequal sample sizes, where the effects are serious. Results for unequal n's follow the same pattern as for the t-test, with either conservative or liberal results.

It should be noted, however, that these general conclusions have boundary conditions, which depend on sample size or ratio of sample sizes, on the amount of heterogeneity, and on the value of nominal alpha. Ramsey (1980) found that even for the equal sample t-test, robustness depends on certain conditions. For example, with n's greater than 15, the t-test will not exceed a significance level of .06 at a nominal level of .05 regardless of the amount of heterogeneity, but robustness may be achieved with n's as small as five if the ratio of variances in the two populations is 1:4 or less. Also, there is an inverse relationship between nominal alpha level used and sample size needed for robustness.

Mixed-model Repeated Measures

In univariate mixed-model analysis, the RM dimensions are treated as additional design factors. Two assumptions are made for a valid univariate test: (1) equality of covariance matrices across levels of the between-group factor, and (2) uniformity of the common covariance matrix (i.e., equality of the variances of the RM and of the pairwise correlations between these measures). In a RM situation, variances might change between observations, possibly due to treatment effects on each occasion. Also, there is potential for lack of independence between error components of the observations, particularly if the RM factor reflects time.

Huynh and Feldt (1970) demonstrated that uniformity is merely a sufficient and not a necessary condition for validity of within-group F-tests. What is required is that the assumptions stated above be met by the covariance matrices of orthonormalized variates rather than of the original variates. Nevertheless, the majority of the robustness literature in the mixed-model case has focused on violation of the uniformity assumption with the original variates. Some of these studies are reviewed below.

While the studies in this section have a different focus from the rest of this paper, since variances and covariances are equal across groups, they are included as backgound to a study of consequences of assumption violations in a RM study. Also, they provide another indication of the idiosyncratic nature of the behavior of test statistics under different forms of violations.

In a theoretical study, Box (1954) assessed the approximate effects of unequal variances and serial correlations in one factor of a two-way design with one observation per cell. He showed that these conditions reduced the apparent number of degrees of freedom in both the numerator and denominator of the F-ratio and that the effect was to produce a slightly liberal test.

Empirical results for a k-sample RM study (with $k = 3$ and $p = 4$) were obtained by Collier, Baker, Mandeville, and Hayes (1967). They compared Type I error rates for three ANOVA F-tests: unadjusted, adjusted by Box's correction factor, and by Greenhouse and Geisser's conservative lower bound for the correction factor. They considered 15 different patterns of covariance matrices, where both variances and pairwise correlations were varied, although covariance matrices were common across groups.

As expected, their results showed that the F-test for group differences had a close agreement between empirical and nominal alpha, but that the F-test for occasions and group by occasions effects did not. In both cases the unadjusted F was liberal and the adjusted F was fairly robust with Box's correction factor but conservative with the lower bound test. An unexpected finding was that departures from nominal alpha did not significantly decrease, and in some cases actually increased, when sample sizes increased from five to 15. A similar, but smaller study conducted by Mendoza, Toothaker, and Nicewander (1974) upheld the above conclusions.

In an empirical study comparing the mixed-model ANOVA, MANOVA of RM, and analysis of covariance structures (ANCOVST), Scheifley (1974; Scheifley and Schmidt, 1978) considered a one-group RM case with a 2x2 design on the measures. Three covariance matrices were used, where one matrix conformed to the assumptions of each analysis. When the ANOVA assumption of uniformity was not met, all three tests were generally conservative. ANCOVST had the greatest power when a significant difference in means was present in only one of the RM factors and MANOVA of RM had the greatest power when the null for both RM factors and the interaction were false.

Significance level results for the univariate test under violation of uniformity in the above study were not consistent with the previous two empirical studies in this section, where results tended to be liberal. This may be partly due to the fact that the two covariance matrices used to model univariate assumption violation in Scheifley's (1974) study had variances that were fairly close to being equal, while the other two studies had larger discrepancies between variances. Another possibility is that the opposite results were due to the different patterns for the covariance matrices. The first two studies considered successive trials on one RM factor and the covariance matrices had simplex patterns (i.e., successive diagonals had lower values). Due to the two-way factorial structure on the RM in the third study, those covariance matrices had circumplex patterns (i.e., values in successive diagonals first increased and then decreased).

Two-sample MANOVA - Hotelling's T^2

Unlike the mixed-model case, in multivariate analysis the separate repeated measurements are considered as multiple criterion variables. They may have unequal variances and a general pattern of correlations. The assumption is that this general covariance matrix is common across groups. To test for differences among the dependent variables, the original variables are transformed into contrasts of interest. Hotelling's T^2 statistic is the multivariate analog to the t-test, and is the uniformly most powerful test for comparing two groups on p variables (Anderson, 1958, pp. 115-118). Several researchers have found it to behave in a fashion similar to the t-test under heterogeneity conditions.

In an empirical study using Monte Carlo methods with relatively small samples ($N = n_1+n_2$ ranging from 10 to 40), Hopkins and Clay (1963) examined the Type I error rates of

Hotelling's T^2 statistic for testing the equality of two independent mean vectors in the $p = 2$ case. The two populations studied were $N(\underline{0}, \sigma_1^2 I)$ and $N(\underline{0}, \sigma_2^2 I)$, where heterogeneity between covariance matrices was present equally in both canonical dimensions and of the form $\sigma_2^2/\sigma_1^2 = 1.6$ and 3.2. Under these circumstances, they found that with $n_1 = n_2 > 10$, heterogeneity had little effect on test results, but that, as in the univariate case, this robustness does not extend to unequal sample sizes. Everything else being equal, the greater the heterogeneity, the greater the departure of the observed significance level from the nominal alpha level. Furthermore, regardless of the amount of heterogeneity, the T^2 test was conservative if the larger group had more variability and liberal if it had less variability.

Another empirical study of the effect of inequality of covariance matrices and of sample size on the distribution of Hotelling's T^2 statistic was conducted by Holloway and Dunn (1967). They considered both level of significance and power with number of variables ranging from one to 10 and total sample sizes from five to 100. In canonical form, the covariance matrix for one population was equal to the identity, I, and for the other it was either dI or diag(d,1,...,1), with $d = 1$, 1.5, 3, 10, and 100. They confirmed the robustness of T^2 for $p = 2$ as found by Hopkins and Clay and concluded that the actual level of significance increases when any of the following occur: (1) number of variables increases, (2) total sample size with equal groups decreases, or (3) number of heterogeneous dimensions increases (i.e., all $d_i = d$). They also stated that "equal sample sizes help in keeping the level of significance close to the supposed level, but have little effect in maintaining the power of the test" (p. 125). In general, power was often considerably reduced by departures that left the significance level satisfactory.

In a third empirical study of the robustness of T^2, with p = 2, 6, or 10, Hakstian, Roed, and Lind (1979) did not use covariance matrices in canonical form. However, all variances in one population were equal to one and covariances had an irregular pattern. Two distinct matrices were used for a second population, where all elements were greater than in the first by a factor of 1.44 or 2.25. For two variates, robustness was evident with equal sample sizes as small as six. For unequal sample sizes, their results paralleled the previous studies. Additionally, they found that increasing the total sample size while keeping the ratio of sample sizes constant does not help, and may actually hurt, the situation.

In summary, while the T^2 test is robust to covariance matrix heterogeneity with equal n's, it is not robust with unequal n's. The latter is true even for relatively mild departures from equality of the covariance matrices and of sample sizes.

General MANOVA Test Statistics

The MANOVA tests discussed in this section are all functions of the eigenvalues of HE^{-1}, where H and E are the hypothesis and error SSCP matrices. For two groups, the tests are equivalent. Hotelling's T_o^2 is a generalization of the T^2 test, which may be used with more than two groups. The T statistic is often used in place of T_o^2, since they are directly related (i.e., $T_o^2 = df_e T$). Robustness studies of multivariate test statistics for more than two groups have shown that, in general, these test statistics behave in a manner comparable to the univariate F-test.

One of the earliest and most cited theoretical studies of multivariate robustness to heterogeneity of covariance matrices was conducted by Ito and Schull (1964). They investigated the asymptotic distribution of Hotelling's T_o^2 statistic, with one to four variables and two to five groups. For the case of two large samples of equal size, they showed analytically that the test is fairly well behaved, with respect to both significance level and power, in the presence of heterogeneity. Also, for samples of nearly equal size, robustness holds as long as the characteristic roots of $\Sigma_2\Sigma_1^{-1}$ fall in the range (.5,2). For two large samples of unequal size, the departure from a nominal alpha level of .05 increased as: (1) the ratio of sampl

of he

depar

incre

ther

effe

Howe

even

effec

test.

signi

the s

had t

Hotel

Roy's

10), R

covari

B(d) w

one po

for k

represe

k = 3

the th

they w

other

sample sizes ($r = n_1/n_2$) departed from one, (2) the degree of heterogeneity (d = the characteristic roots of $\Sigma_2\Sigma_1^{-1}$) departed from one, or (3) the number of dependent variables increased. For more than two groups and equal samples, there was a tendency to overestimate significance, but the effect was not serious with moderate heterogeneity. However, if one or some of the groups were of unequal size, even moderate heterogeneity conditions produced large effects on the significance level and the power of the test. In all cases with unequal sample sizes, actual significance was greater than .05 if the larger group had the smaller variance and less than .05 if the larger group had the larger variance.

In an empirical study of the robustness properties of Hotelling's T, Wilks' likelihood ratio criterion W, and Roy's largest root R with small equal samples ($n = 5$ or 10), Korin (1972) specified departure from equality of covariance matrices in two ways, symbolized by A(d) and B(d) with $d = 1.5$ or 10. A(d) represents cases where only one population covariance matrix differed (i.e., (I,I,dI) for $k = 3$ and (I,I,I,I,I,dI) for $k = 6$), while B(d) represents cases where two differed (i.e., (I,dI,2dI) for $k = 3$ and (I,I,I,I,dI,2dI) for $k = 6$). Results showed that the three tests were somewhat comparable and that, although they were all liberal, R tended to be more so than did the other two. The discrepancy between nominal and actual values was slight with small violations of covariance homogeneity (d = 1.5), but was pronounced with larger violations (d = 10). This indicates that, unlike the large sample case, with small n even equal samples do not guarantee robustness.

A very extensive Monte Carlo investigation of the performance of six multivariate test criteria under heterogeneity conditions was conducted by Olson (1973, 1974). He considered groups of equal sizes (n = 5, 10, and 50) with both number of variables and of groups equal to 2, 3, 6, and 10. With populations having distributions $N(\underline{0},I)$ or $N(\underline{0},D)$, he used two types of contaminating covariance matrices (i.e., where all canonical dimensions varied equally, D = dI, or where only one dimension varied, $D = \text{diag}(pd-p+1,1,\ldots,1)$), with d = 4, 9, or 36. For a given value of d, total variability in both matrices, as measured by the trace of D, were equal. Therefore, only the manner in which variability is allocated was varied for a given d, and not the total variability. The latter being varied by different choices of d.

Under various combinations of these factors, Olson examined Type I error rates and power of Roy's largest root, R, two trace-type tests (Hotelling-Lawley's T and Pillai-Bartlett's V), and three determinental tests (Wilks' likelihood ratio, W, Gnanadesikan's criterion, U, and Olson's alternative criterion, S). The U and S tests

tend
favo
furt

test
libe
susp
asym
that
erroi
hypot
were
and W
thoug
robus
equal
overe
exten

subst
four
Chapt
excee
dimen
increa
Also,

tended to be quite conservative and did not respond favorably to violations and so will not be discussed further.

Olson concluded that, although the remaining four tests all tended to be liberal, the R test was far too liberal and should be rejected if any heterogeneity is suspected. For large samples, the V, W, and T tests are asymtotically equivalent and he suggests as a rule of thumb that they may be so considered whenever degrees of freedom error are at least 10p times larger than degrees of freedom hypothesis. For smaller samples, the T, W, and V tests were robust against mild heterogeneity, but in general, T and W did not fare as well. Findings showed that even though it tended to be liberal, the V test was the most robust under the conditions examined. These results with equal samples uphold Korin's (1972) conclusions of overestimation of significance level for small samples and extend them to even moderately large samples ($n = 50$).

Although departures from assumptions have substantially different effects on the distributions of the four test statistics to be considered in this study (see Chapter IV), general conclusions for equal samples are that exceedance of nominal alpha may be decreased by reducing dimensionality, p, or number of groups, k. However, increasing sample size with equal n's does not always help. Also, even though the percentage exceedance tended to be greater at larger nominal alpha, Olson (1974) found that "different proportions of contamination showed their effects in much the same way at all three significance levels" (i.e., for .01, .05, and .10) (p. 898). In general, exceedance rates increased with greater heterogeneity, but they "tended to increase more as d increased from 1 to 4 and from 4 to 9 than as it increased from 9 to 36" (p. 898). Furthermore, regardless of p and k, effects were relatively minor when only one canonical dimension varied (D = C(d)) but severe when they all did (D = dI).

For situations where D = dI, larger n's corresponded to lower exceedance rates for R, T, and W whereas for V, rates either decreased or increased as necessary to converge to T and W for large n. This is due to the fact that, for small n, V was significantly better than the other tests in many of the cases. It should be noted that, for equal samples, when D = dI "effects of kurtosis and heterogeneity tend to be in opposite directions, the former yielding conservative rates and the latter producing too many significant results" (Olson, 1974, p. 901).

With respect to power, differences among the R, T, V, and W statistics were typically small. However, the R statistic tended to have slightly higher power if differences in the population mean vectors were confined to one of the s dimensions, while the V statistic had a slight advantage if the differences were equally pronounced in all the s dimensions. Furthermore, holding the noncentrality parameter constant, increasing the number of groups tended to decrease power, while increasing group size had no consistent effect on power.

Another Monte Carlo study on the significance levels of R, T, W, and V test criteria with equal n's, but where heterogenetity was modeled on the original covariance matrices and not on the canonical dimensions, was conducted by Ceurvorst (1980). He considered a variety of situations that included varying the number of dependent variables (2 and 3), number of groups (2, 3, and 6), degrees of freedom error (18, 60, and 180), and both type and degree of heterogeneity. For differences of type, he considered inequality of variance alone, of correlations alone, and of both together, with combinations of three variances (1, 4, and 9) and three correlations (.2, .5, and .8).

For heterogeneity of correlation he found only mild liberal exceedance rates for the four test statistics using a .05 nominal alpha. The observed significance levels were always less than .09 and proved to be fairly robust in most cases. Results for heterogeneity of variance confirmed previous results for canonical forms, plus indicated that the effects did not depend on the magnitude of the common within-group correlation(s) for any of the cases considered. Comparisons among the test criteria showed that they were consistently ordered R-T-W-V from highest to lowest exceedance rate of the nominal alpha. The most serious discrepancies occured when $k = 6$ and five groups had variances equal to unity, while the sixth had variances equal to nine.

When both heterogeneity of variance and of correlation were present, results differed depending on the relative size of variances and correlations. If groups with the largest variances had the largest correlations (LVLC), violations became increasingly more serious than for heterogeneity of variance (HV) alone. If groups with the largest variances had the smallest correlations (LVSC), the reverse was true, with violations being less serious. Comparisons of the criteria under LVSC conditions were similar to the HV situations with the V test being uniformly most robust, followed in order by W, T, and R. Under the LVLC conditions, no criteria was uniformly best. When only one variance differed, R was often the best choice, but it was the worst when all variances differed. Also, when R was best, the other tests generally had exceedance rates that were .07 or less.

Pillai and Sudjana (1975) studied the effects of unequal covariance matrices on the R, T, V, and W statistics in the exact case by deriving central and noncentral distributions and applying them in a numerical study with $n = 5$, 15, and 40. Considering $p = 2$, they stated that low heterogeneity produces modest changes in the powers of the test statistics, but that changes become pronounced as heterogeneity increases. None of the four statistics showed an advantage over the rest.

In summary, the discrepancy between actual and nominal alpha tends to decrease with lower degrees of heterogeneity, and with smaller number of variables and of groups. It appears that, for two equal samples, neither the significance level nor the power of Hotelling's T^2 is seriously affected by heterogeneity, but that this is not necessarily true for unequal n's. For more than two groups of large equal samples, robustness may be achieved with moderate departures from homogeneity, but even moderate heterogeneity produces large effects on both significance level and power when samples are unequal. For several small or moderately large groups, even equal samples do not protect against departure from nominal significance levels, with test criteria tending to be liberal. Consequences of violation of the homogeneity assumption through a contaminating covariance matrix is generally worse if all canonical variances differ by an equal amount than when only one differs. The case of only some equally discrepant variances falls between the two extremes. In general, Roy's largest root, R, appears to be the worst of the invariant tests and Pillai-Bartlett's trace, V, the best, with respect to both robustness and power.

CHAPTER IV

METHOD

Previous work exploring the robustness of MANOVA test criteria to violation of homogeneous covariance matrices across groups has dealt only with fixed-effects between-group tests in a one-way classification. In the present research, the effect of violating the assumption of homoscedasticity was considered in a repeated measures (RM) situation with data from ordered time points and a fixed-effects, one-way design over the subjects. The purpose of the study was twofold: (1) to compare the robustness of multivariate test statistics for between-group and within-group tests, and (2) to analyze the behavior of within-groups tests under various conditions with respect to both robustness and power.

When data in the p-variate response measures reflect the passage of time, and assuming no group by measures interaction, overall within-group tests encompass all p-1 degrees of freedom (df) for trend, thereby testing the null hypothesis of no trend in the data or, equivalently, of equality of occasion means. Subsequent tests may be confined to any p-q-1 degrees of freedom (df) remaining after a $q \leq p$ degree trend is hypothesized.

In this chapter, details are presented about the use of covariance matrices in canonical form for RM analyses, the parameters and procedures for the study, and the Monte Carlo techniques that were used.

Reduction to Canonical Form

The assumption of homoscedastisity for tests of between-group differences relates to population covariance matrices for the original score vectors. For simplicity, canonical forms of the covariance matrices are typically used in MANOVA robustness studies (see Chapter III).

For MANOVA of RM, the score vectors are linearly transformed to reflect the design on the measures. The transformed vectors remain multivariate normal if the original vectors are multivariate normal (Finn, 1974, p. 62) and the assumption now relates to the transformed covariance matrices.

With time ordered data, the transformation consists of a matrix of normalized orthogonal polynomial coefficients. When such a matrix is applied to populations with covariance matrices reduced to the canonical forms I , dI, or $C(d) = diag(d,1,\ldots,1)$, the transformation does not alter I or dI. Although C(d) becomes a general matrix, it is reduced to C(d) when diagonalized. Therefore, the same underlying violation is modeled for both between-group and within-group tests when the original covariance matrices are in canonical form.

mu

nu

ho

so

be

no

pa

to

vi

co

Tes

set

gro

wer

p-v

tre

ove

hig

nul

of a

Test

stat

larg

trac

Parameters of the Study

A major problem in any study of robustness of multivariate test statistics comes from the seemingly vast number of ways in which the assumption of covariance homogeneity can be violated and the many factors that have some bearing on levels of robustness. Therefore, it becomes necessary to specify these factors and nonconforming populations in terms of some relevant parameters and to choose particular levels of each in order to have a systematic coverage of different forms of violations under various conditions. The parameters considered in this study are described in this section.

Tests of Multivariate Hypotheses. For each simulated data set, tests of a between-group hypothesis and two within-group hypotheses were conducted. The hypotheses tested were: (1) the null of no between-group differences on p-variate mean vectors, using k-1 df, (2) the null of no trends across the p-variate data, using p-1 df, for an overall test on occasions, and (3) the null of no trend higher than linear, using p-2 df. Rejection of the second null hypothesis, but not of the third implies the existence of a linear trend across time.

Test Criteria. For each hypothesis, four multivariate test statistics, defined in Table 2-4, were calculated: Roy's largest root, R, Hotelling-Lawley trace, T, Pillai-Bartlett trace, V, and Wilks' likelihood ratio, W.

Number of Measures, p. Experiments were simulated with p = 4 or 5 response measures. This enabled both within-group tests to be multivariate. Since a multivariate test for linearity uses SSCP matrices of order p-2, the smallest value for p that allows for such a test is four. The SSCP matrices for hypothesis and error were: (1) of order-4 or 5 for between-group tests, (2) of order-3 or 4 for overall occasion tests of no trends, and (3) of order-2 or 3 for tests of no trends higher than linear.

Number of Groups, k. The simulated experiments had simple one-way fixed designs on the independent factor with k = 2, 3, or 6 groups.

Group Size, n. Small to moderately large experiments were simulated, with n = 10, 20, or 50 in each group. In all cases, groups of equal size were considered.

Type of Heterogeneity. The identity matrix, I, was used to model homogeneous populations. For heterogeneity conditions, two populations with covariance matrices equal to I and dI were used. This type of diffuse structure was chosen for the contaminating matrix because it is the kind of violation that typically produces the most severe departures from nominal significance levels.

Degree of Heterogeneity, d. This factor relates to the size of the violation (i.e., to how much more variable one distribution is relative to another). Small to large violations were modeled, with d = 2, 4, or 9. For homogeneity conditions, d = 1.

Significance Level, α. The probablity of making a Type I error was considered at the .01, .05, and .10 nominal alpha levels. For a given nominal level, (100α)% of the values in a test statistic's distribution will exceed the appropriate critical value under a true null with no assumption violation. Hence, a dependent variable in the Monte Carlo experiments was the empirical estimate of a statistic's percentage exceedance of its critical value at significance level alpha, given a true null and heterogeneous covariance matrices. (The phrase _percentage exceedance_ is used throughout the thesis to refer to the percentage of replications of a statistic that exceed a critical value).

Power. This is equal to 1 - P(Type II error). Nominal power relates to the percent of values in a test statistic's distribution that will exceed the critical value under a true alternative with no assumption violation. A second dependent variable in the Monte Carlo experiments was the empirical estimate of actual power (i.e., the percentage exceedance given a true alternative and heterogeneous covariance matrices). This was conducted at all three nominal alpha levels.

Power is a function of the discrepancy between central and noncentral distributions for a test statistic. The MANOVA noncentrality parameter (ncp) is a standardized measure of the distance between group means in the population (Olson, 1974) and may be defined as the sum of the eigenvalues, g_j ($j = 1,\ldots,p$), or trace of a matrix G, where

$$G = FV^{-1}.$$

V is the population covariance matrix and

$$F = \sum_{i=1}^{k} n_i(\underline{\mu}_i - \underline{\mu})(\underline{\mu}_i - \underline{\mu})',$$

where $\underline{\mu}_i$ is the population mean vector for the i^{th} of k groups and $\underline{\mu}$ is the grand mean vector in the population. When data are ordered according to time, the ncp for tests of within-group hypotheses incorporates the time dimension. This is done by representing the elements of the covariance matrix and the means in the above equations as functions of time (Morrison, 1972).

Since power depends on the common covariance matrix V, no theoretical power values exist under heterogeneous conditions, and the choice for V is open. Therefore, the noncontaminated covariance matrix (I in canonical form) is typically used for V in order to calculate the ncp. In this way, a comparison can be made between a test's ability to detect differences when assumptions are violated to its ability to do so when they they are met.

Procedures

Monte Carlo techniques were used to generate either 10,000 or 2,000 replications of multivariate data sets distributed $N(\underline{0},I)$. Each data set represented a particular combination of k and n with $p = 5$ measures across time. Using these data, critical values were calculated for tests of three multivariate hypotheses using four test statistics at three nominal alpha levels. The data in each set were then transformed seven times to calculate: (1) actual significance levels in three central heterogeneous cases for between-group and within-group tests, (2) nominal power in a noncentral homogeneous case for within-group tests, and (3) actual power in three noncentral heterogeneous cases for within-group tests. All calculations were performed a second time on the same data sets using only the first four measures to simulate conditions with $p = 4$. Since noncentral situations refer only to tests of within-group differences, in these cases, the null hypotheses of no group by occasion interaction and of no group differences remained true.

A FORTRAN V program was written to generate, transform, and analyze the data. A detailed description of the computational procedures appears in Appendix A. These procedures guided the creation of the computer program, which also appears in Appendix A. The remainder of this section describes the determination of critical values, the design for the study, the analysis procedures, and the interpretation of computed significance levels and power values.

Determination of Critical Values

Critical values for the multivariate test criteria and the combinations of p, k, n, and alpha levels used in the study were not all available in published tables. Also, tabled values have generally been obtained analytically rather than empirically. Therefore, values used in the study were empirically determined via Monte Carlo techniques.

Using three nominal significance levels, critical values were calculated such that $(100\alpha)\%$ of the N noncontaminated replications (where N = 10,000 or 2,000) under a true null would be judged significant using that critical value. This was accomplished by taking the arithmetic average of the $(N\alpha)$th and $(N\alpha+1)$th smallest of the N values for W and the corresponding largest of the N values for R, T, and V. Values thus obtained will be referred to as Monte Carlo critical values to distinguish them from tabled values.

Design for the Study

The design for the study is given in Table 4-1, where combinations of k and n used for all levels of p and d are denoted by an X in part (a). Hypotheses tested under central and noncentral conditions with four statistics at three nominal levels are indicated in part (b). The matrix in part (c) shows how the two types of conditions from (a) and (b) were combined to create the necessary statistics.

Table 4-1

Design for the Study

a) Number of measures (p), of groups (k), and equal sample sizes (n) under heterogeneity conditions (d). *

Condition	d	k: / n: / p	k=2, n=10	20	50	k=3, n=10	20	50	k=6, n=10	20	50
Homogeneity	1	5		X		X	X	X		X	
		4		X		X	X	X		X	
Heterogeneity	2	5		X		X	X	X		X	
		4		X		X	X	X		X	
	4	5		X		X	X	X		X	
		4		X		X	X	X		X	
	9	5		X		X	X	X		X	
		4		X		X	X	X		X	

* X indicates conditions replicated 2,000 times.
Conditions replicated 10,000 were with k = 3 and n = 20.

b) Statistics calculated under central and noncentral conditions for various hypotheses at three nominal alpha levels and for every combination of factors indicated in (a). *

Condition	Hypothesis	Alpha: .01 — Statistic: R T V W	.05 — R T V W	.10 — R T V W
Central	B			
	C			
	L			
Noncentral	C			
	L			

* Hypotheses tested were: between-group differences, B, within-group test of trends, C, within-group test of trends higher than linear, L; using test statistics: Roy's largest root, R, Hotelling-Lawley trace, T, Pillai-Bartlett trace, V, and Wilks' likelihood ratio, W.

Table 4-1 (con't.)

c) Empirical values derived from each replicated data set by crossing elements from conditions on covariance matrices in (a) and conditions on hypotheses in (b).

		Condition on Covariance Matrices	
		Homogeneity	Heterogeneity
Condition on Hypotheses	Central	Monte Carlo critical values	Actual significance levels
	Noncentral	Nominal power values	Actual power values

For the first part of the study, five-variate vector scores from a population distributed $N(\underline{0},I)$ were generated for 10,000 replications of one situation with three equal groups of size 20. Four-variate situations were simulated by using the same data and dropping the fifth measure in each vector score.

For the second part of the study, new sets of 2,000 replications from the same population were generated for each of the five combinations of k and n indicated in Table 4-1(a). Equal cell sizes were used throughout the study and the same procedures followed for every combination of p, k, and n, regardless of the number of replications.

Calculated statistics from the data in each set of N replications under homogeneous conditions were used to determine Monte Carlo critical values for all combinations of multivariate tests, test statistics, and nominal alpha levels shown under the central case of Table 4-1(b). Regardless of the number of groups represented, score vectors for only one group in each case were transformed to simulate data that might arise from populations distributed $N(\underline{0}, dI)$. The data sets represented central heterogeneous conditions, and all test statistics were recalculated.

Each resulting value was then compared to the corresponding Monte Carlo critical value for the three alpha levels considered. Actual significance levels (i.e., empirical Type I errors under heterogeneity) were determined by counting the number of values in each replication that were: (1) greater than the corresponding critical value for R, T, and V statistics, and (2) less than the corresponding critical value for W statistic, and then dividing by N, the number of replications.

To investigate the power of multivariate within-group tests under true alternatives for the occasions, the original noncontaminated data sets (with $d = 1$) were transformed to reflect a given curvilinear trend across time. Under homogeneity and an alternative condition for within-group tests only, the above calculations were again performed to determine Monte Carlo nominal power values for tests of within-group differences.

The final step in the process was to add the curvilinear trend to the heterogeneous data sets and repeat the calculations to determine actual powers for within-group tests under noncentral heterogeneous conditions. By comparing these values to those for nominal power, the effects of heterogeneity on the power of within-group tests under an alternative hypothesis could be evaluated.

Comparison of Tests

In order to empirically determine whether between-group and within-group test statistics respond differentially to identical heterogeneity conditions under true null hypotheses, one experimental situation with $k = 3$ and $n = 20$ was replicated 10,000 times. The large number of replications was used in order to insure relatively small standard errors.

The main interest in a comparison between actual significance levels for the group and occasion tests was examined from two perspectives. First, tests were compared within a given p to simulate practical analyses where tests of both hypotheses are performed on the same data set. However, discrepancies between actual significance levels evidenced here might occur because group and occasion tests are based on SSCP matrices of order-p and p-1, respectively. Therefore, a second comparison was made between the group tests with $p = 4$ and the occasion tests with $p = 5$, so that both would be based on order-4 SSCP matrices.

Analysis of Within-group Tests

Using the same 10,000 replications, comparisons of actual significance levels were made between the two sets of within-group tests for general trends and for trends higher than linear. With the data modified to reflect true alternatives for within-group hypotheses, the power to reject the null under heterogeneity was also evaluated.

The second stage of the research was an attempt to examine the effects of heterogeneity on within-group tests when the number of groups and of equal sample sizes are varied. Both robustness and power were considered with 2,000 replications each for five combinations of k and n. Tests of between-group differences in the central case were also made in order to determine if discrepancies between these tests and within-group tests were sensitive to changes in number of groups and sample size.

Interpretation of Obtained Probability Values

The critical values and probability levels for significance (Type I error) and power were obtained via Monte Carlo methods and are therefore subject to sampling error. To take this error into account, the standard error (S.E.) of a proportion for a sample size equal to the number of replications was employed.

The S.E. for a proportion depends on the true value of the proportion, P, and is equal to $(P(1-P)/N)^{1/2}$, where N

equals the number of replications. Since the true value of P (i.e., nominal alpha) is known, this formula may be used to calculate the S.E. at the three nominal alpha levels considered. These are given in Table 4-2.

Table 4-2

Standard Errors for Nominal Alpha Levels and Number of Replications Used in the Study

Alpha	N = 2,000	N = 10,000
.01	.0022	.0010
.05	.0049	.0022
.10	.0067	.0030

Monte Carlo Techniques

The methods for exploring the issues of robustness in this study involved the use of simulated data generated by computer algorithms. Through the analysis of a large number of samples under known population parameters, one can investigate the properties of statistics by observing their resulting distributions. These empirical distributions obtained under heterogeneity are then compared to the nominal distributions obtained under homogeneity for the statistics in question. The FORTRAN program was used to generate either 10,000 or 2,000 samples of vector observations for each experimental condition and to perform the required data transformations and analyses. The procedures followed are specified in this section.

The required data were 5x1 vector observations, normally distributed with known mean vector and covariance matrix. The generation and transformation procedures consisted of three steps:

1) Generate a set of independent random observations uniformly distributed on the interval 0 to 1.

2) Combine the uniform variates to create a set of observations normally distributed with mean vector zero and covariance matrix equal to the identity.

3) Transform these observations to obtain the desired structure with mean $\underline{\mu}$ and covariance matrix V.

Each step will be considered separately.

Random Number Generation

Hammersley and Handscomb (1964) stated that "the essential feature common to all Monte Carlo computations is that at some point we have to substitute for a random variable a corresponding set of actual values, having the statistical properties of the random variable" (p. 25). These values are called random numbers. In practice, what is actually produced via computer programs are a set of pseudo-random numbers calculated sequentialy from a completely specified algorithm. This algorithm is devised in such a way that a statistical test should not detect any significant departure from randomness.

The subroutine GGUBS from the International Mathematical and Statistical Library was used to obtain a sequence of uniform random numbers $U_1,\dots,U_n$ distributed U(0,1). This routine uses a congruential generator based on the following relation

$$X_i = aX_{i-1} \pmod{m}$$

where $a = 7^5$ and $m = 2^{+31} - 1$. Once the procedure is started by an initial seed value, each X_i is determined from the previous value. The constant terms a and m are chosen so as to maximize the period of the generator, since a sequence repeats itself when a value for X_i reappears.

The numbers $U_i = X_i/2^{31}$ are a pseudo-random sequence in the interval 0 to 1. They are independent of each other and behave as if they were random.

Creation of Normal Deviates

Several approaches are available to create independent normal deviates from uniform random numbers. A simple approach to program is based on the Central Limit Theorem (CLT) and uses a summation of a fixed number of values, where this number may be as low as 12 for reasonable approximations. However, the procedure "is very slow and it does not adequately sample in the extreme tails of the normal distribution" (Lehman, 1977, p. 148).

The method used in this study for generating normal deviates from independent random numbers, which was devised to be reliable in the tails, was suggested by Box and Muller (1958). They cite a detailed comparison with several other methods, including the Central Limit summation, and state that their approach gives higher accuracy and compares favorably in terms of speed. The procedure uses a pair of random numbers U_1 and U_2 from the same distribution on the interval (0,1) to generate a pair of normal deviates from the same normal distribution, N(0,1). The following transformations are used:

$$z_1 = (-2\log_e U_1)^{1/2}\cos 2\pi U_2$$

$$z_2 = (-2\log_e U_1)^{1/2}\sin 2\pi U_2$$

The resulting values are a pair of independent random variables, normally distributed with zero mean and unit variance.

Vectors of five such variables taken together represent 5x1 observational vectors, which are multivariate normal and distributed $N(\underline{0}, I)$ (Anderson, 1958, pp. 19-27). Observations of this form were used to simulate the central case with homogeneous covariance matrices.

Transformation to Desired Structure

The first step to determine a vector with specified variances and intercorrelations among the variables is to factor a known covariance matrix V into a lower triangular matrix T such that

$$V = TT'$$

This is the square root method or Cholesky factorization of a symmetric positive-definite matrix, V (Bock, 1975, p. 85). Then, a transformation of a vector of normal deviates $\underline{z}$,

$$y = T\underline{z} + \underline{\mu}$$

produces a normally distributed vector y with the desired characteristics, since

$$\mathrm{Var}(y) = T(\mathrm{Var}(\underline{z}))T' = TT' = V$$

when $\mathrm{Var}(\underline{z}) = I$. The only effect due to adding a known vector of means $\underline{\mu}$ is to change the point of central tendency for the distribution of $\underline{y}$.

In the present study, where $V = dI$,

$$T = d^{1/2}I$$

and therefore,

$$\underline{y} = (d^{1/2}I)\underline{z} + \underline{\mu}$$
$$= d^{1/2}\underline{z} + \underline{\mu}$$

was the transformation used for one group to simulate data from heterogeneous populations in the noncentral case. Other transformations used the above equation with (1) $\underline{\mu} = \underline{0}$ for the central heterogeneous case, and (2) $d = 1$ for the noncentral homogeneous case.

After generating the data, the program performed the required multivariate tests, calculated the critical values, and tabulated the proportion of times the values of each statistic exceeded its critical value for a given nominal significance level when: (1) a null hypothesis was true, and (2) an alternative hypothesis was true. Obtained proportions were the actual Type I error rates and powers, respectively, for the statistics. Multiplying these obtained values by 100 produces percentage exceedance rates under heterogeneity.

CHAPTER V

RESULTS

The results of the study are presented in this chapter in four sections. The first two sections are based on 10,000 replications of experiments with $k = 3$ and $n = 20$ and deal first with comparisons of multivariate between-group and within-group tests with respect to robustness and then with the power of within-group tests under heterogeneity of group covariance matrices. The latter two sections present the effects of varying sample size and number of groups first on the robustness and then on the power of within-group tests under heterogeneity conditions. Results in these latter sections are based on 2,000 replications for each of five combinations of k and n.

Critical values for each set of N replications (where $N = 10{,}000$ or 2,000) under central homogeneous conditions were obtained empirically through Monte Carlo methods and are tabled in Appendix B. Actual significance levels under central heterogeneous conditions and powers under noncentral conditions were calculated by determining the number of times obtained test statistics exceeded the corresponding critical values and then dividing by the number of replications. These empirical values were multiplied by 100 and are reported in this chapter in terms of percentage exceedance rates of the Monte Carlo critical values.

Comparison of Tests on Robustness

The objective for this portion of the study was to determine whether heterogeneity of group covariance matrices produces differential effects for multivariate tests of between-group and within-group differences. The question could be phrased: Given no interaction effects and no main effects for either group or occasions in the populations from which the data are sampled, are there differences in the frequency with which rival test statistics indicate a significant effect for tests of between-group and within-group hypotheses under heteroscedastic conditions? A secondary question relates to differences between two within-group hypotheses (i.e, of no trends in the occasion means and of no trends higher than linear).

The situation considered was that of three equal groups of size 20 with either five or four measures across occasions. The procedures for central conditions, which were detailed in Chapter IV, were followed.

Since the data were randomly generated using computer algorithms, random error in the data must be considered. To insure that this error be small, 10,000 replications were used. Given known parameters (i.e., nominal alpha levels), the standard error of a proportion with 10,000 replications (see Chapter IV) may be used to calculate 95% probability intervals around the known parameters instead of confidence intervals around the sample estimates. This produces the following intervals for the three nominal levels considered:

$$.01 \pm .0020$$
$$.05 \pm .0043$$
$$.10 \pm .0059$$

Expressed in terms of percentages, to correspond with tabled values, the 95% probability intervals are:

(0.80, 1.20)	at	.01
(4.57, 5.43)	at	.05
(9.41,10.59)	at	.10

Thus, obtained percentage values within these intervals may be considered to be within sampling error of nominal percentages.

Critical values were estimated with Monte Carlo methods and, therefore, are subject to error. Since exceedance rates were derived from the same data sets transformed to heterogeneous conditions, the deviations from nominal levels in the following tables reflect only added error due to heterogeneity.

As far as possible, parameters used in this part of the study will be discussed separately in terms of their effects on the percentage exceedance of Monte Carlo critical values for the three hypotheses under investigation (i.e., of no between-group differences, B, of no trend over occasions, C, and of no trend higher than linear, L). Table 5-1 contains the actual percentage exceedance rates (i.e., empirical Type I error times 100) for central heterogeneous situations.

Significance Level, α. Percentage exceedance rates for all three hypotheses tested increased with larger nominal alpha levels, except where obtained values were within 95% probability intervals of the nominals. Although the patterns were similar, increases in exceedance rates were greatest for the between-group tests, B, and lowest for the within-group tests of no trends higher than linear, L.

However, when tests for a given hypothesis are considered with respect to standard errors, which also increase with alpha level, different amounts of heterogeneity showed consistent effects regardless of significance level. For example, at all three alpha levels, departures from the nominal for tests of B ranged from about one standard error with the V statistic at d = 2 to over 50 times the standard error with the R statistic at d = 9. Departures for the within-group tests were typically around one standard error with all test statistics at d = 2, and never exceeded 13 standard errors for tests of C and eight standard errors for tests of L at d = 9. The larger numerical values for exceedance rates as alpha increases is apparently a function of corresponding larger standard errors.

Table 5-1

Percentage Exceedance Rates of Monte Carlo Critical Values
for Multivariate Tests Under True Null Hypotheses*

			α = .01				α = .05				α = .10			
p	H_0	d	R	T	V	W	R	T	V	W	R	T	V	W
5	B	2	1.57	1.27	1.11	1.21	6.42	5.50	5.24	5.32	11.77	10.68	10.17	10.38
		4	3.59	2.31	1.56	2.01	10.84	7.85	6.15	7.00	17.42	13.48	11.72	12.70
		9	7.88	4.65	2.47	3.67	18.18	11.55	7.75	9.81	26.11	18.06	13.18	15.81
	C	2	1.30	1.18	1.11	1.12	5.06	5.15	4.93	5.05	10.52	10.34	10.26	10.38
		4	1.54	1.40	1.34	1.40	5.93	5.83	5.71	5.79	11.74	11.47	11.21	11.40
		9	2.17	1.98	1.83	1.93	7.52	7.44	7.16	7.30	13.79	13.18	12.67	12.95
	L	2	1.04	1.15	1.11	1.13	4.93	4.92	4.89	4.90	10.10	10.19	10.37	10.25
		4	1.34	1.28	1.24	1.23	5.51	5.27	5.27	5.30	10.74	10.85	10.97	10.91
		9	1.82	1.69	1.50	1.63	6.71	6.39	6.37	6.38	12.28	12.07	12.05	12.00
4	B	2	1.63	1.48	1.28	1.39	6.22	5.71	5.35	5.47	11.21	10.56	10.30	10.34
		4	3.65	2.49	1.83	2.19	10.02	7.82	6.48	7.20	15.74	12.83	11.61	12.28
		9	6.89	4.37	2.56	3.68	15.82	10.98	7.96	9.47	21.98	16.22	13.60	15.20
	C	2	1.09	1.09	1.04	1.07	5.01	4.99	5.00	4.93	10.13	10.09	10.07	10.07
		4	1.37	1.35	1.33	1.34	5.73	5.65	5.69	5.61	10.60	10.39	10.30	10.40
		9	1.94	1.80	1.72	1.72	6.92	6.81	6.65	6.69	12.00	11.52	11.54	11.56
	L	2	1.18	1.12	1.07	1.09	5.17	5.23	5.23	5.29	10.07	10.04	10.04	10.05
		4	1.28	1.24	1.23	1.24	5.45	5.45	5.48	5.50	10.66	10.78	10.75	10.75
		9	1.71	1.53	1.46	1.50	5.94	6.02	6.05	6.09	11.44	11.35	11.23	11.31

* From 10,000 replications of k = 3 equal groups of size n = 20 with p measures under d degree of heterogeneity. Hypotheses tested were: between-group differences, B; within-group trends, C; within-group trends higher than linear, L; using test statistics: Roy's largest root, R; Hotelling-Lawley trace, T; Pillai-Bartlett trace, V; and Wilks' likelihood ratio, W.

Number of Measures, p. Percentage exceedance rates were generally larger with five dependent variates than with four. Whenever this was not the case, discrepancies between corresponding exceedance rates at $p = 4$ and 5 were less than twice the standard error for a difference of two proportions. The smallest differences between exceedance rates for corresponding tests at the two levels of p occurred for the L tests. This may be due to the relatively small departures from nominal levels for tests of L, regardless of the number of variates.

Degree of Heterogeneity, d. In general, tabled values tended to be within 95% probability intervals of nominal values with low heterogeneity and, in all cases, the percentage exceedance rates increased with d. The effects of greater heterogeneity were the most pronounced for the B tests, where actual Type I error departed by as much as .16 from a nominal .10 level. However, discrepancies between actual and nominal values were less than .04 for the C tests and .02 for the L tests at a nominal .10 level.

Test Statistic. Considering low heterogeneity, percentage exceedance rates tended to fall within 95% probability intervals of nominal values with the V or W statistics when testing the between-group hypothesis, while they did so with all four statistics when testing either within-group hypothesis. As expected from previous research on the robustness of between-group tests (e.g., Olson, 1973), the four tests statistics were ordered V-W-T-R from best to worst when testing B. Differences between actual Type I errors for the V and R statistics for B were always greater than twice the standard error of a difference, reaching as high as .13 with high heterogeneity and five variates.

While results for tests of C generally followed the same ordering from best to worst statistic, those for tests of L did not. However, differences in departures from nominal levels among the statistics for tests of both within-group hypotheses were negligible, generally being less than twice the standard error of a difference and only once reaching a .01 difference. Except for tests of L, the effect of greater heterogeneity increased the differences between the best and worst statistic. This increase was considerably more pronounced for tests of between-group differences than for within-group tests of trends.

Tests of Multivariate Hypotheses. Exceedance rates for within-group tests tended to be within 95% probability intervals of nominal levels only under low heterogeneity. For between-group tests, this tended to be true only when the V or W statistics were used. To evaluate robustness in terms of acceptable Type I error, results were considered too liberal if they exceeded .015, .06, and .12 at nominal levels of .01, .05, and .10, respectively. Using these criteria when k = 3 and n = 20, only the between-group tests with T, V, and W statistics would be considered robust under low heterogeneity. For within-group tests, robustness would extend to all four statistics and to moderate heterogeneity ($d = 4$).

To summarize the differential effects of heterogeneity on tests of the three hypotheses and to examine more explicitly the differences among them, Table 5-2 provides differences in the actual percentage exceedance rates. The first two sets of rows relate to tests with a given p to simulate practical analyses with tests for both B and C performed on the same data sets. In the third set of rows the comparisons between B and C control for the size of the SSCP matrices from which the tests are derived, so that both sets of tests are based on matrices of order-4.

In the lower half of the table are presented similar comparisons for the two within-group tests. As before, the first two sets of rows relate to within-group tests with the same initial p, while the third set compares the tests with equal SSCP matrices of order-3.

The differences portray the extent to which departure from nominal levels were typically greater for tests on B than on C. Regardless of which set of comparisons was considered, the differences followed similar general patterns. The discrepancies in exceedance rates between tests on B and C tended to be less than two standard errors of a difference of two proportions when $d = 2$ or when the V statistic was used.

Table 5-2

Differences in Percentage Exceedance Rates
Under True Null Hypotheses*

		$\alpha = .01$				$\alpha = .05$				$\alpha = .10$			
$H_o(p)$	d	R	T	V	W	R	T	V	W	R	T	V	W
B(5)-C(5)	2	.27	.09	.00	.09	1.36	.35	.31	.27	1.25	.34	-.09	.00
	4	2.05	.91	.22	.61	4.91	2.02	.44	1.21	5.68	2.01	.51	1.30
	9	5.71	2.67	.64	1.74	10.66	4.11	.59	2.51	12.32	4.88	.51	2.86
B(4)-C(4)	2	.54	.39	.24	.32	1.21	.72	.35	.54	1.08	.47	.23	.27
	4	2.28	1.14	.50	.85	4.29	2.17	.79	1.59	5.14	2.44	1.31	1.88
	9	4.95	2.57	.84	1.96	8.90	4.17	1.31	2.78	9.98	4.70	2.06	3.64
B(4)-C(5)	2	.33	.30	.17	.27	1.16	.56	.42	.42	.69	.22	.04	-.04
	4	2.11	1.09	.49	.79	4.09	1.99	.77	1.41	4.00	1.36	.40	.88
	9	4.72	2.39	.73	1.75	8.30	3.54	.80	2.17	8.19	3.04	.93	2.25
C(5)-L(5)	2	.26	.03	.00	-.01	.13	.23	.04	.15	.42	.15	-.11	.13
	4	.20	.12	.10	.17	.42	.56	.44	.49	1.00	.62	.24	.49
	9	.35	.29	.33	.30	.81	1.05	.79	.92	1.51	1.11	.62	.95
C(4)-L(4)	2	-.09	-.03	-.03	-.02	-.16	-.24	-.23	-.36	.06	.05	.03	.02
	4	.09	.11	.10	.10	.28	.20	.21	.11	-.06	-.39	-.45	-.35
	9	.23	.27	.26	.22	.98	.79	.60	.60	.56	.17	.31	.25
C(4)-L(5)	2	.05	-.06	-.07	-.06	.08	.07	.11	.03	.03	-.10	-.30	-.18
	4	.03	.07	.09	.11	.22	.38	.42	.31	-.14	-.46	-.67	-.51
	9	.12	.11	.22	.09	.21	.42	.28	.31	-.28	-.55	-.51	-.44

* Differences were calculated from values in Table 5-1 for hypotheses of: between-group differences, B; within-group trends, C; within-group trends higher than linear, L; at d degree of heterogeneity, using test statistics: Roy's largest root, R; Hotelling-Lawley trace, T; Pillai-Bartlett trace, V; and Wilks' likelihood ratio, W. Differences with equal measures, p, relate to tests based on the same data sets, while those with unequal measures relate to tests based on the same size SSCP matrices of order-(p-1).

For a given statistic, differences in percentage exceedance rates decreased as either nominal alpha or heterogeneity decreased. Consistently the smallest differences between B and C tests occurred with the V statistic, typically being less than two percentage points. The largest occurred with the R statistic, where differences were as high a 12 percentage points. These patterns reflect that between-group tests tend toward robustness when homogeneity is low or when the V test statistic is used and that actual significance levels for the B tests increase considerably from V to R, while they remain relatively stable across the four statistics for the C tests.

Differences between the two within-group tests did not follow the patterns of the B and C differences. The discrepancies in percentage exceedance rates between tests on C and L were less than two standard errors of a difference with both $d = 2$ and 4, as well as in over half the cases with $d = 9$. Regardless of the alpha level, these differences were typically negligible and rarely exceeded one percentage point.

Power of Within-group Tests of Trends

The power of the tests to reject the null was evaluated under a homogeneous ($d = 1$) and three heterogeneous conditions. The original 10,000 data sets for three equal groups of size 20 were transformed to reflect the same quadratic trend over the five time points for each vector score. Since all the groups were equally transformed, this provides a situation with neither interaction nor between-group main effects, but with a within-group main effect. The percentage of rejections for the null hypotheses of no trend (C) and of no trend higher than linear (L) were determined.

As shown in Table 5-3, power values were quite stable across the four statistics within a given heterogeneity condition and alpha level. This is fairly consistent with previous findings for power of between-group tests under heterogeneity (e.g., Olson, 1973), where differences across the test statistics, although sometimes present, were relatively minor.

Regardless of heterogeneity, power was always larger at larger nominal levels. This trend follows what is expected under general homogeneity conditions, since "...setting alpha larger makes for relatively more powerful tests of H_0" (Hays, 1973, p. 359).

Within each nominal alpha level, power decreased as heterogeneity increased. For example, with p = 5, power of the C test at nominal .01 went from over 90% under homogeneity to around 30% with a high degree of heterogeneity (d = 9). At .05 and .10, power dropped from 98% and 99% to slightly over 50% and 65%, respectively. This downward trend was remarkably consistent among all

Table 5-3

Percentage Exceedance Rates for Within-group Tests
Under True Alternatives*

			$\alpha = .01$				$\alpha = .05$				$\alpha = .10$			
p	H_0	d	R	T	V	W	R	T	V	W	R	T	V	W
5	C	1	92.05	92.33	92.24	92.46	98.11	98.31	98.28	98.31	99.27	99.36	99.34	99.35
		2	79.78	79.78	79.36	80.00	92.51	93.30	93.19	93.25	96.56	96.84	96.75	96.83
		4	57.95	57.55	56.56	57.47	78.02	79.35	78.96	79.28	87.01	87.00	86.79	86.99
		9	31.76	30.82	29.52	30.74	52.98	53.98	52.99	53.54	66.14	66.01	65.17	65.73
	L	1	93.89	94.38	94.44	94.44	98.71	98.85	98.89	98.88	99.55	99.56	99.57	99.57
		2	82.43	83.57	83.52	83.62	94.35	94.85	95.02	94.92	97.63	97.83	97.87	97.83
		4	60.14	61.60	61.64	61.73	80.87	81.59	81.75	81.70	88.53	89.28	89.48	89.33
		9	31.63	32.83	32.60	32.72	54.51	55.35	55.40	55.37	67.51	68.33	68.59	68.46
4	C	1	85.99	86.62	86.52	86.50	95.81	96.31	96.37	96.27	98.08	98.18	98.19	98.20
		2	71.05	72.04	71.89	71.99	88.02	88.65	88.70	88.60	93.60	93.73	93.83	93.83
		4	48.80	49.39	49.02	49.26	71.36	72.70	72.52	72.49	81.77	82.05	82.08	82.11
		9	25.43	25.41	24.95	25.20	46.64	47.46	46.95	46.96	59.22	59.38	59.23	59.47
	L	1	47.70	48.35	48.28	48.28	72.77	73.37	73.57	73.55	82.86	83.20	83.22	83.23
		2	33.77	34.23	34.24	34.22	59.09	59.39	59.76	59.67	71.65	72.10	72.19	72.12
		4	20.99	21.24	21.10	21.18	42.00	42.47	42.85	42.82	56.06	56.41	56.34	56.31
		9	10.57	10.68	10.56	10.57	26.10	26.28	26.28	26.38	37.87	38.04	37.98	38.05

* From 10,000 replications of k = 3 equal groups of size n = 20 with p measures under d degree of heterogeneity (d = 1 reflects homogeneity). Hypotheses tested were: within-group trends, C; within-group trends higher than linear, L; using test statistics: Roy's largest root, R; Hotelling-Lawley trace, T; Pillai-Bartlett trace, V; and Wilks' likelihood ratio, W.
Mean vectors were (0 .4 .8 .5 .1) for p = 5 and (0 .4 .8 .5) for p = 4.

test statistics for both hypotheses using both values of p. The only difference among the four conditions was one of magnitude.

With five variates, power for the subsequent L tests tended to be slightly better than for the corresponding C tests. However, with four variates, the reverse was true, with a dramatic loss in power occurring between the C and corresponding L tests (e.g., going from 86% to 48% under homogeneity at the .01 nominal level). Comparing $p = 5$ to $p = 4$, power dropped only slightly for the C test, but significantly for the L test.

The reason for the substantial reduction in power for the L test with four variates seems to be due to the nature of the transformation used to create an alternative hypothesis condition. While the curve was strongly curvilinear with five measures, a linear trend serves as a reasonable approximation of the data when only the first four measures were used in the analyses (see Figure 5-1).

To test the above hypothesis and further explore effects of heterogeneity on power, a second trend transformation was used that resulted in more pronounced curvilinearity at four time points (see Figure 5-2).

Power results for this second curve are presented in Table 5-4. Comparing tables 5-3 and 5-4 shows that both Monte Carlo nominal values and obtained values under homogeneity were quite similar in all cases when $p = 5$, as

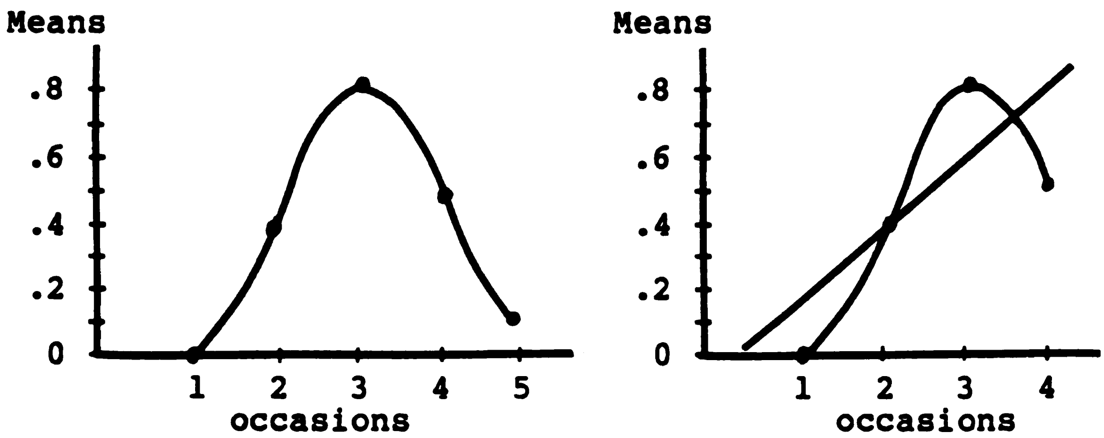

Figure 5-1. Trend transformation for power results of Table 5-3 with mean vectors:
(0 .4 .8 .5 .1) for p = 5
(0 .4 .8 .5) for p = 4

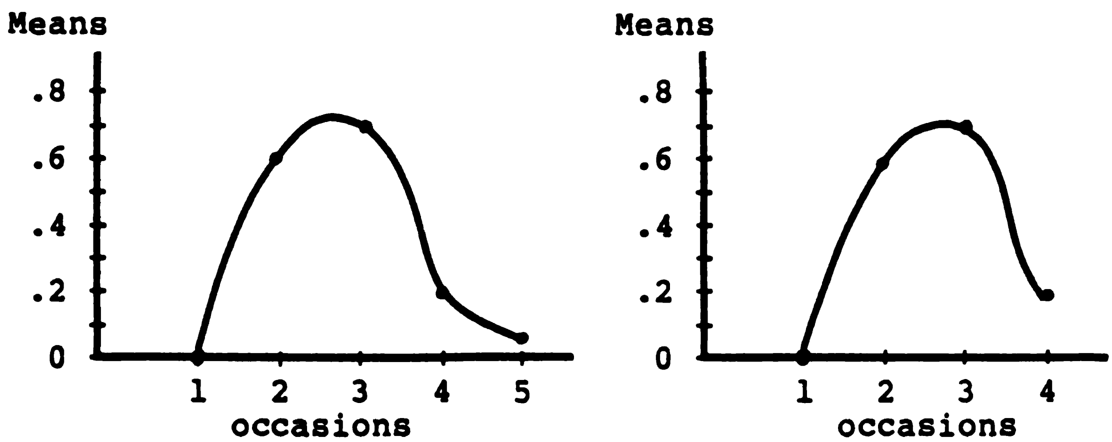

Figure 5-2. Second trend transformation for power results of Table 5-4 with mean vectors:
(0 .6 .7 .2 .05) for p = 5
(0 .6 .7 .2) for p = 4

Table 5-4

Percentage Exceedance Rates for Within-group Tests with Modified True Alternatives*

			α = .01				α = .05				α = .10			
p	H_o	d	R	T	V	W	R	T	V	W	R	T	V	W
5	C	1	92.35	91.15	91.20	91.35	98.00	98.40	98.35	98.40	99.30	99.40	99.45	99.45
		2	79.60	78.35	78.00	78.55	92.45	93.20	93.25	93.25	96.60	96.55	96.50	96.55
		4	57.35	54.55	53.85	53.35	78.40	79.20	79.25	79.20	86.75	87.45	87.40	87.50
		9	29.70	27.10	25.75	26.35	51.20	52.20	50.95	51.55	65.45	66.35	65.50	65.95
	L	1	92.10	92.75	92.40	92.70	98.75	98.75	98.70	98.75	99.55	99.55	99.55	99.55
		2	79.70	81.50	81.20	81.40	94.30	94.25	94.15	94.30	97.30	97.35	97.40	97.25
		4	56.60	58.25	57.85	58.35	81.85	81.55	81.20	81.35	88.30	89.15	89.25	89.10
		9	27.60	29.10	28.55	29.10	53.80	53.35	53.40	53.55	67.95	68.45	68.10	68.20
4	C	1	85.35	84.20	84.90	84.50	95.35	95.70	95.85	95.75	98.10	98.35	98.30	98.35
		2	69.65	67.50	68.40	67.95	87.15	88.25	88.50	88.40	93.90	94.35	94.20	94.40
		4	47.00	44.75	45.10	44.95	69.30	70.05	70.70	70.25	82.30	83.10	83.00	83.10
		9	23.75	22.05	21.70	21.85	44.65	45.50	45.45	45.40	58.35	58.40	58.00	58.50
	L	1	86.20	85.30	84.95	85.15	96.35	96.50	96.45	96.50	98.20	98.35	98.35	98.35
		2	70.50	69.15	68.55	68.90	90.10	90.10	90.15	90.10	94.70	94.70	94.90	94.85
		4	46.80	45.20	44.60	45.00	73.70	74.00	74.05	73.95	83.45	83.50	83.75	83.60
		9	23.50	22.65	22.15	22.40	46.50	46.85	46.95	46.75	58.35	58.50	58.65	58.55

* From 10,000 replications of k = 3 equal groups of size n = 20 with p measures under d degree of heterogeneity (d = 1 reflects homogeneity). Hypotheses tested were: within-group trends, C; within-group trends higher than linear, L; using test statistics: Roy's largest root, R; Hotelling-Lawley trace, T; Pillai-Bartlett trace, V; and Wilks' likelihood ratio, W.
Mean vectors were (0 .6 .7 .2 .05) for p = 5 and (0 .6 .7 .2) for p = 4.

well as for tests of C when $p = 4$. However, exceedance rates for the L test were considerably higher with the second trend transformation than with the first when $p = 4$. These results were consistent with those for five variates using the first transformation. There was a slight gain in power from the C to the L test regardless of size of p. Also, as in the first case for the C test, there were slight reductions in power when going from $p = 5$ to 4 with both tests.

Robustness Under Various Conditions

Having shown that multivariate tests of between-group and within-group differences respond differently to violations of homogeneity, the second stage of this research was an attempt to evaluate the effects of heterogeneity on within-group tests based on different levels of k and n. The design for this part of the study allows for an assessment of robustness and power under heterogeneity when: (1) sample size is varied (with equal groups of 10, 20, and 50), while holding the number of groups constant at three, and (2) the number of groups is varied ($k = 2$, 3, and 6), while holding sample size constant at 20 per group. Results tabled in this section deal with tests of the hypothesis of no within-group trends, C.

The data in this and the following section are based on 2,000 replications each of five combinations of k and n.

With a reduction in the number of generated data sets comes an increase in standard errors. Therefore, using the same procedure as before, 95% probability intervals for the three nominal levels considered now become:

$$.01 \pm .0044$$
$$.05 \pm .0096$$
$$.10 \pm .0131$$

In terms of the tabled values, which are expressed in percentage form, these intervals are:

(0.56, 1.44)	at	.01
(4.04, 5.96)	at	.05
(8.69, 11.31)	at	.10

Sample Size and Robustness

Table 5-5 gives the percentage exceedance rates of within-group tests of trends over occasions, C, for experimental conditions with k = 3 and equal groups of size 10, 20, and 50. With samples of size 10, actual significance levels were within the 95% probability intervals of nominal values only if heterogeneity was low (d = 2). Increasing the sample size to 20 brought improved results (e.g., exceedance rates were also within 95% probability intervals with d = 4 in all cases with four variates and about half the time with five variates). For large samples of 50, values were additionally within these intervals about half the time with d = 9.

When outside the confidence intervals, empirical significance levels were all liberal. However, excluding results with d = 9, Type I errors did not exceed .02, .08,

Table 5-5

Percentage Exceedance Rates Under a True Null
for Tests of Trends with k = 3*

			α = .01				α = .05				α = .10			
p	n	d	R	T	V	W	R	T	V	W	R	T	V	W
5	10	2	1.00	1.05	.90	1.10	5.40	5.65	5.50	5.25	10.80	11.05	10.40	10.55
		4	1.50	1.70	1.50	1.80	7.90	7.55	6.60	6.70	13.40	13.30	12.60	12.30
		9	3.60	2.95	2.40	3.15	11.35	11.00	8.60	9.15	17.70	17.15	14.15	15.90
	20	2	1.20	1.10	1.05	1.10	4.75	5.05	4.80	4.90	10.90	10.75	10.70	10.65
		4	1.65	1.45	1.30	1.35	6.10	5.85	5.65	5.75	11.90	11.90	11.50	11.60
		9	2.35	1.75	1.60	1.65	8.05	7.85	7.50	7.55	14.55	14.25	13.40	13.65
	50	2	.95	.95	1.00	.95	5.45	5.30	5.25	5.25	9.90	10.40	10.50	10.45
		4	.95	.95	1.00	1.00	5.70	5.65	5.65	5.65	11.55	11.80	11.75	11.70
		9	1.10	1.05	1.10	1.10	6.80	6.10	6.25	6.15	13.40	13.70	13.65	13.70
4	10	2	1.25	1.20	1.30	1.30	5.90	5.45	5.85	5.65	10.55	10.70	10.80	10.70
		4	1.65	1.65	1.65	1.75	6.50	6.20	6.70	6.45	11.70	11.85	11.70	11.60
		9	2.65	2.75	2.15	2.65	8.65	8.15	7.70	7.95	14.95	14.45	13.75	13.95
	20	2	1.35	1.05	1.05	1.05	4.95	5.05	5.20	5.05	10.25	10.15	9.95	10.05
		4	1.30	1.25	1.40	1.25	5.65	5.55	5.45	5.55	11.00	11.30	11.05	11.15
		9	1.90	1.70	1.70	1.65	7.20	7.10	6.95	7.00	12.90	13.25	12.95	13.10
	50	2	.95	1.10	1.10	1.10	5.25	4.80	4.85	4.80	9.90	9.60	9.65	9.60
		4	1.00	1.10	1.10	1.10	5.40	5.20	5.20	5.20	10.85	10.70	10.75	10.75
		9	1.10	1.10	1.10	1.10	6.25	5.80	5.85	5.80	11.35	10.85	10.80	10.80

* From 2,000 replications of k = 3 equal groups of size n with p measures under d degree of heterogeneity. Hypothesis tested was of no within-group trends, C, using test statistics: Roy's largest root, R; Hotelling-Lawley trace, T; Pillai-Bartlett trace, V; and Wilks' likelihood ratio, W.

and .14 at nominal levels of .01, .05, and .10, respectively, and were typically much lower. With $d = 9$, they never exceeded .04, .11, and .18 at the three nominal levels.

Considering the results in terms of acceptable robustness limits (i.e., .015, .06, and .12), cases with 10 subjects per group would be robust only with $d = 2$, while cases with 20 subjects per group would be robust with both $d = 2$ and 4. With sample sizes of 50, robustness extends to conditions with high heterogeneity ($d = 9$) when only four variates were analyzed.

For a given sample size, the other factors in this study behaved in the same manner as previously described for 10,000 replications where $k = 3$ and $n = 20$. In general, departure from the nominal significance level increased as heterogeneity, number of variates, or nominal significance level increased. The main difference across conditions with different sample sizes was one of degree.

With respect to the multivariate test statistics, in only about half the cases did the R statistic produce the greatest exceedance rates and the V statistic the smallest. But, where this was not the case, the two values were typically within sampling error of the nominals and their difference was within one standard error of a difference of two proportions. Even when the R statistic had larger exceedance rates than the V statistic, the differences among the four statistics were considerably less pronounced than is typical for between-group tests.

Comparison with Tests of Trends Higher than Linear. For tests of the second within-group hypothesis of no trends higher than linear, L, percentage exceedance rates followed the patterns of the overall within-group tests. Obtained values were either within 95% probability intervals of nominal values or liberal. In most cases, significance levels for the L tests were lower than those for the C tests. However, differences between them were typically negligible. Values for tests of L are tabled in Appendix D.

Comparison with Between-group Tests. Percentage exceedance rates obtained in this study under a true null for between-group tests were consistent with previously established results. These values are tabled in Appendix C. The differences between the B and C tests followed the patterns described earlier in this chapter regardless of sample size. The only difference was one of degree. Discrepancies between the tests were generally smallest with large samples and low heterogeneity, or with the V and W statistics. They were greatest with small samples and high heterogeneity, or with the R and T statistics.

Number of Groups and Robustness

Table 5-6 gives the percentage exceedance rates of within-group tests of trends over occasions when sample

Table 5-6

Percentage Exceedance Rates Under a True Null
for Tests of Trends with n = 20*

			α = .01				α = .05				α = .10			
p	k	d	R	T	V	W	R	T	V	W	R	T	V	W
5	2	2	.95	1.10	.90	1.15	5.15	5.20	4.95	5.15	10.70	10.35	10.35	10.35
		4	1.35	1.05	1.15	1.25	5.85	5.95	6.10	5.85	12.80	11.90	11.65	11.95
		9	1.75	1.75	1.65	1.95	7.75	7.25	6.95	7.15	15.20	14.05	13.00	13.85
	3	2	1.20	1.10	1.05	1.10	4.75	5.05	4.80	4.90	10.90	10.75	10.70	10.65
		4	1.65	1.45	1.30	1.35	6.10	5.85	5.65	5.75	11.90	11.90	11.50	11.60
		9	2.35	1.75	1.60	1.65	8.05	7.85	7.50	7.55	14.55	14.25	13.40	13.65
	6	2	1.05	1.00	1.05	1.00	5.10	4.80	4.90	4.80	9.90	10.10	10.15	10.10
		4	1.60	1.30	1.30	1.30	5.95	5.65	5.70	5.60	11.00	11.15	11.45	11.40
		9	2.35	2.20	2.20	2.20	8.15	7.60	7.50	7.60	13.35	13.05	13.30	13.05
4	2	2	1.05	.90	.85	.90	4.95	4.60	4.40	4.65	10.55	9.95	10.10	9.95
		4	1.15	1.15	1.35	1.05	5.70	5.55	4.95	5.25	11.55	10.90	10.75	10.60
		9	1.50	1.40	1.70	1.45	7.05	6.70	6.00	6.30	12.20	11.50	11.10	11.20
	3	2	1.35	1.05	1.05	1.05	4.95	5.05	5.20	5.05	10.25	10.15	9.95	10.05
		4	1.30	1.25	1.40	1.25	5.65	5.55	5.45	5.55	11.00	11.30	11.05	11.15
		9	1.90	1.70	1.70	1.65	7.20	7.10	6.95	7.00	12.90	13.25	12.95	13.10
	6	2	1.10	1.15	1.15	1.15	4.90	5.00	4.90	4.95	10.30	10.05	10.05	10.05
		4	1.55	1.35	1.35	1.35	5.40	5.20	5.20	5.15	11.40	10.50	10.45	10.50
		9	2.10	1.90	1.90	1.90	6.35	6.55	6.40	6.45	11.95	11.60	11.55	11.60

* From 2,000 replications of k equal groups of size n = 20 with p measures under d degree of heterogeneity. Hypothesis tested was of no within-group trends, C, using test statistics: Roy's largest root, R; Hotelling-Lawley trace, T; Pillai-Bartlett trace, V; and Wilks' likelihood ratio, W.

size is held constant at 20 and the number of groups varied, with k = 2, 3, or 6. The values tended to be within 95% probability intervals of the nominal alpha with both low and moderate heterogeneity regardless of number of groups. The major exception was with five variates at .10 alpha, where values tended to fall outside the probability intervals with moderate heterogeneity.

When d = 9 values were all liberal. However, Type I errors were .02, .08, and .15 for corresponding .01, .05, and .10 nominal levels. Results would be considered robust with low and moderate heterogeneity in all cases, as well as in almost half the cases with high heterogeneity.

An unexpected finding from this set of results was that the impact of heterogeneity did not appear to be greatest with the larger number of groups. In about half the cases, the largest exceedance rates occurred with k = 3. The remaining cases were split with the largest departures occurring about equally with either k = 2 or 6. It might be assumed that this result was due to the high level of robustness of the C tests, since over half of the values in Table 5-6 were within 95% confidence intervals of the nominal value. However, even when only considering values for d = 9, which were outside these intervals, in more than half of the cases the largest departures still occurred with k = 3, while the rest tended to occur with k = 6.

It appears that the impact of heterogeneity on within-group tests of trends is lessened by increasing the size of equal samples but that, for a given n, decreasing the number of groups may not help.

With a given number of groups, the effects of the other factors being examined were not always evident. This is probably due to the fact that actual values tended to be within sampling error of nominal alpha except for high heterogeneity. Still, some patterns emerged. In general, differences among the four statistics were still relatively minor, never exceeding two percentage points. Departure from nominal alpha typically increased as heterogeneity and alpha levels increased, although the latter reflects larger standard errors at higher alpha levels. The effect of decreasing p was evident only when d = 9, where lower exceedance rates were associated with the smaller number of variates.

Comparison with Tests of Trends Higher than Linear. As was evident when sample size was varied, differences in actual significance levels between the two within-group tests were minimal. The discrepancies between most of the corresponding exceedance rates rarely exceeded one percentage point. Percentage exceedance rates for tests of non-linearity when number of groups was varied are tabled in Appendix D.

Comparison with Between-group Tests. Unlike the results for different sample sizes, when number of groups was varied, the B and C tests responded differently not only in terms of degree but also in kind. For the B tests, departures from nominal alpha consistently increased when there were more groups (except if heterogeneity was low). It appears that, for this situation with equal groups of size 20, within-group tests under any level of heterogeneity were similar in robustness to between-group tests under low heterogeneity.

Power Under Various Conditions

The effects of sample size and number of groups on the power of within-group tests of trends under heterogeneity conditions were assessed with the same 2,000 replications used to study robustness. The data were transformed to model the alternative hypothesis situation in Figure 5-1.

Power trends for both within-group tests were similar to those previously defined for the same transformation with 10,000 replications. The tests for non-linearity had slightly better power than the overall tests for trends with five variates but had considerably lower power with four variates. Since this was consistent across the experimental conditions considered, only the results for the test of trends will be discussed. Holding other factors constant, power values were extremely stable across the four test statistics for a given condition. With samples of size 10, there was slightly more variability among the test statistics but, even here, the discrepancies were not noteworthy. Therefore, results presented in this section are percentage exceedance rates for tests of trends averaged over the four statistics. Complete tables for both sets of within-group tests are included in Appendix E.

Sample Size and Power

Table 5-7 gives the average percentage exceedance rates for three groups of size 10, 20, and 50 under noncentral conditions. For a given sample size, the results were consistent with those from the first part of the study. With respect to varying the size of equal samples, n had a considerable effect on power, which decreased as n did even under homogeneity (d = 1). This effect was compounded as heterogeneity was introduced.

With n = 50, where within-group tests were robust, the effect of heterogeneity was negligible, particularly if alpha was greater than or equal to .05, in which case exceedance rates were still over 90% with d = 9. Although lower, power values with n = 20 were reasonable under low and moderate heterogeneity, where robustness was achieved. However, with only 10 subjects per group, power tended to be poor even with low heterogeneity, where tests were robust. This was particularly so at a .01 nominal level, where power was low even under homogeneity.

Table 5-7

Average Percentage Exceedance Rates Under True Alternatives for Tests of Trends with k = 3*

		p = 5			p = 4		
	alpha:	.01	.05	.10	.01	.05	.10
n	d						
10	1	45.04	70.38	81.61	38.19	64.65	78.01
	2	33.10	58.21	71.16	27.69	51.64	67.25
	4	21.85	44.53	58.19	18.81	38.89	52.39
	9	14.98	30.93	43.00	12.05	26.69	38.95
20	1	91.33	98.26	99.51	85.10	95.75	98.14
	2	77.93	92.31	96.46	69.24	87.96	93.86
	4	55.71	77.65	86.63	45.81	70.53	82.16
	9	28.84	52.58	65.04	23.40	44.61	58.99
50	1	100.00	100.00	100.00	100.00	100.00	100.00
	2	99.95	100.00	100.00	99.53	100.00	100.00
	4	97.95	99.74	99.94	95.61	99.26	99.65
	9	76.15	93.06	96.34	67.25	88.64	93.88

* From 2,000 replications of k = 3 equal groups of size n with p measures under d degrees of heterogeneity (d = 1 reflects homogeneity). Hypothesis tested was of no within-group trends, C. Tabled values were averaged over four test statistics: Roy's largest root, R; Hotelling-Lawley trace, T; Pillai-Bartlett trace, V, and Wilks' likelihood ratio, W.

Number of Groups and Power

Table 5-8 gives the average percentage exceedance rates for equal groups of size 20 with two, three, and six group designs under noncentral conditions. Effects within a given condition were again consistent with results from the first part of this study. With respect to number of groups, power was best with k = 6 and worst with k = 2.

For six groups, empirical power was above 90% in all cases except for d = 9 at .01 alpha with four variates. The powers with six groups and high heterogeneity were consistently higher than those for three groups with low heterogeneity.

The effect of heterogeneity was considerable with both two and three groups, particularly at .01 alpha. However, in both these cases, power was reasonable with d = 2 and 4, where tests achieved robustness, as long as a nominal level of .01 is not considered.

Total Sample Size and Power

Considering all five combinations of k and n together, power under low and moderate heterogeneity seems to be a function of total sample size, N. As shown in Figure 5-3 (power curves of values in Tables 5-7 and 5-8, which were averaged over four test statistics), for d = 2 and 4 power increases as total N increases until, with N = 120 and 150, the curves are indistinguishable. However, with d = 9, heterogeneity appears to have a greater impact on power

Table 5-8

Average Percentage Exceedance Rates Under True Alternatives for Tests of Trends with n = 20*

		p = 5			p = 4		
	alpha:	.01	.05	.10	.01	.05	.10
k	d						
2	1	70.65	87.19	93.01	63.14	82.00	90.04
	2	47.31	70.46	80.89	40.58	63.60	75.19
	4	26.10	47.96	61.86	22.45	41.84	54.88
	9	13.55	27.99	39.78	10.96	24.14	35.03
3	1	91.33	98.26	99.50	85.10	95.75	98.15
	2	77.93	92.31	96.46	69.24	87.96	93.86
	4	55.71	77.65	86.63	45.81	70.53	82.16
	9	28.84	52.58	65.04	23.40	44.61	58.99
6	1	100.00	100.00	100.00	99.88	100.00	100.00
	2	99.81	100.00	100.00	99.44	99.90	100.00
	4	99.31	99.83	99.94	98.11	99.29	99.75
	9	93.04	97.34	98.86	86.84	94.81	97.51

* From 2,000 replications of k equal groups of size n = 20 with p measures under d degrees of heterogeneity (d = 1 reflects homogeneity). Hypothesis tested was of no within-group trends, C. Tabled values were averaged over four test statistics: Roy's largest root, R; Hotelling-Lawley trace, T; Pillai-Bartlett trace, V, and Wilks' likelihood ratio, W.

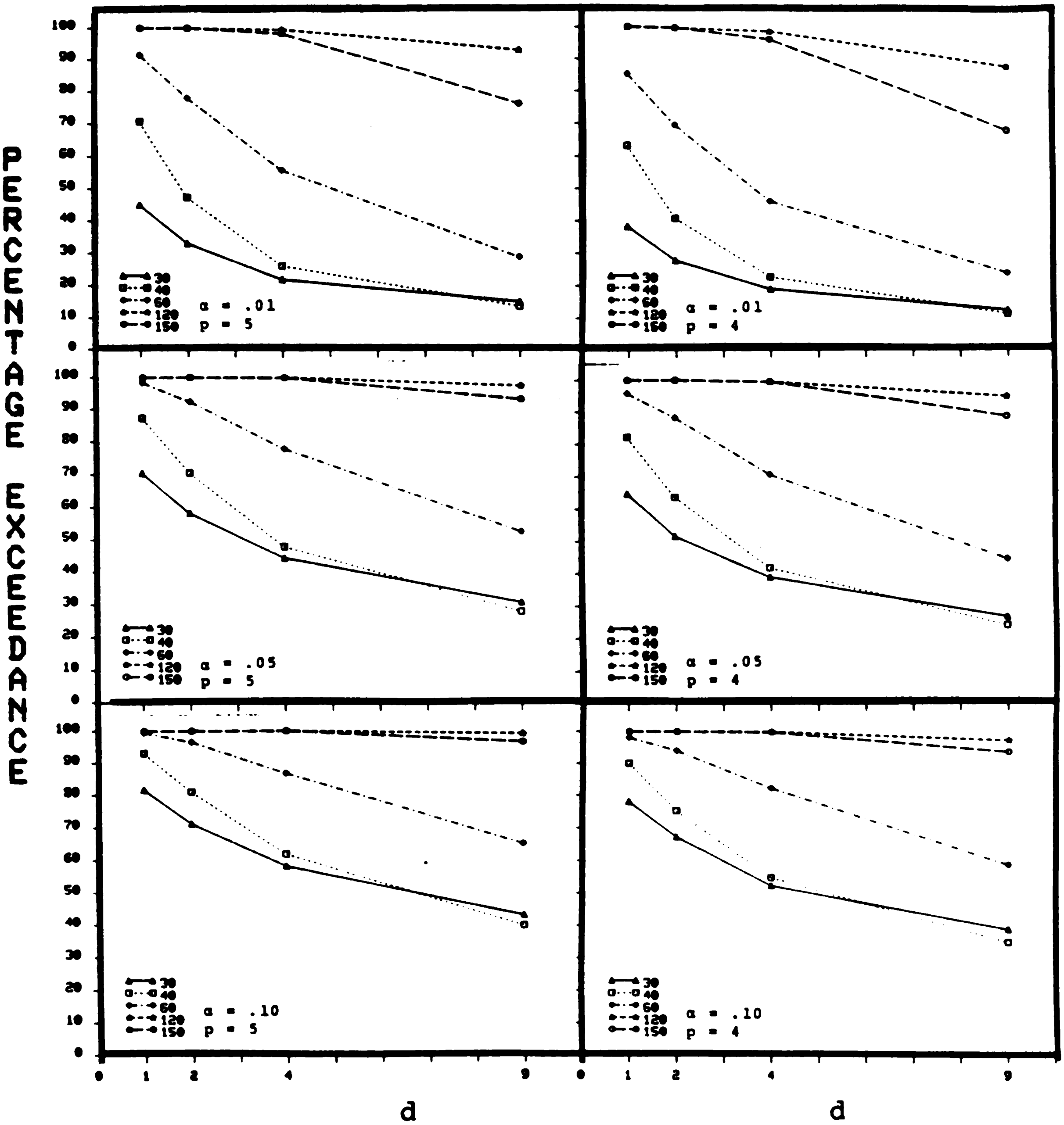

Figure 5-3. Power curves averaged over four test statistics for different total sample sizes N, where:

N =	30	40	60	120	150
k =	3	2	3	6	3
n =	10	20	20	20	50

30
40
60
120
150

when a larger percentage of vector scores were heterogeneous, even though the group sizes were larger.

This latter result is most likely due to the manner in which heterogeneity was allocated across the groups, coupled with the analysis being performed. The test for within-group main effects assumes that the curves of the k groups are parallel (i.e., that no group by occasion interaction exists). The test then is used to evaluate whether the curves are constant (i.e., if there is any trend across the occasions). Hence, this test compares the means of each measure over the total number of subjects, N.

In all cases, only one group was drawn from a heterogeneous population. Therefore, since the 120 vector scores came from six groups of size 20, only 20 vectors (17%) were heterogeneous. The 150 vector scores were from three groups of size 50, so that 50 vectors (33%) were heterogeneous. The higher proportion of discrepant vectors in the latter situation may have produced the reverse effect at d = 9 than would be expected based on N. This result was consistent across all four test statistics and three alpha levels (see Appendix E).

The identical phenomenon occurred with small total samples. When N = 30 (k = 3 and n = 10), only 10 (33%) of the vector scores were heterogeneous, while half of the vectors were heterogeneous when N = 40 (k = 2 and n = 20). This reversal of power levels was consistently evidenced across all but the V statistic, where power was slightly higher with N = 40 at the .01 and .05 nominal levels.

When the percentage of heterogeneous vectors was held constant at 33% (three situations with k = 3), power decreased steadily as N decreased. Unfortunately, given the present data, the effects on power if total N is held constant while varying the percentage of heterogeneous vectors could not be evaluated. For the conditions examined, the results indicate that, for low or moderate heterogeneity, total N dictates the level of power but that, for high heterogeneity, the percent of discrepant vector scores has the greater impact.

CHAPTER VI

DISCUSSION

The results presented in the previous chapter provide an indication that multivariate tests of between-group and within-group differences are not equally subject to the effects of heterogeneity of covariance matrices. Conclusions based on these results will be presented in this chapter, followed by guidelines for the researcher analyzing repeated measures studies with time ordered data and suggestions for future research.

Conclusions

Under the conditions considered in this study, it appears that multivariate tests for trends over occasions in repeated measures designs with equal groups are not as sensitive to violations of the assumption of homoscedasticity across groups as are tests for between-group differences. In most cases, within-group tests are extremely well behaved, while between-group tests tend to be robust when two groups are involved or if heterogeneity is low.

This difference is most likely due to the manner in which the mean square hypothesis (HE^{-1}) is formed for the tests of the two hypotheses. The elements of the hypothesis matrix, H, for between-group tests consists of sums of squares and cross products, while for within-group tests, these elements consist of squared sums of means and products of means. It is likely that this difference produced the differential effect that heterogeneity had on multivariate tests of the two hypotheses.

Within-group tests for general trends tended to be robust even with heterogeneity of $d = 4$. Although such a level of heterogeneity was considered moderate in this thesis, covariance matrices that differ by a factor of four would indicate a dramatic discrepancy from a practical perspective. Hence, in most practical situations, where groups are equal and heterogeneity is present, multivariate within-group tests should be robust. Conclusions for between-group tests with equal samples uphold previous findings that these tests tend to be robust when covariance matrices differ only by a factor of two.

Differences among the four test statistics considered were evident, with Pillai-Bartlett's trace, V, typically showing the least departure from nominal levels and Roy's largest root, R, the most. However, discrepancies between the V and R statistics were relatively minor for within-group tests but pronounced for between-group tests. Even under low heterogeneity ($d = 2$), the R statistic on between-group tests tended to be liberal unless sample size was at least 50 per group. However, for within-group tests, R was robust at $d = 2$, as well as in over half the cases at $d = 4$.

For within-group tests, increasing the number of groups did not produce a consistent effect on actual significance level, but changes in the sample size and, to a small degree, the number of occasions did. With four variates the tests were robust even under high heterogeneity ($d = 9$) with equal samples of at least 50. With equal samples of 20, robustness was achieved with all four statistics under moderate heterogeneity ($d = 2$) when $p = 4$, and also when $p = 5$ except for Roy's largest root. This remained true for a constant group size of 20 regardless of whether there were three or six groups. Only with two groups of size 20 and four variates did robustness tend to extend to high heterogeneity.

Actual significance levels for the within-group tests of trends higher than linear were typically closer to nominal alpha levels than for general tests of trends, although differences were minor. It is expected that, given a larger number of time points, a slight increase in robustness would be achieved with each succeeding multivariate test for higher order trends. However, given the fact that departures from nominal levels were typically more severe with five variates than with four, a word of caution is in order. It is likely that, with a significant increase in the number of time points, the overall test of trends would produce a too liberal test. Therefore, if the initial departure from nominal levels is large enough, no worthwhile gains in robustness may result with successive tests for higher order trends. Also, tests of between-group differences become increasingly more liberal with increases in the number of variates. The combined effect on both tests would therefore need to be considered.

The four multivariate test statistics evaluated in this study behaved almost as one with respect to their power to reject a null of no trend in the data. Heterogeneity affected them equally (i.e., power was reduced as heterogeneity increased). At a given heterogeneity level, power of the second test for trends higher than linear was slightly greater than that of the overall test for trends when a fairly strong curvilinear trend was present. As would be expected, the subsequent test lost power dramatically when the trend tended toward linearity.

Decreasing the number of subjects per group compounded the effect of heterogeneity on the power of within-group tests (i.e., the smaller the group, the lower the power, even under homogeneity). However, power was still reasonable with equal groups of 20 or more, where the within-group tests were robust.

Guidelines for the Researcher

The analysis of RM data may be undertaken with either univariate or multivariate statistical tests. It is known (e.g., Davidson, 1972) that, as long as there are more subjects per group then there are groups, multivariate tests are the preferred choice when the univariate assumption of uniformity is violated. With RM data from equally spaced time points, this assumption is generally violated, since autocorrelation is typically present.

If more than one group is involved and the researcher further suspects that the assumption of a common covariance matrix across groups may be violated, the results from this study would indicate that the problem may not be too serious for within-group tests, even for violations as large as variances differing by a multiple of four. However, previous research has shown that this is not the case for between-group tests.

Given this discrepancy, the question arises about how to analyze and interpret repeated measures data when heterogeneity across groups is suspected. Of particular concern are situations where group by occasion interactions exist. Since a test of this hypothesis must precede tests of both main effects hypotheses, the problem may be considerable. The mean square hypothesis for interaction consists of the (p-1)x(p-1) submatrices of the hypothesis and error matrices used for the between-group test. It may therefore be presumed that violations would cause effects similar to the between-group tests and results from between-group robustness studies should apply.

Hence, for tests of interactions, Pillai-Bartlett's V statistic should be employed and results interpreted with caution. If the number of measures is relatively small and equal groups have been maintained, then results may be considered valid. However, if this is not the case, an assessment of whether greater heterogeneity was present in a smaller or larger group should be attempted. In the former case, results would be liberal and, in the latter case, conservative.

If it can be assumed that there are no group by occasion interactions, then a two-stage analysis would be recommended. The RM dimension may be tested through a multivariate test of within-group trends without excessive concern. An appropriate approach for testing between-group effects would be to use the mean of the RM as the dependent variable in a univariate analysis of group differences. This would eliminate concern for heterogeneity since it has been demonstrated repeatedly that, except when samples are small and unequal, the univariate F-test is robust against this violation.

Suggestions for Future Research

An aspect that needs to be considered is unequal sample sizes. Although it is quite likely that this would not produce any radical departures from nominal significance levels for within-group tests, it might do so for interaction tests, since the latter are based on submatrices of hypothesis and error matrices of between-group tests. Also, with equal samples, consideration should be given to a study of the effects of varying number of groups while holding total N constant, thereby varying the n/N ratio. This would be particularly useful in further exploring the effects of heterogeneity on the power of within-group tests.

Additionally, noncentrality structure would be a relevant factor for inclusion in a study of the power of within-group tests of trends. Different results in terms of power are likely to occur depending on whether noncentrality is concentrated in the first canonical variate, or spread equally over all canonical variates.

Since the RM dimension in the present study represented the same measure repeatedly taken over time, a polynomial representation was used to transform the data for within-group tests. This transformation uses the regression model matrix (see Chapter II), which is a matrix of normalized orthogonal polynomial coefficients. If the measures are taken to correspond to a factorial classification, a treatment contrasts and interaction representation would be used. Since the various transformations each partition the source of variation for occasions in a different manner, it is possible that they may react differently to violation of homoscedasticity across groups.

The regression model matrix is orthogonal. Design matrices may be orthogonal, such as a Helmert contrast matrix used with hierarchical designs over the measures, or they may be nonorthogonal, such as a paired differences matrix used in profile analyses. It might be assumed that all orthogonal transformations would behave in like manner but that nonorthogonal ones would not. However, this assumption needs to be tested.

Although findings from this study are of a preliminary nature, they provide strong evidence that, at least for equal samples, within-group tests of trends in a repeated measures design are fairly robust to violations of homoscedasticity that might occur in practical situations. Furthermore, these tests maintain reasonable power under heterogeneity, except for small sample sizes.

APPENDICES

APPENDIX A

COMPUTATIONAL DETAILS AND COMPUTER PROGRAM

The data generation and analysis were performed on a CDC (Control Data Corporation) Cyber 750 computer at Michigan State University. This 60 bit word machine uses the Scope/Hustler operating system. The program, which was coded in FORTRAN V by Jeff Glass, uses some Compass assembly language to decrease field lengths and thereby reduce costs. All computation was done in double precision.

In this appendix, the following are listed or described: (1) the subroutines used from package programs, (2) the actual values input by the user, (3) the steps followed in the computer program, (4) the procedures performed to check the operation of the computer program, and (5) the complete listing of the computer program.

Subroutines from Published Sources

a) Subroutines taken from the International Mathematical and Statistical Library (IMSL, 1979):

   GGUBS To generate uniform pseudo-random numbers.

   VMULFF For matrix multiplication.

VMULFM	For matrix multiplication of the transpose of a matrix A by a matrix B.
VMULFP	For matrix multiplication of a matrix A by the transpose of a matrix B.
LINV2F	To compute the inverse of a matrix.

b) Subroutines taken from the EISPACK library (The Argonne eigensystem package, 1972):

TRED2	To determine eigenvalues of a symmetric tridiagonal matrix.
IMTQL2	To reduce a positive-definite matrix to a triadiagonal matrix for input into TRED2.

User Input Values

a) Seed values for data generation. A different value was used for each combination of k and n. These are listed in the Checking Procedures section in this appendix.

b) Defining parameters for N, k, n, and p.

c) Matrices of normalized othogonal polynomial coefficients for calculations using P(pxp), where p = 5 or 4.

$$P1 = \begin{bmatrix} .44721 & -.63246 & .53452 & -.31623 & .11952 \\ .44721 & -.31623 & -.26726 & .63246 & -.47809 \\ .44721 & 0 & -.53452 & 0 & .71714 \\ .44721 & .31623 & -.26726 & -.63246 & -.47809 \\ .44721 & .63246 & .53452 & .31623 & .11952 \end{bmatrix}$$

$$P2 = \begin{bmatrix} .5 & -.67082 & .5 & -.22361 \\ .5 & -.22361 & -.5 & .67082 \\ .5 & .22361 & -.5 & -.67082 \\ .5 & .67082 & .5 & .22361 \end{bmatrix}$$

d) A vector of constants to be added to the second through fifth measures for use in power calculations. For the transformation used in all cases, this was:

POWER = (.4 .8 .5 .1)

For the modified transformation used only with 10,000 replications, this was:

POWER = (.6 .7 .2 .05)

Program Steps

The program consists of five main components (RM, DATA, SORT, NTABLE, and OTABLE). Rm is the central part of the program and consists of seven subroutines. Additionally, a sixth component STATS was used to output the mean and variance of the generated data and to test the data for normality. Comments throughout the program explain the steps and provide information for its use at other installations.

The actual steps used in the computer program are described below. Variable names are the same as those used in the program, except for the following:

N	=	COUNT	number of replications
K	=	GROUP	number of groups
n	=	SUBJECT	number of subjects per group
p	=	MEASURE	number of repeated measures
Z.	=	ZBAR	matrix of group means

A. Data Generation

1. Use IMSL subroutine GGUBS to generate N replications of nkp uniform (0,1) pseudo-random numbers using double precision (where p = 5 and N = 2,000 or 10,000).

2. Transform these values to normal deviates N(0,1) using

Box-Muller transformation (see Chapter IV).

3. Array resulting values in N matrices Z of dimension (nk x p) such that

$$Z = \begin{bmatrix} Z_1 \\ \cdot \\ \cdot \\ \cdot \\ Z_k \end{bmatrix} \sim N(\underline{0}, I) \quad \text{with } Z_j \ (n \times p)$$

B. <u>Computation of Test Statistics</u>

<u>For p = 5</u>, compute for each data set:

1. (k x p) matrix of means

$$Z. = \begin{bmatrix} z._1' \\ \cdot \\ \cdot \\ \cdot \\ z._k' \end{bmatrix}$$

where $z._j' = (z._j^{(1)} \ z._j^{(2)} \ z._j^{(3)} \ z._j^{(4)} \ z._j^{(5)})$

2. Preliminary SSCP matrices for:

Total	T	=	Z'Z
Groups	G	=	nZ.'Z.

3. SSCP for calculations of test statistics for:

Constant	C	=	(n/k)Z.'UU'Z. where U = unit vector
Between	B	=	G − C
Error	E	=	T − G

4. Transform C and E matrices by P(pxp), a matrix of normalized orthogonal polynomial coefficients with resulting elements x_{ij} (i,j = 1,...,p). The transformed matrices are:

CTRAN = P'CP   ETRAN = P'EP

5. Take the lower right (p-1)x(p-1) submatrices from (B-4) to be used for tests of occasion trends (elements x_{ij}, i,j = 2,...,p), and label:

   CRM   ERM

6. Take the lower right (p-1)x(p-1) submatrices of (B-4) to be used for tests of non-linearity (elements x_{ij}, i,j = 3,...,p), and label:

   CL   EL

7. Compute HE^{-1} for tests of:

group differences	$HB = BE^{-1}$
occasion trends	$HC = (CRM)(ERM)^{-1}$
non-linearity	$HL = (CL)(EL)^{-1}$

8. Compute the eigenvalues for each test in (B-7) and label:

EIGB	(p values)
EIGC	(p-1 values)
EIGL	(p-2 values)

9. Using the formulas for the R, T, V, and W test statistics (see Chapter II) and the eigenvalues from (B-8), compute each statistic for the C, B, and L tests. Each resulting list of either 10,000 or 2,000 values is labeled according to the combination of test and test statistic used:

RB	TB	VB	WB
RC	TC	VC	WC
RL	TL	VL	WL

For $p = 4$:

10. Drop the last row and column of the B, C, and E matrices in (B-3) and repeat steps (B,1-9).

C. Determination of Monte Carlo Critical Values

1. Sort the 24 lists of test statistics in (B-9) (12 for $p = 5$ and 12 for $p = 4$), placing the values in rank order.
2. Using $\alpha = .01$, .05, and .10 and N = COUNT, calculate for each list:
   a) the average of the ($N\alpha$)th and ($N\alpha+1$)th largest values to be the critical values for the R, T, and V statistics (result - 54 critical values).
   b) the average of the ($N\alpha$)th and ($N\alpha+1$) smallest values to be the critical values for the W statistics (result - 18 critical values).

D. Determination of Actual Significance Levels

Repeat this step three times for $d = 2$, 4, and 9.

1. Transform the Z matrix from (A-3) so that one group has larger variances, with resulting (pxl) score vectors:

$$\mathbf{z}_{ij} = d^{1/2} z_{ij} \sim N(\underline{0}, dI) \qquad \begin{array}{l} i = 1,\ldots,n \\ j = 1,\ldots,p \end{array}$$

$$\mathbf{z}_{ij} = z_{ij} \sim N(\underline{0}, I) \qquad \begin{array}{l} i = n+1,\ldots,nk \\ j = 1,\ldots,p \end{array}$$

2. Repeat (B,1-10) to get 24 lists of test statistics under a true null and a given heterogeneity condition.

3. Compare each value in each list to the corresponding critical values from (C-2) at each alpha level and count the number of times each value is:

   a) greater than its corresponding critical value for R, T, and V statistics.

   b) less than its corresponding critical value for W statistic.

4. For each list and alpha level, divide the resulting values from (D-3) by the number of repetitions (COUNT) to get actual Type I error rates.

E. Determination of Monte Carlo Nominal Powers

1. Transform the Z matrix from (A-3) so that all groups reflect a polynomial trend across the time points by adding a constant to each measure. The resulting (px1) score vectors are:

$$z_{ij} = z_{ij} + \mu \sim N(\mu, I) \quad \begin{array}{l} i = 1,\ldots,nk \\ j = 1,\ldots,p \end{array}$$

2. Repeat (B,1-10) to get 24 lists of test statistics under a true alternative and no violation to homogeneity.

3. Repeat (D,3-4) to get Monte Carlo nominal power values.

F. Determination of Actual Powers

Repeat this step three times for d = 2, 4, and 9:

1. Transform the Z matrix from (A-3) so that one group has larger variances and all groups reflect a polynomial trend across timepoints. The resulting (px1) score vectors are:

$$z_{ij} = d^{1/2} z_{ij} + \underline{\mu} \sim N(\underline{\mu}, dI) \qquad \begin{matrix} i = 1,\ldots,n \\ j = 1,\ldots,p \end{matrix}$$

$$z_{ij} = z_{ij} + \underline{\mu} \sim N(\underline{\mu}, I) \qquad \begin{matrix} i = n+1,\ldots,nk \\ j = 1,\ldots,p \end{matrix}$$

2. Repeat (B,1-10) to get 24 lists of test statistics under a true alternative and a given heterogeneity condition.

3. Repeat (D-3,4) to get actual powers under heterogeneity conditions.

Checking Procedures

Using the IMSL subroutine GTNOR to test for normality of the generated data, the following results were obtained for 1,000 replications of 300 data points with six randomly choses initial seeds:

Seed	Chi-square	Probability
444,852,461	5.78	.76
9,458,577,882	6.20	.72
11,261,152,461	14.68	.10
2,344,743,849	5.18	.82
2,341	5.08	.83
112,623,455	9.06	.43

For each experimental condition, the mean and variance of the generated N(0,1) data points were calculated. The initial seed used to generate the data for the study with the corresponding means and variances were as follows:

COUNT	k	n	Seed	Mean	Variance
10,000	3	20	739,604,919	1.1×10^{-4}	.99901
2,000	3	10	740,848,519	-3.0×10^{-3}	.99707
	2	20	837,616,087	-1.6×10^{-3}	.99895
	3	20	739,604,919	-1.5×10^{-3}	.99870
	6	20	344,148,214	-1.4×10^{-4}	1.00005
	3	50	208,577,315	-9.8×10^{-6}	1.00000

A final check on the calculation of the SSCP matrices was performed by outputing the B, C, E, CTRAN, ETRAN, CRM, ERM, CL and EL matrices from one replication with $k = 3$ and $n = 20$. The results were hand checked to ascertain that the program was correctly calculating these matrices.

To check the results of the Monte Carlo critical values, the parameters needed to find tabled values were determined. Tabled values were found in 92 cases (21.3%). These were in fairly close agreement with the calculated values.

```
          IDENT    RM              ASSEMBLY-LANGUAGE DRIVER PROGRAM
          ENTRY    RM
RM        BSS      0
          RJ       =XXRM
          ENDRUN
          ENTRY    ABORT           ERROR TERMINATION -
ABORT     BSS      1
          RJ       =XENDOUT        CLOSE OUTPUT FILES IN CASE OF ERROR
          ABORT
          END      RM
      SUBROUTINE XRM
C************
C* RM - REPEATED MEASURES TESTING.
C*
C* THIS PROGRAM IS THE HEART OF THE RM PROGRAM SET. IT CALCULATES TEST
C* STATISTICS R, T, V, AND W, WRITING THE RESULTS TO FILES FOR FURTHER
C* PROCESSING BY OTHER PROGRAMS.
C*
C************
C* TRANSPORTABILITY NOTE:
C* THESE PROGRAMS WERE WRITTEN AS CLOSE TO STANDARD FTN5 THAT ACTUAL
C* CONSIDERATIONS OF COST WOULD ALLOW.
C* 1) SEVERAL PROGRAMS (THE DATA GENERATION AND STATISTICS PROGRAMS,
C*    THE SORT PROGRAM, AND THIS ONE) HAVE ASSEMBLY-LANGUAGE MAIN
C*    PROGRAMS TO REDUCE EXECUTION COST. THESE ASSEMBLY-LANGUAGE
C*    ROUTINES MAY BE REPLACED BY FTN5 PROGRAMS IF NEED BE.
C* 2) STANDARD FTN5 I/O WAS JUDGED TO BE TOO EXPENSIVE FOR MULTIPLE
C*    RUNS OF 10000 CASES, SO A NON-STANDARD I/O PACKAGE CALLED FASTIO
C*    WAS USED IN THE AFOREMENTIONED PROGRAMS.
C*    CONVERSION TO STANDARD FTN5 I/O WOULD BE STRAIGHTFORWARD. FOR
C*    THIS PROGRAM, ROUTINES SETOUT, OUTPUT, ENDOUT, AND GETDATA HANDLE
C*    ALL THE I/O, AND ONLY THOSE ROUTINES WOULD NEED MODIFICATION.
C* 3) THE GETREG SUBROUTINE AND THE CCL REGISTERS R1 AND R2
C*    (USED IN ROUTINE GETDATA) ARE MSU SYSTEM FUNCTIONS WHICH ALLOW
C*    FTN5 PROGRAMS TO COMMUNICATE WITH THE USER (AND OTHER PROGRAMS)
C*    BY A MEANS OTHER THAN THROUGH FILES. SHOULD THESE PROGRAMS NEED
C*    TO BE TRANSPORTED, THE USER COULD PROVIDE HIS OWN SUBROUTINE
C*    GETREG WHICH WOULD RETURN THE SAME VALUES, BY SOME OTHER (FTN5
C*    STANDARD) DEVICE.
C************
C* INPUT CONDITIONS:
C* 1) THE TEST DATA RESIDES ON LOCAL FILE DFILE .
C* 2) CCL REGISTER R1 IS SET TO THE VALUE OF 0 FOR AN ACTUAL-VALUE RUN;
C*    OR TO 1 FOR A NOMINAL-VALUE RUN.
C* 3) CCL REGISTER R2 IS NON-ZERO FOR A POWERS RUN; ZERO OTHERWISE.
C*
C* OUTPUT CONDITIONS:
C* 1) DFILE, R1, AND R2 ARE UNCHANGED.
C* 2) LOCAL FILES TAPE1 THROUGH TAPE24 CONTAIN THE TEST STATISTICS:
C*
C*          R        T        V        W
C* P=5 B  TAPE1    TAPE2    TAPE3    TAPE4
C*     C  TAPE5    TAPE6    TAPE7    TAPE8
C*     L  TAPE9    TAPE10   TAPE11   TAPE12
C* P=4 B  TAPE13   TAPE14   TAPE15   TAPE16
C*     C  TAPE17   TAPE18   TAPE19   TAPE20
C*     L  TAPE21   TAPE22   TAPE23   TAPE24
C************
C* OPERATION OF RM PROGRAMS:
C* 1) THE DATA-GENERATION PROGRAM DATA IS RUN.
C* 2) THE DATA-STATISTICS PROGRAM STATS IS RUN.
C* 3) CCL REGISTER R1 IS SET TO 1; R2 IS SET TO 0.
C* 4) THIS PROGRAM RM IS RUN.
```

```
C* 5) FILES TAPE1 THROUGH TAPE24 ARE SORTED; THE SORTED FILES ARE
C*    PLACED ON FILES TAPE25 THROUGH TAPE48 IN THE SAME ORDER
C*    ( TAPE1 IS SORTED ONTO TAPE25; TAPE17 IS SORTED ONTO TAPE41;
C*    TAPE<N> IS SORTED ONTO TAPE<N+24> ).
C* 6) OUTPUT PROGRAM NTABLE IS RUN.
C* 7) R1 IS SET TO D.
C* 8) THIS PROGRAM RM IS RUN.
C* 9) OUTPUT PROGRAM OTABLE IS RUN.
C* 10) STEPS 7,8,9 ARE REPEATED FOR AS MANY VALUES OF D AS ARE DESIRED.
C* 11) R1 IS SET TO 1; R2 IS SET NON-ZERO.
C* 12) THIS PROGRAM RM IS RUN.
C* 13) OUTPUT PROGRAM OTABLE IS RUN.
C* 14) STEPS 7,8,9 ARE REPEATED FOR AS MANY VALUES OF D AS ARE DESIRED.
C*
C* MISCELLANEOUS INFORMATION:
C* 1) LOCAL FILES DFILE AND NVALUE SHOULD *NEVER* BE RETURNED.
C* 2) LOCAL FILE STATFIL MAY BE RETURNED AFTER STEP 2.
C* 3) LOCAL FILES TAPE1 THROUGH TAPE24 ARE NOT NEEDED AFTER STEP 5,
C*    AND MAY BE RETURNED.
C* 4) LOCAL FILES TAPE25 THROUGH TAPE48 ARE NOT NEEDED AFTER STEP 6,
C*    AND MAY BE RETURNED.
C* 5) TAPE100 IS USED FOR DEBUG PURPOSES.
C************
C* PARAMETERS COMMON AMONG THE RM PROGRAMS:
C*
C* COUNT   - THE NUMBER OF CASES IN THE TEST
C* GROUP   - THE NUMBER OF GROUPS PER CASE
C* SUBJECT - THE NUMBER OF SUBJECTS PER GROUP
C* MEASURE - THE NUMBER OF TESTS OR MEASURES PER SUBJECT
C************
C* CODING CONVENTIONS:
C* COMMENT LINES BEGINNING WITH 'C*' DENOTE INFORMATIONAL COMMENTS,
C* THIS. COMMENT LINES BEGINNING WITH 'C ' DENOTE DEBUGGING CODE THAT
C* MAY BE USEFUL IN THE FUTURE, ETC.
C************
C* ROUTINES USED:
C* VMULFF, VMULFM, VMULFP, LINV2F - FROM IMSL.
C* TRED2, IMTQL2 - FROM EISPACK.
C************
      IMPLICIT REAL (A-Z)
      INTEGER COUNT, MEASURE, SUBJECT, GROUP
      PARAMETER ( COUNT=10000, MEASURE=5, SUBJECT=20, GROUP=3 )
      LOGICAL FIRST,SECOND
      PARAMETER ( FIRST=.TRUE., SECOND=.FALSE. )
      INTEGER ITERATE, I, J, K, IERR
      COMMON /DATA/ Z(GROUP*SUBJECT,MEASURE), ZBAR(GROUP,MEASURE),
     +   T(MEASURE,MEASURE), G(MEASURE,MEASURE),
     +   C(MEASURE,MEASURE), B(MEASURE,MEASURE), E(MEASURE,MEASURE),
     +   CTRAN(MEASURE,MEASURE), ETRAN(MEASURE,MEASURE),
     +   CRM(MEASURE-1,MEASURE-1), ERM(MEASURE-1,MEASURE-1),
     +   CLIN(MEASURE-2,MEASURE-2), ELIN(MEASURE-2,MEASURE-2),
     +   HB(MEASURE,MEASURE), HC(MEASURE-1,MEASURE-1),
     +   SCR1(MEASURE,MEASURE), SCR2(MEASURE*MEASURE+3*MEASURE),
     +   EIGB(MEASURE), EIGC(MEASURE-1), EIGL(MEASURE-2)
      REAL P1(MEASURE,MEASURE), P2(MEASURE-1,MEASURE-1), U(GROUP,1)
      COMMON /ITERATE/ ITERATE
      DATA (P1(1,I),I=1,MEASURE)
     +   /.44721, -.63246,  .53452, -.31623,  .11952/
      DATA (P1(2,I),I=1,MEASURE)
     +   /.44721, -.31623, -.26726,  .63246, -.47809/
      DATA (P1(3,I),I=1,MEASURE)
```

```
     +   /.44721, 0.0,      -.53452, 0.0,      .71714/
      DATA (P1(4,I),I=1,MEASURE)
     +   /.44721, .31623,  -.26726, -.63246, -.47809/
      DATA (P1(5,I),I=1,MEASURE)
     +   /.44721, .63246,   .53452,  .31623,  .11952/
      DATA (P2(1,I),I=1,MEASURE-1)/.5, -.67082,  .5, -.22361/
      DATA (P2(2,I),I=1,MEASURE-1)/.5, -.22361, -.5,  .67082/
      DATA (P2(3,I),I=1,MEASURE-1)/.5,  .22361, -.5, -.67082/
      DATA (P2(4,I),I=1,MEASURE-1)/.5,  .67082,  .5,  .22361/

C* SET THE UNIT VECTOR.
      DO 5 I=1,GROUP
5        U(I,1)=1.0

C* INITIALIZE OUTPUT FILES
      CALL SETOUT

C* BEGIN!
      DO 100 ITERATE=1,COUNT
         CALL GETDATA( Z )

C* COMPUTE ZBAR -- THE MEAN OF MEASURES ACROSS GROUPS
         DO 10 K=1,MEASURE
            DO 20 I=1,GROUP
               SUM=0.0
               DO 30 J=1,SUBJECT
30                SUM=SUM+Z( (I-1)*SUBJECT+J, K )
               ZBAR(I,K)=SUM/SUBJECT
20             CONTINUE
10          CONTINUE


C* DEBUG PRINT...
C     WRITE(100,*) ' ZBAR='
C     WRITE(100,'(1X,5F15.5)') ((ZBAR(I,J),J=1,MEASURE),I=1,GROUP)

C* T = Z'Z
         CALL VMULFM( Z, Z, GROUP*SUBJECT, MEASURE, MEASURE,
     +      GROUP*SUBJECT, GROUP*SUBJECT, T, MEASURE, IERR )
         IF (IERR .NE. 0) THEN
C           PRINT*, 'ERROR - IN Z''Z - IERR=', IERR
C           PRINT*,'ON ITERATION ', ITERATE
            CALL ABORT
            ENDIF

C* G = ZBAR'ZBAR (MULTIPLICATION BY SUBJECT TO FOLLOW)
         CALL VMULFM( ZBAR, ZBAR, GROUP, MEASURE, MEASURE, GROUP,
     +      GROUP, G, MEASURE, IERR )
         IF (IERR .NE. 0) THEN
C           PRINT*, 'ERROR - IN ZBAR''ZBAR - IERR=', IERR
C           PRINT*,'ON ITERATION ', ITERATE
            CALL ABORT
            ENDIF

C* C = ZBAR'U (MORE TO FOLLOW)
         CALL VMULFM( ZBAR, U, GROUP, MEASURE, 1, GROUP, GROUP, SCR2,
     +      MEASURE, IERR )
         IF (IERR .NE. 0) THEN
C           PRINT*, 'ERROR - IN ZBAR''U - IERR=', IERR
C           PRINT*,'ON ITERATION ', ITERATE
            CALL ABORT
```

```
            ENDIF

C* C = (ZBAR'U)U' (MORE TO FOLLOW)
         CALL VMULFP( SCR2, U, MEASURE, 1, GROUP, MEASURE, GROUP, SCR1,
     +      MEASURE, IERR )
         IF (IERR .NE. 0) THEN
C           PRINT*, 'ERROR - IN (ZBAR''U)U'' - IERR=', IERR
C           PRINT*,'ON ITERATION ', ITERATE
            CALL ABORT
            ENDIF

C* C = (ZBAR'U'U)ZBAR (MULTIPLICATION BY SUBJECT/GROUP TO FOLLOW)
         CALL VMULFF( SCR1, ZBAR, MEASURE, GROUP, MEASURE, MEASURE,
     +      GROUP, C, MEASURE, IERR )
         IF (IERR .NE. 0) THEN
C           PRINT*, 'ERROR - IN (ZBAR''JJ'')ZBAR - IERR=', IERR
C           PRINT*,'ON ITERATION ', ITERATE
            CALL ABORT
            ENDIF

C* G = SUBJECT*G ; C = SUBJECT/GROUP*C ; B = G-C ; E = T-G
         DO 40 J=1,MEASURE
            DO 40 I=1,MEASURE
               G(I,J)=FLOAT(SUBJECT) * G(I,J)
               C(I,J)=FLOAT(SUBJECT)/GROUP * C(I,J)
               B(I,J)=G(I,J) - C(I,J)
               E(I,J)=T(I,J) - G(I,J)
40          CONTINUE

C* DEBUG PRINT...
C     WRITE(100,*) ' C='
C     WRITE(100,'(1X,5F15.5)') ((C(I,J),J=1,5),I=1,5)
C     WRITE(100,*) ' B='
C     WRITE(100,'(1X,5F15.5)') ((B(I,J),J=1,5),I=1,5)
C     WRITE(100,*) ' E='
C     WRITE(100,'(1X,5F15.5)') ((E(I,J),J=1,5),I=1,5)

         CALL COMPUTE( P1, MEASURE )
         CALL RESULT( EIGB, MEASURE, RB, TB, VB, WB )
         CALL RESULT( EIGC, MEASURE-1, RC, TC, VC, WC )
         CALL RESULT( EIGL, MEASURE-2, RL, TL, VL, WL )

C* WRITE THE EIGENVALUES TO THE (UNSORTED) OUTPUT FILES.
         CALL OUTPUT(FIRST,RB,TB,VB,WB,RC,TC,VC,WC,RL,TL,VL,WL)

         CALL COMPUTE( P2, MEASURE-1 )
         CALL RESULT( EIGB, MEASURE-1, RB, TB, VB, WB )
         CALL RESULT( EIGC, MEASURE-2, RC, TC, VC, WC )
         CALL RESULT( EIGL, MEASURE-3, RL, TL, VL, WL )

C* WRITE THE EIGENVALUES TO THE (UNSORTED) OUTPUT FILES.
         CALL OUTPUT(SECOND,RB,TB,VB,WB,RC,TC,VC,WC,RL,TL,VL,WL)

100      CONTINUE

C* CLOSE OUTPUT FILES.
      CALL ENDOUT

C* MAKE SURE RM HASN'T OVERWRITTEN ITSELF; CLOSE DEBUG OUTPUT FILE.
C     WRITE(100,*) ' P1='
C     WRITE(100,'(1X,5F15.5)') ((P1(I,J),J=1,5),I=1,5)
```

```
C     WRITE(100,*) ' P2='
C     WRITE(100,'(1X,5F15.5)') ((P2(I,J),J=1,4),I=1,4)
C     WRITE(100,*) ' U='
C     WRITE(100,'(1X,5F15.5)') U
C     REWIND(100)

      RETURN
      END

      SUBROUTINE COMPUTE( P, LENGTH )
C************
C* COMPUTE PERFORMS SEVERAL REPETITIVE COMPUTATIONS. THE ONLY
C* DIFFERENCE AMONG THE REPETITIONS IS THE VALUE OF THE ARRAY
C* P AND THE VALUE OF LENGTH. RESULTS ARE RETURNED THROUGH /DATA/ .
C************
C* COMPUTATIONAL NOTE:
C* IF AN ERROR IS DETECTED IN INVERTING A MATRIX, THE INVERSE
C* IS SET TO 0. THIS FORCES ALL THE EIGENVALUES COMPUTED LATER
C* TO BE 0 ALSO.
C************
      IMPLICIT REAL(A-Z)
      INTEGER MEASURE, SUBJECT, GROUP
      PARAMETER ( MEASURE=5, SUBJECT=20, GROUP=3 )
      INTEGER ITERATE, I, J, K, IERR, LENGTH, OPT, IDIGIT
      COMMON /DATA/ Z(GROUP*SUBJECT,MEASURE), ZBAR(GROUP,MEASURE),
     +   T(MEASURE,MEASURE), G(MEASURE,MEASURE),
     +   C(MEASURE,MEASURE), B(MEASURE,MEASURE), E(MEASURE,MEASURE),
     +   CTRAN(MEASURE,MEASURE), ETRAN(MEASURE,MEASURE),
     +   CRM(MEASURE-1,MEASURE-1), ERM(MEASURE-1,MEASURE-1),
     +   CLIN(MEASURE-2,MEASURE-2), ELIN(MEASURE-2,MEASURE-2),
     +   HB(MEASURE,MEASURE), HC(MEASURE-1,MEASURE-1),
     +   SCR1(MEASURE,MEASURE), SCR2(MEASURE*MEASURE+3*MEASURE),
     +   EIGB(MEASURE), EIGC(MEASURE-1), EIGL(MEASURE-2)
      REAL P(LENGTH,LENGTH)
      CHARACTER*10 C1,C2
      COMMON /ITERATE/ ITERATE

C* CTRAN = P'C (MORE TO FOLLOW)
      CALL VMULFM( P, C, LENGTH, LENGTH, LENGTH, LENGTH, MEASURE,
     +   SCR1, MEASURE, IERR )
      IF (IERR .NE. 0) THEN
C        PRINT*, 'ERROR - IN P''C - LENGTH=', LENGTH, ' IERR=', IERR
C        PRINT*,'ON ITERATION ', ITERATE
         CALL ABORT
         ENDIF

C* CTRAN = (P'C)P
      CALL VMULFF( SCR1, P, LENGTH, LENGTH, LENGTH, MEASURE, LENGTH,
     +   CTRAN, MEASURE, IERR )
      IF (IERR .NE. 0) THEN
C        PRINT*, 'ERROR - IN (P''C)P - LENGTH=', LENGTH, ' IERR=', IERR
C        PRINT*,'ON ITERATION ', ITERATE
         CALL ABORT
         ENDIF

C* ETRAN = P'E (MORE TO FOLLOW)
      CALL VMULFM( P, E, LENGTH, LENGTH, LENGTH, LENGTH, MEASURE,
     +   SCR1, MEASURE, IERR )
      IF (IERR .NE. 0) THEN
C        PRINT*, 'ERROR - IN P''E - LENGTH=', LENGTH, ' IERR=', IERR
C        PRINT*,'ON ITERATION ', ITERATE
```

```
         CALL ABORT
         ENDIF

C* ETRAN = (P'E)P
      CALL VMULFF( SCR1, P, LENGTH, LENGTH, LENGTH, MEASURE, LENGTH,
     +   ETRAN, MEASURE, IERR )
      IF (IERR .NE. 0) THEN
C        PRINT*, 'ERROR - IN (P''E)P - LENGTH=', LENGTH, ' IERR=', IERR
C        PRINT*,'ON ITERATION ', ITERATE
         CALL ABORT
         ENDIF

C* DEBUG PRINT...
C     WRITE(100,*) ' CTRAN='
C     WRITE(100,'(1X,5F15.5)')  ((CTRAN(I,J),J=1,LENGTH),I=1,LENGTH)
C     WRITE(100,*) ' ETRAN='
C     WRITE(100,'(1X,5F15.5)')  ((ETRAN(I,J),J=1,LENGTH),I=1,LENGTH)

      DO 20 J=2,LENGTH
         DO 20 I=2,LENGTH
            CRM(I-1,J-1)=CTRAN(I,J)
            ERM(I-1,J-1)=ETRAN(I,J)
20       CONTINUE

C* HB = B E-INVERSE - SCR1 CONTAINS E-INVERSE
      IDIGIT=0
      CALL LINV2F( E, LENGTH, MEASURE, SCR1, IDIGIT, SCR2, IERR )
      IF (IERR .NE. 0) THEN
         CALL INT2CHR( ITERATE, C2 )
         CALL INT2CHR( IERR, C1 )
         CALL REMARK('ERROR - E-INV - IERR='//C1//' ITERATE='//C2)
C        CALL ABORT
         DO 50 J=1,MEASURE
            DO 50 I=1,MEASURE
50             SCR1(I,J)=0.0
         ENDIF

      CALL VMULFF( B, SCR1, LENGTH, LENGTH, LENGTH, MEASURE, MEASURE,
     +   HB, MEASURE, IERR )
      IF (IERR .NE. 0) THEN
C        PRINT*, 'ERROR - IN B''E - LENGTH=', LENGTH, ' IERR=', IERR
C        PRINT*,'ON ITERATION ', ITERATE
         CALL ABORT
         ENDIF


C* HC = CRM ERM-INVERSE - SCR1 CONTAINS ERM-INVERSE
      IDIGIT=0
      CALL LINV2F( ERM, LENGTH-1, MEASURE-1, SCR1, IDIGIT, SCR2, IERR )
      IF (IERR .NE. 0) THEN
         CALL INT2CHR( ITERATE, C1 )
         CALL INT2CHR( IERR, C2 )
         CALL REMARK('ERROR - ERM -INV - IERR='//C2//' ITERATE='//C1)
C        CALL ABORT
         DO 40 J=1,MEASURE
            DO 40 I=1,MEASURE
40             SCR1(I,J)=0.0
         ENDIF

      CALL VMULFF( CRM, SCR1, LENGTH-1, LENGTH-1, LENGTH-1, MEASURE-1,
     +   MEASURE-1, HC, MEASURE-1, IERR )
```

```
      IF (IERR .NE. 0) THEN
C        PRINT*,'ERROR - IN CRM ERM-INV - LENGTH=',LENGTH,' IERR=',IERR
C        PRINT*,'ON ITERATION ', ITERATE
         CALL ABORT
         ENDIF

C* CALL EISPACK ROUTINES TRED2 AND IMTQL2 TO DO EIGENVALUES.
C* AFTER TRED2, SCR1 WILL CONTAIN Z
C*              EIGB WILL CONTAIN D
C*              SCR2 WILL CONTAIN E
      CALL TRED2( MEASURE, LENGTH, HB, EIGB, SCR2, SCR1 )
      CALL IMTQL2( MEASURE, LENGTH, EIGB, SCR2, SCR1, IERR )
      IF (IERR .NE. 0) THEN
         CALL INT2CHR( ITERATE, C1 )
         CALL INT2CHR( IERR, C2 )
         CALL REMARK('ERROR - HB IMTQL2 - IERR='//C2//' ITERATE='//C1)
         CALL ABORT
         ENDIF

      CALL TRED2( MEASURE-1, LENGTH-1, HC, EIGC, SCR2, SCR1 )
      CALL IMTQL2(MEASURE-1, LENGTH-1, EIGC, SCR2, SCR1, IERR)
      IF (IERR .NE. 0) THEN
         CALL INT2CHR( ITERATE, C1 )
         CALL INT2CHR( IERR, C2 )
         CALL REMARK('ERROR - HC IMTQL2 - IERR='//C2//' ITERATE='//C1)
         CALL ABORT
         ENDIF

C* DEBUG PRINT...
C     WRITE(100,*) ' EIGC='
C     WRITE(100,*) (EIGC(I),I=1,LENGTH-1)

C*
C* PERFORM LINEAR TRENDS STATISTICS.
C*
      DO 10 J=3,LENGTH
         DO 10 I=3,LENGTH
            CLIN(I-2,J-2)=CTRAN(I,J)
            ELIN(I-2,J-2)=ETRAN(I,J)
10       CONTINUE

C* HLIN = CLIN ELIN-INVERSE - SCR1 CONTAINS ELIN-INVERSE
C* STORE HLIN IN THE HC ARRAY, SINCE THE DATA IN HC WILL NOT
C* BE REUSED.
C*
C* DEBUG PRINT...
C     WRITE(100,*) ' CLIN='
C     WRITE(100,'(1X,3F15.5)') ((CLIN(I,J),J=1,LENGTH-2),I=1,LENGTH-2)
C     WRITE(100,*) ' ELIN='
C     WRITE(100,'(1X,3F15.5)') ((ELIN(I,J),J=1,LENGTH-2),I=1,LENGTH-2)

      IDIGIT=0
      CALL LINV2F( ELIN, LENGTH-2, MEASURE-2, SCR1, IDIGIT, SCR2, IERR )
      IF (IERR .NE. 0) THEN
         CALL INT2CHR( ITERATE, C1 )
         CALL INT2CHR( IERR, C2 )
         CALL REMARK('ERROR - ELIN -INV - IERR='//C2//' ITERATE='//C1)
C        CALL ABORT
         DO 30 I=1,MEASURE
            DO 30 J=1,MEASURE
30             SCR1(J,I)=0.0
```

```
            ENDIF

         CALL VMULFF( CLIN, SCR1, LENGTH-2, LENGTH-2, LENGTH-2, MEASURE-2,
     +      MEASURE-2, HC, MEASURE-2, IERR )
         IF (IERR .NE. 0) THEN
C           PRINT*,'ERROR-IN CLIN ELIN-INV. LENGTH=',LENGTH,' IERR=',IERR
C           PRINT*,'ON ITERATION ', ITERATE
            CALL ABORT
            ENDIF

         CALL TRED2( MEASURE-2, LENGTH-2, HC, EIGL, SCR2, SCR1 )
         CALL IMTQL2(MEASURE-2, LENGTH-2, EIGL, SCR2, SCR1, IERR)
         IF (IERR .NE. 0) THEN
            CALL INT2CHR( ITERATE, C1 )
            CALL INT2CHR( IERR, C2 )
            CALL REMARK('ERROR - HL IMTQL2 - IERR='//C2//' ITERATE='//C1)
            CALL ABORT
            ENDIF

C* DEBUG PRINT...
C        WRITE(100,*) ' EIGL='
C        WRITE(100,*)  (EIGL(I),I=1,LENGTH-2)
         RETURN
         END

         SUBROUTINE RESULT( EIGEN, LENGTH, R, T, V, W )
C*************
C* RESULT CALCULATES SEVERAL STATISTICS (R, T, V, AND W) BASED
C* ON THE EIGENVALUES IN ARRAY EIGEN .
C*************
         IMPLICIT REAL(A-Z)
         INTEGER LENGTH, I, ITERATE
         DIMENSION EIGEN(LENGTH)
         COMMON /ITERATE/ ITERATE
C        INTEGER DEBUG
C        DATA DEBUG/10/

         V=T=0.0
         W=1.0

         DO 10 I=1, LENGTH
            VALUE=EIGEN(I)
            T=T + VALUE
            V=V + VALUE/(1.0 + VALUE)
            W=W/(1.0 + VALUE)
10          CONTINUE

         R=EIGEN(LENGTH)/(1.0 + EIGEN(LENGTH))

C        IF (DEBUG .GT. 0) THEN
C           DEBUG=DEBUG-1
C           WRITE(100,*) 'EIGENVALUES=',EIGEN
C           WRITE(100,*) 'STATS=',R,T,V,W
C           ENDIF
         RETURN
         END

         SUBROUTINE GETDATA( Z )
C*************
C* GETDATA RETURNS THE NEXT SET OF VALUES TO BE ANALYZED BY RM.
C*
```

```
C* THE DATA IS READ FROM LOCAL FILE DFILE, WHICH IS INITIALLY
C* REWOUND.
C************
      IMPLICIT REAL(A-Z)
      INTEGER MEASURE, SUBJECT, GROUP
      PARAMETER ( MEASURE=5, SUBJECT=20, GROUP=3 )
      REAL Z(GROUP*SUBJECT,MEASURE), POWER(2:MEASURE)
      INTEGER FET(8), BUF(2049), EOP
      INTEGER I, J, D, POWERON
      LOGICAL FIRST
      CHARACTER DC*3
      DATA FIRST/.TRUE./, POWER/.4, .8, .5, .1/

C* IF THIS IS THE FIRST CALL TO GETDATA, INITIALIZE THE DATA FILE,
C* CHECK CCL REGISTER 1 FOR THE D PARAMETER, AND CHECK R2 FOR THE
C* POWER PARAMETER.
      IF (FIRST) THEN
         CALL FILEC( 'DFILE', FET, 8, BUF, 2049 )
         CALL REWINDF( FET )
C*
C* FTN5 STANDARD CODE - IF RM NEEDS TO BE TRANSPORTED,
C* DELETE THE PRECEDING FILEC AND REWINDF CALLS, AND USE
C* THE FOLLOWING CODE:
C*
C        OPEN(999,FILE='DFILE')
C        REWIND(999)
C*

         FIRST=.FALSE.

         CALL GETREG( 'R1', D )
         IF (D .NE. 1) THEN
            DSQRT=SQRT(FLOAT(D))
            CALL INT2CHR( D, DC )
            CALL REMARK( 'RM CALLED WITH D='//DC )
            ENDIF

         CALL GETREG( 'R2', POWERON )
         IF (POWERON .NE. 0) THEN
            CALL REMARK( ' CALCULATING POWERS' )
            ENDIF
         ENDIF

      CALL READW( FET, Z, GROUP*MEASURE*SUBJECT, EOP )
C*
C* FTN5 STANDARD CODE - IF RM NEEDS TO BE TRANSPORTED,
C* DELETE THE PRECEDING READW CALL, AND USE
C* THE FOLLOWING CODE:
C*
C        READ(999,*,IOSTAT=EOP) Z
C*

      IF (EOP .EQ. 0) THEN
C* MULTIPLY ONLY THE FIRST GROUP OF Z BY THE SQUARE ROOT OF D.
         IF (D .NE. 1) THEN
            DO 10 J=1,MEASURE
               DO 10 I=1,SUBJECT
10                Z(I,J)=DSQRT * Z(I,J)
            ENDIF

C* ADD CONSTANTS TO ALL GROUPS.
```

```
         IF (POWERON .NE. 0) THEN
            DO 20 J=2,MEASURE
               DO 20 I=1,SUBJECT*GROUP
20                Z(I,J)=Z(I,J) + POWER(J)
            ENDIF

      ELSE
         CALL REMARK( 'UNEXPECTED *EOP ON READING DATA.' )
         CALL ABORT
         ENDIF

C     WRITE(100,*) ' DATA='
C     WRITE(100,'(1X,5F15.5)') ((Z(I,J),J=1,MEASURE),I=1,GROUP*SUBJECT)
      RETURN
      END

      SUBROUTINE SETOUT
C*************
C* SETOUT INITIALIZES THE OUTPUT FILES THAT THE UNSORTED TEST
C* STATISTICS WILL BE WRITTEN TO.
C* FILES USED ARE 'TAPE1' THROUGH 'TAPE24' .
C*************
      IMPLICIT INTEGER(A-Z)
      COMMON /IO/ FET( 8, 24 )
      DIMENSION BUF( 513, 24 )
      CHARACTER UNITLFN*7

      UNITLFN='TAPE'

C* RETURN EACH FILE BEFORE ANY FURTHER PROCESSING.
      DO 10 I=1,24
         CALL INT2CHR( I, UNITLFN(5:) )
         CALL FILEC( UNITLFN, FET(1,I), 8, BUF(1,I), 513 )
         CALL RETF( FET(1,I) )
         CALL FILEC( UNITLFN, FET(1,I), 8, BUF(1,I), 513 )
10       CONTINUE
C*
C* FTN5 STANDARD CODE - IF RM NEEDS TO BE TRANSPORTED,
C* DELETE THE DO LOOP, AND USE THE FOLLOWING CODE:
C*
C        DO 10 I=1,24
C        OPEN( I )
C        CLOSE( I, STATUS='DELETE' )
C        OPEN( I )
C  10    CONTINUE
C*
      RETURN
      END

      SUBROUTINE OUTPUT( FIRST, RB, TB, VB, WB, RC, TC, VC, WC,
     +   RL, TL, VL, WL )
C*************
C* OUTPUT WRITES THE EIGENVALUES TO THE OUTPUT FILES.
C* WILL BE WRITTEN TO.
C* FILES USED ARE 'TAPE1' THROUGH 'TAPE24' .
C*
C* FIRST = .TRUE. IFF THIS SET OF EIGENVALUES WAS OBTAINED WITH
C* LENGTH = MEASURE; FIRST = .FALSE. IFF LENGTH = MEASURE-1 .
C*************
      IMPLICIT REAL(A-Z)
      INTEGER BEGIN
```

```
      LOGICAL FIRST
      COMMON /IO/ FET( 8, 24 )

      IF (FIRST) THEN
         BEGIN=0
      ELSE
         BEGIN=12
         ENDIF

      CALL WRITEW( FET(1,1+BEGIN), RB, 1 )
      CALL WRITEW( FET(1,2+BEGIN), TB, 1 )
      CALL WRITEW( FET(1,3+BEGIN), VB, 1 )
      CALL WRITEW( FET(1,4+BEGIN), WB, 1 )
      CALL WRITEW( FET(1,5+BEGIN), RC, 1 )
      CALL WRITEW( FET(1,6+BEGIN), TC, 1 )
      CALL WRITEW( FET(1,7+BEGIN), VC, 1 )
      CALL WRITEW( FET(1,8+BEGIN), WC, 1 )
      CALL WRITEW( FET(1,9+BEGIN), RL, 1 )
      CALL WRITEW( FET(1,10+BEGIN), TL, 1 )
      CALL WRITEW( FET(1,11+BEGIN), VL, 1 )
      CALL WRITEW( FET(1,12+BEGIN), WL, 1 )

C*
C* FTN5 STANDARD CODE - IF RM NEEDS TO BE TRANSPORTED,
C* DELETE THE WRITEW CALLS ABOVE, AND USE THE FOLLOWING CODE:
C*
C     WRITE( BEGIN+1, * ) RB
C     WRITE( BEGIN+2, * ) TB
C     WRITE( BEGIN+3, * ) VB
C     WRITE( BEGIN+4, * ) WB
C     WRITE( BEGIN+5, * ) RC
C     WRITE( BEGIN+6, * ) TC
C     WRITE( BEGIN+7, * ) VC
C     WRITE( BEGIN+8, * ) WC
C     WRITE( BEGIN+9, * ) RL
C     WRITE( BEGIN+10,* ) TL
C     WRITE( BEGIN+11,* ) VL
C     WRITE( BEGIN+12,* ) WL
C*
      RETURN
      END

      SUBROUTINE ENDOUT
C************
C* ENDOUT CLOSES THE FILES THAT THE EIGENVALUES WERE WRITTEN TO.
C************
      IMPLICIT INTEGER(A-Z)
      COMMON /IO/ FET( 8, 24 )

      DO 10 I=1,24
10       CALL WRITEOR( FET(1,I) )
C*
C* FTN5 STANDARD CODE - IF RM NEEDS TO BE TRANSPORTED,
C* DELETE THE DO LOOP CALL ABOVE, AND USE THE FOLLOWING CODE:
C*
C        DO 10 I=1,24
C  10    REWIND( I )
C*
      RETURN
      END
*EOS00  LINE=672 SEC=1
```

```
          IDENT    DATA
**************
* DATA-GENERATION PROGRAM. CREATES COUNT COLLECTIONS OF DATA, EACH WITH
* MEASURE*GROUP*SUBJECT ITEMS, USING THE BOX-MUELLER (SP?) METHOD.
*
* DATA IS WRITTEN TO LOCAL FILE DFILE .
*
* SEE THE COMMENT SECTION FOR PROGRAM RM FOR DETAILED INFORMATION
* ABOUT THE OPERATION OF THESE PROGRAMS.
**************
          ENTRY    DATA
DATA      BSS      0
          RJ       =XXDATA
          ENDRUN
          END      DATA
      SUBROUTINE XDATA
C************
C* ROUTINES USED:
C*
C* GGUBS - FROM IMSL.
C************
      IMPLICIT INTEGER(A-Z)
      PARAMETER (COUNT=10000, MEASURE=5, GROUP=3, SUBJECT=20)
      REAL ARRAY(MEASURE*GROUP*SUBJECT)
      REAL R, THETA, AVERAGE, PIX2
      DIMENSION FET(8), BUF(2049)
      DOUBLE PRECISION DSEED
      DATA DSEED/739604919.0D0/

      PIX2=8.0*ATAN(1.0)

      CALL FILEC( 'DFILE', FET, 8, BUF, 2049 )
      CALL RETF( FET )
      CALL FILEC( 'DFILE', FET, 8, BUF, 2049 )
C*
C* FTN5 STANDARD CODE - DELETE THE PRECEDING FILEC AND RETF CALLS,
C* AND USE THE FOLLOWING CODE:
C*
C       OPEN(1,FILE='DFILE')
C       CLOSE(1,STATUS='DELETE')
C       OPEN(1,FILE='DFILE')
C*


      AVERAGE=0.0

      DO 10 J=1, COUNT
         CALL GGUBS( DSEED, MEASURE*GROUP*SUBJECT, ARRAY )

         DO 20 I=1,MEASURE*GROUP*SUBJECT/2
            R=SQRT( -2.0 *  LOG(ARRAY(2*I-1)) )
            THETA=PIX2 * ARRAY(2*I)

            ARRAY(2*I-1)=R * COS( THETA )
            ARRAY(2*I)=R * SIN( THETA )

            AVERAGE=AVERAGE+ARRAY(2*I)+ARRAY(2*I-1)
20          CONTINUE


         CALL WRITEW( FET, ARRAY, MEASURE*GROUP*SUBJECT )
C*
C* FTN5 STANDARD CODE - DELETE THE PRECEDING WRITEW CALL,
```

```
C* AND USE THE FOLLOWING CODE:
C*
C       WRITE(1,*) ARRAY
C*
10       CONTINUE

      AVERAGE=AVERAGE/(GROUP*MEASURE*SUBJECT*COUNT)

      CALL WRITEOR( FET )
C*
C* FTN5 STANDARD CODE - DELETE THE PRECEDING WRITEOR CALL,
C* AND USE THE FOLLOWING CODE:
C*
C       REWIND 1
C*


C*
C* WRITE THE AVERAGE OF THE DATA GENERATED TO LOCAL FILE
C* STATFIL . THIS AVERAGE WILL BE NEEDED IN ORDER TO COMPUTE
C* THE VARIANCE OF THE DATA.
C*
      CALL WNBF( FET )
      CALL FILEC( 'STATFIL', FET, 8, BUF, 65 )
      CALL RETF( FET )
      CALL WRITEW( FET, AVERAGE, 1 )
      CALL WRITEOR( FET )
C*
C* FTN5 STANDARD CODE - DELETE THE PRECEDING FILEC AND RETF CALLS,
C* AND USE THE FOLLOWING CODE:
C*
C       OPEN(2,FILE='STATFIL')
C       CLOSE(2,STATUS='DELETE')
C       OPEN(2,FILE='STATFIL')
C       WRITE(1,*) AVERAGE
C       REWIND 1
C*

      RETURN
      END
*EOSOO  LINE=103 SEC=1
```

```
          IDENT    STATS
**************
* DATA-STATISTICS PROGRAM. COMPUTES THE VARIANCE OF THE DATA, USING
* THE AVERAGE ALREADY CALCULATED BY THE DATA-GENERATION PROGRAM.
* BOTH THE AVERAGE AND THE VARIANCE ARE WRITTEN TO LOCAL FILE STATFIL .
*
* INPUT CONDITIONS:
* DATA IS ON FILE DFILE ; THE AVERAGE IS ON LOCAL FILE STATFIL .
*
* SEE THE COMMENT SECTION OF PROGRAM RM FOR MORE DETAILED EXPLANATION
* OF THE FUNCTIONING OF THESE PROGRAMS.
**************
          ENTRY    STATS
STATS     BSS      0
          RJ       =XXSTATS
          ENDRUN
          END      STATS
      SUBROUTINE XSTATS
      IMPLICIT INTEGER(A-Z)
      PARAMETER (COUNT=10000, MEASURE=5, GROUP=3, SUBJECT=20)
      REAL ARRAY(MEASURE*GROUP*SUBJECT), AVERAGE, VAR
      DIMENSION FET(8), BUF(2049)
C     REAL OBSC(100), STAT(3), CSOBS(100)

C* READ THE DATA AVERAGE ALREADY COMPUTED.
      CALL FILEC( 'STATFIL', FET, 8, BUF, 65 )
      CALL REWINDF( FET )
      CALL READW( FET, AVERAGE, 1, EOP )
      CALL WNBF( FET )

      CALL FILEC( 'DFILE', FET, 8, BUF, 2049 )
      CALL REWINDF( FET )

C     OPEN(1,FILE='RAWDATA')
C     REWIND 1
C        DO 5 I=1,100
C  5        CSOBS(I)=OBSC(I)=0.0

      VAR=0.0
      DO 10 J=1, COUNT
         CALL READW( FET, ARRAY, MEASURE*GROUP*SUBJECT, EOP )
         IF (EOP .LT. 0) THEN
            CALL REMARK( 'UNEXPECTED *EOP ON DATA FILE.' )
            RETURN
            ENDIF

         DO 20 I=1,MEASURE*GROUP*SUBJECT
20          VAR=VAR + (ARRAY(I)-AVERAGE)*(ARRAY(I)-AVERAGE)

C           IF (J .NE. COUNT) THEN
C              STAT(3)=0
C           ELSE
C              STAT(3)=1
C              ENDIF
C           K=100
C     CALL GTNOR(ARRAY,GROUP*MEASURE*SUBJECT,K,STAT,OBSC,CSOBS,IERR)
C           IF (IERR .NE. 0) WRITE(1,*) ' IERR=',IERR

10       CONTINUE

      VAR=VAR/(COUNT*GROUP*MEASURE*SUBJECT)

C     WRITE(1,*) STAT
```

```
C     REWIND 1

C*
C* WRITE THE AVERAGE AND VARIANCE OF THE DATA GENERATED TO LOCAL
C* FILE STATFIL .
C*
      CALL WNBF( FET )
      CALL FILEC( 'STATFIL', FET, 8, BUF, 65 )
      CALL RETF( FET )
      CALL WRITEW( FET, AVERAGE, 1 )
      CALL WRITEW( FET, VAR, 1 )
      CALL WRITEOR( FET )

      RETURN
      END
*EOSOO  LINE=78 SEC=1
```

```
          IDENT    SORT
***************
* SORT - SHELL-METZNER SORT OF TEST STATISTICS.
*
* CALLING SEQUENCE:
*
*   SORT,INLFN,OUTLFN.
*
* SEE THE COMMENT SECTION OF PROGRAM RM FOR A DETAILED EXPLANATION
* OF THE FUNCTIONING OF THESE PROGRAMS AND THEIR INTERACTION.
***************
          ENTRY    SORT
SORT      BSS      0
          SA1      PLIST            PLIST CONTAINS INLFN AND OUTLFN
          RJ       =XXSORT
          ENDRUN
PLIST     BSS      0
          CON      2
          CON      3
          DATA     0
          END      SORT
      SUBROUTINE XSORT( INLFN, OUTLFN )
      IMPLICIT INTEGER(A-Z)
      PARAMETER (COUNT=10000)
      REAL ARRAY(COUNT), T
      DIMENSION FET(8), BUF(513)

      CALL FILEC( INLFN, FET, 8, BUF, 513 )
      CALL REWINDF( FET )
      CALL READW( FET, ARRAY, COUNT, EOP, LEVEL, NGET )
      CALL WNBF( FET )
C*
C* FTN5 STANDARD CODE - DELETE THE PRECEDING FILEC, REWINDF, READW,
C* AND WNBF CALLS, CHANGE INLFN AND OUTLFN TO CHARACTER VARIABLES,
C* AND USE THE FOLLOWING CODE:
C*
C       OPEN(1,FILE=INLFN)
C       REWIND 1
C       READ(1,*,IOSTAT=EOP) ARRAY
C*
      IF (EOP .NE. 0) THEN
         CALL REMARK( 'UNEXPECTED *EOP ON READING.' )
         ENDIF

      M=NGET

10    CONTINUE
      M=M/2

      IF (M .EQ. 0) THEN
         CALL FILEC( OUTLFN, FET, 8, BUF, 513 )
         CALL RETF( FET )
         CALL FILEC( OUTLFN, FET, 8, BUF, 513 )
         CALL WRITEW( FET, ARRAY, NGET )
         CALL WRITEOR( FET )
C*
C* FTN5 STANDARD CODE - DELETE THE PRECEDING FILEC, RETF, WRITEW,
C* AND WRITEOR CALLS, AND USE THE FOLLOWING CODE:
C*
C       OPEN(2,FILE=OUTLFN)
C       REWIND 2
C       WRITE(2,*) ARRAY
C       REWIND 2
```

```
C*

            RETURN
            ENDIF

C* ELSE...
         K=NGET-M
         J=1

20       CONTINUE
         I=J

30       CONTINUE
         L=I+M

         IF (ARRAY(I) .GT. ARRAY(L)) THEN
            T=ARRAY(I)
            ARRAY(I)=ARRAY(L)
            ARRAY(L)=T

            I=I-M

            IF (I .GE. 1) GOTO 30
            ENDIF

         J=J+1

         IF (J .GT. K) GOTO 10
         GOTO 20
         END
*EOS00  LINE=93 SEC=1
```

```
      PROGRAM NTABLE( OUTPUT )
C******************
C* TABLE READS THE SORTED LISTS OF TEST STATISTICS, AND CALCULATES
C* THE VALUES FOR ALPHA = 0.01, 0.05, AND 0.10 .
C*
C* FOR THE EXPECTED SIZE OF 10,000 , THESE VALUES ARE CALCULATED
C* BY AVERAGING THE 100TH AND 101ST, THE 500TH AND THE 501ST, AND
C* THE 1000TH AND THE 1001ST ELEMENTS, RESPECTIVELY, FOR W, AND
C* THE 9900TH AND 9901ST, 9500TH AND 9501ST, AND 9000TH AND 9001ST
C* ELEMENTS FOR THE R, T, AND V TESTS.
C*
C* CALLING SEQUENCE:
C*
C*    NTABLE,OUTLFN.
C*
C* WHERE OUTLFN IS THE FILE TO WHICH THE TABLED OUTPUT IS TO BE WRITTEN.
C* DEFAULT: OUTPUT
C*
C* TAPE25 THROUGH TAPE48 CONTAIN THE SORTED TEST STATISTICS:
C*
C*              R         T          V          W
C*
C* P=5  B    TAPE25    TAPE26     TAPE27     TAPE28
C*      C    TAPE29    TAPE30     TAPE31     TAPE32
C*      L    TAPE33    TAPE34     TAPE35     TAPE36
C* P=4  B    TAPE37    TAPE38     TAPE39     TAPE40
C*      C    TAPE41    TAPE42     TAPE43     TAPE44
C*      L    TAPE45    TAPE46     TAPE47     TAPE48
C*
C******************
C* NTABLE IS USED TO PRODUCE THE OUTPUT FOR THE NOMINAL-VALUE RUN.
C* IT ALSO WRITES THE NOMINAL VALUES TO FILE NVALUE FOR USE BY THE
C* OBSERVED-VALUE TABLE-GENERATION PROGRAM.
C******************
C* SEE THE COMMENT SECTION OF PROGRAM RM FOR A DETAILED EXPLANATION OF
C* THE INTERACTION OF THESE PROGRAMS.
C******************
      IMPLICIT INTEGER(A-Z)
      PARAMETER (COUNT=10000)
      PARAMETER (A1=COUNT/100, A2=COUNT/20, A3=COUNT/10)
      PARAMETER (B1=COUNT-A1, B2=COUNT-A2, B3=COUNT-A3)
      REAL ARRAY(COUNT)
      DIMENSION FET(8), BUF(513)
      CHARACTER UNITLFN*7, TITLE*80
      REAL N1(24), N2(24), N3(24)

      CALL NOMSG

      CALL FILEC( 'ZZZZIN', FET, 8, BUF, 65 )
      CALL CONNECF( FET, 0 )
      CALL WRITEH( FET, ' NTABLE - PLEASE ENTER A TITLE -', 4 )
      CALL READH( FET, TITLE, 8, EOP )
C*
C* FTN5 STANDARD CODE - DELETE THE PRECEDING FILEC, CONNECF, WRITEH,
C* AND READH CALLS, AND SUBSTITUTE SOME OTHER METHOD OF READING IN
C* A TITLE FROM THE USER.
C*

      L=LNB(TITLE)
      DO 5 I=L,1,-1
         P=(80-L)/2+I
         TITLE(P:P)=TITLE(I:I)
5        CONTINUE
```

```
      TITLE(:(80-L)/2)=' '

      UNITLFN='TAPE'
      DO 10 I=1,24
         UNIT=I+24
         CALL INT2CHR( UNIT, UNITLFN(5:) )

         CALL FILEC( UNITLFN, FET, 8, BUF, 513 )
         CALL REWINDF( FET )
         CALL READW( FET, ARRAY, COUNT, EOP )
         CALL WNBF( FET )
C*
C* FTN5 STANDARD CODE - DELETE THE PRECEDING FILEC, REWINDF, READW,
C* AND WNBF CALLS, AND USE THE FOLLOWING CODE:
C*
C        OPEN( 1,FILE=UNITLFN )
C        REWIND 1
C        READ(1,*,IOSTAT=EOP) ARRAY
C*
         IF (EOP .EQ. 0) THEN
C*
C* I MOD 4 = 0 INDICATES A W TEST;
C* OTHERWISE IT IS AN R, T, OR V TEST.
C*
            IF (MOD(I,4) .EQ. 0) THEN
               N1(I)=( ARRAY(A1)+ARRAY(A1+1) )/2
               N2(I)=( ARRAY(A2)+ARRAY(A2+1) )/2
               N3(I)=( ARRAY(A3)+ARRAY(A3+1) )/2

            ELSE
               N1(I)=( ARRAY(B1)+ARRAY(B1+1) )/2
               N2(I)=( ARRAY(B2)+ARRAY(B2+1) )/2
               N3(I)=( ARRAY(B3)+ARRAY(B3+1) )/2
               ENDIF


         ELSE
            CALL REMARK( 'UNEXPECTED *EOP ON READ OF '//UNITLFN )
            N1(I)=N2(I)=N3(I)=0.0
            ENDIF

10       CONTINUE

      WRITE(*,'(A)') '1'
      WRITE(*,'(1X,A)') TITLE

      WRITE(*,*) ' '
      WRITE(*,'(T6,A,T15,A,2X,A,T31,4(A,13X)')
     +   'ALPHA', 'P', 'TEST', 'R', 'T', 'V', 'W'

      WRITE(*,*) ' '
      WRITE(*,100)
     +   '0.01', '5', 'B', (N1(I),I=1,4), 'C', (N1(I),I=5,8),
     +   'L', (N1(I),I=9,12)
      WRITE(*,110)
     +   '4', 'B', (N1(I),I=13,16), 'C', (N1(I),I=17,20),
     +   'L', (N1(I),I=21,24)

      WRITE(*,100)
     +   '0.05', '5', 'B', (N2(I),I=1,4), 'C', (N2(I),I=5,8),
     +   'L', (N2(I),I=9,12)
```

```
      WRITE(*,110)
     +   '4', 'B', (N2(I),I=13,16), 'C', (N2(I),I=17,20),
     +   'L', (N2(I),I=21,24)

      WRITE(*,100)
     +   '0.10', '5', 'B', (N3(I),I=1,4), 'C', (N3(I),I=5,8),
     +   'L', (N3(I),I=9,12)
      WRITE(*,110)
     +   '4', 'B', (N3(I),I=13,16), 'C', (N3(I),I=17,20),
     +   'L', (N3(I),I=21,24)

C*
C* WRITE THE NOMINAL VALUES TO NVALUE.
C*
      CALL FILEC( 'NVALUE', FET, 8, BUF, 513 )
      CALL RETF( FET )
      CALL FILEC( 'NVALUE', FET, 8, BUF, 513 )
      CALL WRITEW( FET, N1, 24 )
      CALL WRITEW( FET, N2, 24 )
      CALL WRITEW( FET, N3, 24 )
      CALL WRITEOR( FET )
C*
C* FTN5 STANDARD CODE - DELETE THE PRECEDING FILEC, RETF, WRITEW,
C* AND WRITEOR CALLS, AND USE THE FOLLOWING CODE:
C*
C       OPEN( 2,FILE='NVALUE' )
C       REWIND 2
C       WRITE(2,*) N1
C       WRITE(2,*) N2
C       WRITE(2,*) N3
C       REWIND 2
C*

      STOP
100   FORMAT( T6,A,T15,A,3(T19,A,T20,4F14.5,/) )
110   FORMAT( T15,A,3(T19,A,T20,4F14.5,/) )
      END
*EOSOO  LINE=161 SEC=1
```

```
      PROGRAM OTABLE( OUTPUT )
C*******************
C* OTABLE READS THE LISTS OF TEST STATISTICS AND CALCULATES THE
C* ACTUAL VALUES FOR ALPHA = 0.01, 0.05, AND 0.10 BY EMPIRICALLY
C* FINDING THE PROPORTION OF STATISTICS EXCEEDING THE NOMINAL VALUES.
C*
C* FOR THE EXPECTED SIZE OF 10,000 , THESE VALUES ARE CALCULATED
C* BY AVERAGING THE 100TH AND 101ST, THE 500TH AND THE 501ST, AND
C* THE 1000TH AND THE 1001ST ELEMENTS, RESPECTIVELY.
C*
C* CALLING SEQUENCE:
C*
C*    OTABLE,OUTLFN.
C*
C* WHERE OUTLFN IS THE FILE TO WHICH THE TABLED OUTPUT IS TO BE WRITTEN.
C* DEFAULT: OUTPUT
C*******************
C* REFER TO THE COMMENT SECTION OF PROGRAM RM FOR A DETAILED EXPLANATION
C* OF THE INTERACTION OF THESE PROGRAMS; REFER TO THE COMMENT SECTION
C* OF PROGRAM NTABLE FOR INFORMATION ON MAKING THIS PROGRAM TRANSPORT-
C* ABLE (THE PROCEDURE IS ALMOST EXACTLY THE SAME AS FOR NTABLE).
C*******************
      IMPLICIT INTEGER(A-Z)
      PARAMETER (COUNT=10000)
      REAL ARRAY(COUNT)
      DIMENSION FET(8), BUF(2049)
      CHARACTER UNITLFN*7, TITLE*80, ANSWER
      REAL N1(24), N2(24), N3(24), O1(24), O2(24), O3(24)
      LOGICAL SHOWB

      CALL NOMSG

      CALL FILEC( 'ZZZZIN', FET, 8, BUF, 65 )
      CALL CONNECF( FET, 0 )
      CALL WRITEH( FET, ' OTABLE - PLEASE ENTER A TITLE -', 4 )
      CALL READH( FET, TITLE, 8, EOP )
      L=LNB(TITLE)
      DO 5 I=L,1,-1
         P=(80-L)/2+I
         TITLE(P:P)=TITLE(I:I)
5        CONTINUE
      TITLE(:(80-L)/2)=' '

      CALL WRITEH( FET, ' OTABLE - PRINT B TEST?', 3 )
      CALL READH( FET, ANSWER, 1, EOP )
      SHOWB=ANSWER(:1) .EQ. 'Y'

      CALL FILEC( 'NVALUE', FET, 8, BUF, 2049 )
      CALL REWINDF( FET )
      CALL READW( FET, N1, 24, EOP )
      CALL READW( FET, N2, 24, EOP )
      CALL READW( FET, N3, 24, EOP )
      IF (EOP .LT. 0) THEN
         CALL REMARK('UNEXPECTED *EOP ON NVALUE FILE.' )
         CALL ABORT
         ENDIF

      UNITLFN='TAPE'

      OPEN(1,FILE='RAWDATA')
      REWIND 1

      DO 10 I=1,24
```

```
         CALL INT2CHR( I, UNITLFN(5:) )

         CALL FILEC( UNITLFN, FET, 8, BUF, 513 )
         CALL REWINDF( FET )

         CALL READW( FET, ARRAY, COUNT, EOP )
         CALL WNBF( FET )
         IF (EOP .EQ. 0) THEN

            WRITE(1,*) UNITLFN, ARRAY(1)

            I1=I2=I3=0

C*
C* I MOD 4 = 0 INDICATES THAT THE FILE CONTAINS VALUES FROM THE
C* W TEST, SO THE TEST IS REVERSED: OTABLE MUST CHECK FOR VALUES
C* THAT ARE LESS THAN THE NOMINAL VALUE, NOT GREATER THAN.
C*
            IF (MOD(I,4) .EQ. 0) THEN
               DO 20 J=1,COUNT
                  IF ( ARRAY(J) .LT. N3(I) ) THEN
                     I3=I3+1

                     IF ( ARRAY(J) .LT. N2(I) ) THEN
                        I2=I2+1

                        IF ( ARRAY(J) .LT. N1(I) ) THEN
                           I1=I1+1
                           ENDIF
                        ENDIF
                     ENDIF
20                CONTINUE

C*
C* ELSE THE FILE CONTAINS VALUES FROM A R, T, OR V TEST.
C*
            ELSE
               DO 30 J=1,COUNT
                  IF ( ARRAY(J) .GT. N3(I) ) THEN
                     I3=I3+1

                     IF ( ARRAY(J) .GT. N2(I) ) THEN
                        I2=I2+1

                        IF ( ARRAY(J) .GT. N1(I) ) THEN
                           I1=I1+1
                           ENDIF
                        ENDIF
                     ENDIF
30                CONTINUE
               ENDIF

            O1(I)=I1/FLOAT(COUNT)
            O2(I)=I2/FLOAT(COUNT)
            O3(I)=I3/FLOAT(COUNT)

         ELSE
            CALL REMARK( 'UNEXPECTED *EOP ON READ OF '//UNITLFN )
            O1(I)=O2(I)=O3(I)=0.0
            ENDIF
```

```
10      CONTINUE

      WRITE(1,*) 'O1=',O1
      WRITE(1,*) 'O2=',O2
      WRITE(1,*) 'O3=',O3

      WRITE(*,'(A)') '1'

      WRITE(*,'(1X,A)') TITLE

      WRITE(*,*) ' '
      WRITE(*,'(T6,A,T15,A,2X,A,T31,4(A,13X)')
     +   'ALPHA', 'P', 'TEST', 'R', 'T', 'V', 'W'
      WRITE(*,*) ' '

      IF (SHOWB) THEN
         WRITE(*,100)
     +      '0.01', '5', 'B', (O1(I),I=1,4), 'C', (O1(I),I=5,8),
     +      'L', (O1(I),I=9,12)
         WRITE(*,110)
     +      '4', 'B', (O1(I),I=13,16), 'C', (O1(I),I=17,20),
     +      'L', (O1(I),I=21,24)

         WRITE(*,100)
     +      '0.05', '5', 'B', (O2(I),I=1,4), 'C', (O2(I),I=5,8),
     +      'L', (O2(I),I= 9,12)
         WRITE(*,110)
     +      '4', 'B', (O2(I),I=13,16), 'C', (O2(I),I=17,20),
     +      'L', (O2(I),I=21,24)

         WRITE(*,100)
     +      '0.10', '5', 'B', (O3(I),I=1,4), 'C', (O3(I),I=5,8),
     +      'L', (O3(I),I= 9,12)
         WRITE(*,110)
     +      '4', 'B', (O3(I),I=13,16), 'C', (O3(I),I=17,20),
     +      'L', (O3(I),I=21,24)

      ELSE
         WRITE(*,100)
     +      '0.01', '5', 'C', (O1(I),I=5,8),
     +      'L', (O1(I),I=9,12)
         WRITE(*,110)
     +      '4', 'C', (O1(I),I=17,20),
     +      'L', (O1(I),I=21,24)

         WRITE(*,100)
     +      '0.05', '5', 'C', (O2(I),I=5,8),
     +      'L', (O2(I),I= 9,12)
         WRITE(*,110)
     +      '4', 'C', (O2(I),I=17,20),
     +      'L', (O2(I),I=21,24)

         WRITE(*,100)
     +      '0.10', '5', 'C', (O3(I),I=5,8),
     +      'L', (O3(I),I= 9,12)
         WRITE(*,110)
     +      '4', 'C', (O3(I),I=17,20),
     +      'L', (O3(I),I=21,24)
         ENDIF
      STOP
100   FORMAT( T6,A,T15,A,3(T19,A,T20,4F14.5,/) )
```

```
 110    FORMAT( T15,A,3(T19,A,T20,4F14.5,/) )
        END
*EOSOO  LINE=187 SEC=1
```

APPENDIX B

MONTE CARLO CRITICAL VALUES

The values in the following tables were determined under conditions of homogeneity and true null hypotheses. The tables were generated by the computer program written for this study. Values in the first table were used in determining actual significance levels and powers for 10,000 replications with k = 3 equal groups of size n = 20 and p measures. Values in the remaining tables were used in determining actual significance levels for 2,000 replications of the corresponding five combinations of k equal groups of size n. The hypotheses tested at three nominal alpha levels were:

B = between-group differences
C = within-group trends
L = within-group trends higher than linear

The test statistics used were:

R = Roy's largest root
T = Hotelling-Lawley trace
V = Pillai-Bartlett trace
W = Wilks' likelihood ratio

Table B-1

Monte Carlo Critical Values for 10,000 Replications with k = 3 and n = 20

ALPHA	P	TEST	R	T	V	W
0.01	5	B	.29748	.49304	.35917	.65905
		C	.21515	.26976	.20988	.78932
		L	.18695	.22451	.18255	.81699
	4	B	.26094	.40354	.31401	.70192
		C	.18757	.22685	.18358	.81548
		L	.15433	.18074	.15280	.84691
0.05	5	B	.23765	.37667	.29602	.71652
		C	.16293	.18695	.15624	.84298
		L	.13508	.15335	.13180	.86760
	4	B	.20652	.30743	.25208	.75682
		C	.13515	.15213	.13127	.86791
		L	.10385	.11464	.10222	.89770
0.10	5	B	.20872	.32239	.26275	.74786
		C	.13428	.15148	.13076	.86893
		L	.10842	.11927	.10559	.89380
	4	B	.18140	.26363	.22107	.78567
		C	.10952	.12156	.10774	.89212
		L	.08007	.08620	.07930	.92066

Table B-2

Monte Carlo Critical Values for 2,000 Replications
with k = 3 and n = 10

ALPHA	P	TEST	R	T	V	W
0.01	5	B	.53545	1.25236	.66298	.41601
		C	.42160	.66821	.38710	.61309
		L	.36360	.53917	.33503	.65774
	4	B	.48346	1.01626	.59320	.47011
		C	.36661	.56467	.34922	.64663
		L	.30948	.44764	.30053	.69648
0.05	5	B	.44937	.92470	.53044	.50591
		C	.32701	.45314	.29353	.69039
		L	.27214	.35114	.25202	.74533
	4	B	.39334	.74214	.46911	.56274
		C	.27732	.37156	.25615	.73542
		L	.21232	.26543	.20792	.79097
0.10	5	B	.39952	.76459	.48227	.54870
		C	.27756	.35781	.24814	.74014
		L	.22555	.27932	.21205	.78485
	4	B	.35537	.62687	.42039	.60098
		C	.22439	.27599	.20772	.78702
		L	.16391	.19107	.15748	.84065

Table B-3

Monte Carlo Critical Values for 2,000 Replications
with k = 2 and n = 20

ALPHA	P	TEST	R	T	V	W
0.01	5	B	.34340	.50937	.32607	.66774
		C	.30204	.41337	.28295	.71466
		L	.26384	.35165	.25351	.74351
	4	B	.30278	.42565	.30001	.70105
		C	.25855	.34089	.25086	.74716
		L	.21301	.27535	.21321	.78518
0.05	5	B	.27026	.34624	.24630	.74701
		C	.23348	.29027	.21786	.77803
		L	.19251	.23149	.18389	.81425
	4	B	.23248	.28844	.21809	.77825
		C	.19183	.23453	.18849	.81033
		L	.14905	.17455	.14793	.85173
0.10	5	B	.23284	.28805	.21639	.77896
		C	.18995	.23051	.18281	.81496
		L	.15478	.17803	.14927	.84981
	4	B	.19325	.23445	.18664	.81110
		C	.15692	.18489	.15403	.84457
		L	.11112	.12495	.11054	.88924

Table B-4

Monte Carlo Critical Values for 2,000 Replications
with k = 3 and n = 20

ALPHA	P	TEST	R	T	V	W
0.01	5	B	.29842	.48918	.35070	.66234
		C	.21377	.27422	.21171	.78701
		L	.19239	.23179	.18764	.81258
	4	B	.25618	.40140	.31066	.70302
		C	.18758	.23400	.18623	.81193
		L	.15721	.19199	.16227	.83836
0.05	5	B	.23578	.37212	.29158	.72123
		C	.16182	.18500	.15411	.84475
		L	.13092	.14981	.13025	.87023
	4	B	.20725	.30725	.25319	.75580
		C	.13667	.15228	.13055	.86852
		L	.10207	.11313	.10141	.89841
0.10	5	B	.20961	.32782	.26555	.74499
		C	.13191	.14621	.12685	.87265
		L	.10571	.11603	.10365	.89588
	4	B	.18420	.26972	.22469	.78222
		C	.10691	.11629	.10412	.89607
		L	.07937	.08576	.07868	.92109

Table B-5

Monte Carlo Critical Values for 2,000 Replications
with k = 6 and n = 20

ALPHA	P	TEST	R	T	V	W
0.01	5	B	.21560	.43440	.36468	.67156
		C	.10540	.11724	.10383	.89558
		L	.09101	.09914	.08992	.90983
	4	B	.20003	.36618	.30561	.71746
		C	.09068	.09890	.09017	.91001
		L	.07343	.07904	.07344	.92666
0.05	5	B	.17965	.36608	.31655	.71138
		C	.08016	.08734	.08017	.91975
		L	.06604	.07040	.06584	.93422
	4	B	.16148	.30003	.26445	.75457
		C	.06712	.07221	.06732	.93250
		L	.05084	.05370	.05087	.94905
0.10	5	B	.16269	.32807	.28637	.73667
		C	.06783	.07211	.06702	.93288
		L	.05502	.05795	.05486	.94519
	4	B	.14390	.26571	.23559	.77941
		C	.05418	.05770	.05457	.94543
		L	.03948	.04063	.03905	.96096

Table B-6

Monte Carlo Critical Values for 2,000 Replications with k = 3 and n = 50

ALPHA	P	TEST	R	T	V	W
0.01	5	B	.11607	.16125	.14473	.85858
		C	.09285	.10090	.09130	.90861
		L	.07806	.08467	.07806	.92194
	4	B	.10709	.14106	.12804	.87400
		C	.08215	.08669	.07970	.92028
		L	.07061	.07582	.07030	.92971
0.05	5	B	.09463	.12903	.11860	.88346
		C	.06224	.06574	.06145	.93839
		L	.05102	.05413	.05134	.94861
	4	B	.08505	.11040	.10205	.89938
		C	.05213	.05554	.05242	.94746
		L	.04089	.04300	.04128	.95875
0.10	5	B	.08161	.11256	.10367	.89740
		C	.05141	.05339	.05066	.94929
		L	.04084	.04238	.04065	.95936
	4	B	.07087	.09318	.08718	.91392
		C	.04241	.04428	.04235	.95761
		L	.03125	.03213	.03112	.96888

APPENDIX C

SIGNIFICANCE LEVELS FOR BETWEEN-GROUP TESTS

The following tables are actual significance levels expressed as percentage exceedance rates of Monte Carlo critical values for multivariate tests of between-group differences, B, calculated under heterogeneity levels, d. Values are based on 2,000 replications of five combinations of k equal groups of size n with p = 4 or 5 measures. In the first table k = 3 and sample size varies while in the second table n = 20 and the number of groups varies. The test statistics used were:

R = Roy's largest root
T = Hotelling-Lawley trace
V = Pillai-Bartlett trace
W = Wilks' likelihood ratio

Table C-1

Percentage Exceedance Rates Under a True Null
for Between-group Tests with k = 3*

			.01				.05				.10			
p	n	d	R	T	V	W	R	T	V	W	R	T	V	W
5	10	2	1.80	1.50	1.05	1.10	6.35	6.25	5.30	5.95	12.60	11.70	9.60	11.00
		4	3.95	3.15	.95	1.55	12.45	9.70	5.40	7.30	20.20	17.15	10.25	13.95
		9	11.05	7.45	1.35	2.55	22.75	15.80	5.40	9.70	32.10	24.20	8.45	15.90
	20	2	1.55	1.50	1.40	1.25	7.25	5.50	5.55	5.70	11.90	10.50	10.20	10.25
		4	4.00	2.45	1.75	2.30	11.35	8.50	6.95	8.00	17.05	12.85	11.70	12.55
		9	8.10	5.25	2.70	4.20	17.75	12.25	8.55	10.75	25.10	16.85	12.90	15.20
	50	2	1.90	1.40	1.35	1.35	5.60	5.50	5.45	5.35	11.85	10.05	10.20	10.00
		4	3.95	2.40	2.10	2.30	9.80	6.85	6.20	6.55	16.70	12.30	12.30	12.05
		9	7.15	3.50	2.90	3.25	15.65	8.80	7.85	8.10	22.70	14.70	13.60	13.95
4	10	2	1.35	1.30	.85	.95	6.25	5.75	5.20	6.10	11.10	10.95	10.05	10.20
		4	3.75	2.65	.90	1.80	10.85	8.70	5.85	7.40	15.85	14.75	10.30	12.75
		9	8.20	6.30	1.45	2.80	18.20	13.40	5.75	9.80	24.40	19.20	10.30	15.05
	20	2	2.05	1.35	1.30	1.35	6.95	6.40	5.85	5.95	11.20	10.20	9.75	10.00
		4	4.75	3.10	2.40	2.70	10.10	8.35	6.75	7.60	14.45	12.05	11.20	11.75
		9	8.45	5.15	2.75	4.10	14.95	10.85	7.70	9.30	20.35	15.00	12.60	14.05
	50	2	1.25	1.05	1.00	1.00	5.40	4.90	4.95	4.90	10.90	10.85	11.05	10.95
		4	2.55	1.75	1.50	1.65	8.30	6.10	6.10	6.10	15.45	12.40	12.40	12.40
		9	5.25	2.85	2.65	2.75	12.65	8.80	8.25	8.60	21.00	14.60	13.85	14.25

* Explanatory remarks appear on p. 143

Table C-2

Percentage Exceedance Rates Under a True Null
for Between-group Tests with n = 20*

			.01				.05				.10			
p	k	d	R	T	V	W	R	T	V	W	R	T	V	W
5	2	2	1.15	1.00	1.00	.95	5.30	5.30	5.65	5.45	10.90	11.20	10.75	11.00
		4	1.70	1.35	1.35	1.30	6.70	6.60	6.40	6.80	13.25	12.45	11.85	12.40
		9	2.35	1.95	1.95	1.95	8.85	9.25	8.05	8.95	15.25	14.55	12.75	13.95
	3	2	1.55	1.50	1.40	1.25	7.25	5.50	5.55	5.70	11.90	10.50	10.20	10.25
		4	4.00	2.45	1.75	2.30	11.35	8.50	6.95	8.00	17.05	12.85	11.70	12.55
		9	8.10	5.25	2.70	4.20	17.75	12.25	8.55	10.75	25.10	16.85	12.90	15.20
	6	2	1.80	1.20	1.50	1.30	6.90	5.60	5.40	5.45	12.20	10.35	10.45	10.45
		4	6.40	2.75	1.90	2.20	15.85	7.95	6.55	7.45	21.60	13.80	12.00	13.00
		9	18.80	7.40	3.80	5.35	31.15	14.60	9.50	12.25	39.30	20.30	15.65	18.05
4	2	2	1.05	1.10	.90	1.10	5.35	5.35	5.40	5.55	10.95	10.60	10.40	10.35
		4	1.45	1.40	1.35	1.45	6.15	6.00	6.15	5.85	13.15	12.55	11.65	12.05
		9	1.95	1.75	1.40	1.55	8.05	7.65	7.40	7.35	15.00	13.65	12.35	13.05
	3	2	2.05	1.35	1.30	1.35	6.95	6.40	5.85	5.95	11.20	10.20	9.75	10.00
		4	4.75	3.10	2.40	2.70	10.10	8.35	6.75	7.60	14.45	12.05	11.20	11.75
		9	8.45	5.15	2.75	4.10	14.95	10.85	7.70	9.30	20.35	15.00	12.60	14.05
	6	2	1.55	1.25	1.20	1.35	6.25	5.70	5.15	5.35	12.05	10.10	10.30	10.20
		4	4.50	2.65	2.25	2.55	12.85	7.25	6.35	6.65	20.15	13.30	12.30	12.95
		9	14.15	5.95	4.05	5.15	25.95	13.35	9.20	11.70	34.15	19.25	16.00	17.95

* Explanatory remarks appear on p. 143

APPENDIX D

SIGNIFICANCE LEVELS FOR WITHIN-GROUP TESTS OF NON-LINEARITY

The following tables are actual significance levels expressed as percentage exceedance rates of Monte Carlo critical values for multivariate within-group tests of the null hypothesis of no trends higher than linear, L, calculated under heterogeneity levels, d. Values are based on 2,000 replications of five combinations of k equal groups of size n with p = 4 or 5 measures. In the first table k = 3 and sample size varies while in the second table n = 20 and the number of groups varies. The test statistics used were:

R = Roy's largest root
T = Hotelling-Lawley trace
V = Pillai-Bartlett trace
W = Wilks' likelihood ratio

Table D-1

Percentage Exceedance Rates Under a True Null
for Within-group Tests of Non-linearity with k = 3*

			.01				.05				.10			
p	n	d	R	T	V	W	R	T	V	W	R	T	V	W
5	10	2	1.30	1.30	1.15	1.35	5.10	5.50	5.30	5.50	10.60	10.65	10.65	10.90
		4	1.70	1.65	1.60	1.50	6.85	6.90	6.60	6.85	12.40	12.25	11.65	12.10
		9	2.75	2.55	2.35	2.45	9.20	9.55	8.10	9.20	15.55	14.70	13.60	14.50
	20	2	.95	1.25	1.25	1.25	5.40	5.20	5.20	5.30	10.60	10.60	10.60	10.40
		4	1.25	1.45	1.40	1.40	6.20	5.95	5.75	5.80	11.90	11.60	11.45	11.40
		9	1.55	1.55	1.40	1.55	7.95	7.40	6.95	7.25	13.75	13.05	12.70	12.85
	50	2	1.00	.85	.90	.85	5.45	5.35	5.40	5.40	10.75	10.85	10.80	10.85
		4	1.00	.95	.90	.90	6.15	5.80	5.80	5.80	11.55	11.15	11.15	11.15
		9	1.20	1.00	1.00	1.00	6.85	6.25	6.20	6.20	12.50	12.35	12.25	12.25
4	10	2	1.05	.85	1.00	1.00	5.20	4.95	4.75	4.90	10.10	10.55	10.85	10.60
		4	1.55	1.35	1.40	1.45	5.65	5.45	5.10	5.20	10.65	10.80	11.10	10.85
		9	2.00	1.80	1.60	1.70	6.85	6.35	6.15	6.20	12.45	12.55	12.50	12.50
	20	2	1.25	1.05	1.00	1.00	5.55	5.40	5.35	5.40	10.15	10.05	10.20	10.05
		4	1.70	1.45	1.30	1.30	5.80	5.65	5.50	5.55	11.45	11.40	11.40	11.35
		9	2.20	1.75	1.65	1.75	6.40	6.30	6.15	6.30	12.15	11.80	1.70	11.80
	50	2	.85	.80	.85	.85	5.50	5.35	5.35	5.35	10.10	10.40	10.35	10.35
		4	.70	.65	.60	.65	5.90	5.70	5.70	5.70	10.75	10.65	10.65	10.65
		9	.85	.85	.85	.85	6.15	5.80	5.80	5.80	11.45	11.25	11.20	11.25

* Explanatory remarks appear on p. 146

Table D-2

Percentage Exceedance Rates Under a True Null
for Within-group Tests of Non-linearity with n = 20*

			.01				.05				.10			
p	k	d	R	T	V	W	R	T	V	W	R	T	V	W
5	2	2	1.00	.85	.95	.85	5.00	5.05	5.05	5.15	10.65	11.00	10.90	10.90
		4	.95	.95	1.20	.90	5.35	5.50	5.40	5.50	11.85	11.50	11.15	11.40
		9	1.20	1.30	1.40	1.45	6.65	6.55	6.15	6.60	12.60	12.55	12.55	12.50
	3	2	.95	1.25	1.25	1.25	5.40	5.20	5.20	5.30	10.60	10.60	10.60	10.40
		4	1.25	1.45	1.40	1.40	6.20	5.95	5.75	5.80	11.90	11.60	11.45	11.40
		9	1.55	1.55	1.40	1.55	7.95	7.40	6.95	7.25	13.75	13.05	12.70	12.85
	6	2	1.05	.90	.90	.90	5.25	5.50	5.45	5.50	9.75	9.90	9.75	9.80
		4	1.25	1.25	1.25	1.25	5.70	5.95	6.00	5.95	10.35	9.85	9.80	9.80
		9	2.15	1.95	1.95	1.95	6.85	6.75	6.55	6.70	11.90	11.60	11.45	11.50
4	2	2	1.00	1.00	1.05	1.05	5.10	5.20	5.10	5.15	10.90	10.70	10.70	10.65
		4	1.15	1.00	1.15	1.05	5.50	5.25	5.20	5.15	11.95	11.65	11.55	11.60
		9	1.40	1.20	1.10	1.10	6.10	6.00	5.95	6.05	12.30	11.85	11.80	11.85
	3	2	1.25	1.05	1.00	1.00	5.55	5.40	5.35	5.40	10.15	10.05	10.20	10.05
		4	1.70	1.45	1.30	1.30	5.80	5.65	5.50	5.55	11.45	11.40	11.40	11.35
		9	2.20	1.75	1.65	1.75	6.40	6.30	6.15	6.30	12.15	11.80	1.70	11.80
	6	2	.95	.90	.90	.90	5.15	4.95	4.95	4.95	10.55	10.50	10.55	10.55
		4	1.15	1.00	1.00	1.00	5.70	5.75	5.80	5.75	10.90	10.80	10.80	10.80
		9	1.65	1.65	1.65	1.65	5.85	5.80	5.90	5.90	11.15	10.90	11.10	11.00

* Explanatory remarks appear on p. 146

APPENDIX E

POWER VALUES FOR WITHIN-GROUP TESTS

The following tables include nominal powers under homogeneity (where d = 1) and actual powers under three heterogeneous conditions (where d = 2, 4, or 9). Values are expressed as percentage exceedance rates of Monte Carlo critical values for tests of two multivariate within-group hypotheses: (1) of no trends over the occasions, C, and (2) of no trends higher than linear, L. These values are based on 2,000 replications of five combinations of k equal groups of size n with p = 4 or 5 measures. The mean vectors used to transform the RM data to reflect a polynomial trend were (0 .4 .8 .5 .1) for p = 5 and (0 .4 .8 .5) for p = 4. The test statistics used were:

R = Roy's largest root
T = Hotelling-Lawley trace
V = Pillai-Bartlett trace
W = Wilks' likelihood ratio

The averages in Tables 5-7 and 5-8 were calculated from the corresponding values in these tables.

Table E-1

Percentage Exceedance Rates Under True Alternatives
for Within-group Tests of Trends with k = 3*

			.01				.05				.10			
p	n	d	R	T	V	W	R	T	V	W	R	T	V	W
5	10	1	43.95	47.70	39.75	48.75	69.65	71.55	70.00	70.30	81.00	82.50	80.95	82.00
		2	31.90	35.35	29.10	36.05	57.95	60.05	56.25	58.60	70.80	71.95	70.50	71.40
		4	22.75	23.95	17.45	23.25	45.00	46.55	42.10	44.45	58.95	60.00	55.55	58.25
		9	16.70	17.00	10.80	15.40	34.00	33.25	27.00	29.45	45.35	45.00	38.65	43.00
	20	1	91.60	91.05	91.30	91.35	98.00	98.30	98.40	98.35	99.50	99.50	99.55	99.50
		2	78.60	77.55	77.65	77.90	91.25	92.50	92.90	92.60	96.00	96.60	96.70	96.55
		4	57.50	55.25	54.90	55.20	76.55	78.25	77.75	78.05	85.95	86.80	86.75	87.00
		9	30.90	28.30	27.70	28.45	52.45	52.85	52.45	52.55	64.45	65.20	65.15	65.35
	50	1	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		2	99.95	99.95	99.95	99.95	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		4	97.85	97.95	98.00	98.00	99.70	99.75	99.75	99.75	99.90	99.95	99.95	99.95
		9	76.05	76.20	76.10	76.25	93.10	93.10	93.00	93.05	96.30	96.35	96.35	96.35
4	10	1	38.25	38.50	36.75	39.25	63.00	64.15	65.90	65.55	76.50	78.25	78.90	78.40
		2	27.35	28.20	26.55	28.65	50.00	51.75	52.35	52.45	65.10	67.45	67.85	68.60
		4	19.05	19.25	17.20	19.75	38.65	39.15	38.50	39.25	51.60	52.70	52.40	52.85
		9	13.35	12.45	10.35	12.05	27.15	27.15	25.80	26.65	39.25	39.45	37.85	39.25
	20	1	85.35	84.90	85.10	85.05	95.15	95.90	96.10	95.85	98.15	98.10	98.15	98.15
		2	69.75	68.90	69.40	68.90	86.95	88.10	88.45	88.35	93.50	93.95	94.00	94.00
		4	46.60	45.00	46.25	45.40	68.75	71.10	71.20	71.05	81.55	82.35	82.20	82.55
		9	24.10	23.20	23.15	23.15	43.40	45.30	44.85	44.90	58.35	59.30	59.05	59.25
	50	1	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		2	99.45	99.55	99.55	99.55	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		4	95.00	95.80	95.85	95.80	99.30	99.25	99.25	99.25	99.65	99.65	99.65	99.65
		9	65.95	67.70	67.70	67.65	88.70	88.60	88.60	88.65	93.40	94.00	94.10	94.00

*Explanatory remarks appear on p. 149

Table E-2

Percentage Exceedance Rates Under True Alternatives
for Within-group Tests of Trends with n = 20*

			.01				.05				.10			
p	k	d	R	T	V	W	R	T	V	W	R	T	V	W
5	2	1	69.40	71.00	69.55	72.65	86.35	87.45	87.30	87.65	92.65	93.10	93.10	93.20
		2	46.25	47.70	46.30	49.00	68.95	70.80	70.90	71.20	81.35	80.70	80.65	80.85
		4	25.95	26.25	25.30	26.90	47.35	48.20	47.65	48.65	62.10	61.70	61.65	62.00
		9	14.05	13.70	12.65	13.80	28.75	28.15	27.10	27.95	41.30	39.85	38.50	39.45
	3	1	91.60	91.05	91.30	91.35	98.00	98.30	98.40	98.35	99.50	99.50	99.50	99.50
		2	78.60	77.55	77.65	77.90	91.25	92.50	92.90	92.60	96.00	96.60	96.70	96.55
		4	57.50	55.25	54.90	55.20	76.55	78.25	77.75	78.05	85.95	86.80	86.75	87.00
		9	30.90	28.30	27.70	28.45	52.45	52.85	52.45	52.55	64.45	65.20	65.15	65.35
	6	1	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		2	99.85	99.80	99.80	99.80	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		4	99.50	99.25	99.25	99.25	99.80	99.80	99.85	99.85	99.90	99.95	99.95	99.95
		9	93.30	92.85	93.00	93.00	97.70	97.25	97.15	97.25	99.00	98.85	98.80	98.80
4	2	1	62.40	63.50	62.85	63.80	81.45	82.30	82.00	82.25	89.60	90.05	90.35	90.15
		2	39.50	40.85	40.45	41.50	63.30	63.75	63.75	63.60	74.75	75.25	75.40	75.35
		4	22.15	22.65	22.50	22.50	42.35	41.95	41.40	41.65	54.85	54.90	55.00	54.75
		9	11.25	11.05	10.70	10.85	24.30	24.45	23.80	24.00	35.95	34.95	34.50	34.70
	3	1	85.35	84.90	85.10	85.05	95.15	95.90	96.10	95.85	98.15	98.15	98.15	98.15
		2	69.75	68.90	69.40	68.90	86.95	88.10	88.45	88.35	93.50	93.95	94.00	94.00
		4	46.60	45.00	46.25	45.40	68.75	71.10	71.20	71.05	81.55	82.35	82.20	82.55
		9	24.10	23.20	23.15	23.15	43.40	45.30	44.85	44.90	58.35	59.30	59.05	59.25
	6	1	99.90	99.90	99.85	99.85	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		2	99.40	99.45	99.45	99.45	99.90	99.90	99.90	99.90	100.00	100.00	100.00	100.00
		4	98.10	98.15	98.05	98.15	99.25	99.30	99.30	99.30	99.75	99.75	99.75	99.75
		9	86.75	87.00	86.65	86.95	95.15	94.70	94.70	94.70	97.45	97.50	97.55	97.55

*Explanatory remarks appear on p. 149

Table E-3

Percentage Exceedance Rates Under True Alternatives
for Within-group Tests of Non-linearity with k = 3*

			.01				.05				.10			
p	n	d	R	T	V	W	R	T	V	W	R	T	V	W
5	10	1	50.60	53.95	54.65	55.80	74.30	77.75	78.05	78.30	84.75	85.90	86.30	86.50
		2	37.95	41.11	41.30	42.30	61.30	64.75	65.70	66.05	74.20	75.45	75.95	75.95
		4	24.60	26.45	26.65	27.30	48.60	51.00	50.45	51.65	60.45	61.50	61.35	62.20
		9	15.90	16.95	15.45	16.95	32.25	34.20	32.20	34.40	44.55	45.25	43.70	44.90
	20	1	92.10	92.95	92.90	93.15	98.95	98.80	98.80	98.80	99.65	99.60	99.60	99.60
		2	79.70	80.90	81.00	81.10	94.10	94.60	94.60	94.70	97.70	98.00	98.00	97.95
		4	56.25	57.90	57.65	58.05	80.60	80.50	80.50	80.70	88.10	89.15	89.15	89.20
		9	29.00	30.10	29.40	30.25	54.40	54.40	54.20	54.65	67.10	67.95	67.60	67.55
	50	1	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		2	99.95	99.95	99.95	99.95	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		4	98.45	98.40	98.45	98.40	99.85	99.90	99.90	99.90	99.95	99.95	99.95	99.95
		9	80.45	80.70	80.70	80.75	94.20	94.00	94.00	94.00	96.75	96.95	96.90	96.90
4	10	1	16.35	16.05	17.00	16.90	39.25	39.50	39.45	39.75	53.55	54.90	55.25	54.80
		2	12.00	11.80	12.20	12.40	31.05	31.30	31.20	31.40	45.10	45.85	46.30	46.20
		4	8.85	8.40	8.15	8.40	23.80	24.00	24.15	24.15	35.70	36.00	36.15	36.15
		9	6.60	5.75	5.30	5.65	17.80	17.25	17.10	17.25	27.25	27.45	27.85	27.55
	20	1	44.70	41.90	41.50	41.80	72.60	72.75	72.75	72.75	82.60	82.85	83.15	82.85
		2	30.00	28.85	28.05	28.50	58.20	58.75	58.60	58.65	71.10	71.45	71.50	71.40
		4	19.15	17.55	17.05	17.30	40.60	40.75	40.90	40.90	54.80	55.00	55.15	55.05
		9	10.10	9.15	8.65	8.85	25.70	25.55	25.30	25.50	36.80	36.45	36.45	36.40
	50	1	92.50	92.85	92.90	92.85	99.00	98.95	98.95	98.95	99.65	99.70	99.70	99.70
		2	80.25	80.60	80.65	80.75	95.50	95.60	95.60	95.60	97.95	98.05	98.05	98.05
		4	56.35	56.55	56.95	57.00	84.05	83.70	83.60	83.65	90.85	91.40	91.45	91.40
		9	28.80	26.90	27.35	27.35	58.00	57.80	57.75	57.80	69.85	69.75	69.70	69.70

*Explanatory remearks appear on p. 149

Table E-4

Percentage Exceedance Rates Under True Alternatives
for Within-group Tests of Non-linearity with n = 20*

			.01				.05				.10			
p	k	d	R	T	V	W	R	T	V	W	R	T	V	W
5	2	1	73.25	74.05	75.40	75.20	89.10	90.05	90.60	90.35	94.40	95.15	95.45	95.25
		2	48.90	50.30	51.65	51.25	73.55	74.70	75.40	75.05	83.55	84.55	84.45	84.60
		4	26.35	26.75	27.75	27.20	50.30	52.10	53.40	52.70	63.65	65.35	65.95	65.60
		9	13.00	13.25	13.55	13.45	28.50	28.85	29.00	28.85	41.00	42.35	42.30	42.45
	3	1	92.10	92.95	92.90	93.15	98.95	98.80	98.80	98.80	99.65	99.60	99.60	99.60
		2	79.70	80.90	81.00	81.10	94.10	94.60	94.60	94.70	97.70	98.00	98.00	97.95
		4	56.25	57.90	57.65	58.05	80.60	80.50	80.50	80.70	88.10	89.15	89.15	89.20
		9	29.00	30.10	29.40	30.25	54.40	54.40	54.20	54.65	67.10	67.95	67.60	67.55
	6	1	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		2	99.95	99.95	99.95	99.95	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00
		4	99.45	99.40	99.50	99.40	99.95	99.95	99.95	99.95	99.95	99.95	99.95	99.95
		9	92.75	93.05	93.20	93.05	97.75	97.85	97.80	97.85	99.10	99.00	99.00	99.00
4	2	1	28.95	28.10	28.85	28.45	53.20	53.75	54.15	53.70	68.70	68.55	68.65	68.60
		2	16.75	16.15	16.20	16.35	36.10	36.05	36.00	36.00	53.40	54.00	53.95	53.95
		4	9.15	8.25	8.50	8.60	22.70	22.70	22.45	22.45	36.65	36.60	36.75	36.50
		9	4.65	4.25	4.35	4.35	13.85	13.30	13.25	13.20	24.45	23.65	23.60	23.60
	3	1	44.70	41.90	41.50	41.80	72.60	72.75	72.75	72.75	82.60	82.85	83.15	82.85
		2	30.00	28.85	28.05	28.50	58.20	58.75	58.60	58.65	71.10	71.45	71.55	71.40
		4	19.15	17.55	17.05	17.30	40.60	40.75	40.90	40.90	54.80	55.00	55.15	55.05
		9	10.10	9.15	8.65	8.85	25.70	25.55	25.30	25.50	36.80	36.45	36.45	36.40
	6	1	91.00	91.05	90.90	90.95	97.05	97.15	97.10	97.10	98.60	98.70	98.70	98.70
		2	85.40	85.30	85.10	85.25	94.25	94.15	94.25	94.15	97.10	97.35	97.35	97.35
		4	72.15	71.35	71.25	71.30	87.40	87.30	87.40	87.35	92.65	92.65	92.60	92.60
		9	47.95	48.20	48.05	48.10	68.85	68.85	69.05	68.85	79.75	80.20	80.25	80.20

*Explanatory remarks appear on p. 149

BIBLIOGRAPHY

BIBLIOGRAPHY

Anderson, T.W. (1958) An Introduction to Multivariate Statistical Analysis. New York: John Wiley and Sons.

Bock, R.D. (1963) Multivariate analysis of repeated measures. In C.W. Harris (ed.) Problems in Measuring Change (pp. 85-103). Madison, Wisconsin: University of Wisconsin Press.

Bock, R.D. (1975) Multivariate Statistical Methods in Behavioral Sciences. New York: McGraw-Hill Book Company.

Box, G.E.P. (1954) Some theorems on quadratic forms applied in the study of analysis of variance problems, II. Effects of inequality of variance and of correlation between errors in the two-way classification. Annals of Mathematical Statistics, 25, 484-498.

Box, G.E.P. and Muller, M.E. (1958) A note on the generation of random normal deviates. Annals of Mathematical Statistics, 29, 610-611.

Ceurvorst, R.W. (1980) Robustness of MANOVA under heterogeneity of variance and correlation. Unpublished doctoral dissertation. Arizona State University.

Collier, R.O., Baker, F.B., Mandeville, G.K., and Hayes, T.F. (1967) Estimates of test size for several test procedures based on conventional variance ratio in the repeated measures design. Psychometrika, 32, 339-353.

Davidson, M.L. (1972) Univariate versus multivariate tests in repeated measures experiments. Psychologocal Bulletin, 77, 446-452.

Finn, J.D. (1974) A General Model for Multivariate Analysis. New York: Holt, Rinehart and Winston, Inc.

Glass, G.V., Peckham, P.D., and Sanders, J.R. (1972) Consequences of failure to meet assumptions underlying the fixed-effects analysis of variance and covariance. _Review of Educational Research_, _42_, 237-288.

Greenhouse, S.W., and Geisser, S. (1959) On methods in the analysis of profile data. _Psychometrika_, _24_, 95-112.

Hakstian, A.R., Roed, J.C. and Lind, J.C. (1979) Two-sample T^2 procedure and the assumption of homogeneous covariance matrices. _Psychological Bulletin_, _86_, 1255-1263.

Hammersley, J.M. and Handscomb, D.C. (1964) _Monte Carlo Methods_. New York: Barnes and Noble, Inc.

Harris, R.J. (1975) _A Primer of Multivariate Statistics_. New York: Academic Press.

Hays, W.L. (1973) _Statistics for the Social Sciences_. New York: Holt, Rinehart, and Winston, 1973.

Holloway, L.N. and Dunn, O.J. (1967) The robustness of Hotelling's T^2. _Journal of the American Statistical Association_, _62_, 124-136.

Hopkins, J.W. and Clay, P.P.F. (1963) Some empirical distributions of bivariate T^2 and homoscedasticity criterion M under unequal variance and leptokurtosis. _Journal of the Americal Statistical Association_, _58_, 1048-1053.

Huynh, H. and Feldt, L.S. (1970) Conditions under which mean square ratios in repeated measuresments designs have exact F-distributions. _Journal of the American Statistical Association_, _65_, 1582-1589.

Ito, K. (1962) A comparison of the powers of two multivariate analysis of variance tests. _Biometrika_, _49_, 455-462.

Ito, K. (1969) On the effect of heteroscedasticity and nonnormality upon some multivariate test procedures. In P.R. Krishnaiah (ed.), _Multivariate Analysis II_ (pp.87-120) New York: Academic Press.

Ito, P.K. (1980) Robustness of ANOVA and MANOVA test procedures. In P.R. Krishnaiah (ed.) _Handbook of Statistics, Vol. I_ (pp. 199-236) North Holland Publishing Company.

Ito, K. and Schull, W.J. (1964) On the robustness of the T_0^2 test in multivariate analysis of variance when variance-covariance matrices are not equal. *Biometrika*, *51*, 71-82.

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

Lehman, R.S. (1977) *Computer Simulation and Modeling*. Hillsdale, New Jersey.

Mendoza, J.L., Toothaker, L.E. and Nicewander, W.A. (1974) A Monte Carlo comparison of the univariate and multivariate methods for the groups by trials repeated measures design. *Multivariate Behavioral Research*, *9*, 165-177.

Morrison, D.F. (1972) The analysis of a single sample of repeated measurements. *Biometrics*, *28*, 55-71.

Morrison, D.F. (1976) *Multivariate Statistical Methods*. New York: McGraw Hill Book Company, 1976.

Olson, C.L. (1973) A Monte Carlo investigation of the robustness of multivariate analysis of variance. Unpublished doctoral dissertation, University of Toronto.

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

Pillai, K.C.S. and Sudjana. (1975) Exact robustness of tests of two multivariate hypotheses based on four criteria and their distribution problems under violations. *The Annals of Statistic*, *3*, 617-636.

Potthoff, R.F. and Roy, S.N. (1964) A generalized multivariate analysis of variance model useful especially for growth curve problems. *Biometrika*, *51*, 313-326.

Ramsey, P.H. (1980) Exact Type I error rates for robustness of Student's t test with unequal variances. *Journal of Educational Statistics*, *5*, 337-349.

Scheifley, V. (1974) Analysis of repeated measures data: A simulation study. Unpublished doctoral dissertation, Michigan State University.

Scheifley, V. and Schmidt, W. (1978) Analysis of repeated measures data: A simulation study. Multivariate Behavioral Research, 13, 347-362.

Scheffé, H. (1959) The Analysis of Variance. New York: John Wiley and Sons.

Tatsuoka, M.M. (1971) Multivariate Analysis: Techniques for Educational and Psychological Research. New York: John Wiley and Sons.

Timm, N.H. (1975) Multivariate Analysis with Applications in Education and Psychology Research. Monteray, California: Brooks Cole Publishing Company.

Timm, N.H. (1980) Multivariate analysis of variance of repeated measures. In P.R. Krishnaiah (ed.) Handbook of Statistics, Vol I (pp. 41-87). New York: North Holland Publishing Company.